*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 3, Ser. V (May - June 2019), PP 45-52*
*www.iosrjournals.org*

# Credit Card Fraud Detection System Based On Machine Learning Techniques

[#1]Gurram Sai Kumar, *B.Tech Student,*
[#2]Madala Venkaiah Naidu, *B.Tech Student,*
[#3]Dr.Madugula Sujatha, *Professor,*
*Dept of CSE, Jyothishmathi Institute Of Technology & Science, Karimnagar,T.S., India.*

***Abstract:*** *Due to rapid growth in cashless transaction, the chances of number of fraudulent transactions can also increasing. A Fraud transaction can be identified by analyzing various behaviors of credit card customers from previous transaction history datasets. If any deviation is noticed in spending behavior from available patterns, it is possibly of fraudulent transaction. Data mining and machine learning techniques are widely used in credit card fraud detection.In this paper is introduced best data mining algorithm called "machine learning algorithm", which is used to detect the credit card fraud, so initially use this algorithm and it is one of the standard model. Then, secondly apply the hybrid methods namely, "AdaBoost and majority vote method". Use this model efficacy, which is evaluated, and then use the credit card data set it is publicly available one. The financial institution included the real-world credit card data set, so it is taking and then analyzed..*

***Keywords:*** *AdaBoost, classification, machine learning, hybrid method, credit card, fraud detection, predictive modeling, voting.*

---

Date of Submission: 12-06-2019 Date of acceptance: 28-06-2019

---

## I. Introduction

With the developments in the information technology and improvements in the communication channels, fraud is spreading all over the world, resulting in huge financial losses. There have been several researches done in the field of fraud detection, with various methods employed in detection and prevention [1]. Methods such as decision tree learning, support vector machines, neural networks, expert systems and artificial immune systems have been explored and identified [2] for fraud detection. The scope of this paper has been reduced to only credit card application fraud and risk based on decision tree induction using ensemble learning techniques and genetic algorithms. Due to sensitivity of customers' financial information, getting clean data is hard for mining applications. The dataset used is obtained from the UCI (University of California, Irvine) machine learning repository-German Credit Card dataset[3] and Australian Credit Card dataset for the present paper.

Fraud prevention is a subject that always brought interest from financial institutions, since the advent of new technologies as telephone, automatic teller machines (ATM)and credit card systems have leveraged the volume of fraud loss of many banks [4]. In this context, fraud prevention, with a special importance of fraud automatic detection, arises as an open field for application of all known classification methods. Classification techniques play a very important role, once it is able to learn from past experience (fraud happened in the past) and classify new instances (transactions) in a fraud group or in a legitimate group.

**CATEGORİES OF ECONOMİC FRAUDS:**

In the latest news of India's biggest banking scam PNB scam in which fraudster obtain $1.8 billion from overseas loan by making SWIFT defenseless to fraud. As per the Russian government, Hackers stole $6 million from one country bank using SWIFT network. Unconstitutional user monitors the traffic of transactions and interrupts it and sends all the money to his account by passing the forged message. Another type of fraud for which financial institutions have to compact with is Push Payment fraud, in which through electronic mail fraudster act as a genuine supplier and take money from the certified user. This is very crucial fraud in which ordinary user get affected accidentally.

This happens because 70% of the population fails to report about frauds. Due to this, fraudster gets motivated and does sharper practice. According to one business news 29% fraudster attempt is being done in India and 18% in another country. As per their survey, many fraudsters will deposit dummy currencies or demonetized currencies in the bank, due to which bank has to face financial troubles[5].

There are more than 20+ families of frauds. Some are known to the users and some are unknown. But several kinds of fraud will damage human assets in large extent. The following sections, discussed the special categories of frauds. Inside Burglary is the one way where business users use their own organization product without intimating it to their higher authority which leads to disloyalty towards the organization.

Backhander which is highly liable for identity crime. In this crime, the person will sell the secret information about the users or about the company to the third party. The third party will then sell this information at a high cost. This is the way where fraudsters earn money.

Skimming most well-known form of frauds. Unlawful user tries to get the information online from the legal users by using some electronics gadgets to copy the content of the user's runtime[6].

Ponzi Scheme is the scheme in which some agents like outlay planner will give you knowledge about how to invest money and how much profit user will get after taking the plan. But actually, these planners are only for taking money for savings without giving any profit to the legitimate users. Like this many common men become dupe under this scheme.

Identity Crime is most dangerous and harmful to the public. Since human's mainly precious thing is his/her uniqueness. Fraudsters use legal user's details and steal money from various financial institutes. This leads to bad bang on user's life.

Phishing is the exceedingly used techniques which can be done online by creating fake websites which looks original and asks users to fill their personal or financial testimonial.The online users get butt of the phishing attacks.

Smart cards deception in which impostor will get both the way to complete their unlawful task of using their smart cards for the immoral thing. Card stealing and getting information is the common thing which fraudster will do with cards[7]. This will leads great financial failure.

Fund Transfer Scam in which ordinary user receives an electronic message for a payment, but during a transaction that amount get hacked by the third party and transfer this money into outside country account. For which authorized user has to pay a huge penalty.

Financial assertion fraud comes under the group in which user will come to know about fraud always after being a victim of it[8].

This kind of fraud has to be carefully handled. Since a large amount of data is stored in databases in every organization. In this state of the art, it mainly focused on financial statement fraud.

There are many frauds which we usually not taking it seriously. Due to which fraudsters get chance to make it more vulnerable. Listed here some of this kind of frauds

Dummy Prizes in which many online users get messages of winning some huge amount and they have been asking for their banking details for the transaction. But some users unknowingly send their banking details and turn out to be a victim of such fraud. International Raffle fraud[9] is luck by chance fraud. Fraudsters sell some tickets of the lottery and making fool to the users that they will get it back after winning the ticket.

Bill and remission fraud which generally happen in the programmed system. When users cannot buy anything online but they are getting a bill of which they are not buying. Because of computerized payment mode there, amount gets deducted from their account. The user is not able to understand that some fallacious has happened.

Blackmailing is one of the common fraud fraudsters always do. Just like trial and error basis. Due to this apprehension near about 80% users fulfill fraudster demands.

Tax Evasion is done by not only fraudsters but also it is done by any users who are not going to pay tax to the government. The people who are persistently not paying tax, they come under fraudster category.

## II. Related Work

There is an elongated list of frauds, which are not only financial but also identity frauds. According to the analysis every large or small organization is having banking associations. It means that finance plays a key role in industries. Healthcare, banking, insurance, educational institutes and much more organization are doing a transaction of personal as well as financial information. This organization is dealing with large or big databases to store transaction history. Data comes from a mixture of paths, some are alike or some are unlike data. The basic concern comes with this is how to supervise these data and how to provide security of real dataset [10]. Supplementary with this the basic subject come in front to all of us is how to classify this data. If data is large, cataloging of data based on its similarity so that it results to planned data. According to author structured data will get by finding out rare category data [11]. This paper deals with finding rare category by applying active learning algorithm. One more important thing while dealing with big data, maintaining safety measures is a great concern. Now at this point, it is better to understand the fraudster's outlook. Fraudster basically targeting large data for the reason that this large data is open to or prone to attacks[12]. So our basic necessity is to find out vulnerabilities and the component of data which require more security.

There are basically three things which should be handled carefully.

1. User's vital data
2. Cataloging of data
3. Sanctuary of data

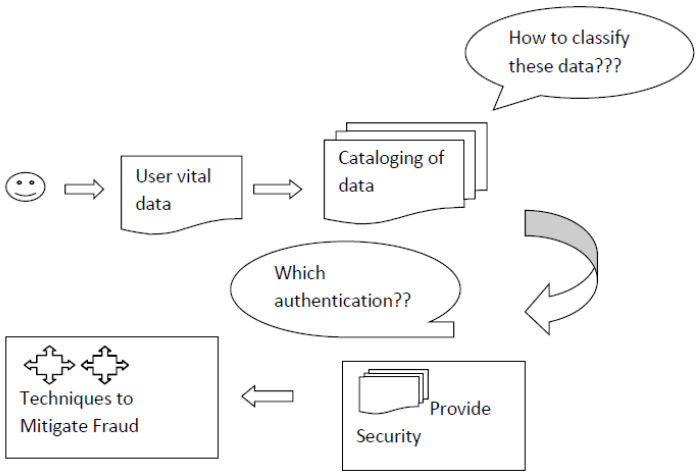


**Figure: 1.** Three Points of Vulnerability

Extending further discussion shows some techniques which have been used for different kinds of frauds. Due to the complexity of the data, visual analytics is very much required [13]. With Visual analytics, it can detect a wide range of fraud. This article largely focuses on the fraudulent transaction with a financial institution. The person behind it discussed some problem complexity like scalability, context density, changes in data, false positive and negative, time analysis. This paper uses the scoring system by which we can compare past data with present data and with the result of comparison we can detect suspicious fraud.

Multifarious data get classified with different machine learning techniques. One of these techniques is majority voting [14] which was used for data classification and got good results. This technique is applied to frauds which are performed on credit cards. Author has proposed one fusion method which is called Adaboost.

Credit card fraud becomes so popular, as today world is the digital world. Human wants everything in hand by sitting at home. This increases the use of e-commerce, by which attackers or fraudster being paid more chance to attempt frauds. Generally, fraudsters use some strategies to commit fraud. It is the necessity to recognize their strategy to stop further frauds. In paper [15] where author specifies some opponent learning techniques to find out their future scheduling. Since the use of different types of cards gets fluctuate every time. The behavior of every user changes according to the time which makes the situation more difficult to predict the fraud. This can be also reduced by applyingsome data mining techniques [16] and results prove that K nearest algorithm works better than Naïve Bayes in data mining techniques. As credit cards are connected with the banking sector and due to fraud in financial sectors it affects the whole economic system as well as it affects the emotions of the users who are connected with this sector[17].

Many types of research discussed a range of machine learning techniques like the genetic algorithm, fuzzy logic to reduce frauds in the financial sector. One best approach is to find out the sequential use of credit card and find out the supreme behavior of the users using DRL(Deep reinforcement Learning)[18] But as fraudster becoming more intellectual they are using different strategies to consign fraud. So for a continuous change in fraud prototype, it's time to improve the training on different datasets.

## III. Machine Learning Techniques For Credit Risk

Machine learning methods have frequently been used in the analysis of the credit scoring system because they require fewer assumptions and deliver higher analytical accuracy. The ensemble model is one of the commonly-used algorithms in current machine learning techniques. The tree-based ensemble model using bagging and boosting is especially popular. Given that the hierarchical tree structure can model non-linear associations, this method is typically used for regression and classification, and is likely to perform well for complex, independent variables. Random Forest (RF) is an ensemble machine learning method that uses multiple trees as classifiers using bagging. After taking the majority vote over all classifiers, the RF method combines information across all trees to reveal variable importance. Boosting methods such as Adaptive Boosting (AdaBoost) and Gradient Boosting Machine (GBM) are another kind of ensemble method with strong similarity to RF. Ensemble models have been applied to many financial studies, such as studies concerning credit risk, customer profit, stock prices and automated trading[19].

In practice, the ensemble model can combine many weak, simple models to obtain a stronger ensemble prediction. In real credit risk applications, many results shows that traditional, single-prediction models have

lower prediction accuracy and are less robust than ensemble models, especially in high-dimensional or large sample data sets. In this research, ensemble models were the majority concerned.

### 3.1 Random Forest

Bagging is one of the ensemble algorithms in machine learning used to improve the stability and accuracy of machine learning algorithms. Random forest (RF), proposed by Breiman, is one such algorithm. RF can also help identify the truly relevant predictor variables so that feature selection can be conducted by the model. Furthermore, some results also illustrate the importance of the choice of the number of variables in each tree and it is found to be optimal with respect to prediction accuracy in empirical studies.

### 3.2eXtreme Gradient Boosting

eXtreme Gradient Boosting (XGBoost), derived by Chen and He (2015), is one kind of GBM model. Both XGBoost and GBM follow the principle of gradient boosting, but there are differences in modelling details. Specifically, XGBoost uses a more regularized model formalization to control over-fitting, which grants better performance. XGBoost has used second derivative information, and ordinary GBM only uses first-order derivatives. XGBoost models greatly optimize the traditional gradient boosting model, and is one of the fastest learning algorithm of gradient boosting algorithm[20].

### 3.3 Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) a gradient boosting framework that uses tree-based learning algorithms. It is highly efficient and scalable, and can support many different GBM algorithms. This method was developed by Microsoft Research Asia. LightGBM has been shown to be several times faster than existing implementations of gradient boosting trees, due to its fully greedy tree-growth method, histogram-based memory and computation optimization. LightGBM adds a maximum depth limit on the leaf-wise algorithm to ensure high efficiency and prevent overfitting[21].

## IV. Proposed Algorithm

### AdaBoost.M1

Boosting is another kind of ensemble algorithm for improving the accuracy of any given learning algorithm, and it means that a weak learning algorithm better than random guessing in a Probability Approximately Correct (PAC) model can be boosted into a strong learning algorithm. The Adaptive Boosting (AdaBoost) algorithm solved many of the practical difficulties with the earlier boosting algorithms. AdaBoost.M1 is used to extend AdaBoost to multi-class cases in generalization.

AdaBoost, also known as Adaptive Boosting is used as part of implementation method to boost the performance of decision tree and it is implemented in WEKA (Waikato Environment for Knowledge Analysis) as AdaBoost.M1 [21]. This boosting algorithm can be applied to any classifier's learning algorithm. AdaBoost algorithm in a pseudo code form is given in Figure 1[22]. This algorithm creates an ensemble of classifiers, with each having a weighted vote which is function of βt [16].

**Input** : Training set $S = \{x_i, y_i\}$, i = 1,...,$N$, and $y_i \in \mathbb{C}$, $\mathbb{C} = \{c_1,...,c_m\}$; $T$ : number of iterations; $I$ : Weak leaner

**Output** : Boosted classifier:

$$H(x) = \arg\max_{y \in \mathbb{C}} \sum_{t=1}^{T} \ln\left(\frac{1}{\beta_t}\right)[h_t(x) = y]$$

where $h_t, \beta_t$ are the induced classifiers (with $h_t(x) \in \mathbb{C}$) and their assigned weights respectively

1: $D_1(i) \leftarrow 1/N$ for $i = 1, \ldots, N$

2: **for** $t = 1$ to $T$ **do**

3: $h_t \leftarrow I(S, D_t)$

4: $\varepsilon_t \leftarrow \sum_{i=1}^{N} D_t(i)[h_t(x_i) \neq y_i]$

5: **if** $\varepsilon_t > 0.5$ **then**

6: $T \leftarrow t - 1$

7: **return**

8: **end if**

9: $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

10: $D_{t+1}(i) = D_t(i) \cdot \beta_t^{1-[h_t(x_i) \neq y_i]}$ for $i = 1, \ldots, N$

11: Normalise $D_{t+1}$ to be a proper distribution

12: **end for**

**Figure 2:** Pseudocode for AdaBoost.M1

## V. Experimental Results

The implementation of the proposed solution will be limited only to credit approval risk. Based on this scope, dataset containing credit approval is obtained for experimental studies. Two forms of experimental results are provided, which are, experimental results of decision tree without any boosting techniques and experimental results of decision tree together with AdaBoost.M1[23].

Fraud Type : Wrong CVV

| ID | Card Number | User Name | Bank Name | Fraud Amount | Web Site | Date |
|---|---|---|---|---|---|---|
| 5 | 350881406571 | Praniti | Canara Bank | 18000 | Flipkart | 31/10/2018 18:34:55 |
| 9 | 536470266101 | Roshan | Indian Bank | 10000 | Flipkart | 01/11/2018 11:55:17 |
| 10 | 537785904513 | Shanmukh | Indian Bank | 18000 | Flipkart | 01/11/2018 12:02:32 |
| 21 | 537785904513 | Shanmukh | Indian Bank | 18000 | Flipkart | 01/11/2018 12:21:12 |
| 24 | 646597512025 | Sujan | SBI Bank | 18000 | Flipkart | 01/11/2018 12:35:34 |
| 25 | 642855074991 | Ashwin | SBI Bank | 10000 | Flipkart | 01/11/2018 12:38:27 |

**Fig3.** Fraud Result with Wrong CVV

Fraud Type : Expired Card

| ID | Card Number | User Name | Bank Name | Fraud Amount | Web Site | Date |
|---|---|---|---|---|---|---|
| 2 | 536470266101 | Roshan | Indian Bank | 10000 | Flipkart | 31/10/2018 18:32:54 |
| 6 | 320622743637 | Sanjay | Corporation Bank | 10000 | Flipkart | 01/11/2018 11:28:27 |
| 7 | 320622743637 | Sanjay | Corporation Bank | 10000 | Flipkart | 01/11/2018 11:30:20 |
| 11 | 537785904513 | Shanmukh | Indian Bank | 10000 | Flipkart | 01/11/2018 12:03:33 |
| 16 | 537785904513 | Shanmukh | Indian Bank | 18000 | Flipkart | 01/11/2018 12:08:28 |
| 22 | 537785904513 | Shanmukh | Indian Bank | 18000 | Flipkart | 01/11/2018 12:22:09 |
| 32 | 649942232755 | Shivaji | SBI Bank | 35000 | Flipkart | 01/11/2018 13:38:08 |

**Fig4.** Fraud Result with Expired Card

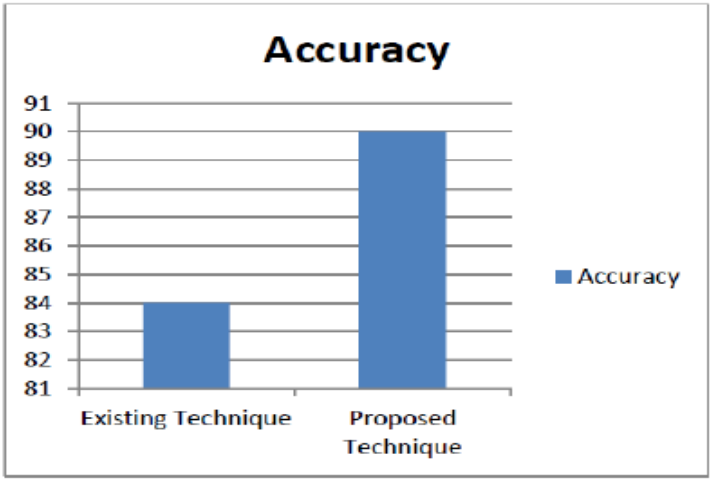


**Figure 5:** Accuracy Comparison

***Comparing the Results***

| | KNN | Random Tree | Proposed Algorithm |
|---|---|---|---|
| **Accuracy** | 0.9691 | 0.9432 | 0.9824 |
| **Sensitivity** | 0.8835 | 0 | 0.9767 |
| **Specificity** | 0.9711 | 0 | 0.9824 |
| **Limitations** | Cannot detect the fraud at the time of transaction | Not suitable for Randomness in dataset | Not Applied for non-Linear data |

TABLE: COMPARISION OF ALGORITHMS

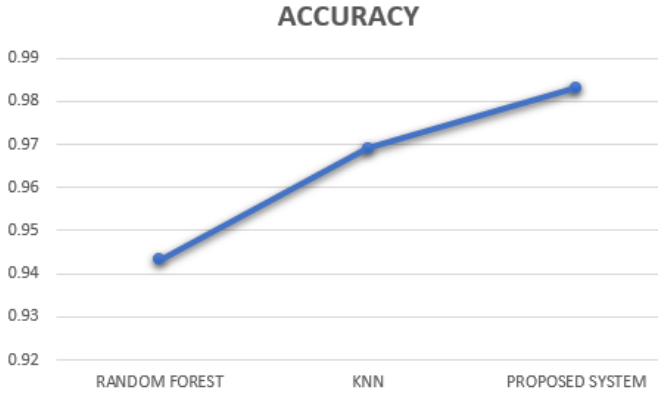


**FIG 6:** ACCURACY COMPARISION

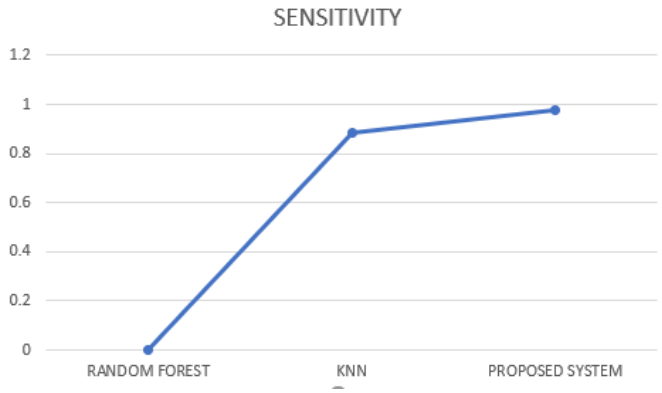


**FIG 7:** SENSITIVITY COMPARISION

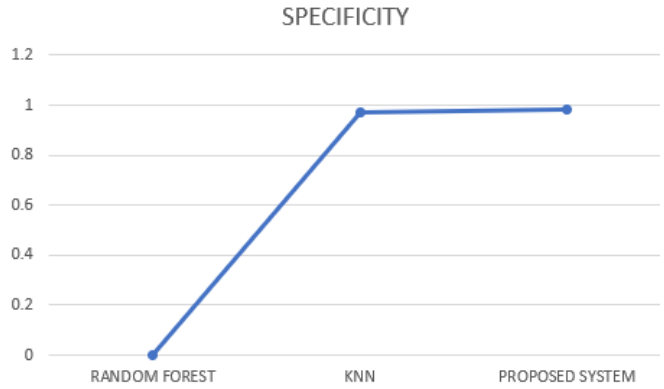


**FIG 8:** SPECIFICITY COMPARISION

## VI. Conclusion

The Data mining, best concept of machine learning algorithm is used for credit card fraud in this proposed system is proposed. Then, the number of standard models such as NB, SVM, and DL is used for evaluation terms. The credit card data is available in publically, it is used for evaluation that is, use the standard models and hybrid models. The hybrid models such as AdaBoost and majority voting, this models are combination methods, also. The MCC metrics are only calculates the performance measures and it takes the account, and it predicts the true or false outcomes of credit card transaction. The best MCC score majority voting is used the majority voting. The financial institution gives the credit card data set for evaluation.In this proposed concept is enhanced to online learning models. Use the online learning to enable the rapid detection of credit card fraud. The proposed system is help to detect and before prevent the fraudulent transaction and activities, so to reduce the number of losses in financial industry.

## References

[1]. Leite R. A., Gschwandtner T., Miksch S., Kriglstein S., Pohl M., Gstrein E., Kuntner J.: EVA: Visual Analytics to Identify Fraudulent Events. IEEE Transactions on Visualization and Computer Graphics Vol. 24, Issue: 1, (2018)
[2]. Randhawa K., Loo1 C. K., Seera M., Lim C. P., Nandi A. K.: Credit Card Fraud Detection Using AdaBoost and Majority Voting. IEEE Access, (2018)
[3]. Lin H., Gao S., Gotz D., Du F., He J., Cao N.: RCLens: Interactive Rare Category Exploration and Identification, IEEE Transactions on Visualization and Computer Graphics, (2017)
[4]. Chen Y., Wu C.: Big Data-Based Fraud Detection Method for Financial Statements of Business Groups, 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), (2017)
[5]. Zeager M. F., Sridhar A., Fogal N., Adams S., Brown D. E., Beling P. A.: Adversarial Learning in Credit Card Fraud Detection, International Conference on Systems and Information Engineering Design Symposium (SIEDS), (2017)
[6]. Awoyemi J. O., Adetunmbi A. O., Oluwadare S. A.: Credit Card Fraud Detection Using Machine Learning Techniques: a Comparative Analysis, International Conference on Computing Networking and Informatics (ICCNI), (2017)
[7]. Mubarek M., AdaliE.,: Multilayer Perceptron Neural Network Technique for Fraud Detection, International Conference on Computer Science and Engineering (UBMK), (2017)
[8]. Malini N., Pushpa M.: Analysis on Credit Card Fraud Identification Techniques Based on KNN and Outlier Detection, Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), (2017)
[9]. El Bouchti A., Chakroun A., Abbar H., Okar C.: Fraud Detection in Banking Using Deep Reinforcement Learning, Seventh International Conference on Innovative Computing Technology (INTECH), (2017)
[10]. ] Vergara L., Salazar A., Belda J., Safont G., Moral S., Iglesias S.: Signal Processing on Graphs for Improving Automatic Credit Card Fraud Detection, International Carnahan Conference on Security Technology (ICCST), (2017)
[11]. Rizki A., Surjandari I., WayastiR.,: Data Mining Application to Detect Financial Fraud in Indonesia's Public Companies, 3rd International Conference on Science in Information Technology (ICSITech), (2017)
[12]. Rahmawati D., Sarno R., Fatichah C., Sunaryono D.: Fraud Detection on Event Log of a Bank Financial Credit Business Process Using Hidden Markov Model Algorithm, 3rd International Conference on Science in Information Technology (ICSITech), (2017)
[13]. Liu J., Tian J., Cai Z., Zhou Y., Luo R., Wang R.: A Hybrid Semi-Supervised Approach for Financial Fraud Detection, International Conference on Machine Learning and Cybernetics (ICMLC), (2017)
[14]. Saia R., Boratto L., Carta S.,: Multiple Behavioral Models: A Divide and Conquer Strategy to Fraud Detection in Financial Data Streams, 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), (2015)
[15]. Modic D., Anderson R.: It's All Over But the Crying: The Emotional and Financial Impact of Internet Fraud, IEEE Security & Privacy, (2015)
[16]. Li X., Xu W., Tian X.: How to Protect Investors? A GA-Based DWD Approach for Financial Statement Fraud Detection, IEEE International Conference on Systems, Man, and Cybernetics (SMC), (2014)
[17]. Humaid E., Barhoum T.,: Water Consumption Financial Fraud Detection: A Model Based on Rule Induction, International Conference on Information and Communication Technology (PICICT), (2013)
[18]. Panigrahi P.: A Framework for Discovering Internal Financial Fraud Using Analytics, International Conference on Communication Systems and Network Technologies (CSNT), (2011)

[19]. Kou,Y., Lu,C., Sirwongwattana, S., Huang,Y. Survey of Fraud Detection Techniques. *International Conference on Networking, Sensing & Control*, 749-754, 2004.
[20]. Delamaire, L., Abdou, H., Pointon, J., Credit Card Fraud and Detection Techniques: A Review. *Banks and Banks Systems*, 4(2), 57-68, 2009.
[21]. Rocha, B.C., Sousa Junior, R, Identifying Bank Frauds Using Crisp-DM and Decision Trees. *International Journal of Computer Science & Information Technology*, 2(5), 162-169, 2010.
[22]. Witten, I.H., Frank, E., Hall, M.A., Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.), Morgan Kaufmann, 2011.
[23]. J. Ross Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, Inc., 1993.

**AUTHOR'S PROFILE:**

**[1]. GURRAM SAI KUMAR** Pursuing his B.Tech Final Year in Computer Science and Engineering Department at Jyothishmathi Institute Of Technology & Science, Karimnagar ,T.S., India.

[2]. **MADALA VENKAIAH NAIDU** Pursuing his B.Tech Final Year in Computer Science and Engineering Department at Jyothishmathi Institute Of Technology & Science, Karimnagar ,T.S., India.

**[3]. Dr. MADUGULA SUJATHA** Working as Associate Professor in Department of Computer Science and Engineering at Jyothishmathi Institute Of Technology & Science, Karimnagar ,T.S., India. Her Interested areas are Data Mining, Machine Learning and Artificial Intelligence.

1Gurram Sai Kumar. " Credit Card Fraud Detection System Based On Machine Learning Techniques." IOSR Journal of Computer Engineering (IOSR-JCE) 21.3 (2019): 45-52.