Engineering and Technology Journal e-ISSN: 2456-3358

Volume 10 Issue 02 February-2025, Page No.- 3735-3739

DOI: 10.47191/etj/v10i02.01, I.F. – 8.482

© 2025, ETJ

Fraud Detection in Health Insurance Claims Based on Artificial Intelligence (AI)

Aditya Kurniawan[1]*, Sri Widodo[2], Joni Maulindar[3]

[1,2]Faculty of Health Science, Universitas Duta Bangsa Jl. K.H. Samanhudi No. 93, Sondakan Surakarta, Central Java, Indonesia 57147

[3]Faculty of Computer Science, Universitas Duta Bangsa, Jl. Bhayangkara No.55, Tipes, Serengan, Surakarta, Central Java, Indonesia 57154

ABSTRACT: Health insurance claim fraud has become a major issue that impacts healthcare delivery worldwide. The majority of health insurance claim fraud is perpetrated by insurance companies, clients, and service providers. Service providers, insurance clients, and insurance companies are all involved in health insurance claim fraud. There is an urgent need to design a decision support system (DSS) that can automatically detect fraud and handle claims accurately. The goal of this project is to develop an artificial intelligence (AI)-based machine learning model that can identify fraud in health insurance claims. Deep learning is the technique employed. Eighty-six percent accuracy was achieved.

KEYWORDS: AI, DSS, Deep learning, Fraud.

I. INTRODUCTION

An extensive invoice detailing the specific services that the patient or patients got at the time of healthcare delivery is called a claim, and it is submitted to a health insurance company by the service provider [1]. Three criteria are used to classify healthcare fraud, specifically. (1) fraud by providers (physicians and hospitals), (2) fraud by beneficiaries (patients), and (3) fraud by insurance companies [2]. This issue in health insurance programs worldwide is caused by a variety of fraud techniques. These include, among other things, bribery, impersonation, ganging, illegal cash exchange for prescriptions, careless use of medication services, insurance carrier fraud, reimbursement fraud, recoding of services, insurance customer fraud, upcoding of goods and services, duplicate billing, unbundling/creative billing of claims, medically unnecessary services (bill padding), excessive services (bill padding), and billing for services never rendered (identity theft and ghost billing) [3] [4] [5] [6]. Thus, an intelligent fraud detection system is required.

Data mining techniques have been applied extensively in fraud detection in recent years to minimize errors brought on by expert judgment, particularly in financial fraud detection [7]. These scams are concealed in vast volumes of data, and professional analysis occasionally falls short of accounting for everything. This issue is resolved by using data mining techniques. Support Vector Machine is one of the data mining techniques utilized [8]. The current autoencoder (AE) clustering, local outlier factor (LOF), isolation forest (IF), and K-means methods are outperformed by the use of the NN technique. While the current techniques, AE, IF, LOF, and K-Means, offer 97%, 98%, and 99.75% accuracy, respectively, the suggested NN-based fraud detection system achieves 99.87% accuracy [9]. Undersampling, oversampling, and SMOTE techniques can be used to balance unbalanced fraud detection datasets. With an area under the curve (AUC) of 91.37% produced by the oversampling technique, machine classifier results are superior. An accuracy of 99.96% is achieved while using ANN with backpropagation on real value transactions, which are used as training and testing data points. Because this algorithm can identify transactions in real time, banks are able to identify fraudulent activity and halt ongoing transactions [10]. On tiny data sets, Random Forest produced good results; nonetheless, some issues, such as imbalanced data, persist. The voting process, for instance, presumes that each base classifier has an equal weight, even though some may be more significant than others [11]. XGBoost, Random Forest, and Decision Tree algorithms were used on a dataset of 284,808 credit card records. With the maximum accuracy of 99.962%, the XGBoost algorithm demonstrated the best performance. The decision tree's performance was not very good. With an accuracy of 99.923% and a random forest algorithm performance of 99.957%, decision trees had very poor performance. When it comes to detecting fraud, ANNs are more accurate than support vector machines and logistic regression [12].

The goal of this project is to create an artificial intelligence (AI)-based program that can identify health insurance claim fraud. Machine learning, particularly machine learning using ensemble and deep learning, is used to detect fraud in health insurance claims. To create a model with the highest accuracy and precision ratings, both forms of machine learning will be configured and modified. The paper's notable contributions are as follows: evaluating the health insurance claims fraud detection system using data from National Health Insurance clients; designing, developing, and implementing a decision support system (DSS) that combines knowledge representation, business intelligence, and false claims detection models for processing claims; creating an intelligent fraud detection system's user-friendly graphical user interface (GUI); and analyzing data mining and machine learning techniques that use deep learning to detect fraud.

II. METHOD

There is currently a dearth of research on expert systems-based automatic health insurance claim fraud detection. In order to compare the accuracy results of a study, it is required to obtain the findings of earlier investigations that are relevant to this one. Since fraud is commonly confused with abuse and waste, studying the field of health insurance claim fraud necessitates a distinct definition of fraud. Fraud and abuse, on the other hand, are instances in which third-party insurance firms are reimbursed or healthcare services are paid for but not rendered. Further examples of fraud and abuse include when medical professionals take bribes, when patients seek treatment that could endanger them (e.g., using medications to feed an addiction), and when they prescribe services that are known to be unneeded [13][14]. The intentional act of lying, hiding, or distorting facts that leads to the payment of medical benefits to a person or group is known as health insurance fraud.

Account audits and investigative investigations are used to discover health insurance fraud. A thorough account audit can reveal questionable policyholders and providers. It is ideal to audit each claim separately. However, there is no realistic way to audit every claim. Furthermore, it is challenging to audit providers in the absence of any concrete smoking-related indicators. Creating a select list and reviewing and auditing the patients and providers on it would be a sensible strategy. The audit shortlist can be created using a variety of analytical methods.

The most widely used fraud detection approaches, according to the literature, include statistical techniques, data mining, machine learning, and artificial intelligence. When compared to other classifiers, the most economical model, which employed the Naïve-Bayes algorithm, failed to produce a clear picture of the decision. It created a subsample of 20 claims with 400 objects, of which 50% were classified as fraudulent and the remaining 50% as legal. The decision image in contrast to alternative classifiers [15]. In the battle against fraud, a new field of study has emerged: the integration of multiple traditional methodologies. This method can be either supervised, unsupervised, or both, depending on which one is used for classification. Either approach can be applied as a pre-processing phase to alter the data before categorization [16], or, to a lesser degree, the algorithm's component steps can be merged to produce something fundamentally unique. Solutions for certain issue domains can be customized through the use of hybrid approaches. It is possible to target many aspects of performance, such as computing efficiency, ease of use, and classification skill.

Neural networks and fuzzy logic are used to automatically evaluate and categorize medical claims. The healthcare industry uses the idea of data warehousing for data mining to create an electronic fraud detection tool that evaluates service providers based on behavioral heuristics and contrasts them with other service providers of a similar nature. To find uncommon occurrences in pathology insurance data, the Australian Health Insurance Commission has researched online discount learning algorithms [17].

To detect false claims, researchers in Taiwan created a process mining-based detection algorithm that methodically finds procedures based on clinical pathways. Despite the discovery of several anomalies and irregularities, the capacity to detect suspicious claims was significantly limited to the detection of health insurance claims fraud in relation to payer-fixed price providers. For a private health insurance program in Chile, an application to identify medical fraud and abuse was created using neural networks [17]. The novel aspect of this approach is explained by its real-time claim processing capability. In order to identify questionable claims and maybe dishonest people, association rule mining was used to analyze billing trends among particular specialist groups.

It is evident from the foregoing description that the current state of study is an analysis of the incidence of fraud based on the research that has been conducted. Research on the development of intelligent systems to detect fraud intelligently is still rarely done. Expert systems used to detect fraud use conventional methods, K-Means, Naïve-Bayes, and C4.5. The weaknesses of the K-Means, Naïve-Bayes and C4.5 methods are that modelling the uncertainty of the calculation process is still debated; for data with more than 2 pieces, several data processing must be done. While the weakness of the Support Vector Machine method is that it can decrease the accuracy obtained. This decrease is because the data used is a lot of error value and does not vary. In addition, the data used does not go through a preprocessing process first. The research conducted is to develop an application to detect fraud in artificial intelligence-based health insurance claims. The method used is deep learning. Additionally, experimental results have demonstrated that the deep learning approach can achieve rapid and stable network convergence. The stages in developing smart applications for fraud detection in insurance claims based on artificial intelligence can be explained in Figure 1.

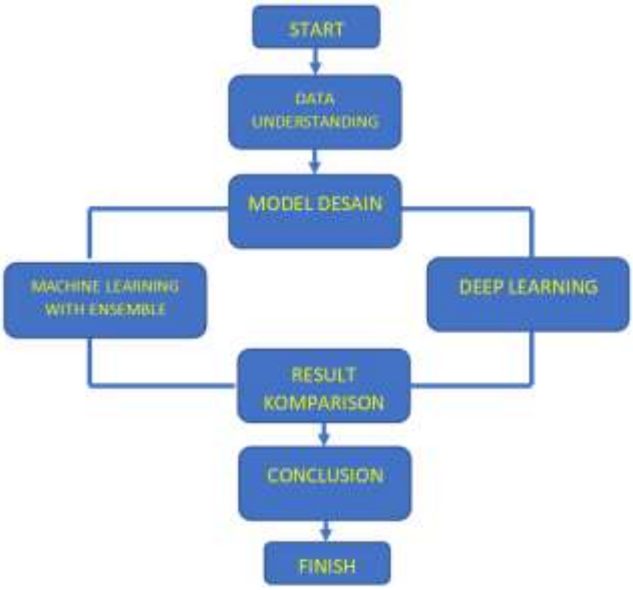


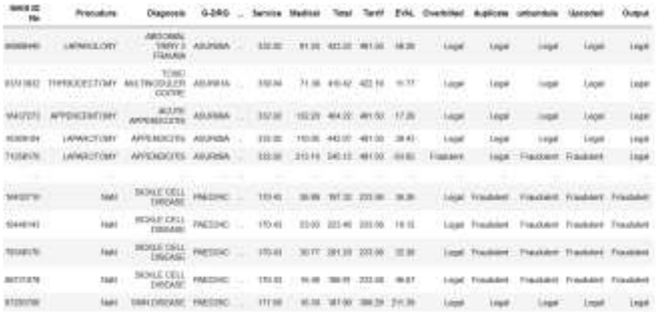
Figure 1. Research method

1. Data Understanding
   Many financial transactions that contain fraud are analyzed to detect the fraud. To build this machine learning model using Kaggle data.
2. Model Design
   Deep learning, a type of machine learning, is used to detect fraud in electronic financial transactions. Deep learning is one of the various machine learning techniques that use artificial neural networks (ANN). In deep learning, there are three different learning styles: supervised, semi-supervised, and unsupervised. Supervised data is included in the processed data.
   a. Data Source. The BankSim dataset, a bank payment simulation based on the Multi Agent-Based Simulation (MABS) idea, was employed in this investigation.
   b. Data Preparation. The first data preparation is to convert character data into numeric, and the second is to remove the transaction origin code with the aim that the model can be implemented in other countries or cities.
   c. Smote. SMOTE will create new data points from the minority class using other instances so that the resulting sample is not an exact copy but similar to the owned instances. The smotes to be used are random oversampling (ROS) and random undersampling (RUS) and one synthesis algorithm, FSMOTE.
   d. Divide the testing and training data. Training data is used to train the algorithm, whereas testing data is utilized to evaluate how well the previously taught algorithm performs while uncovering previously unknown data. Twenty percent is the testing data, and eighty percent is the training data.
   e. Classification Model. The methods that will be used to identify the best model are Extra Trees Classification, Random Forest Classification, AdaBoost Classification, Stochastic Gradient Boosting, and Bagged Decision Trees. The model will also be simulated with several numbers of trees and random states.
   f. Selection of the best accuracy. From the experiments conducted, the model with the best accuracy will be analysed and selected, which will then be used to run the testing data.
   g. Run the model on the testing data.
   The model with the best accuracy is run on the testing data. The results obtained will be analyzed and then implemented.

III. RESULTS

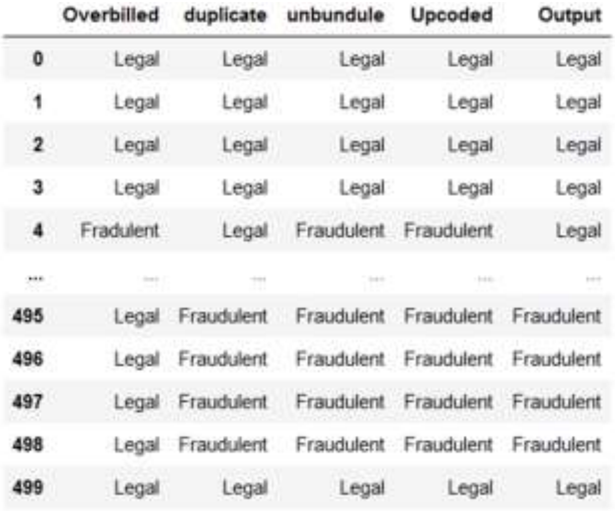
The implementation of fraud detection using the deep learning method is implemented using the Python 3 language with Jupyter Notebook tools and assistance from several open-source libraries that help in the formation of source code. First, import data that will be used for training classification models along with modules that will be used to help the classification process. The data used for training is shown in Tables 1 and 2.

Table 1. Insurance Claim Data



[illegible]	Procedure	Diagnosis	G-DRG	...	Service	Medical	Total	Tarif	EVAL	Overbilled	duplicate	unbundule	Upcoded	Output
[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Legal	Legal	Legal	Legal
[illegible]	THYROIDECTOMY	TOXIC MULTINODULER GOITRE	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Legal	Legal	Legal	Legal
[illegible]	APPENDICECTOMY	ACUTE APPENDICITIS	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Legal	Legal	Legal	Legal
[illegible]	LAPAROTOMY	APPENDICITIS	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Legal	Legal	Legal	Legal
[illegible]	LAPAROTOMY	APPENDICITIS	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Fradulent	Legal	Fraudulent	Fraudulent	Legal
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
[illegible]	NaN	SICKLE CELL DISEASE	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
[illegible]	NaN	SICKLE CELL DISEASE	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
[illegible]	NaN	SICKLE CELL DISEASE	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
[illegible]	NaN	SICKLE CELL DISEASE	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
[illegible]	NaN	[illegible] DISEASE	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Legal	Legal	Legal	Legal	Legal

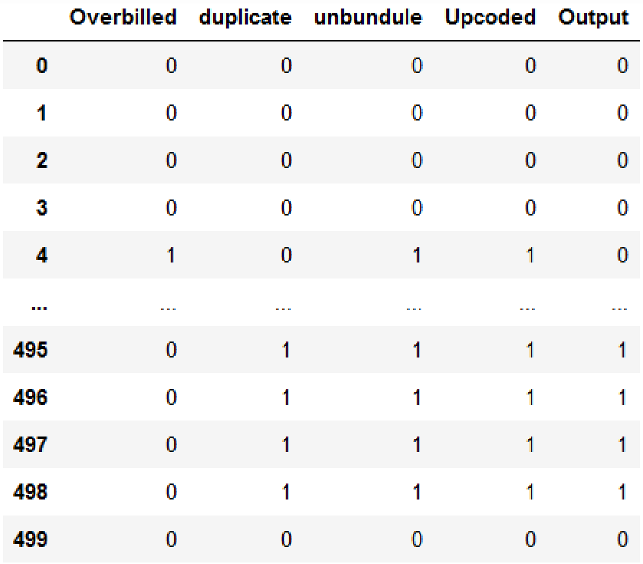
Table 2. Training Used for Classification Process



	Overbilled	duplicate	unbundule	Upcoded	Output
0	Legal	Legal	Legal	Legal	Legal
1	Legal	Legal	Legal	Legal	Legal
2	Legal	Legal	Legal	Legal	Legal
3	Legal	Legal	Legal	Legal	Legal
4	Fradulent	Legal	Fraudulent	Fraudulent	Legal
...	...	...	...	...	...
495	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
496	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
497	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
498	Legal	Fraudulent	Fraudulent	Fraudulent	Fraudulent
499	Legal	Legal	Legal	Legal	Legal

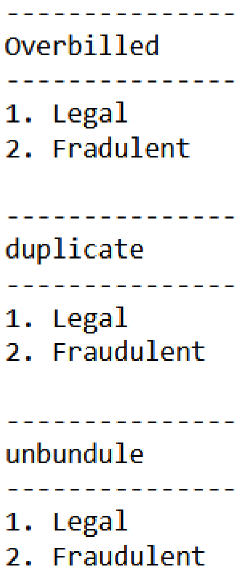
Next, the string value is converted into a discrete value. Conversion is done to change the attributes in a class to be represented by numbers; for example, in the age class, there are 3 attributes, namely age 20-40, 40-50, and 50-60 years. These attributes are converted into the number 0, representing the age of 20-40, 1 representing 40-50, and number 2 representing 50-60 years. This also applies to other classes. The results of the string-to-discrete conversion are shown in Table 3.

Table 3. Results of String to Discrete Data Conversion



	Overbilled	duplicate	unbundule	Upcoded	Output
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	1	0	1	1	0
...	...	...	...	...	...
495	0	1	1	1	1
496	0	1	1	1	1
497	0	1	1	1	1
498	0	1	1	1	1
499	0	0	0	0	0

After that, handling of missing values and outliers will interfere with the classification process. Previously, we will check for attribute deviations (outliers) in each symptom class according to the dataset column described in the previous section. The image of the attribute division is shown in Figure 2.



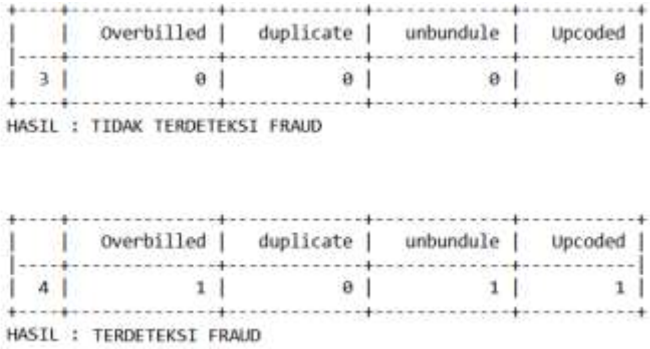
```
---------------
Overbilled
---------------
1. Legal
2. Fradulent

---------------
duplicate
---------------
1. Legal
2. Fraudulent

---------------
unbundule
---------------
1. Legal
2. Fraudulent
```

Figure 2. Class Attributes

After the data is ready for classification, the model for classification will be tested using a dataset that is divided into a training set and a testing set. Testing is done by repeating the model test by choosing different testing set sizes to get the best accuracy level, which will then be used as a prediction model for the input data. The results of repeated testing using a test sample size of 10% to 90% are shown in Figure 3, while the test results per input data are shown in Figure 4.



```
+----+------------+-----------+-----------+----------+
|    | Overbilled | duplicate | unbundule |  Upcoded |
|----+------------+-----------+-----------+----------|
|  3 |          0 |         0 |         0 |        0 |
+----+------------+-----------+-----------+----------+
HASIL : TIDAK TERDETEKSI FRAUD
```

```
+----+------------+-----------+-----------+----------+
|    | Overbilled | duplicate | unbundule |  Upcoded |
|----+------------+-----------+-----------+----------|
|  4 |          1 |         0 |         1 |        1 |
+----+------------+-----------+-----------+----------+
HASIL : TERDETEKSI FRAUD
```

Figure 3. Fraug Detection Results

Figure 3, shows the graph of the relationship between the number of testings sets and the accuracy rate. It can be seen that the highest accuracy rate is obtained using a testing set size of 20% and 80% training data.

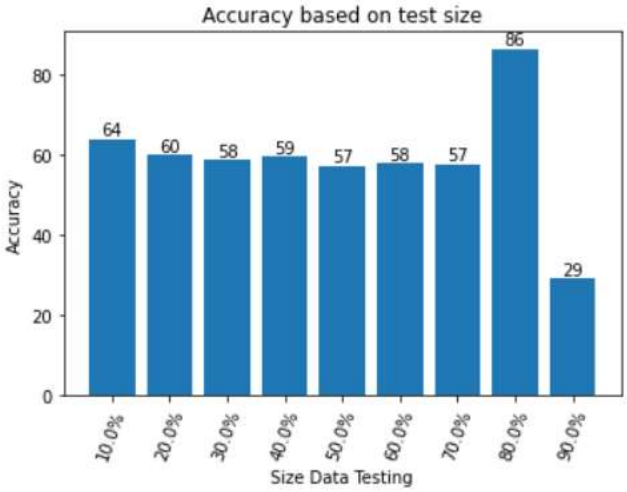


Figure 4. Graph of Testing Results Based on the amount of Training Data

CONCLUSION

From the tests conducted, it can be concluded that the machine learning method using deep learning is very suitable for detecting fraud in insurance claims. This can be seen from the accuracy value that reaches 86% using 80% training data and 20% testing data.

ACKNOWLEDGMENT

We would like to thank Universitas Duta Bangsa Surakarta for funding our research through an Applied Research Grant in 2024.

REFERENCES

1. R. A. Sowah *et al.*, "Decision Support System (DSS) for Fraud Detection in Health Insurance Claims Using Genetic Support Vector Machines (GSVMs)," *J. Eng. (United Kingdom)*, vol. 2019, no. January 2007, 2019, doi: 10.1155/2019/1432597.
2. I. Technology, *Proceedings of the Eighth Australasian Data Mining Conference (AusDM'09)*, vol. 101. 2009.
3. P. V. Mohan, S. Dixit, A. Gyaneshwar, U. Chadha, K.

Srinivasan, and J. T. Seo, “Leveraging Computational Intelligence Techniques for Defensive Deception: A Review, Recent Advances, Open Problems and Future Directions,” *Sensors*, vol. 22, no. 6, 2022, doi: 10.3390/s22062194.

4. E. Kirkos, C. Spathis, and Y. Manolopoulos, “Data Mining techniques for the detection of fraudulent financial statements,” *Expert Syst. Appl.*, vol. 32, no. 4, pp. 995–1003, 2007, doi: 10.1016/j.eswa.2006.02.016.
5. H. Joudaki *et al.*, “Using data mining to detect health care fraud and abuse: a review of literature,” *Glob. J. Health Sci.*, vol. 7, no. 1, pp. 194–202, 2015, doi: 10.5539/gjhs.v7n1p194.
6. H. Joudaki *et al.*, “Improving fraud and abuse detection in general physician claims: A data mining study,” *Int. J. Heal. Policy Manag.*, vol. 5, no. 3, pp. 165–172, 2016, doi: 10.15171/ijhpm.2015.196.
7. C. W. Lu Wang, “Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map, Knowledge-Based Systems, Volume 121, 2017, Pages 99-110, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2017.0,” *sciencedirect*, vol. 121, pp. 99–110, 2017, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095070511730028X
8. J. Zhang, J. Yao, L. Wang, Y. Chen, and Y. Pan, “A financial fraud detection model based on organizational impression management strategy,” *J. Phys. Conf. Ser.*, vol. 1616, no. 1, 2020, doi: 10.1088/1742-6596/1616/1/012093.
9. F. Zamachsari and N. Puspitasari, “Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 203–212, 2021, doi: 10.29207/resti.v5i2.2952.
10. K. S. M. and A. A. K. C. Dubey, “Credit Card Fraud Detection using Artificial Neural Network and BackPropagation.” Madurai, India, pp. 268–273, 2020. doi: doi: 10.1109/ICICCS48265.2020.9120957.
11. S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, “Random forest for credit card fraud detection,” *ICNSC 2018 - 15th IEEE Int. Conf. Networking, Sens. Control*, pp. 1–6, 2018, doi: 10.1109/ICNSC.2018.8361343.
12. X. Yu, X. Li, Y. Dong, and R. Zheng, “A Deep Neural Network Algorithm for Detecting Credit Card Fraud,” in *Proceedings - 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2020*, China, 2020, pp. 181–183. doi: 10.1109/ICBAIE49996.2020.00045.
13. V. R. and G. Anuradha, “Fraud detection in health insurance using data mining techniques.” Mumbai, India, pp. 1–5, 2015. doi: 10.1109/ICCICT.2015.7045689.
14. A. C. Nugraha and M. I. Irawan, “Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost),” *J. Sains dan Seni ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.107032.
15. R. A. D. and G. D. S. Viaene, “A case study of applying boosting naive Bayes to claim fraud diagnosis.” pp. 612–620, 2004. doi: doi: 10.1109/TKDE.2004.1277822.
16. S. Vadlakonda and S. Siddam, “ML REGRESSION-BASED PREDICTIVE MODELING FOR DISEASE OUTBREAK THRESHOLD ESTIMATION,” vol. 15, pp. 435–443, 2024, doi: 10.47750/jett.2024.15.05.43.
17. P. Roddick, John & Li, Jiuyong & Christen, Peter & Kennedy, “Data Mining and Analytics 2008.” 2008.