International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

The Effect of Best First Search Optimization on Credit Card Fraudulent Transaction Detection

Sapan Sahu, Shikha Agrawal, Raju Baraskar

Abstract*: *In the digital world, Recently growth of online shopping site for purchasing clothes, electronic items, glossary etc and online transaction for transfer money is increasing day by day . At the same time, criminals have become able to doing fault and earning money through wrong ways .that's why fraud grows. With the development of Machine Learning in the field of Computer Science and Engineering, its application in the different domain also in fields like Medical, Marketing, Telecommunication, finance, etc. The reason for the popularity of Machine Learning in these domains is due to its high accuracy prediction. That's why over many years, machine learning has been used in fraud detection. With the advancement of technology in online transactions, fraud is the greatest issue for businesses and has become difficult to recognize than the traditional form of this crime. Historically, the area of Fraud Detection is interrelated to Data Mining & Text Mining. Due to the sudden growth of fraud whose outcome is loss of trillions of rupees worldwide every year, various modern techniques in detecting fraud were proposed that are progressed without interruption and applied to many business fields. Bank frauds worth ₹2.05 trillion happened in the last 11 years, among which there were overall 53,334 fraud issues in the usage of RBI data. The principle purpose behind this write up is to review different methods in identifying frauds corresponding to the unusualness in the transactions. The supervised and unsupervised machine learning algorithms will be used to identify fraud and the best first search optimization will be analyzed to compare both results, i.e., before and after optimization.

Keywords* : *Machine-Learning; Fraud Detection, Supervised Learning Algorithm, Unsupervised Algorithm, Best First Search

I. INTRODUCTION

The word fraud means taking services, goods & currency in a wrong way & raising a difficulty on the world today. The deals of Frauds with cases connecting mainly criminal purposes that are difficult to recognize. Fraud detection is one of the major challenges for most organizations particularly in banking, finance, retail, and electronic commerce .Fraud detection is a top priority for banks and financial institutions, which can be addressed through machine learning.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Sapna Sahu*, department of computer science, university of institute of technology Rajiv Gandhi proudyogiki vishwavidyalaya , Bhopal, India. Email: sapnasahu965@gmail.com
Dr shikha agarawal, department of computer science, university of institute of technology Rajiv Gandhi proudyogiki vishwavidyalaya , Bhopal, India**.** Email: shikha@rgtu.net
Dr. Raju Baraskar, department of computer science, university of institute of technology Rajiv Gandhi proudyogiki vishwavidyalaya , Bhopal, India. Email: rajubaraskar@rgtu.net

1.1 Methods of Fraud Detection

Fraud detection is an extremely complex function that must be performed where there is no organization capable of guaranteeing a 100% satisfaction rate. It is likely that all existing methods may involve fraudulent transactions and not guarantee results. Consider the features of a superior fraud detection method:

a) Must be capable to accurately recognize fraud.
b) Fraud cases must be detected quickly.
c) In any case, a real transaction cannot be considered a fraud.

II. BACKGROUNDS

2.1 Machine Learning Algorithm

Machine Learning tends to train machines on how to handle the data more effectively. Many times after showing the data, we never understand the pattern or remove information from the data. In this case, machine learning is applied with a lot of datasets available; the requirement for machine learning is on the rise. The goal of a machine learning algorithm is to learn automatically without any human intervention. Machine learning is the core subarea of Artificial Intelligence as learning is at the core of intelligence.

1 pattern reorganization
2 Games
3 Data mining
4 Robotic
5 Language Processing

2.2 Learning

The process of the convert to experiment into experience or knowledge is known as learning. Learning can be broadly classified into three categories, as indicated below, depending on the nature of the learning data and the intervention between student and the environment. Machine learning methods have supervised and unsupervised. In supervised Machine Learning, we first train and then we test . For training labeled data is used. Another one is supervised ML. Both types of learning have a variety of approaches such as neural network, decision tree , Bayesian classifier , etc

1) Supervised Learning
2) Unsupervised Learning

2.3 Supervised Algorithm

Supervised learning such as the name indicates the presence of an administrator as a teacher (as shown in figure 1).
Really supervised learning is learning in which we teach and train the machine using well-labeled data, which means that some data is already labeled with the correct answer.

Supervised automatic learning algorithm: it is possible to apply what has been learned in the past from new data using coding.

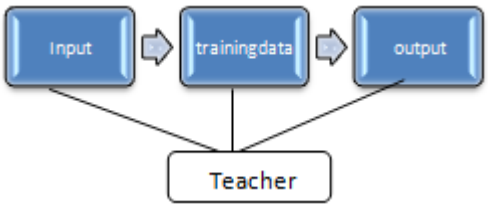

Figure 1: Supervised Algorithm

Example: predict future events These algorithms make predictions on a given set of data samples. The entry contains information on training and presents tags known as spam or non-spam. A model is prepared through a training process in which it makes forecasts and a correction is made when the forecast is incorrect. This training process continues until the model reaches the level of precision required in the training data (as shown in Figure 2.2). types are - classification and regression. Example algorithm: Logistic Regression and Support Vector Machine

• **Regression:** A regression problem is when the yield variable is a real value, for example, "dollars" or "weight". & salary.

• **Classification:** A classification problem is when the yield variable is a class, its result always Boolean values .for example fraud or not fraud, decease and not deceased . Classification algorithms are mainly used in a fraud detection procedure

2.4 Unsupervised Algorithm

The unsupervised machine learning algorithm is used when training information is not classified or labeled.

Unsupervised learning is automatic training that uses information that is not classified or labeled and that allows the algorithm to act on that information without a teacher. Here the machine's task is to group unordered information based on similarities, patterns, and differences without any prior data formation.

Unsupervised learning is the point where you only have input data (X) and there are no corresponding yield variables.

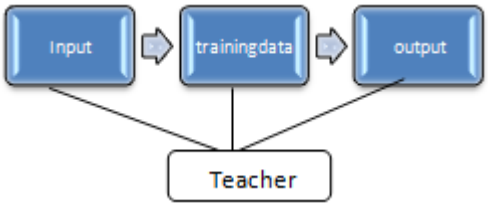

Figure 2: Unsupervised Algorithm

Problems of unsupervised learning can be grouped into clustering and association problems.

• **Clustering**: a grouping problem is a point where you want to find native sets in the data, for example by grouping customers based on the buying behavior.

• **Association**: a problem of learning association rules is the point where you want to find rules that explain most of the data, for example, people who buy X continue to buy Y.

III. MACHINE LEARNING TOOL

It's simply to use interface and researchers can without difficulty to use them because of free open-source software. Popular is data mining and machine learning tool are WEKA, ORANGE, and ANACONDA, etc. We used weka and anaconda navigator.

3.1 Weka Tool Anaconda Navigator

The WEKA tool was created in New Zealand by the University of Waikato which uses the Java language and consists of several data mining algorithms. This tool performs data mining activities by collecting machine learning techniques applied directly to data sets . The WEKA tool provides tools for data preprocessing, classification, grouping, regression, visualization and association rules. This tool is an open-source software in which WEKA uses the ARFF file format that identifies different things using special tags.

3.2 Anaconda Navigator

Anaconda Navigator is a free desktop graphical user interface for python. It includes condo packages and allows to launch the application and easily manage , environments and channels without the use of command-line commands. It is a cross-platform GUI available for Windows, macOS, and Linux. It is written in python.

IV. RELATED WORK

In this paper, analysts Anuruddha Thennakoon (2019) et al. [17] centers around four principle extortion openings regarding genuine exchanges. Every extortion is tended to through a progression of AI models and the best strategy is chosen through an assessment. This assessment gives a total manual to select an ideal calculation as for the kind of misrepresentation and we show the assessment with a satisfactory presentation measure. Another significant key zone tended to in our venture is the identification of charge card extortion continuously. For this, they utilize prescient examination performed by actualized AI models, they can infer that there is a major effect when utilizing resampling systems to accomplish similarly higher classifier execution. AI models that have caught the four extortion models (hazardous client focus, obscure web address, ISOResponse code, the exchange above $ 100) with the most astounding precision rates are LR, NB, LR, and SVM. Moreover, the models demonstrated exactness paces of 74%, 83%, 72%, and 91% individually. Since the created AI models have a medium degree of precision, they center around improving gauge levels to procure better conjectures and an API module to choose whether a given exchange is credible or false.

According to Jianrong Yao (2018) et al.[7] ,Fraud in financial statements has become a complicated problem for both public and government auditors, so different data mining methods have been used to detect fraud in financial statements to support decisions.

Interested parties The aim of this study is to suggest an Improved financial fraud detection model combining feature selection and automatic learning classification. The author stated that the random forest has overcome four other methods: SVM, DT, ANN, LR. Where they find RF does the best work.

According to Pedro Shiguihara-JuáreZ (2018)et al.[9] Supervised automatic learning techniques for fraud detection have been applied. They proposed a method to generate a probabilistic graphical model for fraud detection, using domain-related restrictions. The reported result is that they have achieved an accuracy of 99.272% and we have outperformed the other basic techniques of probabilistic graphical models.

As indicated by Tanupriya Choudhury et al [17], A model for foreseeing whether the submitted Mastercard is false or not dictated by in excess of 150 traits for each guest, which have recently been prepared with an informational index. Since the information utilized for the motivations behind this record were exceptionally lopsided, distinctive inspecting strategies were utilized to adjust the preparation information. The analysis demonstrates great execution alongside the exactness of extortion identification. The most astounding territory under the bend (AUC) is accomplished through strategic relapse (0.9375) with a precision of 99.75 percent that is prepared with adjusted information beneath the example.

As indicated by Priyanka S. (2017)et al.[11] The bank division plays the fundamental capacity of in the nation's economy. Clients are the bank's fundamental resource. Thusly, it's important to concentrate on the issues that banks face. So they are taking a shot at client maintenance and extortion recognition. In this work, a fake neural system calculation is actualized for order purposes. For the grouping, they utilized two arrangements of information, information from bank clients and German credit information. The counterfeit neural system is utilized for order purposes. This calculation gives 72% and 98% precision to information set1 and information set2 separately. From the outcomes, it demonstrates that the created model works productively for two informational collections.

According to Riya Roy Thomas George (2017)et al.[12] they focus on detecting automotive fraud using the machine learning technique In the study, select a model of over 500 data & the data is divided into train and test data. You can see it with respect to algorithms, Decision Tree and Random Forest Algorithms; They perform improved than the naive Bayes.

According to Wen-Fang YU et al.[16] A credit card fraud detection model that uses a typical value detection mining based on the sum of the distance in the detection of credit card fraud and proposes this survey procedure & its experimental procedure. lastly, this resulting process is accurate to predict fraudulent transactions through an atypical emulation experiment of mining the credit card data set of a particular commercial bank. The research shows with the intention of typical mining can identify credit card fraud improved than cluster based anomalies when anomalies are much smaller than common data.

Author	Year	Application domain	Technique	Data set
Anuruddha Thennakoon	2019	Real-time Credit Card Fraud Detection	Supervised	Credit card transaction dataset
JIanrong Yao, Jie Zhang, Luwang	2018	Financial statement fraud	Hybrid Random forest	Financial statement dataset
Hongu Wang, ping Zhu Xueqiang Sujuan Qin	2018	Credit Card Fraud Detection	Clustering Set partitioning	Credit Card transaction Dataset
Deepti Dighe	2018	Credit Card Fraud transaction	Neural network	Balanced dataset
Richard A. Bauder Taghi M Khoshgoftaar	2017	Health care	Hybrid Supervised Unsupervised	Medicare dataset
Sudan chen	2016	Fraudulent financial statement	Hybrid	Financial Statement dataset
Wen-Fang YU , Na Wang	2009	Credit Card Fraud Detection	Outlier detection mining	commercial bank dataset

Table1. Summarize table of literature survey.

V. METHODOLOGY

5.1 ARCHITECTURE OF PROPOSED WORK

The overall process of analysis of best search optimization credit card fraudulent transaction detection is based on the following six steps (Show in figure 3.)

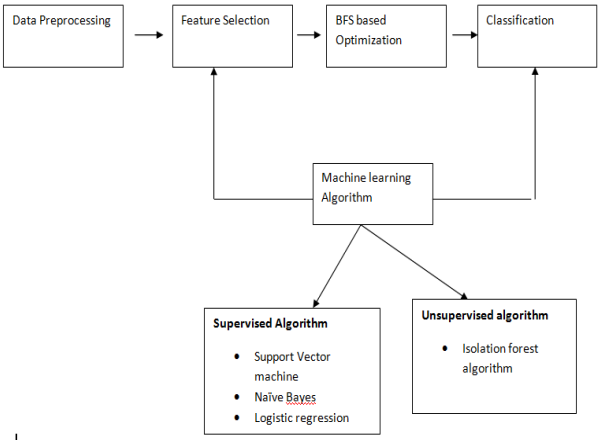

Figure3 :- ARCHITECTURE Of PROPOSED WORK

1. Data collection
2. Data pre-processing
3 Feature Selection
4 BFS Based Optimization
5 The Classification Using Machine Learning Algorithm

Step1 . Data collection

The data set sourced from the Kaggle website [20]. We use a European data set with 284807 transactions, which is for two days in 2013. The above data set has 492 fraudulent transactions that are labeled with 1 and others with 0. In fact, the percentage of fraudulent and non-fraudulent transactions is 0.17%, which shows that the data set is highly unbalanced. Based on customer privacy, the original features of this dataset are not available and contain 28 features that are the result of the PCA mapping feature plus two unassigned features called transaction time and amount.

Step2 Data Preprocessing

Once the data has been selected, it must be pre-processed with the indicated steps:

1. Format the data to fitting for ML (organized way)
2. Delete the data to eliminate not complete parameters.
3. Additional data sampling to reduce the execution time of algorithms and memory requirements.

Data cleanup in this phase implies that filtering is based on the following variables:

Not enough data

The amount of data necessary for the ML algorithms can vary from thousands to millions, depending on the complexity of the problem and the algorithm chosen.

Non-representative data

The selected sample must be an accurate representation of all data since unrepresentative data can train an algorithm so that it does not generalize well in the new test data.

Lower quality data

Abnormal values, errors, and noise can be eliminated to better adapt to the model. Missing functions, such as the age of 10% of the public, can be completely ignored or the average value for the missing component can be assumed.

Step 3 Feature Selection

The feature selection is an interesting technique that can be launched before the data classification task. Feature Selection is a technique which is used to reduce the dimensionality of data or eliminate the irrelevant features and improve the predictive accuracy. Attribute selection involves searching through all possible combinations of attributes in the data to find which subset of attributes works best for prediction. The Best first search is an important AI search strategy that allows backtracking along the search path. The attributes selected through Best First Search based CFS were V2, V3, V4, V5, V7, V9, V10, V11, V12, V14, V16, V17, V18, V21 out of 30 attributes. For feature selection & optimization in this research work, the BFS Technique is used Best First Search algorithm.

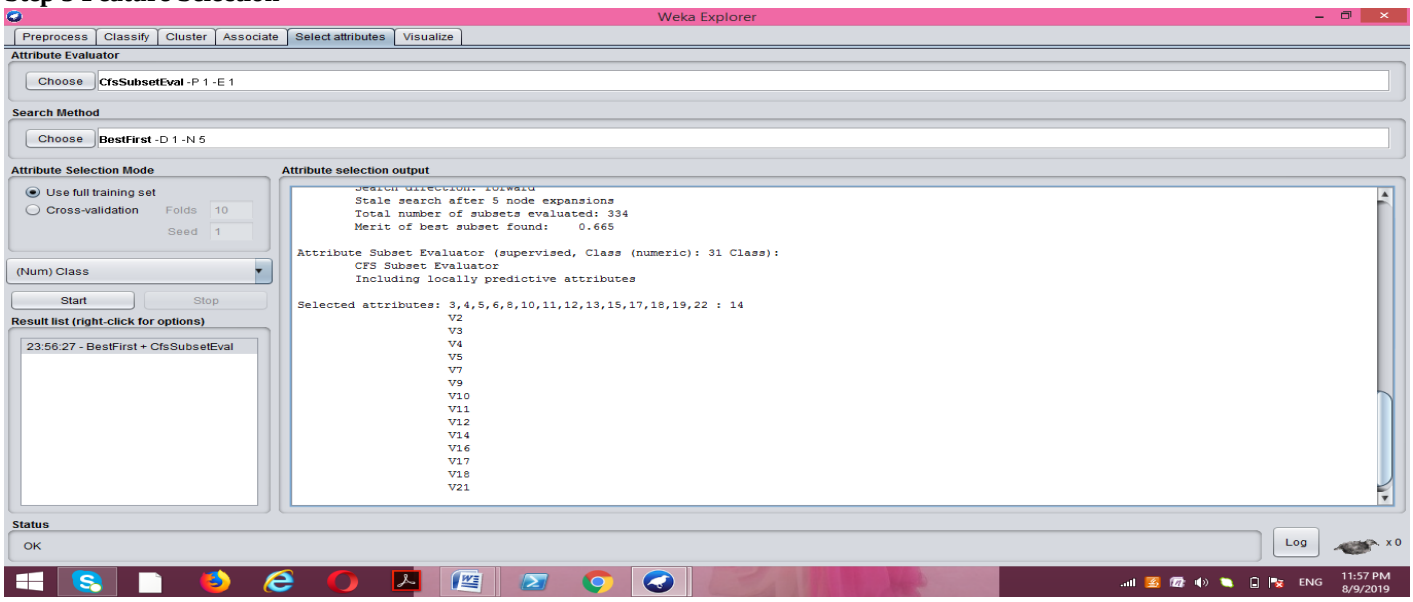

Figure 4:- Select attribute

Step 4 Best First Search Optimization

Best-first search in its most common form is a simple heuristic search algorithm. "Heuristic" here relates to a general solving the problem rule or a number of rules that cannot secure the better solution or still any solution but serves as useful guidance for solving the problem[23]. A graph-based search algorithm is a Best-first search (Dechter and Pearl, 1985), meaning that the search space could be described as a series of nodes linked by routes. Searches the space of feature subsets by greedy hill-climbing improved with a rollback capacity. Setting the number of repeated non-improving nodes allowed controls the level of rollback done. The best first search is able to start with the null set of features & find forward, or start with the full set of features & search backward, or start at any location and search in both directions (by considering all possible single attribute additions and deletions at a given point).

Algorithm of the best first search

Best-first search[24] in its most basic form consists of the following algorithm (adapted from Pearl, 1984): The algorithm Best First Search is represented here in pseudo-code:

```
Begin
Open: = [start];
Closed: = [ ];
While open! = [ ] do
  Begin
  Remove leftmost state from open, call it x;
  if x is a goal then return(success)
  else
    Begin
    Generate children of x; put x on closed;
    Eliminate children of x
      Already in open or closed; put remaining children on right end of open;
    End
  End;
```

Step5 The Classification Using Machine Learning Algorithm

Classification is the process of forecasting the type of data points provided. Classes are sometimes referred to as goals / labels or categories. Predictive classification modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y). The classification belongs to the category of supervised learning in which the objectives also provide input data. There are many applications in classification in many areas, such as credit approval, medical diagnosis, objective marketing, etc.

Here we have the types of classification algorithms in Machine Learning:

1. Logistic Regression
2. Support Vector Machines
3. Naive Bayes Classifier

K-Fold Cross-Validation

Cross-validation of K-Fold is where a given data set is divided into a number K of sections / folds in which each fold is used as a test set at a given point [21]. Take the 10-fold cross-validation scenario (K = 10). Here, the data set is divided into 10 folds. In the first iteration, the first fold is used to test the model and the rest is used to train the model. In the second iteration, the second fold is used as a test set, while the rest acts as a training set. This process is repeated until each fold of the 10 folds has been used as a test set.

5 EVALUATION METRICS

In machine learning, we train the model with the training data, and then we check the generalization capability of the model. In simple terms, we examine how the model performs when tested on data that was unseen. So how do we measure the performance of the model? We use evaluation metrics for evaluating the performance of the model depending on the nature of the problem (whether it is a regression or classification). In this section, we will only discuss the evaluation metrics related to the confusion matrix.

5.1 CONFUSION MATRIX

This research, a confusion matrix is used to find overall performance and get model building time. In this Experiment after performing classification, it generates a confusion matrix on the basis of the parameter which is –to the classification problem.

		Predicted Class	
		N	P
Actual Class	N	TP	FP
	P	FN	TP

Figure5 : Confusion Matrix l

It is the most commonly used evaluation metrics in predictive analysis mainly because it is very easy to understand and it can be used to compute other essential metrics such as accuracy, recall, precision, etc. It is an NxN matrix that describes the overall performance of a model when used on some dataset, where N is the number of class labels in the classification problem. For binary classification, we have a 2x2 confusion matrix (as shown in figure3)

A confusion matrix is composed of statistics such as Negative (N) , Positive (P) , True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) which are calculated using the combination of actual and predicted values.

True Positive (TP) is a case where the actual value was positive (e.g., fraud) and the predicted value is also positive.

False Positive (FP) is a case where the actual value was negative (e.g., normal) but the predicted value is positive.

True Negative (TN) is a case where the actual value was negative (e.g., normal) and the predicted value is also negative.

False Negative (FN) is a case where the actual value was positive (e.g., fraud) but the predicted value is negative.

VI. RESULT

We used 67% of the data is used for training and 33% used for the testing set. Data was unbalanced by using a Best First optimization technique. So we used Accuracy, F1 Score, Precision, and Recall. .

Result Analysis of Supervised Machine learning. Algorithm Performance Evolution of proposed methodology without optimization Figure (7)

Table2. Show Results obtained for 3 supervised machine learning based classifier with input given without applying optimization.

Model	Accuracy	Precision	Recall	F1 score
SVM	0.97	0.95	0.98	0.97
LR	0.99	0.81	0.57	0.67
NB	0.97	0.058	0.84	0.10

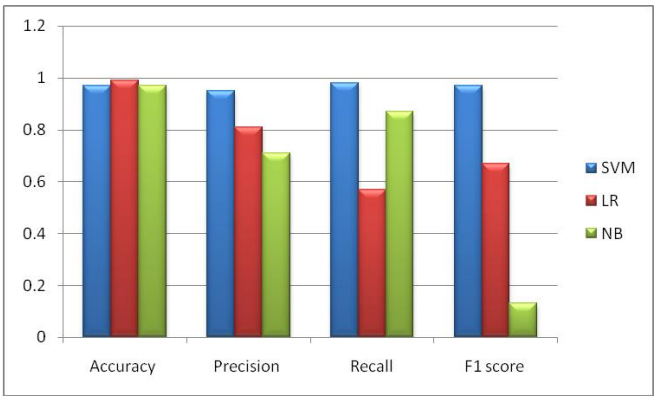

Figure 6:- Results of Supervised Learning Without *optimization*

Table 3 Show Results obtained for 3 supervised machine learning based classifier with input given with applying optimization.

Model	Accuracy	Precision	Recall	F1 score
SVM	0.97	0.95	0.98	0.97
LR	0.99	0.81	0.57	0.67
NB	0.97	0.058	0.84	0.10

Performance Evolution of proposed methodology with optimization Figure (7)

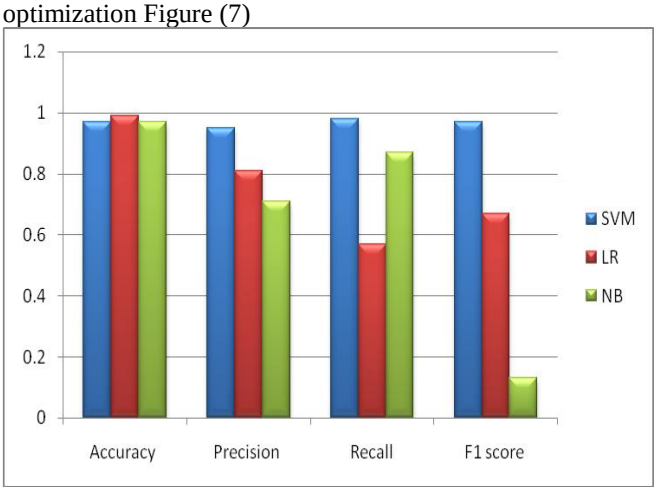

Figure 7:- Results of Supervised Learning With optimization

Based on classification accuracy, precision, recall, and F1 score the models were evaluated. CFS with Best first search was applied in this proposed method. Using this model, the prediction accuracy of 99% is achieved. From table 2 & table 3 . it is clear that in both the cases before & after optimization LR outperforms SVM & NB . ML algorithm for credit card fraud detection .Also, we changes were found in the accuracy & after applying BFS as before & after applying BFS as an optimization technique.

Table 4 Show Results obtained for unsupervised machine learning based classifier with input given without applying optimization

Model	Accuracy	Precision	Recall	F1 score
Isolation Forest	0.99	0.02	0.02	0.02

Performance Evolution of proposed methodology without optimization Figure (8)

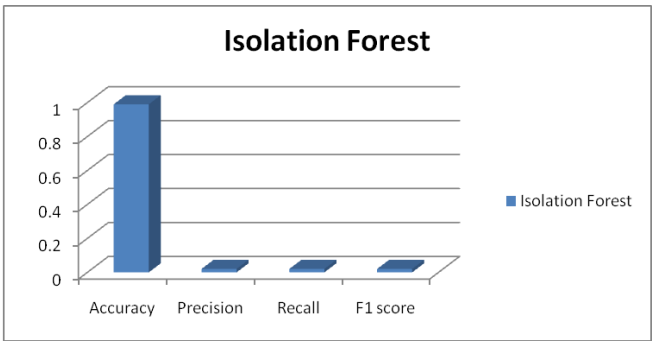

Figure 8:- Results of an Unsupervised Learning Without optimization

Table 5 Show Results obtained for unsupervised machine learning based classifier with input given without applying optimization

Model	Accuracy	Precision	Recall	F1 score
Isolation Forest	0.99	0.02	0.02	0.02

Performance Evolution of proposed methodology with optimization Figure (8)

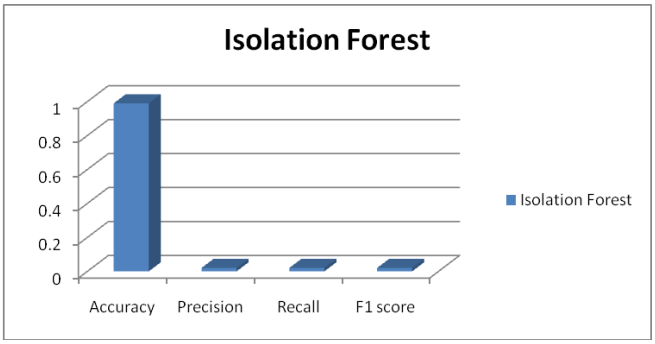

Figure7:- Results of an Unsupervised Learning With optimization

Table3:- RESULT SUMMARY

Classifier	Best Search Optimization	Accuracy	Precision	Recall	F1 score
SVM	Before	0.97	0.95	0.98	0.97
	After	0.97	0.95	0.98	0.97
LR	Before	0.99	0.81	0.57	0.67
	After	0.99	0.81	0.57	0.67
NB	Before	0.97	0.058	0.84	0.10
	After	0.97	0.071	0.84	0.10
Isolation Forest	Before	0.99	0.020	0.02	0.02
	After	0.99	0.020	0.02	0.02

Summarizes the results of the thesis, discusses Machine learning and best first search optimization. Machine learning approaches give good classification result. In this chapter, we do experiment and analysis the performance of different classifier of classification accuracy and after optimization on credit card fraud detection data set. we analysis of the effect of optimization of best-first search on credit card data set and using various machine learning algorithm and their compare result .All operation performed on a python environment and graph are plotted using Microsoft excel for better visualization of a compared result.

VII. CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

The main goal of our research was to identify fraudulent transactions and we achieved a decent accuracy on the dataset. Input data of a highly unbalanced and 30 attributes were present in the dataset. To tackle this problem, we used the best search algorithm through Weka tool which was used for optimizing the unbalanced dataset. It selected 14 out of 30 attributes and so it was advantageous. In the last results before and after optimization are compared. It is shown that there is a minor difference in after optimization reading. We analyzed that the best first search method isn't good for optimization. Due to the imbalance in the dataset, the best first search algorithm doesn't perform well in both supervised and unsupervised learning's. It is shown that unsupervised learning is not performed up to the mark . The fraud transaction in a supervised and unsupervised algorithm is detected in this research. From the experimental result, it has been concluded that Logistic regression has an accuracy of 99% , SVM shows the accuracy of 97% while naïve Bayes is 97% accurate but the best results are obtained by logistic regression with precise accuracy of 99%. The results obtained thus conclude that logistic regression shows the most precise and high accuracy of 98.6% in detecting credit card fraud. The Logistic regression algorithm will perform better with a larger number of training data. The SVM algorithm still suffers from the imbalanced dataset problem and requires more preprocessing to give better results as seen previously, it could have performed better if more preprocessing had been done on the data

7.2 FUTURE WORK

This chapter summarizes the main results of the thesis, openly discusses its issues and presents future research directions. However, this research took only the numerical variables as input and did not include the original features and their background information. In the future, a study can be done on attribute name which isn't given in European data set.

In the future, different classification can be performed by using an algorithm such as greedy search optimization and genetic algorithm to get better results. The data is collected in the research work is only small dataset in future data could be collected as a large data set and apply an algorithm for better prediction.

ACKNOWLEDGMENT

I heartily thank our research guide, Dr. Shikha Agrawal , Dr. Raju Baraskar ,Department of computer science for guidance and suggestions during this project work.

REFERENCES

1. J. O. Sinayobye, F. Kiwanuka and S. Kaawaase Kyanda, "A State-of-the-Art Review of Machine Learning Techniques for Fraud Detection Research," 2018 IEEE/ACM Symposium on Software Engineering in Africa (SEiA), Gothenburg, 2018, pp. 11-19.
2. Ayon Dey / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179 ..... Engineering and Technology, ISSN: 2319 1163, Volume 03, Special.
3. Yufeng Kou, Chang-Tien Lu, S. Sirwongwattana and Yo-Ping Huang, "Survey of fraud detection techniques," IEEE International Conference on Networking, Sensing and Control, 2004, Taipei, Taiwan, 2004, pp. 749-754 Vol.2.
4. S. Rajora et al., "A Comparative Study of Machine Learning Techniques for Credit Card Fraud Detection Based on Time Variance," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1958-1963.
5. N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp.255-258. doi: 10.1109/AEEICB.2017.7972424
6. "Supervised and Unsupervised Machine Learning Algorithms," Machine Learning Mastery, 22-Sep-2016. [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. [Accessed: 08-Aug-2019].
7. J. Yao, J. Zhang, and L. Wang, "A financial statement fraud detection model based on hybrid data mining methods," 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 2018.
8. H. Wang, P. Zhu, X. Zou and S. Qin, "An Ensemble Learning Framework for Credit Card Fraud Detection Based on Training Set Partitioning and Clustering," 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, 2018, pp. 94-98.

9. P. Shiguihara-Juarez and N. Murrugarra-Llerena, "A Bayesian Classifier Based on Constraints of Ordering of Variables for Fraud Detection," 2018 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI), 2018.
10. L. Peng and R. Lin, "Fraud Phone Calls Analysis Based on Label Propagation Community Detection Algorithm," 2018 IEEE World Congress on Services (SERVICES), San Francisco, CA, 2018, pp. 23-24.
11. P. S. Patil and N. V. Dharwadkar, "Analysis of banking data using machine learning," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 876-881.
12. R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, 2017, pp. 1-6.
13. L. Lei, "Card Fraud Detection by Inductive Learning and Evolutionary Algorithm," 2012 Sixth International Conference on Genetic and Evolutionary Computing, Kitakushu, 2012, pp. 384-388.
14. Tao Guo and Gui-Yang Li, "Neural data mining for credit card fraud detection," 2008 International Conference on Machine Learning and Cybernetics, Kunming, 2008, pp. 3630-3634.
15. R. R. Popat and J. Chaudhary, "A Survey on Credit Card Fraud Detection Using Machine Learning," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018.
16. W. Yu and N. Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum," 2009 International Joint Conference on Artificial Intelligence, Hainan Island, 2009, pp. 353-356.

a. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2019, pp. 488-493.

17. S. Dhankhad, E. Mohammed and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, 2018, pp. 122-125

AUTHORS PROFILE

Ms. Sapna Sahu, is current pursuing Dual Degree Integrated Post Graduation Programme (B.E. + M.TECH) in Computer Science and Engineering from University Institute of Technology RGPV, Bhopal (M.P.),India. Her research areas are Fraud Detection, Credit Card Fraud Detection using machine learning . She has done major project in Android based web appliction i.e. Mobile Location tracking

Dr Shikha Agrawal is an Assistant Professor in Department of Computer Science & Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (MP) India. She obtained B.E., M.Tech. and Ph.D in Computer Science & Engineering from Rajiv Gandhi Proudyogiki Vishwavidalaya Bhopal. She has more than fifteen years of teaching experience. Her area of interest is Artificial Intelligence, Soft Computing and Particle Swarm Optimization and Database. She has published more than 40 research papers in different reputed international journals and 10 chapters. For her outstanding research work in Information Technology, she has been awarded as "Young Scientist" by Madhya Pradesh Council of Science and Technology, Bhopal. Her other extraordinary achievements include "ICT Rising Star of the Year Award 2015" in International Conference on Information and Communication Technology for Sustainable Development (ICT4SD - 2015), Ahmedabad, India and Young ICON Award 2015 in Educational category by Dainik National News Paper Patrika, Bhopal, India. She got recognition of IEEE as a Senior member. She is also member of various academic societies such as IEEE, ISTE,CSI,ACM & CSTA.

Dr Raju Barskar is an Assistant Professor in Department of Computer Science & Engineering at University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (MP) India. He obtained B.E. from SATI Vidisha(M.P.), then M.Tech. and Ph.D in Computer Science & Engineering from Maulana Azad National Institute of Technology, Bhopal. He has more than twelve years of teaching experience. His area of interest is Network Security, Vehicular Ad hoc networks, Image Processing, Parallel, Algorithm Pattern matching algorithm and Data Mining etc. He has published more than 20 research papers in different reputed international journals and 02 chapters. He is also member of various academic societies such as IEEE etc.