Physical Education, Health and Social Sciences

https://journal-of-social-education.org E-ISSN: 2958-5996

P-ISSN: 2958-5988

Artificial Intelligence-Based Fraud Detection Technique Using Unsupervised Machine learning for Real Estate Property Insurance

Fadia Shah[1], Faiza Shah[2], Yasir Shah[3], Saman Ikram Abbasi[4], Aftab Hussain Tabasam[5]

[1] Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan. Email:fadiashah13@yahoo.com

[2] School of Political Science and Public Administration, Henan Normal University, China. Email: faizashah55@gmail.com

[3] School of Business, Zhengzhou University, Zhengzhou 450001, China. Email: yasirshah_pk@yahoo.com

[4] Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad. Email: 2173114@szabist-isb.pk

[5] Business Administration, University of Poonch Rawalakot, Pakistan. Email: aftabtabasam@upr.edu.pk

DOI: https://doi.org/10.63163/jpehss.v3i3.829

Abstract

When it comes to money, insurance fraud is a serious issue and still a challenging area of interest. This study proposes that Artificial Intelligence (AI) and its branch of unsupervised machine learning can effectively detect such fraud. A dataset of 1,070,994 property records with 32 structured data fields, location, value, and relevant features, was chosen using clustering algorithms. In addition, Principal Component Analysis (PCA) and an autoencoder for retaining essential information were technically used in this study. The top ten properties exhibiting the highest fraud scores were identified. This study includes other factors to identify anomaly scores for unusual patterns. They are extremely low or high values. The results were significantly high, at almost 97%. This system opens up a direction for researchers to identify ambiguities or fraudulent claims in insurance systems.

Keywords: Insurance Fraud Detection, Unsupervised learning, Machine Learning, Data Anomaly.

Introduction

Financial systems are always at a high risk of fraud, especially the data of insurance companies or agents. Fraud in the insurance industry is always frequent. Owing to the importance of the insurance sector, the legal framework of many countries gives vital importance to insurance systems. Insurance corporations have extended approximately 40 billion euros in compensation for damages. This is the reason why this sector is highly sensitive and why there are more chances of fraudulent activities in this sector. The data are presented so neatly that the claims seem very original. Numerous methods have been proposed by researchers to solve this problem and identify the data with doubts [1]. This is important for maintaining the integrity of the industry. System reliability must be sustained to ensure that claims are properly compensated. Otherwise, either the deserving case will not be facilitated or the wrong claims can be facilitated. Fraud can occur at various stages, including during the application process, claim handling by insurers, and the

involvement of third-party claimants and experts. Insurance brokers and company employees can also commit insurance fraud. Common frauds include exaggerating claims, lying on an insurance application, submitting claims for harm or injuries that never materialized and staging accidents. Fraudulent insurance claimants include organized fraudsters who commit large-scale financial crimes using false business practices, professionals, and technicians who charge for services that are not provided or inflate service rates [2]. People in the general public who want to pay their deductible or see a chance to earn a little money.

Insurance companies that have not yet adopted high-tech technologies should move quickly. False or inflated reports of property damage, burglary, theft, detail, and intentional damage claims are also examples of property insurance fraud. Furthermore, the growing complexity and volume of data in the digital age make it more challenging [3] for these traditional methods to detect fraud efficiently. Therefore, there is a pressing need to develop more advanced and automated techniques to identify and prevent fraudulent activities.

Machine learning works in two ways: supervised machine learning for fraud detection faces challenges such as the need for labelled data and the identification of new types of fraud. The second is unsupervised machine learning, in which there are no labelled data. This approach is chosen for insurance fraud detection because it identifies anomalies without labelled data, detects unknown fraud types, and handles large datasets [4]. Various algorithms have been implemented to provide solutions. The approach automatically discovers patterns and anomalies in large datasets. There was no prior knowledge of fraudulent cases.

To detect fraud, insurance claims primarily rely on data. Existing rule-based systems, called expert systems [5], automatically identify potentially fraudulent claims. However, they are suitable for obsolete systems and cannot handle complex fraud nowadays. To achieve more advancements in the fraud detection ecosystem, businesses must possess advanced solutions. For this purpose, the proposed approach is promising, as it can autonomously analyze data and find patterns without being labelled in the data. Unsupervised algorithms can analyze large datasets and detect anomalous patterns without prior knowledge of what fraud looks like. This is why unsupervised machine learning is a valuable approach for detecting new and previously unknown fraud patterns.

A scholar claims that research [6] presents a useful approach to handling data issues, especially fraudulent data, and enhances fraud detection by identifying the most relevant data for property tax. It also calculates the accuracy of the proposed fraud detection system to determine the system reliability. The study is 787% accurate in identifying patterns and findings, which can help other industries with similar problems. This could lead to the development of better and smarter systems. However, the issue in that research was that the dataset was too small, and some of the missing data were based on averages, which may disturb the accuracy.

Another study was conducted by the researcher [7], claiming that the same problem was solved again with the ML approach but with supervised learning. The primary objective of this study is to overcome fraudulent patterns in real estate property insurance. The model to identify unusual patterns and variations in the dataset was based on the identification of unusual patterns within the data. Principal Component Analysis (PCA) was used as part of our methodology. Its purpose is to simplify the data and apply data preprocessing techniques to handle missing information and outliers in the dataset.

Literature Review

Fraud is a broad term that indicates that it is a rare, deliberate, and organized crime that is difficult to trace. A survey conducted in 2019 identified that insurance fraud accounted for approximately 10% of all expenses in Europe [8]. It is also observed that the number of fraud cases in the insurance sector increases by almost 5% every year. The data from Chinese officials also states a very high proportion of financial expenditure; if untreated, the companies have to pay the

individual who claims. According to a previous study [9], fraud detection systems based on ensemble learning are efficient. The XGBoost model enhanced the gradient hosting framework and generated good efficiency. The performance of the proposed method is highly dependent on parameter optimization. The study also claims that standard features alone are not suitable for capturing complex patterns in fraud identification. The model can analyze data using clustering algorithms. Further improvement to detect fraudulent activities in the data, for the same research [10] proposed models that include many supervised algorithms, such as Support Vector Machine (SVM), decision tree, random forest, logistic regression, and many others. The solution proposed by [11] presented several ML models to predict fraud in property insurance claims using actual data from a Brazilian insurance company. They found that, on average, deep neural networks and ensemble-based techniques performed the best. They also used explainable Artificial Intelligence techniques to create a profile of known fraudsters and analyzed the impact of different features on the classification performance. The RF model had the best performance in terms of precision, accuracy, F1 Score, Cohen's Kappa, and MCC, but the deep neural network model had the highest recall. These models can also be transformed into a probabilistic approach, which can also be used to compute the expected return or loss of an insurance policy and the insurance premium accounting for fraud.

As discussed in a study [12], the authors predicted in advance whether the policy claims were made for abuse to reduce claim payments in the insurance industry's property branch and prevent any financial losses. Label Spreading and Self Training semi-supervised machine learning approaches were used to detect abnormal situations owing to the low label ratio of the anomaly class in the dataset. After the data labelling process, a fraud prediction study was conducted using supervised machine learning models, and the semi-supervised learning approach with the highest performance was identified. The datasets include policy demands in 2017 and 2018 and weather information for that location. Accuracy, precision, recall, specificity, and F1 score were evaluated as model success measures, and according to these results, the self-training approach could label data with higher performance than the Label Spreading approach. The ratio of the anomaly class [13], which constitutes a small part of the dataset, was replicated using semi-supervised machine learning models. While developing semi-supervised models, policy information such as coverage, tariff, and channel was used. Policy claims for abuse were estimated using supervised machine learning models, using datasets whose label ratio was replicated separately according to both label spreading and self-training model results. In the developed predictive models, weather data and whether the policy was issued for the first time were essential factors. The results showed that in fraud prediction, the dataset labelled with the self-training approach outperformed the Label Spreading approach, especially in terms of accuracy, recall, and specificity success metrics.

The solution proposed by [14] used ML techniques, specifically the random forest and K-nearest neighbors (KNN) algorithms, to detect vehicle insurance fraud. They compared the performance of these algorithms using various criteria and found that either algorithm could be used to accurately predict fraud in insurance data. The system takes input from users and uses either the random forest or KNN algorithm to classify the claim as fraudulent or not. When using the KNN algorithm, it is advisable to choose an odd value for the number of nearest neighbors (k) to avoid ties during the majority vote when there is an even number of classes.

In [15], the authors conducted a content analysis of 46 academic studies on this topic and found that traditional statistics-based techniques were being replaced by data mining and AI-based methods. This study also identified cost-sensitive and hybrid approaches as promising areas for further research. The adoption of AI in the insurance industry, including for detecting insurance fraud, may raise regulatory challenges, such as managing the pace of technological progress and the extent of regulation, as well as compliance with data protection regulations, such as the General Data Protection Regulation (GDPR).

For instance, [16] proposes an ML-based credit card fraud detection engine that uses a genetic algorithm for feature selection and various classifiers, including LR, DT, RF, ANN, and Naive Bayes. The proposed system was tested using a dataset of European credit card transactions and was found to have high accuracy, outperforming other similar techniques. The system was also tested on a synthetic dataset and was found to have 96% accuracy. The authors suggest further testing on additional datasets in future studies.

For categorical sequence embeddings [17], deep learning is used to classify valid and fraudulent declarations. In this study, we propose deep learning architectures specifically for solving each of the difficulties and claims data. First, it can be difficult to handle claims data using traditional machine learning techniques because they are usually available in an unstructured format. Second, fraud data are very unbalanced because fraudulent cases are quite rare compared to non-fraudulent cases, and each fraud might be viewed as an anomaly. This study examines the performance of traditional machine learning models and our suggested approaches in classifying claims based on real-world data. Our most suitable model produced a ROC AUC score of 0.873 when extracting unstructured categorical sequences connected to outpatient visits. In contrast, the most advanced model produced a lower outcome with an ROC AUC of 0.815. Empirical tests show that we can further improve our model by enhancing the design of neural networks, gathering more training data, and implementing plans to address the class imbalance issue.

To obtain a practical understanding of the behavior of an insured individual using unsupervised deep learning, we propose a novel deep learning methodology. A new variable importance methodology was proposed by [18] and combined with two popular autoencoder and variational autoencoder unsupervised deep learning models. All model dynamics are analyzed to provide the reader with a better understanding of how they may be changed for fraud detection and how the results can be appropriately evaluated. Performance evaluations were conducted qualitatively and quantitatively. This study used two sets of data. The suggested method for fraud variable importance analysis comprises three steps.

Proposed Model

The framework model, shown in Figure 1, involves data collection and preprocessing in the initial stage, wherein relevant data on past insurance claims and property transactions are gathered. The data were then subjected to exploratory analysis and visualization techniques, such as scatter plots and histograms, allowing for a comprehensive understanding of the data distribution, relationships, and identification of potential patterns and anomalies indicative of fraudulent activity. Moreover, the Z-scale calculation standardizes the variables, ensuring fair comparisons across different scales[19,20]. The framework incorporates clustering as a fundamental component. Clustering algorithms, such as the K harmonic mean, are employed to group data points into clusters. Figure 1 shows the unsupervised machine learning architecture for fraud detection in the real estate industry using PCA to identify fraudsters.

Principal Component Analysis (PCA) and clustering algorithms [21,22] are the basis for extracting meaningful features from the dataset. This is a dimensionality reduction technique. This helps identify the most valuable variables in the deduction of the data. The features selected for further processing contribute significantly to the development of a fraud detection model [23]. The clustering algorithms consolidate similar data points by forming the same cluster; in this way, potential fraud and non-fraud cases are identified.

The framework is completed by a detailed validation process. The main purpose is to ensure the efficacy of the model, which proves whether it is reliable. This is important because the known fraud cases should be confirmed and can be identified for evaluation. The model was tested on a separate dataset to ensure that it could effectively identify fraudulent activities after complete algorithm training. Although PCA and clustering algorithms collectively contribute to the model's

accuracy, a thorough performance evaluation is necessary. This evaluation involved metrics such as precision, recall, and F1-score to understand the model's strengths and identify potential areas for improvement.

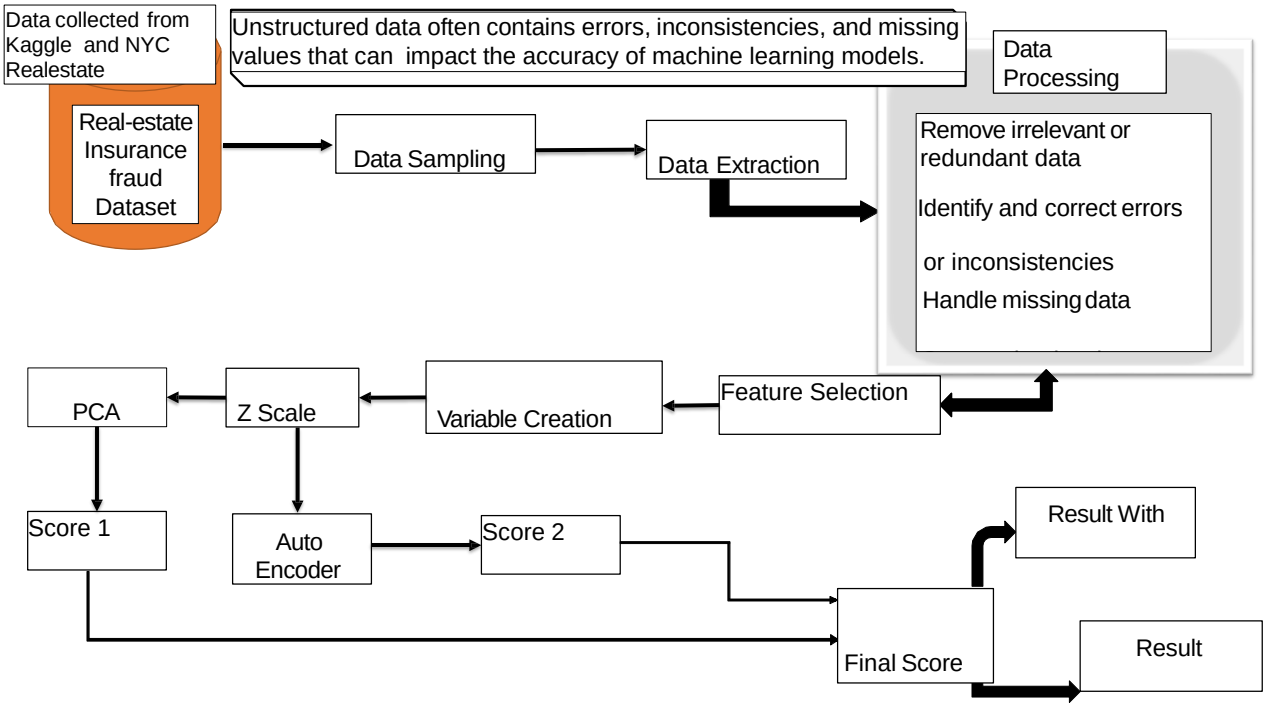

Figure 1: The Proposed Framework

The model also explains the use of AI to detect fraud in the real estate industry. Technically, it begins with data collection. Data sampling ensures that the processing steps for further analysis preserve the dataset. Data extraction was then performed for unsupervised insurance fraud detection. Here, the relevant data were extracted from various sources and collectively considered for analysis. Next, the data processing step is used to ensure the quality and integrity of the data. This is achieved by removing irrelevant or redundant data. Unnecessary data that did not contribute to the fraud detection were removed. Furthermore, the data quality was improved by reducing the noise. Data inconsistencies, such as incorrect entries or conflicting information, were removed in this step. All such anomalies were identified and rectified to maintain data accuracy. The next phase was to handle the missing data points. They are addressed through imputation or other techniques to maintain data completeness. Data bias should be avoided. The data were then converted into a usable format. The data were transformed into a format suitable for analysis. This can be done by converting numerical or categorical variables to the desired form so that they are easily understood by machines. Feature selection involves the selection of the most relevant features. Attributes that contribute significantly to fraud detection are chosen [24]. The key advantage of feature selection is that it reduces dimensionality and focuses only on the essential factors.

The next phase was variable creation to enhance the informativeness of the dataset and capture additional insights related to fraud patterns. The scaling was performed using the Z Scale, PCA, and AutoEncoder. All of these are dimensionality reduction techniques. They streamlined the dataset and extracted essential information. These are purely statistical methods. The Z Scale standardizes the data to have a mean of 0 and a standard deviation of 1, making it easier to compare

different variables. Principal Component Analysis (PCA) identifies and extracts the most critical components from a dataset, reducing its dimensionality while preserving its essential features. An autoencoder is a neural network that learns to compress data into a lower-dimensional representation and reconstructs the original data from the compressed representation.

A random sampling technique was used to select a small subset of observations to ensure that the sampled data were representative of the larger dataset [25]. To address this issue, the sample size was increased by selecting a complete original dataset. This allowed for a more representative sample of the population and increased the accuracy of the results [26]. One of the main issues was ensuring that the sample was representative of the larger dataset. To address this, a random sampling technique was used, and the selected observations were spread across different time periods and geographic regions. Another issue was the presence of missing values and outliers in the dataset. These can significantly impact the analysis results and must be handled appropriately. To address this, data imputation techniques and outlier detection methods were used to ensure that the sampled data were clean and ready for analysis. In data imputation, imputation for missing values is explicitly performed.

Datasets Description

The data used in this study were publicly available datasets produced by the NYC Department of Finance. The dataset consists of property valuations and assessments of NYC real estate, which were used to calculate property taxes and provide exemptions and/or abatements for eligible properties during NYC's 2010/2011 fiscal year. The dataset was created in September 2011 and updated in September 2018. With 1,070,994 property records, the dataset contains 32 data fields, of which 14 are numerical and 18 are categorical. The most critical data fields for identifying potential fraud cases were those related to the location of the property record. The dataset used in 1 contains essential information related to real estate properties in New York City. This is a comprehensive dataset that encompasses various property attributes, such as property values, measurements, and location details. The dataset serves as the foundation for fraud detection in the real estate industry, and its detailed description provides valuable insights into the properties under consideration. In conclusion, the rich and comprehensive dataset in Table 1 plays a crucial role in enhancing fraud detection capabilities within the real estate industry. The dataset contains essential information, including the Building Classification Code (BLDGCL) for NYC building classification, property measurements such as Lot Front Width (LTFRONT) and Lot Depth (LTDEPTH), and the Number of Stories (STORIES) for building structure. Additionally, it includes financial attributes such as Full Market Value (FULLVAL), Actual Land Value (AVLAND), and Actual Total Value (AVTOT) for property valuation analysis. The Zip Code (ZIP) attribute enables geospatial analysis, whereas the Building Front Width (BLDFRONT) and Building Depth (BLDDEPTH) attributes provide insights into architectural characteristics. This dataset is a valuable resource for detecting fraudulent activities in the real estate market and aids in building effective fraud detection models to ensure a transparent and secure industry.

Table 1: Summary Statistics for Data Values

Data Field	Records Val	Mean		Standard Deviation	Min	Max
AVLAND2	282,726	24, 6236		6, 178, 963	3	2, 371, 005, 000
AVTOT2	282,732	713, 911		11, 652, 529	3	4, 501, 180, 002
AVLAND	1,070,994	85, 068		4, 057, 260	0	2, 668, 500, 000
AVTOT	1,070,994	227, 238		6, 877, 529	0	4, 668, 308, 947
BLDFRONT	1,070,994	23 ft		35 ft	0 ft	7,575 ft
BLDDEPTH	1,070,994	39 ft		42 ft	0 ft	9,393 ft

EXLAND2	87,449	351, 236		10, 802, 213	1	2, 371, 005, 000
EXTOT2	130,828	656, 769		16, 072, 510	7	4, 501, 180, 002
EXLAND	1,070,994	36, 424		3, 981, 576	0	2, 668, 500, 000
EXTOT	1,070,994	91, 187		6, 508, 403	0	4, 668, 308, 947
FULLVAL	1,070,994	874, 265		11, 582, 431	0	6, 150, 000, 000
LTFRONT	1,070,994	37 ft		74 ft	0 ft	9,999 ft
LTDEPTH	1,070,994	89 ft		76 ft	0 ft	9,999 ft
STORIES	1,014,730	5.0		8	1	119

Table 1 summarizes the existing data fields of the dataset used in the proposed research analysis. It includes columns for the number of records, mean, standard deviation, minimum, and maximum values for each field. All these fields are valuable data points required for statistical analysis. For example, the "AVLAND2" field has 282,726 records with a mean value of 24,6236 and a standard deviation of 6,178,963, which indicates that there are values different from the normal range of data. Other fields present information on building statistics, such as the frontage, depth, and number of stories. By examining the means of the columns that contain numerical data, one can determine the average value of the data. This is a baseline for comparing the actual values with those of deflected ones. Thus, unusual or extreme values that deviate significantly from the mean can indicate potential fraud or anomalies.

Dimensionality Reduction Calculation via PCA

In unsupervised learning fraud detection for real estate, Principal Component Analysis (PCA) plays a vital role in reducing the dimensionality of the dataset while preserving essential patterns [27]. Through PCA, the original features are transformed into uncorrelated principal components that capture the maximum data variance. This reduction enables a focus on pertinent information, potentially revealing hidden fraud patterns [28]. By reducing and truncating the least significant features, the system efficiency increases. This is a potential approach for using this method. PCA's major advantage of PCA is the improved overall performance and interpretability [29] of the fraud detection system.

Calculation to Reduce dimensions via PCA, discarding all but the top few PCs:

PCA can compute the principal components and their associated explained variance ratios from the qualitative dataset [30,31]. It is possible to plot the variance percentage, which is calculated and explained by each principal component while training the model. The model selects the top eight principal components based on their cumulative explained variance. It is important to note that it accounts for approximately 97.18% of the total variance. The resulting dataset has reduced dimensionality. Each observation is represented by a smaller number of principal components (k components) instead of the original features, which is the conversion of multivalued data to a single-valued component. This dimensionality reduction technique helps capture the most important information and patterns in the data while discarding less significant features, as it is statistically calculated.

Results and Discussion

The results indicate a successful data imputation process using the proposed model. The results are presented in various graphs below. The entire dataset was processed through the model. The system generated an accuracy of 98.7%, which is reliable. This indicates that the model was well-trained, and the unlabeled dataset was identified. The segmentation and clustering of each data record were calculated, and the relevant data were mapped to the same cluster. The counts of

missing values and zeros in various columns of interest demonstrate the effectiveness of the imputation technique, which avoids bias and estimation errors. The results are plotted in the graph below.

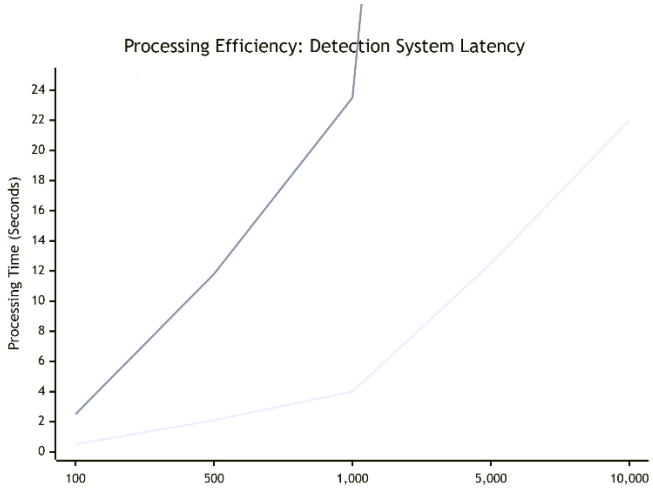

Figure 2 Efficiency calculation

Figure 2 shows a graph. This illustrates the scalability and efficiency of the fraud detection system. The detailed information presented in the figure shows the processing latency against the volumetric data. The system can handle 10,000 insurance claims in the form of data in 22 s.. This highlights the robust performance of the proposed model under load. This also indicates that the system is reliable and can perform complex calculations. This graph also indicates that there are chances of an increase in processing time, confirming the model's practical viability for large-scale deployment in insurance claim verification workflows. However, the results can be captured within a limited time.

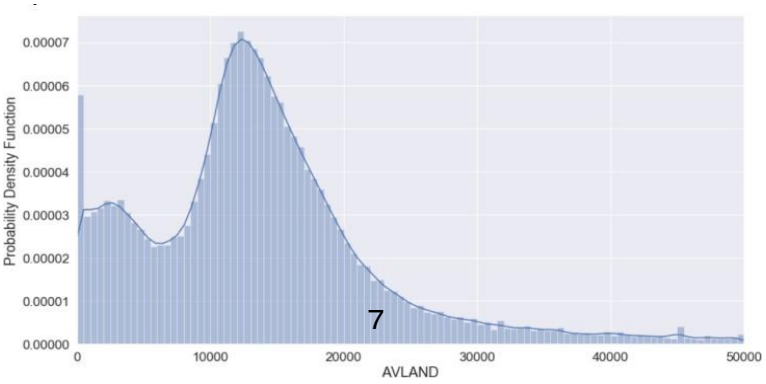

Figure 3: Property values and fraud chances

Figure 3 illustrates the scatter plot. This further illustrates the relationship between property values (FULLVAL), which range up to $2 million. The other component of the graph is an unspecified secondary metric that scales nonlinearly from 0 to 40,000. The highlighted trend demonstrated a strong positive correlation. The secondary value increases progressively with higher property valuation. This pattern potentially represents tax assessments, insurance premiums, or maintenance

costs relative to a property's total market value. It is typical to know that value and fraud activities are highly associated. Higher values indicate a higher probability of fraudulent activities.

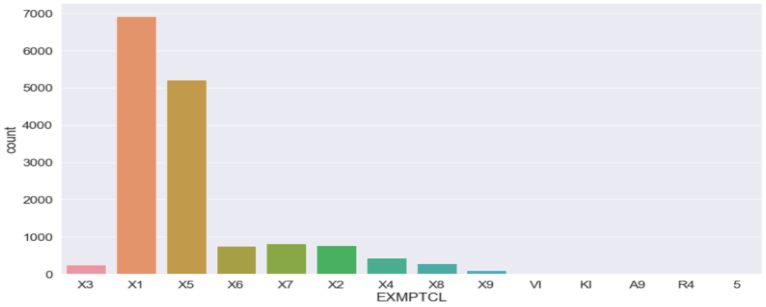

Figure 4 Comparative analysis

The bar chart in Figure 4 displays the frequency distribution of various property classification codes, with the Y-axis representing the count of properties per category (up to 7000) and the X-axis listing the specific codes, such as X3, EXMPTCL, and R4. The visualization reveals that the EXMPTCL tax exemption class is the most prevalent category, significantly outpacing other classifications, thereby highlighting the dominant composition of tax-exempt properties within the analyzed portfolio and jurisdiction.

The results in Figures 4 and 5 highlight the significant impact of the data imputation technique in ensuring the integrity and completeness of the dataset, enabling further analysis and insights to be derived from the data, which helps to enhance the reliability and usability of the dataset, allowing for more robust and accurate data-driven decision-making processes.

Conclusion

The ML-based fraud detection system generated promising results in identifying potential outliers or cases of fraud within a given dataset. The key objective of this study was to develop a fraud detection system that uses unsupervised machine learning techniques. The system was implemented using two different algorithms to calculate the fraud scores for each data point. These scores capture the characteristics of data points that may indicate fraud. Then, a comprehensive assessment of the potential outliers was performed. The most relevant data were chosen to detect fraud and analyze the results, and insights were gained into the characteristics and patterns of the data points that resulted in fraud. The results of the fraud detection system demonstrated its efficacy in identifying potential fraud cases. By pinpointing the top outliers, which indicate data points with the highest potential to be outliers, the system facilitates focused investigation and subsequent actions. Analyzing a subset of the dataset consisting of 45 variables provides further insight into the identified outliers, contributing to a more comprehensive understanding of the potential fraud. The system generated an efficiency of 98.7%, indicating that the system was reliable and the results were standardized.

References

Rehman R, Mazhar A, Faheem A, Matloob I, Khan SA, Shaukat S, Khattak MA, Khan AA, Siddiqui MS, Siddiqui S. AI Driven Framework for need based Insurance Plans Generation and Anomaly detection Using Deep Learning techniques. IEEE Access. 2025 Jun 26.

Razzaq K, Shah M. Next-generation machine learning in healthcare fraud detection: Current trends, challenges, and future research directions. Information. 2025 Aug 25;16(9):730.

Rida B, Alm J. Tax Fraud Detection Using Artificial Intelligence-Based Technologies: Trends and Implications. Journal Risk Financial Manag. 2025;18(9):502.

Kashif H, Naseer F. COMPREHENSIVE ANALYSIS OF FRAUD DETECTION PREVENTION SYSTEMS FOR ACCURACY AND EFFICACY. Spectrum of Engineering Sciences. 2025 Mar 25;3(3):382-401.

Mohammed A. Artificial Intelligence-Powered Cyber Attacks: Adversarial Machine Learning. Authorea Preprints. 2025 Feb 3.

ALAM MA, Alam MK, Mahmud MA. Deep Learning for Early Detection of Systemic Risk in Interconnected Financial Markets: A US Regulatory Perspective. Journal of Computer Science and Technology Studies. 2025 Sep 5;7(9):353-75.

Saeedi M, Jafari F, Chang K. Application of Artificial Intelligence as a Metaverse Technology Tool in Finance. InMetaverse Innovation: Technological, Financial, and Legal Perspectives 2025 Oct 29 (pp. 141-163). Cham: Springer Nature Switzerland.

Ndibe OS. Ai-driven forensic systems for real-time anomaly detection and threat mitigation in cybersecurity infrastructures. International Journal of Research Publication and Reviews. 2025;6(5):389-411.

Alzakari SA, Aljebreen M, Ahmad N, Alahmari S, Alrusaini O, Alqazzaz A, Alkhiri H, Said Y. Explainable artificial intelligence-based cyber resilience in internet of things networks using hybrid deep learning with improved chimp optimization algorithm. Scientific Reports. 2025 Sep 26;15(1):33160.

Fathollahi A. Machine Learning and Artificial Intelligence Techniques in Smart Grids Stability Analysis: A Review. Energies. 2025 Jun 30;18(13):3431.

Rodriguez-Casavilca HM, Mauricio D, Villanueva JM. Evolution of Artificial Intelligence-Based OT Cybersecurity Models in Energy Infrastructures: Services, Technical Means, Facilities and Algorithms. Energies. 2025 Jan;18(19):5163.

COVID-19 and commodity effects monitoring using financial & machine learning models. Yasir Shah, Yumin Liu, Faiza Shah, Fadia Shah, Muhammad Islam Satti, Evans Asenso, Mohammad Shabaz, Azeem Irshad. Scientific African. Volume 21, 2023.

Anomaly Detection In Ioht Using Deep Learning: Enhancing Wearable Medical Device Security. Aftab Arif1 , et al. Migration Letters. Volume: 20, No: S12 (2023), pp. 1992-2006. 2023.

Artificial Intelligence as a Service for Immoral Content Detection and Eradication. Shah, et al. Scientific Programming. 2022.

Pappula-Reddy SP, Kumar S, Bi MH, Singh A, Shrivastav AK, Siddique KH, Chellapilla B. Prediction of Traits Using Artificial Intelligence Machine Learning. Custom-Designed Crop Breeding: Advanced Technologies. 2025 Oct 15;1:333-58.

Ahmed A, Shah A, Ahmed T, Yasin S, Longa FE, Hussaini W, Zubair M. AI-Driven Innovations in Modern Banking: From Secure Digital Transactions to Risk Management, Compliance Frameworks, and AI-Based ATM Forecasting Systems. Journal of Management Science Research Review. 2025 Sep 6;4(3):1145-83.

Razzaq K, Shah M. Next-generation machine learning in healthcare fraud detection: Current trends, challenges, and future research directions. Information. 2025 Aug 25;16(9):730.

Kashif H, Naseer F. COMPREHENSIVE ANALYSIS OF FRAUD DETECTION PREVENTION SYSTEMS FOR ACCURACY AND EFFICACY. Spectrum of Engineering Sciences. 2025 Mar 25;3(3):382-401.

Ikumapayi OJ, Ayankoya BB. AI Powered Forensic Accounting: Leveraging Machine Learning for Real-Time Fraud Detection and Prevention.

Rodríguez Valencia L, Ochoa Arellano MJ, Gutiérrez Figueroa SA, Mur Nuño C, Monsalve Piqueras B, Corrales Paredes AD, Bemposta Rosende S, López López JM, Puertas Sanz E, Levi Alfaroviz A. A Systematic Review of Artificial Intelligence Applied to

Compliance: Fraud Detection in Cryptocurrency Transactions. Journal Risk Financial Manag. 2025 Oct 30;18(11):612.

Mohammed A. Artificial Intelligence-Powered Cyber Attacks: Adversarial Machine Learning. Authorea Preprints. 2025 Feb 3.

Merlano C. Enhancing cyber security through artificial intelligence and machine learning: a literature review. Journal of Cybersecurity. 2024;6:89.

Awadallah A, Eledlebi K, Zemerly MJ, Puthal D, Damiani E, Taha K, Kim TY, Yoo PD, Choo KK, Yim MS, Yeun CY. Artificial intelligence-based cybersecurity for the metaverse: Research challenges and opportunities. IEEE Communications Surveys & Tutorials. 2024 Aug 12;27(2):1008-52.

Saifudin S, Januarti I, Purwanto A. The Role of Artificial Intelligence in the Audit Process and How to Fraud Detections: A Literature Outlook. Journal of Ecohumanism. 2025;4(1):4185-203.

Elumilade OO, Ogundeji IA, Ozoemenam G, Omokhoa HE, Omowole BM. Leveraging financial data analytics for business growth, fraud prevention, and risk mitigation in markets. Gulf Journal of Advanced Business Research. 2025;3(3).

Rodriguez-Casavilca HM, Mauricio D, Villanueva JM. Evolution of Artificial Intelligence-Based OT Cybersecurity Models in Energy Infrastructures: Services, Technical Means, Facilities and Algorithms. Energies. 2025 Jan;18(19):5163.

Gazali MK, Hasikin K, Lai KW, Zamzam AH, Damseh R. State-of-the-art artificial intelligence approaches for anomaly detection and remaining useful life prediction: a review. PeerJ Computer Science. 2025 Sep 26;11:e3056.

Zhang T, Zhang CC, Shi B, Chen Z, Zhao X, Wang Z. Artificial intelligence-based distributed acoustic sensing enables automated identification of wire breaks in prestressed concrete cylinder pipe. Journal of Applied Geophysics. 2024 May 1;224:105378.

Naseer K, Ahmed HN. Effectiveness and reliability of artificial intelligence in fraud detection: A mixed-method study on financial audit. Journal of Management and Informatics. 2025 Apr 26;4(1):706-22.

Kamruzzaman M, Bhuyan MK, Hasan R, Farabi SF, Nilima SI, Hossain MA. Exploring the Landscape: A Systematic Review of Artificial Intelligence Techniques in Cybersecurity. In2024 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI) 2024 Oct 16 (pp. 01-06). IEEE.

Khan W, Ishrat M, Faisal SM. AI-Powered Risk Assessment: Innovations, Challenges, and Future Prospects. Artificial Intelligence for Financial Risk Management and Analysis. 2025:59-92.