IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11610

Fraud Detection in insurance claims using deep learning

M. Raghavendra[1], T. Gopi Chandu[2], P. Sai jayanth[3], N. Surya Tej[4], Dr. M Shivram[5]

Department of computer science and engineering, Jain university, Bengaluru, India

Abstract: As the use of the internet is growing exponentially, more and more businesses such as the financial sector are initiating their services online. Accordingly, financial fraud is increasing in number and forms around the world, which results in financial losses which make financial fraud a major problem. Unauthorized access and immense regular attacks are examples of threats that should be detected by means of financial fraud detection systems. Machine learning and data mining techniques have been extensively used to tackle this problem over the past few years. However, these methods still need to be improved in terms of fast computation, dealing with huge data, and identifying the unknown attack patterns. Therefore, in this paper, deep learning-based method is implemented for the detection of financial fraudulence based on the Long Short-Term Memory (LSTM) technique and Bidirectional encoder representation from transformers (BERT) . This model is aimed at enhancing the present detection techniques as well as enhancing the detection accuracy in the light of big data. To evaluate the proposed model, a real dataset of credit card frauds is utilized and the results are compared with an existing deep learning model named spiking neural network and some other machine learning techniques. The experimental results illustrated a perfect performance of LSTM where it achieved 99.95% of accuracy.

Keywords: Fraud ; fraud detection; deep learning; long short-term memory

INTRODUCTION

The insurance industry is currently adopting effective fraud management. Some people fool the company to get compensation, while others pay a premium.

There are two main categories of scams: hard insurance scams and soft insurance scams. Insurance fraud occurs when people deliberately forge an accident. If a person has a valid insurance claim but has tampered with part of the claim. This is known as soft insurance fraud. Customer satisfaction is improved if the company has a good fraud detection and prevention management system in place. As your satisfaction increases, your billing and settlement costs will be reduced. There are many ways to find a scam.

The most commonly used method is to analyze the data using custom statements. Therefore, complex and time-consuming research is required and various areas of knowledge are involved.

PROBLEM DEFINITION:

In Insurance industries, some of the insurers have been claiming the insurance money by falsifying or altering the parameters of the insurance industry.

For example, In a case of damaged car, statement by the customer is his car is damaged by unknown person but the fact is it is damaged by himself to claim the insurance money or a new one. So likewise few of insurers has been falsifying the insurance parametres and making the big losses to insurance industries.

So here the Problem Definition is about the losses occurred by few individuals who are falsifying the insurance parametres and claiming the insurance money which are affecting the insurance industries.

Objective:

The contribution of this paper is to propose a fraud detection model based on deep learning-based technique (Long Short – Term Memory).

This model is aimed to identify the suspicious financial transaction and alert the relevant authorities about it, to take appropriate action.

As a result, the proposed model may constitute a useful tool for the financial sectors to reduce their potential losses.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11610

RELATED WORK:

TITLE	YEAR OF PUBLICATION	AUTHORS	ALGORITHMS
Deep Learning Detecting Fraud in Credit Card Transactions	2018	Abhimanyu Roy, Jingyi Sun, Robert Mahoney, Loreto Alonzi, Stephen	Ann,Rnn,Lstm
Detecting Credit Card Fraud by ANN and Logistic Regression	2011	Y. Sahin ,E. Duman	ANN and logistic regression
Survey on Anomaly Detection using Data Mining Techniques	2015	Shikha Agrawal, Jitendra Agrawal	Hybrid methods
Study of Hidden Markov Model in Credit Card Fraudulent Detection	2016	Bhusari & Patil	Hidden markov model
Performance of machine learning techniques in the detection of financial frauds	2019	SADI GALI, I.; SAEL, N.; BENABBOU, F.	Machine learning and deep learning techniques
Survey of fraud spotting techniques	2019	Suresh kumar	Problems of various survey techniques
Insurance Fraud Detection using Spiking Neural Network along with NormAD Algorithm .	2021	~Gopikrishna Panda1, Sunil Kumar Dhal1 RabinarayanSatpathy1 Subhendu Kumar Pani2	A spiking neural network is used, and 73% accuracy is obtained.
Automobile Insurance Fraud Detection using supervised classifiers	2020	Arian Dhani	Decision Tree Random Forest Multilayer perceptron
Insurance claim analysis using machine learning algorithms	2019	Rama devu burri Ramesh Reddy Srinivasa Rao	Logistic Model Tree and Random Forest

DOI: 10.17148/IJARCCE.2022.11610

Navie Classification approach for insurance fraud detection	2019	Bhavana bhatra Sheetal Kundra	Naive Bias classification
Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud	2017	Yibo wang Wei Xu	Deep Neural Networks

EXISTING SYSTEM:

In the Research of (Roy, 2018), the productiveness of deep learning methods was evaluated on a dataset of approximately 80 million transactions of credit cards that have been pre-identified as lawful and fraudulent. This study compared the performance of some deep learning algorithms in terms of class imbalance, sensitive analysis of the parameters, and scalability. Many techniques such as Gated Recurrent Unit, Long Short-term Memory, Recurrent Neural Networks and Artificial Neural Networks were used in distributed cloud computing environments. The results showed that the Long Short-term Memory technique achieved the best performance. Also, it can deal with huge complex data.

Another study in (Sahin & Duman, 2011) compared the performance of Artificial Neural Network and Logistic Regression in credit card fraud detection based on a real dataset. The empirical results of the paper accompany an equal performance of these models on training data. Also, it shows the use of Artificial Neural Network on Logistic Regression over the test data. Moreover, Artificial Neural network data needs training to expect output. Thus, the Artificial Neural network uses for classification tasks not for detection fraud or atypical behavior detection tasks

The authors in (Agrawal , 2015) presented different methods for anomaly detection such as Clustering-based Anomaly Detection techniques, Classification based anomaly detection, and their approaches. These approaches gained better results and overcame the disadvantages of different approaches. Moreover, these paths take more time in training, huge cost for computational resources, and have complicated architecture.

Authors in (Bhusari & Patil, 2016) considered the Hidden Markov Model that may help to detect the dis-honest, which carry diverse ranges of the transaction such as low, medium and high as the inspection symbols. This model made the fraud detection systems very simple, not taking a long time although having complex processes. Usually Hidden Markov Model needs training by use of annotated data. Moreover, this model needs manual markup. Therefore, this model is not suitable to detect unfamiliar patterns and unproductive to identify new traits of fraud.

Authors in (Sadi Gali et al., 2019) studied state-of-the-art techniques that can detect different fraudulence. The techniques such as classification, clustering, and regression were examined to identify the contribution of each technique and its productiveness. According to authors, machine-learning techniques have a vital role in fraud detection and being applied to extract and uncover the hidden data. Additionally, the hybrid fraud detection techniques were the most applicable as a result of integrating the strengths of several traditional techniques of detection. However, most hybrid techniques did not work in real time
Authors in (Sureshkumar, 2019) discussed the problems of current techniques used in fraud detection in many areas such as credit card and invasion of detection. The result showed the productiveness of these techniques in some kinds of fraud. Moreover, there are still problems like fraud detection for credit cards. Not many approaches are available in public. Also, Invasion of detection is difficult to test and has less portability because the rules and systems must be specified to the environment being watched. Besides, such systems needs updating to keep them up-to-date with old methods of fraud

Insurance Fraud Detection using Spiking Neural Network along with (Norm AD) Algorithm
~Gopikrishna Panda1, Sunil Kumar Dhal1 RabinarayanSatpathy1 (Subh e and u) Kumar Pani2 In the paper "Paper 4" by Gopi Krishna et al., A spiking neural network is used, and 73% accuracy is resulted.

DOI: 10.17148/IJARCCE.2022.11610

Automobile Insurance Fraud detection using supervised classifiers by(Arian Dhani..) proposed as supervised classifiers method using MLP, DT C4.5, and RF to classify the fraudulent and legitimate claims. Smote is implied in the training data to create the model with reasonable accuracy. The model is validated with testing data, representing real-world data, which is a high imbalanced dataset. The result shows that MLP, DT C4.5, and RF produces a high accuracy. However, RF has the highest performance with 98.5% accuracy, 100% sensitivity, and 98.5% specificity of the model.

The naïve bayes classification method is used in this work for detection. The proposed model is implemented in python language and output are tested in terms of accuracy, execution time. This model gives approximately 5 percent high results as compared to voting classifiers.

PROPOSED SYSTEM:

The section introduces the proposed financial fraud detection model and describes its stages. The model is divided into three stages. They are

1. Pre-applying the model
2. Applying the model
3. Post applying the model

During the pre applying model we collect the data and we clean during this stage and data validation, data normalization. During the applying stage we apply the deep learning algorithms like Long short term memory and Bidirectional encoders representation from transformers. And finally this post applying stage we experiment the results we got from applying the model stage, and lastly we experiment the results.

RESULTS ANALYSIS:

In this section the results of applying the proposed model are explored. As an initial result the model obtained 91% accuracy and 0.59% of loss rate with 15 epoches of 13 each approximately 200 times.

CONCLUSION AND FUTURE SCOPE:

Fraudulence is a problem with far-reaching implications for the financial sector and stakeholders. Compound reliance on come out technology has compounded the issue in recent years. Traditional approaches are ineffective in the huge data age. Therefore, the work developed a model for the identification of fraudulence based on the Long Short-Term Memory (LSTM) and Bidirectional encoders from transformers technique using real data . This training was aimed to better the current detection techniques as well as enhancing the detection accuracy in the light of huge volume of data. It addressed the problem of detection of unknown and deep patterns of fraud by using deep learning techniques to identify patterns quickly and with high accuracy.

Also, the problem of the lack of the existing techniques was addressed by using the proposed model based on the Long short term memory and BERTs technique. Finally, a comparison with other existing machine learning techniques showed that the LSTM technique can achieve perfect performance in addressing fraud detection problems.

As future work, an algorithm can be developed to perform various tasks like, calculate the timing of the fraudulence that occurred in addition to the place of the fraudulence**.**

REFERENCES:

1. Roy, A. (2018). Deep learning detecting fraud in credit card transactions. Stems and Information Engineering Design Symposium (SIEDS), 2018 Sy, IEEE. https://doi.org/10. 1109/SIEDS.2018.8374722
2. Sahin, Y., & Duman, E. (2011). Detecting credit card fraud by ANN and logistic regression. 2011 International Symposium on Innovations in Intelligent Systems and Applications, IEEE.
3. http://libgen.rs/scimag/10.1109%2FSTARTUP.2016.7583942
4. Sureshkumar, B. (2019). Survey of fraud spotting techniques. Journal of the Gujarat Research Society, 21(17), 89–94.
5. Yibo Wang, Wei Xu, “Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud”, 2017, DECS UP 12895
6. K. J. Crocker and S. Tennyson, "Insurance Fraud and Optimal Claims Settlement Strategies: An Empirical Investigation of Liability Insurance Settlements", The Journal of Law and Economics, vol. 45, no. 2, april 2010.
7. E. Belhadji, G. Dionne and F. Tarkhani, A Model for the Detection of Insurance Fraud Geneva Papers on Risk and Insurance Theory, vol. 25, pp. 517-538, may 2012.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 6, June 2022

DOI: 10.17148/IJARCCE.2022.11610

8. Kajia muller, "The Identification of Insurance Fraud – an Empirical Analysis Working papers on Risk Management and Insurance" no: 137, June 2013.
9. Clifton Phuna damminda, Alahakoon, and Vincent phua" Minority Report in Fraud Detection: Classification of Skewed Data". Sigkdd Explorations, Volume – 6, Issue – 1,sep 2011
10. Pérez, J. M, Muguerza J, Arbelaitz, O., Gurrutxaga, I., & Martín, J. I., 2005, "Consolidated Tree Classifier Learning in a Car Insurance Fraud Detection Domain with Class Imbalance",Pattern Recognition and Data Mining. Springer-Verlag. S. Singh et al. (Eds.), 381-389
11. Belhadji, E., G. Dionne, and F. Tarkhani, ―A Model for theDetection of Insurance Fraud, Geneva Papers on Risk andInsurance Theory‖, 25: 517-538, may 2012.
12. Crocker, K. J., and S. Tennyson,‖ Insurance Fraud andOptimal Claims Settlement Strategies: An EmpiricalInvestigation of Liability Insurance Settlements‖ The Journalof Law and Economics, 45(2), April 2010.
13. Kajian Muller, ―The Identification of Insurance Fraud – anEmpirical Analysis Working papers on Risk Management andInsurance‖ no: 137, June 2013
14. Leila Goleiji, M. J Tarokh, "Identification of Influential Features and Fraud Detection in the Insurance Industry using the Data Mining Techniques," Majlesi Journal of Multimedia Processing, Vol. 4, No. 3, September 2015.
15. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113.
16. Behdad, Mohammad, et al. "Nature-inspired techniques in the context of fraud detection." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 (2012): 1273-1290.
17. Konasani, Venkatareddy, Mukul Biswas, and Praveen Krishnan Koleth. "Healthcare fraud management using big data analytics." An Unpublished Report by Trendwise Analytics, Bangalore, India (2012).
18. National Health Care Anti-Fraud Association. "Health Care Fraud–A Serious and Costly Reality For All Americans." April2005 (2007