Financial Statement Fraud Detection Using BBN-Goa

S.Sudha[1], Dr.P.Logeswari[2]

[1] Research Scholar, Department of Computer Science, Sri Krishna Arts & Science College, Coimbatore, Tamilnadu, India.

[2] Research Supervisor, Associate Professor, Department of Computer Science and Information Technology, Jain (Deemed-to-be University), Bengaluru, Karnataka.

Received: 15.07.2024 Revised: 16.08.2024 Accepted: 04.09.2024

ABSTRACT

Financial statement misstatement is characterized as "intentional blackmail by a board of directors to harm a financier or lender by deceiving the financial statements."Even more true is the fact that the amount of fake financial reports filed by certain organizations has increased unimaginably over a decade. To prevent the overwhelming consequences of financial extortion, a strong and reliable strategy for identifying misstatements in financial reports is essential. In this paper, we proposed a hybrid Bayesian Belief structure Genetic Optimization Algorithm (BBN-GOA) system. This algorithm is widely used to extract and reveal secret insights behind huge amounts of data and plays a key role in extortion detection. The proposed strategy provides better results compared to standard classifiers such as SVM and ANN. Presentation measures are evaluated against various classifiers. The accuracy achieved with this strategy is better than other standard classifiers.

Keywords: Financial statement fraud, Bayesian belief network, Genetic Algorithm and classifiers;

1. INTRODUCTION

As per the Relationship of Confirmed Extortion Analysts (ACFE's) ACFE Misrepresentation Inspectors Affiliation Testament, "misrepresentation" is the unlawful hardship of one more's property or cash through duplicity, misdirection, or other out of line implies. Alludes to a deliberate or purposeful demonstration with the expectation of These are designated acts intended to hoodwink and control, bringing about monetary misfortune to the person in question. This broad definition emphasizes the complexity of cheating and emphasizes the intentional nature and use of cheating. It is critical for fraud examiners to understand fraud so that they can identify and address various forms of intentional financial fraud.

1.1 Sorts of financial fraud

There are many kinds of monetary extortion, and exploration has framed a few significant classes. Insurance fraud is prevalent throughout insurance and manifests itself at various stages including application, eligibility, evaluation, claims, and claims. People associated with protection extortion can incorporate purchasers, specialists, intermediaries, insurance agency workers, medical services suppliers, and others inside the protection biological system. Fraudal so includes fraud that undermines the integrity of the insurance system, highlighting the need for industry-wide vigilance and precautions to curb fraud.

Protections and items misrepresentation

Protections and Wares Misrepresentation: The FBI gives a short outline of maybe the most far and wide protections and ware misrepresentation cases in presence today. Models incorporate "market control, high return venture misrepresentation, fraudulent business models, fraudulent business models, prime bank plans, advance charge extortion, and flexible investments." As per another CULS definition, protections extortion incorporates burglary by market control, robbery of protections records, and wire misrepresentation.

Tax evasion

Tax evasion is a cycle wherein culprits stow away or camouflage the returns of deceitful movement, or convert those returns in the process of childbirth or items. This permits offenders to channel illegal assets into the flourishing economy, demolishing monetary establishments and the cash supply, and giving fraudsters unjustifiable financial influence. GAO and Ye likewise describe tax evasion as a cycle in which posses "launder filthy cash" to conceal its unlawful starting points and seem real and "clean".

Financial statement fraud

Financial statement fraud(corporate fraud).A Fiscal report is an important report that reflects the financial status of an organization. It now has a well-balanced shape.

- Falsifying these statements to make the business more favorable
- Improving the presentation of activities
- Reducing tax liabilities
- Endeavoring to overstate execution because of functional strains

Charge card extortion

There are two essential kinds of charge card extortion. Application and social misrepresentation. In application extortion, fraudsters utilize phony or outsider information to buy new cards from guarantors. His four kinds of misrepresentation are mail theft, taken/lost card, fake card, and cardholder non-attendant extortion. Contract misrepresentation is an extraordinary kind of monetary extortion that controls land and home loan reports. The objective is frequently to bring down the worth of the property to persuade a bank to give a credit.

1.2 Financial Fraud Detection

Many financial fraud investigations focus solely on the numerical data in financial reports, often overlooking the potential for textual information. Deliberate concealment and accounting manipulation make it difficult to distinguish between fraudulent data and real numbers. Scientists perceive the worth of printed portrayals and mathematical information like Structure 10-K, featuring the significance of utilizing ignored text based data to recognize monetary extortion . Specifically, the administration conversation and examination (MD&A) segment of monetary reports has shown to be a significant device for identifying budget summary misrepresentation (FSF). In any case, existing exploration on utilizing literary substance in MD&A gives a deliberate, comprehensive, and theoretical rationale to direct misrepresentation recognition endeavors and give distinct and complete printed components for FSF ID. It misses the objective of design. That is the genuine raison d'être of this survey.

Similarly, fraudulent financial statements (FFS) have become a major problem around the world in recent years, fundamentally undermining the security of lenders. This is an inevitable problem given the lack of strong corporate governance in place. With the advancement of data mining innovations, analysts have also applied various techniques and models to identify his FFS. At this stage, financial reporting fraud investigations represent an exceptional opportunity around the world. Monetary misrepresentation massively affects both monetary exchanges and day to day existence. Misrepresentation can sabotage trust in an organization, disintegrate holds and effect the typical cost for most everyday items. Monetary organizations utilize an assortment of misrepresentation counteraction models to resolve this issue. In any case, con artists are versatile and over the long run foster various ways to deal with get through such cautious models. Regardless of monetary foundations' earnest attempts and government oversight, monetary extortion keeps on advancing. Tricksters these days can be an extremely clever, shrewd, and fast bundle.

As information mining innovation keeps on advancing, modern information disclosure techniques are supposed to isolate already dark data from the information. Text mining is the most common way of removing significant arrangements of numbers (organized information) from unstructured text. Text mining can be utilized to look at words and expressions and decide their relationship to different deals factors, for example, whether they are fake or non-deceitful.

1.3 Grouping of Information Digging Methods for Extortion Identification

A visual calculated system for distinguishing monetary bookkeeping misrepresentation utilizing information mining methods is introduced. The order structure introduced in Figure 1 depends on existing information in the two information mining and misrepresentation location research, and is valuable for understanding and carrying out these procedures with regards to distinguishing extortion in monetary bookkeeping. gives an organized way to deal with

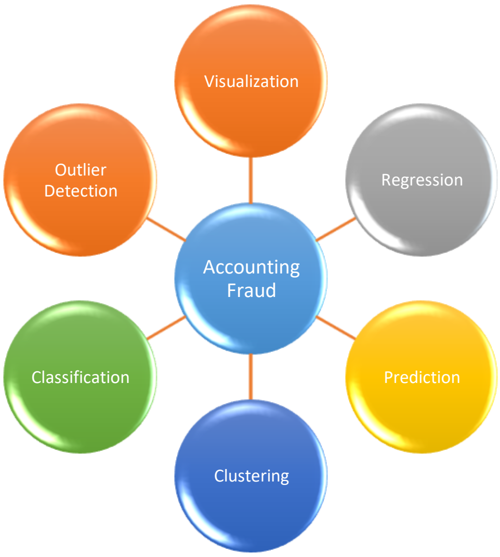

Figure 1. The Graphical Conceptual Framework for Application

A graphical conceptual system for detecting fraud in financial accounting by applying data mining techniques is presented. The classification structure shown in Figure1relies on current knowledge of ideas in data mining research and fraud detection research.

This business locales the issue of outrageous class unevenness in misrepresentation identification. Bayesian Belief Networks(BBNs)improve accuracy while minimizing computational cost. Hybrid fraud detection techniques that combine traditional methods are widely used. It focuses on improving the detection accuracy of false financial statements (FFS) while considering data mining vulnerabilities. Brain organizations (NNs) have been shown to be compelling and have great prescient capacities, and developmental calculations (EAs), encapsulated by hereditary calculations (GAs), have been demonstrated to be powerful in identifying time sensitive examples characteristic of bookkeeping extortion. Great at distinguishing. The proposed hybrid approach, BBN-GA, provides decision support to investors and authorities and contributes to more robust fraud detection.

The previous review did not address this issue of extreme class inequality much. Classification interactions use Bayesian Belief Networks (BBN) to achieve higher accuracy. Additionally, it reduces computing costs. This survey observed that half breed extortion recognition procedures are the most regularly utilized as they join the characteristics of a few conventional identification techniques. Thusly, the motivation behind this study is to work on the exactness of FFS discovery. Information mining methods, for example, inductive learning are not helpless against these assaults. Brain organizations (NNs) have been displayed to have preferable foreknowledge over other NNs in FFS location issues, making them a discretionary answer for grouping issues. Developmental calculations (EAs) are a class of probabilistic inquiry calculations reasonable for performing enhancement and order undertakings inside complex component spaces. One of the principal utilizations of the created calculation for bookkeeping extortion arrangement was introduced by Hoogs et al. I was inhaled into life. (2007). This study utilizes hereditary calculations (GA) to identify transient or time sensitive plans that show bookkeeping extortion.

Therefore, in this study, we proposed a hybrid strategy of Bayesian belief network and genetic algorithm (BBN-GA) method. It can provide funders and governments with an emotional support network.

2. Existing Methodology

Omar N (2017) proposed using artificial neural networks to predict fraudulent financial reports. The increasing frequency of fraudulent financial reporting requires adaptive audit practices. Various techniques with advanced classification and predictive capabilities have been proposed to detect and prevent such fraud. This study centers around building a counterfeit brain organization (ANN)- based prescient model for fake monetary revealing in little cap organizations. Accordingly, there were no indications of false revealing in little and medium-sized undertakings, and ANN accomplished a phenomenal forecast exactness of 94.87%, outflanking past techniques. This study provides valuable insights into the effectiveness of ANN in predicting financial fraud and improves forensic accounting research in Malaysia.

2.2 Support Vector Machines (SVM)

Abdelamid D (2014) et al. proposed programmed bank extortion discovery utilizing support vector machines. Fighting misrepresentation stays a squeezing concern, particularly for banks. They center around three normal instances of Mastercard extortion, tax evasion, and home loan misrepresentation, and propose a structure to identify bank extortion utilizing support vector machines. A hybridized approach of single-class and twofold SVM strategies was tried on different information bases and accomplished 80% precision with single-class SVM. This is a huge improvement. Although thereare challenges with credit scoring databases, it is possible to improve the results by considering the various parameters used in SVM techniques.

2.3 Naïve Bayes (NB)

Balaniuk R (2012)et al. We proposed a gamble based government review plan utilizing a guileless Bayes classifier. Leverage a centralized data platform for government agencies to improve access and compliance. To navigate this data landscape, accounting authorities are increasingly using computer-aided auditing techniques(CAAT) for periodic evaluations of financial and non-financial processes. They proposed using a Naïve Bayes classifier to improve the risk assessment of audit plans and align it with the formal process of the Institute of Internal Auditors. This semi-automated system enables the transition from reactive to proactive auditing, helping to identify and resolve issues early on that indicate fraud, waste, or abuse.

2.4 Decision Tree (DT)

Save P, Tiwarekar P (2017) et al proposed a novel thought for Mastercard misrepresentation discovery utilizing choice trees. Web based shopping and web based banking, worked with by the Web and Visas, have developed essentially alongside an expansion in Visa extortion. To address this, they propose a structure that utilizes choice trees and coordinates the Luhn and Pursue calculations. The Luhn calculation approves the card number, and the location matching principle checks in the event that the stacking address and transport address match. Albeit not convincing, matching addresses improves the probability of authenticity. Deviations from a client's typical exchanges are hailed utilizing exemption discovery to assist with recognizing likely misrepresentation.

3. Proposed Methodology

3.1 Financial Statement Fraud

Accounting fraud involves manipulating reports to make a company appear more profitable, avoid taxes, inflate stock prices, or guarantee loans. These fraudulent activities include creating confidential financial records showing expenses, profits, and loans. Fraudsters falsify facts inorder to increase stock prices, minimize tax liability, attract investors, and obtain loans. The main motive is to project a positive image and mislead stakeholders for economic reasons.

Dataset splitting

To keep up with the trustworthiness of the informational index of 30,000 perceptions, the informational index was parted into preparing, test, and approval sets in the proportion of 70%, 15%, and 15%. The mean vector and covariate lattice are figured utilizing the preparation set. The GA uses the endorsement set to conclude the ideal exceptional case edge. Finally, the accuracy of the model is surveyed using a test set. This approach guarantees vigorous model preparation, powerful thresholding, and precise model assessment across assorted datasets.

Bayesian Belief Network (BBN)

Pearl (1986) first introduced the Bayesian Belief Network(BBN).BBNs are of great importance in dealing with uncertainty and inference challenges and have been broadly applied in fields, for example, in fields such as natural resources, medical diagnostics, and software cost evaluation. It is derived from Bayes theorem and adjusts the probabilities based on the results of important nodes. BBNs are effective at modeling uncertainty by using graphical structures to represent cause-and-effect relationships and derive outcomes. Dynamically update probabilities with new data to facilitate informed decision-making in a variety of areas

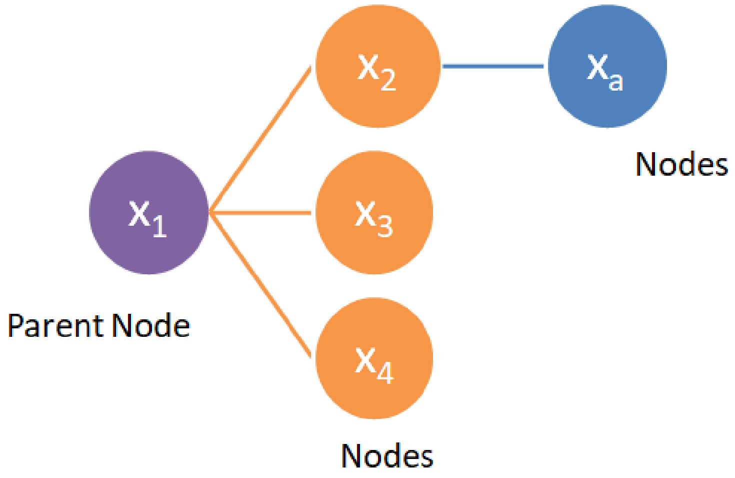

Figure 2.BBN concept diagram

A directed acyclic graph, BBN, consists of nodes connected by edges representing decision variables. See Figure 2 for a conceptual diagram.

Bolt deals with causality and strength. If each node x has a different parent (called Parents(x)),then the conditional probability of obtaining all parents and node x is expressed by the formula: (1).Therefore, we can build a restrictive likelihood table for hub x. The mix of probabilities of n credits () is displayed in Equation(2)

P (x |parents (x)) (1)

$$\mathbf{P}((\boldsymbol{x})) = \mathbf{P}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \prod_{i=1}^{n} \boldsymbol{p}\ (\boldsymbol{x}_i|\boldsymbol{P}_{ai} \qquad \mathbf{(2)}$$

Bayesian Networks

In view of its effortlessness, it's a good idea to begin investigating Bayesian organizations for Visa exchange characterization utilizing the Gullible Bayesian classifier. A naïve Bayesian classifier is a sort of Bayesian organization that expects restricted opportunity of qualities (other than class highlights). The last decision about exchange classes is made after class likelihood assessment [5].

$$\boldsymbol{p}(\boldsymbol{x}_1) = \boldsymbol{x}|\ \boldsymbol{C} = \boldsymbol{c}) = \frac{\boldsymbol{K}_{\boldsymbol{c}+1}}{\boldsymbol{N}+\mathbf{1}} \qquad \mathbf{(3)}$$

This is an informational index with property I and worth x in the preparation informational collection of exchange class c. N is a record of occasions in the preparation dataset This probability estimation requires converting numeric attributes into discrete attributes. Note that most of the attributes used (there are about 24 main attributes) are individual. The disadvantage of this approach is that it expects all detected values to be saved in the training data, making this difficult for many transactions. One possibility is to use a normal distribution [3].

$$\boldsymbol{\mu}_c^i = \frac{1}{n} \times \sum_{j=1\ldots n} \boldsymbol{x}_{cj}^i \qquad \mathbf{(4)}$$

$$\boldsymbol{D}_c^i = \frac{1}{n} \times \sum_{j=1\ldots n} (\boldsymbol{x}_{cj}^i - \boldsymbol{\mu}_c^i)^2 \qquad \mathbf{(5)}$$

$$\boldsymbol{P}(\boldsymbol{X}_i = \boldsymbol{x}|\boldsymbol{C} = \boldsymbol{c}) = \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\mu}_c^i, \boldsymbol{D}_c^i) \qquad \mathbf{(6)}$$

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\mu}_c^i, \boldsymbol{D}_c^i) = \frac{1}{\sqrt{2\times\pi}\times \boldsymbol{D}_c^i} \times \boldsymbol{exp} \frac{-(\boldsymbol{x}-\boldsymbol{\mu}_c^i)^2}{2\times(\boldsymbol{D}_c^i)^2} \qquad \mathbf{(7)}$$

The weight of this approach is to expect an ordinary dissemination of value scores. It's smarter to find class probabilities utilizing bit thickness assessment.

In this situation, we arrive at a Gaussian function in the middle of the set of all property values found in the training data. Such methods provide a more accurate estimate of the true characteristic distribution, but still require that all values be stored.

Bayesian grouping depends on Bayes' factual hypothesis. Bayes hypothesis permits us to work out the converse likelihood. As per Bayes' hypothesis, for instance, on the off chance that H is a speculation and an item, the likelihood that the theory is valid is:

If an object is gullible, a Bayesian classifier makes class conditions suspect of independence. This means that the impact of feature detection on a particular class does not depend on the values of various attributes. This assumption applies to the calculations. If this assumption is correct, the Credulous Bayesian classifier has the highest accuracy rate compared to all other classifiers. However, this assumption is usually not realistic because there may be constraints between attributes. Bayesian belief networks(BBNs) consider the representation of conditions in a subset of attributes. A BBN is a coordinated acyclic graph, where each node corresponds to a feature and each bolt corresponds to a stochastic dependency. If a bolt is drawn from node A to node B, then B and B's parents are descendants of A. In a belief network, each variable has a parent (Han and Camber, 2000).). About each node

$$\boldsymbol{P(x_1, x_2, \ldots, x_n)} = \prod \boldsymbol{P}(\boldsymbol{X_i}|\boldsymbol{Parents}(\boldsymbol{X_i})) \qquad \textbf{(8)}$$

Network configuration can be portrayed deduced or gotten from information. For characterization purposes, one of the hubs can be described as a class hub. The organization can compute the likelihood for every elective.

Genetic Optimization Algorithm (GOA)

Presented by John Holland in 1973, hereditary streamlining calculations are roused by Charles Darwin's standards of regular determination. Intended for ideal arrangements and stochastic inquiries, this calculation joins cycles of chromosome populace, choice, hybridization, and transformation to deliver new posterity. Its viability lies in copying the transformative cycle for canny critical thinking. It is an assortment of steps that a hereditary calculation (GA) takes to show up at an ideal arrangement utilizing a deliberate, hereditarily impacted way to deal with exploring a mind boggling issue space. is. I turned into a specialist at it.

Methods used to choose the best chromosomes inside a cycle incorporate tip top and stochastic expansive determination, rank choice, roulette wheel determination, steady state choice, cutthroat determination, and shortened choice. It will be. The determination cycle guarantees that the best parent is chosen. Hybrid haphazardly takes those focuses and trades the parent substrings to frame her two new kids. The change supervisor refines her GA to track down the ideal arrangement and adjusts a portion of the descendants to shape another chromosome populace. The cycles go on until all relatives have been made.

Figure edge

A threshold is a number that separates fraudulent transactions from non-fraudulent transactions. Therefore, when the most extreme probability measure for a given transaction is not exactly ε,the optimal choice of this value is important to achieve optimal transcription. Transactions are classified as in consistent (fraudulent) and, as a result, are classified as genuine transactions. A genetic optimization calculation (GOA) is utilized to choose the most suitable edge, which frequently ensures an ideal arrangement. Utilizing the attending advance, the worth of the limit still up in the air.

$$\mathbf{p_i} = \underset{(2\pi)2|\Sigma|\frac{1}{2}}{\frac{1}{d}} \ \mathbf{exp}\left[-\frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu})^{\boldsymbol{T}} \textstyle\sum^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})\right] \qquad \textbf{(9)}$$

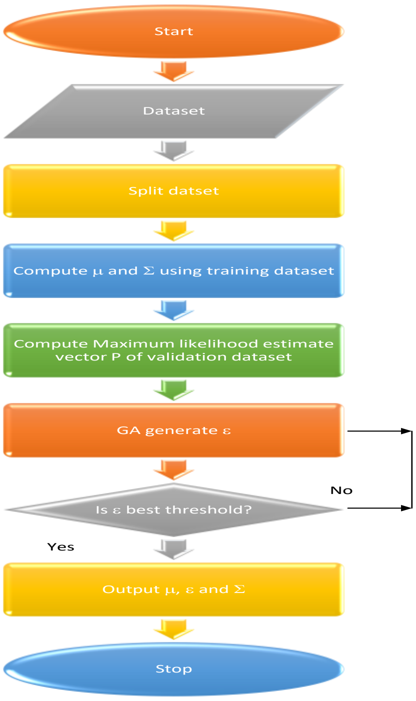
Start
Dataset
Split datset
Compute μ and Σ using training dataset
Compute Maximum likelihood estimate vector P of validation dataset
GA generate ε
No
Is ε best threshold?
Yes
Output μ, ε and Σ
Stop

Figure 3. Proposed Flowchart

The threshold is chosen based on the ultimate goal of increasing the classification accuracy of the actual target in the validation dataset and its expected probability using N (X_validation_set, meu, Z). This process consists of iterations and reviews, and is best done in a GA. A genetic optimization algorithm (GOA) was used to select ε (a real number) that limits the misclassification rate in the validation dataset. The dataset is first partitioned, the training dataset is used to register the mean vector and covariate matrix, and a GA is used to select the optimal threshold for classification.

Confusion Matrix

A rundown of the normal outcomes for a characterization issue is handled utilizing a disarray network. The lattice shows how the order model gets adulterated during the forecast interaction. This gives an exhaustive information on how the model makes mistakes and what sorts of blunders happen.

4. Experiment Results

4.1 Accuracy

Table 1: Comparison Table of Accuracy

No of Data	SVM	ANN	Proposed BBN -GOA
100	**73**	**65**	**79**
200	**89**	**85**	**93**
300	**85**	**81**	**91**
400	**82**	**76**	**89**
500	**72**	**69**	**80**

Comparison of precision values Table 1 portrays the various benefits of existing calculation (SVM, ANN) and the proposed BBN-GOA. Figuring existing calculations (SVM, ANN) with the proposed BBN-GOA. We get improved results. Existing calculation values start at 73-72, 65-69. also, proposed values start at 79-80.

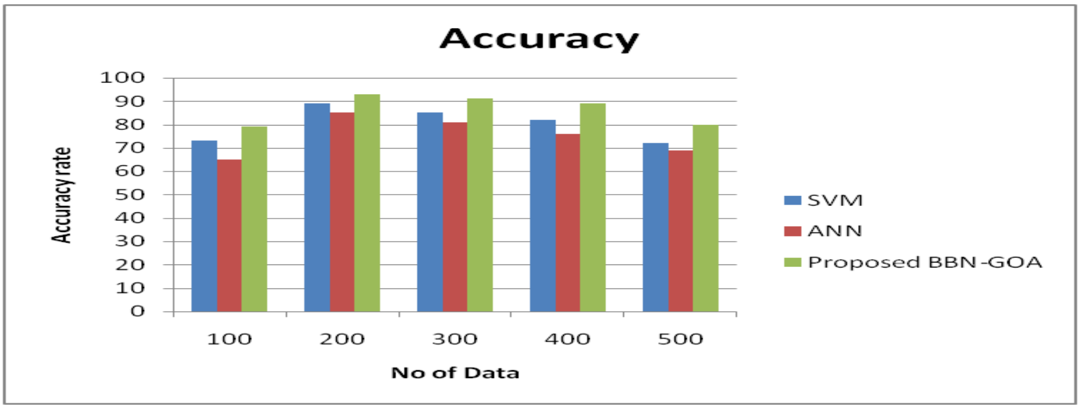

Figure 4: Comparison chart of Accuracy

Figure 4 shows an exactness comparision table appearance Existing 1, Existing 2 (SVM, ANN), and proposed BBN GOA values. The X-hub shows the approval and the y-pivot shows the presentation esteem regarding precision rate. The proposed BBN-GOA esteem beats existing calculations. Existing calculation values start at 73-72, 65-69, and proposed values start at 79-80.

4.2 Sensitivity

Table 2. Comparison Table of Sensitivity

No of Data	SVM	ANN	Proposed BBN - GOA
100	83.48	81.22	86.87
200	84.74	83.52	88.74
300	88.21	84.01	90.55
400	90.48	87.35	92.46
500	93.66	90.65	96.91

Comparison of Sentivity values Table 2 depicts the various upsides of the current and proposed calculations (SVM, ANN). contrasting the exting calculations (SVM, ANN) with the proposed calculation gives improved results. Existing calculation values start from 83. 48 to 93.66, 81.22 to 90.65. also, proposed values start from 86. 87 to 96.91.

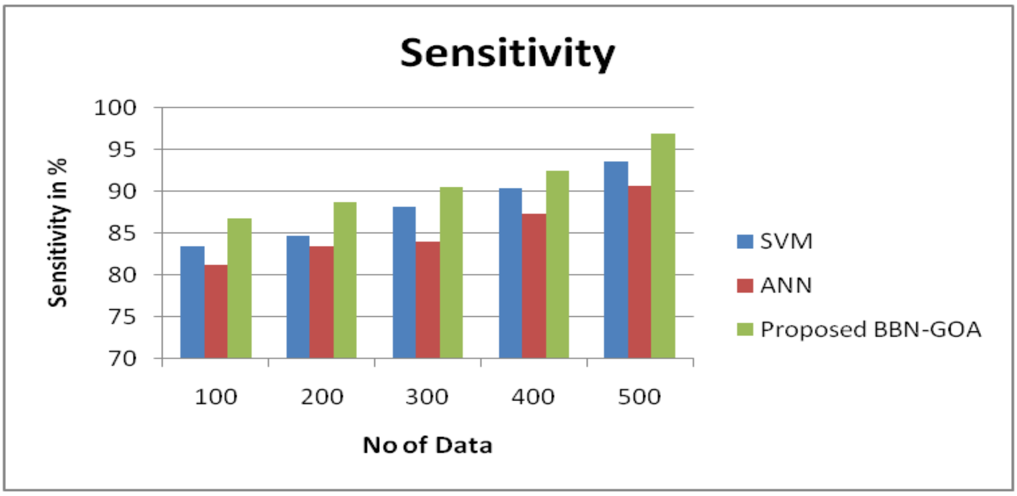

Figure 5: Comparison Chart of Sensitivity

Figure 6 shows a sensitivity comparision table showing Existing 1, Existing 2 (SVM, ANN), and proposed BBN GOA values. The X-axis shows validation and the y-axis shows performance values. Sensitivity in percent. The proposed values are better than existing algorithms. Existing algorithm values start from 83.48 to 93.66, 81.22 to 90.65. and proposed values start from 86.87 to 96.91.

4.3 Specificity

Table 3: Comparison Table of Specificity

No of data	SVM	ANN	Proposed BBN - GOA
100	83.154	83.121	84.745
200	85.649	84.024	87.999
400	90.267	90.135	91.406
500	92.623	92.365	93.91

Comparision of specificity values Table 3 portrays the various benefits of existing calculations (SVM,ANN) and the proposed BBN-GOA. Looking at existing calculation (SVM, ANN) with the proposed BBN-GOA. We get improved results. Existing calculation values from 83.154 to 92.623, 83.121 to 92.365. also, proposed values start from 84.745 to 93.91.

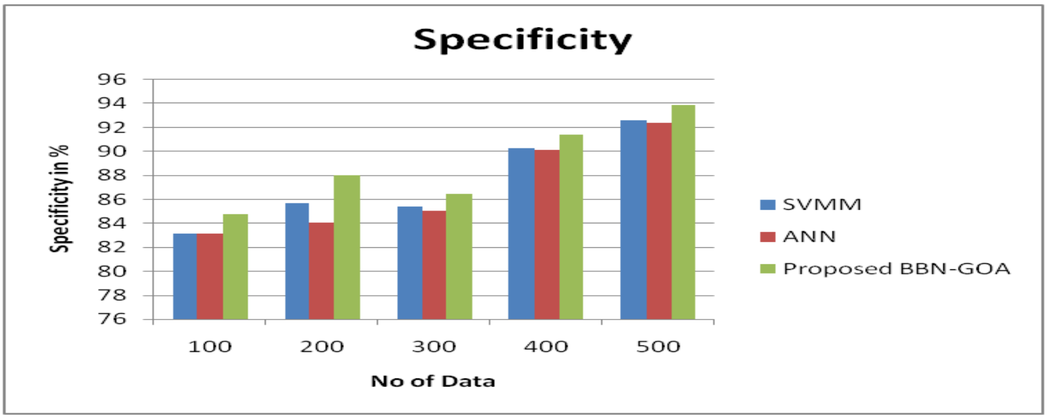

Figure 6: Comparison Chart of Specificity

Figure 6 shows the particularity comparision table. Showing Existing 1, Existing 2 (SVM, ANN), and proposed BBN GOA values. The X-pivot shows approval and the y-hub shows execution values.

Explicitness in percent. The proposed values are superior to existing calculations. Existing calculation values start from 83.154 to 92.623, 83.121 to 92.365, and proposed values start from 84.745 to 93.91.

4.4 Precision

Table 4 : Comparison Table of Precision

No of Data	SVM	ANN	Proposed BBN - GOA
100	70.74	68.02	72.26
200	73.09	70.26	75.23
300	72.89	71.18	75.02
400	75.04	72.11	77.14
500	76.85	74.22	78.99

Comparision of precision values Table 4 portrays the various benefits of existing calculations (SVM, ANN) and the proposed BBN-GOA. Contrasting existing calculations (SVM, ANN) with the proposed BBN-GOA. We get improved results. Existing calculation values start from 83.154 to 92.623. 83.121 to 92.365. also, proposed values start from 84.745 to 93.91.

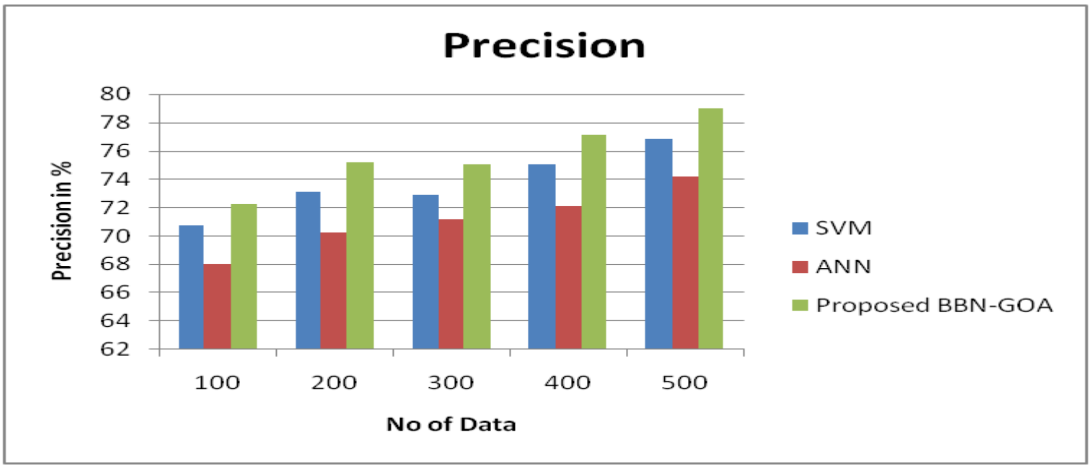

Figure 7 shows the accuracy specificity comparision chart showing existing 1, existing 2 (SVM, ANN), and proposed BBN GOA values. The X-axis shows validation and the y-axis shows performance values. Precision in percent. The proposed values are better than existing algorithms. Existing algorithm values start from 83.154 to 92.623, 83.121 to 92.365, and proposed values start from 84.745 to 93.91.

5. CONCLUSION

Financial reports provide all stakeholders with important information about an organization's current opportunities and prospects. Financial disclosure fraud is the intentional falsification of reports for the purpose of deceiving customers. Financial fraud has become one of the most fundamental problems in the world. Review methodology should now adjust to the rising number of monetary misrepresentation cases. Different information mining strategies to recognize monetary misrepresentation from public monetary exposures. Methodologies utilized incorporate arbitrary timberland calculations, counterfeit brain organizations, strategic relapse, support vector machines, Truck, choice trees (C4.5), Bayesian organizations, packing, stacking, and Adaboost. Data mining techniques can help internal and external evaluators predict financial fraud. This extension successfully identifies fraudulent Bayes ian belief networks using a genetic optimization algorithm (BBN-GOA). In terms of execution, BBN-GOA outperforms fake brain organizations (ANN), support vector machines (SVM), and so on.

REFERENCES

[1] Wei Dong, Shaoyi Liao and Liang Liang (2016), "Financial Statement Fraud Detection Using Text Mining: A Systemic Functional Linguistics Theory Perspective", Pacis 2016 Proceedings. 188. http://aisel.aisnet.org/pacis2016/188.

[2] Dong, Wei, Shaoyi Liao, and Liang Liang. "Financial Statement Fraud Detection using Text Mining: A Systemic Functional Linguistics Theory Perspective." In PACIS, p. 188. 2016.

[3] Fu, Kang, Dawei Cheng, Yi Tu, and Liqing Zhang. "Credit card fraud detection using convolutional neural networks." In International Conference on Neural Information Processing, pp. 483-490. Springer, Cham, 2016.

[4] Rawte, Vipula, and G. Anuradha. "Fraud detection in health insurance using data mining techniques." In Communication, Information & Computing Technology (ICCICT), 2015 International Conference on, pp. 1-5. IEEE, 2015.

[5] Omar N, Johari ZA, Smith M. Predicting fraudulent financial reporting using artificial neural network. Journal of Financial Crime. 2017 May 2.

[6] Abdelhamid D, Khaoula S, Atika O. Automatic bank fraud detection using support vector machines. InThe International Conference on Computing Technology and Information Management (ICCTIM) 2014 Jan 1 (p. 10). Society of Digital Information and Wireless Communication.

[7] Balaniuk R, Bessiere P, Mazer E, Cobbe P. Risk based government audit planning using naïve bayes classifiers. InAdvances in Knowledge-Based and Intelligent Information and Engineering Systems 2012 (pp. 1313-1323). IOS Press.

[8] Save P, Tiwarekar P, Jain KN, Mahyavanshi N. A novel idea for credit card fraud detection using decision tree. International Journal of Computer Applications. 2017;161(13).

[9] Bhavitha, B.K.; Rodrigues, A.P.; Chiplunkar, N.N. Comparative study of machine learning techniques in sentimental analysis. In Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 10–11 March 2017; pp. 216–221.

[10] Carta, S.; Fenu, G.; Recupero, D.R.; Saia, R. Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. J. Inf. Secur. Appl. 2019, 46, 13–22.

[11] Rb, A.; Kr, S.K. Credit card fraud detection using artificial neural network. Glob. Transit. Proc. 2021, 2, 35–41.

[12] Song, R.; Huang, L.; Cui, W.; Vanthienen, J. Fraud Detection of Bulk Cargo Theft in Port Using Bayesian Network Models. Appl. Sci. 2020, 10, 1056.

[13] Dang, T.K.; Tran, T.C.; Tuan, L.M. Machine Learning Based on Resampling Approaches and Deep Reinforcement Learning for Credit Card Fraud Detection Systems. Appl. Sci. 2021, 11, 10004.

[14] Bouchti, E.; Chakroun, A.; Abbar, H.; Okar, C. Fraud detection in banking using deep reinforcement learning. In Proceedings of the 2017 Seventh International Conference on Innovative Computing Technology (INTECH), Luton, UK, 16–18 August 2017; pp. 58–63.

[15] Ahmed, M.; Mahmood, A.N.; Islam, R. A survey of anomaly detection techniques in financial domain. Futur. Gener. Comput. Syst. 2016, 55, 278–288.

[16] Uchhana, N.; Ranjan, R.; Sharma, S.; Agrawal, D.; Punde, A. Literature Review of Different Machine Learning Algorithms for Credit Card Fraud Detection. Int. J. Innov. Technol. Explor. Eng. 2021, 10, 101–108.

[17] Abbasi, A.; Albrecht, C.; Vance, A.; Hansen, J. Metafraud: A meta-learning framework for detecting financial fraud. Mis Q. 2012, 36, 1293–1327.

[18] Kowshalya, G.; Nandhini, M. Predicting Fraudulent Claims in Automobile Insurance. In Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 20–21 April 2018; pp. 1338–1343. 111.

[19] Bauder, R.; Khoshgoftaar, T. Medicare Fraud Detection Using Random Forest with Class Imbalanced Big Data. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 7–9 July 2018; pp. 80–87.

[20] Bartoletti, M.; Pes, B.; Serusi, S. Data Mining for Detecting Bitcoin Ponzi Schemes. In Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT), Zug, Switzerland, 20–22 June 2018; pp. 75–84.

[21] Zhang, W.; He, X. An Anomaly Detection Method for Medicare Fraud Detection. In Proceedings of the 2017 IEEE International Conference on Big Knowledge (ICBK), Hefei, China, 9–10 August 2017; pp. 309–314.

[22] Bauder, R.A.; Khoshgoftaar, T.M. Medicare Fraud Detection Using Machine Learning Methods. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 858–865.

[23] Mitra, S., Mitra, P. & Pal, S.K., (2012). Data mining in soft computing framework: a survey, IEEE Transactions on Neural Networks 13 (1) (2014) 3–14.

[24] P. Ravisankar, V. Ravi, G. Raghava Rao, Bose, Decision Support Systems: Detection of financial statement fraud and feature selection using data mining techniques. Aug 2014,pp 309- 324.

[25] Spathis, C. T. (2012). Detecting false financial statements using published data: some evidence from Greece, Managerial Auditing Journal 17 (4) (2013) 179–191.

[26] Cerda, L.,Sánchez, D.,Vila, M.A., & Serrano, J.M. (2015). Association rules applied to credit card fraud detection, Expert Systems with Applications 36 (2) (2014) 3630–3640.

[27] J.E. Sohl, A.R. Venkatachalam, A neural network approach to forecasting model Selection , Information & Management Jan 2014,pp 297–303.

[28] Baesens, B., Derrig, R.A., & Dedene, G. Viaene, S., (2014). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection, The Journal of Risk and Insurance 69 (3) (2013) 373–421.

[29] Milne.P, Takeuchi J.Yamanishi, K. Williams, (2014). On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms, Data Mining and Knowledge Discovery 8 (3) (2014) 275–300.

[30] Arpit Tiwari and Nishta Hooda (2018) Machine Learning Framework for Audit Fraud Data Prediction, Volume 7, Issue VI, JUNE/2018.