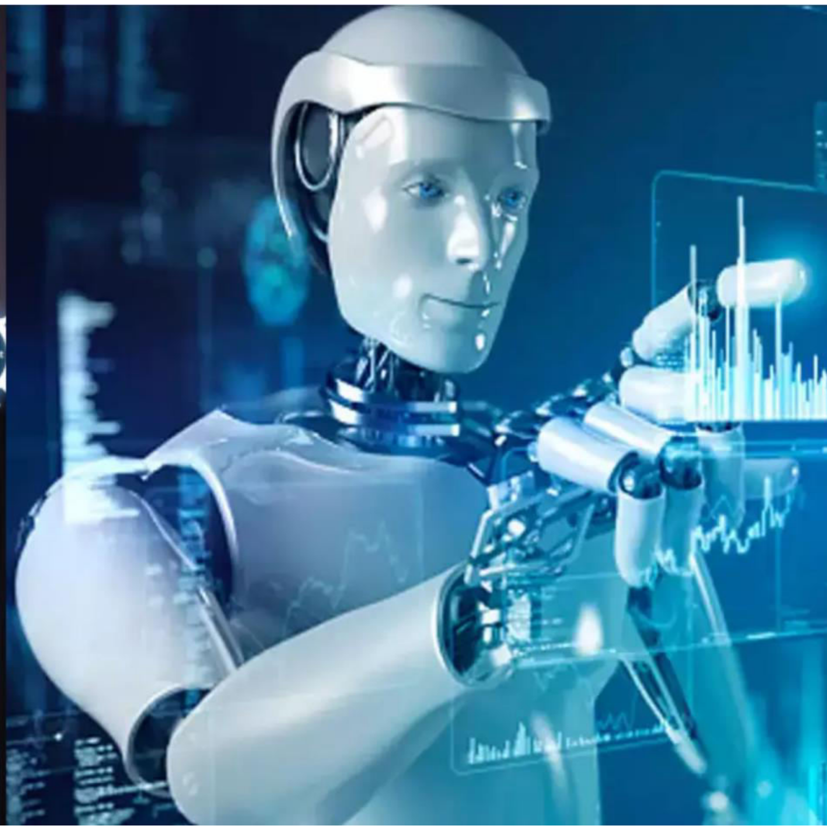
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405398|

Credit Card Fraud Detection using Machine Learning

Ramesh B E[1], Basantha Kumari[2], Shashank M[3], G S Chandu[4], Nithin S L[5], Tharun P[6]

Department of Computer Science and Engineering, SJMIT College, Chitradurga, Karnataka, India[1-6]

ABSTRACT: Credit card fraud detection is presently the most frequently occurring problem in the present world. This is due to the rise in both online transactions and e-commerce platforms. Credit card fraud generally happens when the card was stolen for any of the unauthorized purposes or even when the fraudster uses the credit card information for his use. In the present world, we are facing a lot of credit card problems. To detect the fraudulent activities the credit card fraud detection system was introduced. This project aims to focus mainly on machine learning algorithms. The algorithms used are random forest algorithm and the Adaboost algorithm. The results of the two algorithms are based on accuracy, precision, recall, and F1-score.

I. INTRODUCTION

Credit card fraud is a growing concern in the present world with the growing fraud in the government offices, corporate industries, finance industries, and many other organizations. In the present world, the high dependency on the internet is the reason for an increased rate of credit card fraud transactions but the fraud has increased not only online but also offline transactions. Though the data mining techniques[6] are used the result is not much accurate to detect these credit card frauds. The only way to minimize these losses is the detection of the fraud using efficient algorithms which is a promising way to reduce the credit card frauds. As the use of the internet is increasing[Figure.1], a credit card is issued by the finance company. Having a credit card means that we can borrow the funds. The funds can be used for any of the purposes.

II. RELATED WORK

New methods for credit card fraud detection with a lot of research methods and several fraud detection techniques with a special interest in the neural networks, data mining, and distributed data mining. Many other techniques are used to detect such credit card fraud. When done the literature survey on various methods of credit card fraud detection, we can conclude that to detect credit card fraud there are many other approaches in Machine Learning itself. The research on credit card fraud detection uses both Machine Learning[1][2] and Deep Learning algorithms[7]. In this section, we enhance the work done in two different points:(i) the methods that are readily available for fraud detection, and (ii) The techniques that are available to handle the imbalanced data. To handle the imbalanced data A[11] some of the techniques are available. They are (a) classification methods (b) sampling methods (c) resembling techniques. Here are some of the Machine Learning algorithms that are used for credit fraud detection are support vector machine(SVM), decision trees, logistic regression, gradient boosting, K-nearest neighbor, etc;

III. METHODOLOGY

Credit card fraud detection using machine learning involves a structured methodology that begins with defining the problem of identifying fraudulent transactions within highly imbalanced data, where fraud cases are rare compared to legitimate ones. The process starts with data collection, often from public or proprietary sources, followed by data preprocessing tasks such as cleaning, normalization, feature engineering, and handling class imbalance through techniques like oversampling or anomaly detection. Exploratory Data Analysis (EDA) helps uncover patterns and informs model selection, which may include traditional classifiers like logistic regression, ensemble methods such as Random Forest and XGBoost, or deep learning models like autoencoders and LSTMs for more complex patterns. Model evaluation focuses on precision, recall, F1-score, and ROC-AUC due to the critical impact of false negatives. After model tuning and cross-validation, the system is deployed into a real-time environment with continuous monitoring to adapt to evolving fraud tactics, ensuring robustness and efficiency over time.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)
Impact Factor 8.699
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|
Volume 14, Issue 5, May 2025
|DOI: 10.15680/IJIRSET.2025.1405398|

A. Data Visualization

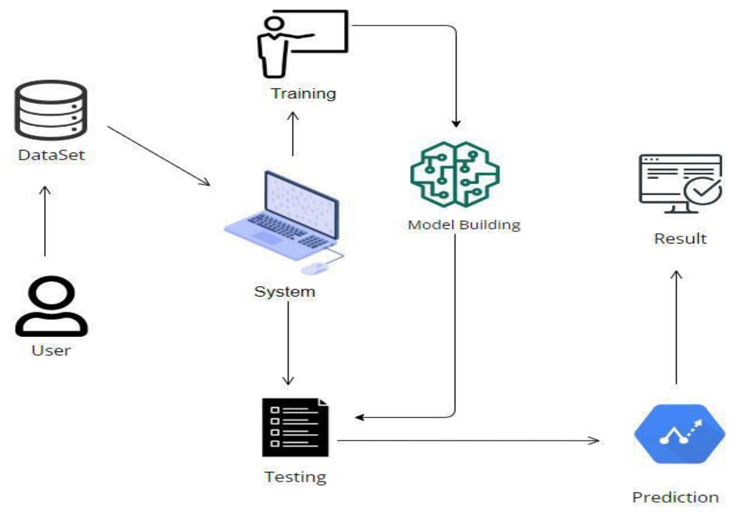


Fig. 1. Process Flow

The detailed architecture diagram for the credit card fraud detection system [Figure. 1.] includes many steps from gathering dataset to deploying model and performing analysis based on results. In this model we take the Kaggle credit card fraud dataset and pre-processing is to be done for the dataset. Now to prepare the model we have to split the data into the training data and the testing data. We use the training data to prepare the Random Forest and the Adaboost models. Then we develop both the models. Finally, the accuracy, precision, recall, and F1-score is calculated for bot the models. Finally the comparison of the credit card fraud transactions more accurately.

B. Random Forest Algorithm

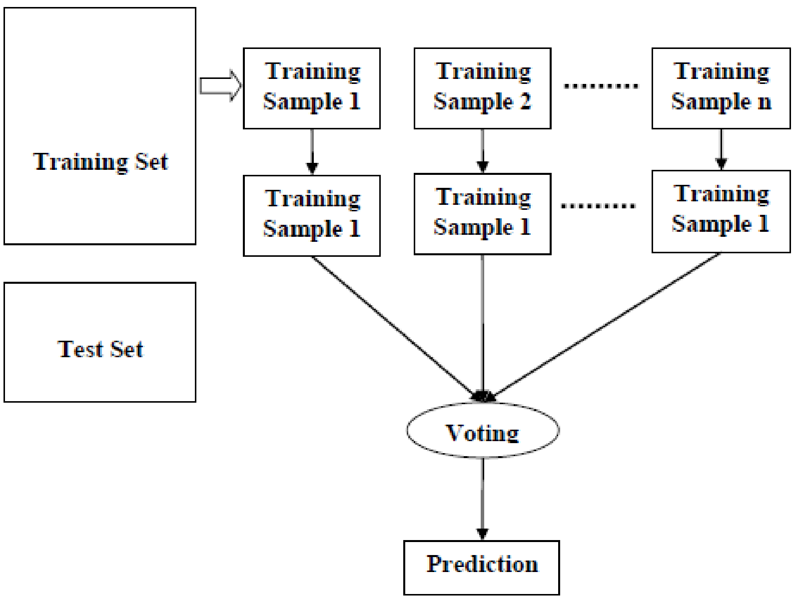


Fig. 2. Random Forest Algorithm

The Random Forest algorithm [Figure. 2.] is one of the widely used supervised learning algorithms. This can be used for both regression and classification purposes. But, this algorithm is mainly used for classification problems. Generally, a forest is made up of trees and similarly, the Random Forest algorithm creates the decision trees on the sample data and gets the prediction from each of the sample data. Then Random Forest algorithm is an ensemble method. This algorithm is better than the single decision trees because it reduces the over-fitting by averaging the result.

Steps for Random Forest Algorithm

1. Take the Kaggle credit card fraud dataset that is trained and randomly select some of the sample data.
2. Using the randomly created sample data now creates the Decision Trees that are used to classify the cases into the fraud and non-fraud cases.
3. The Decision Trees are formed by splitting the nodes, the nodes which have the highest Information gain make it as the root node and classify the fraud and non-fraud cases.
4. Now the majority vote is performed and the decision Trees may result in 0 as output which includes that these are the non-fraud cases.
5. Finally, we find the accuracy, precision, recall, and F1 -score for both the fraud and non-fraud cases.

Random Forest algorithm

Algorithm Random Forest :

To generate c classifiers:

For i=1 to c do
Randomly select the training data D with replacement to produce Di
Create a root node N containing Di and cell Build Tree(N)

End for Majority Vote

Build Tree(N)

Randomly select x% of all the possible splitting features in N
Select the features F that has the highest Information A gain for further splitting
Gain (T,X)=Entropy (T)-Entropy(T,X) Now to calculate the entropy we use,
$(S) = \sum(-Pi \log Pi)$
Create f child nodes For i=1 to f do
Set contents f N to Di Call Build Tree(Ni)
End for

End

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405398|

C. Adaboost Algorithm

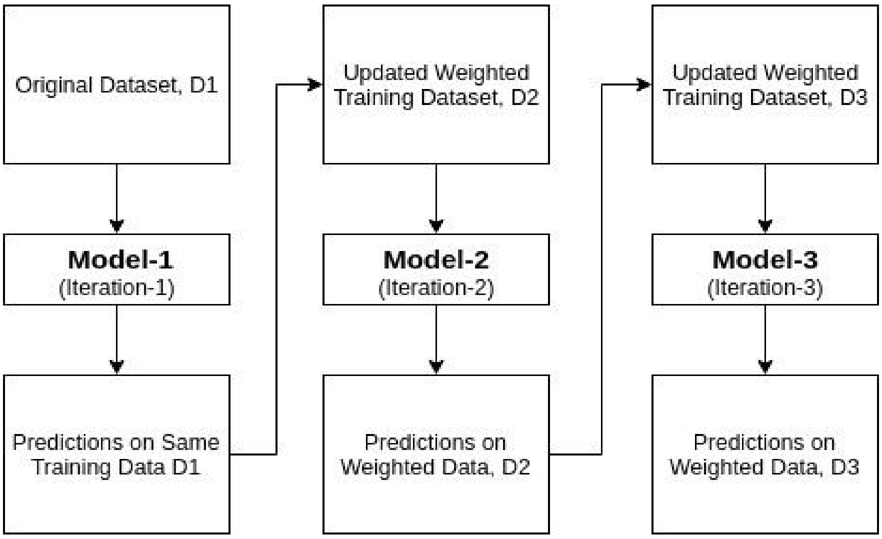


Fig. 3. Adaboost Algorithm

The short name for Adaboost is adaptive boosting. It is best used with weak learners. This Adaboost boosting technique [Figure. 3]combines the multiple weak classifiers into a strong classifier. Adaboost algorithm can be used with short decision trees. The way the Adaboost is created is such that initially at first the nodes are created and the tree is made, then the performance of the tree on each of the instances is checked. Also, a weight is assigned. The training data that is hard to predict is the one that gives more weight. The Adaboost algorithm is a powerful classifier that works well on both the basic and complex problems. The disadvantage of this algorithm is that this algorithm is mostly sensitive to noisy data. This algorithm is also sensitive to outliers.

Steps for Adaboost Algorithm

1. The Kaggle credit card fraud dataset is taken and is trained. Randomly select some of the sample data.
2. Using the randomly created sample data now creates the decision trees sequentially for classifying the fraud and non-fraud cases.
3. The decision trees are formed initially. This can be done by splitting the node based on which has the highest information gain, make it as the root node and classify the fraud and non-fraud cases.
4. Now calculate the error rate, performance, and update the weights of the fraud and non-fraud transactions that are incorrectly classified.
5. Now majority vote is performed and the decision trees may result as output which indicates the non- fraud cases.
6. The decision trees may output 1 which indicates that it is a fraud case.
7. Finally, we find the accuracy, precision, recall, and F1-score for both the fraud and non-fraud cases.

Random Adaboost Algorithm

Algorithm Adaboost :

IINPUTdataset

Initialize weights, w1(n)=1/n Create a decision tree

Select the one that has the lowest Entropy If Incorrectly classified

Calculate Total Error (TE)= sum of up incorrectly

Classified sample weights

Calculate Performance, P = log ^{1-TE}TE

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405398|

For each

Incorrectly classified, increase weights: Weights incorrect =old weight *e^{p} Correctly classified, decrease the weights: Weight correct =old weight *e^{-p} Normalized weight of each sample:

$$\text{Normalized weight} = \frac{dated\ weight}{sum\ of\ updated\ wight}$$

End for

End if

IV. EVALUATION

The confusion matrix and the ROC curve is plotted for both the algorithms. The dataset, when applied for different algorithms, gives different outputs. Firstly we apply the dataset for the random forest model and the results are as below:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	93825
1	0.95	0.77	0.85	162
accuracy			1.00	93987
macro avg	0.97	0.89	0.93	93987
weighted avg	1.00	1.00	1.00	93987

Figure.4 Output for Random Forest

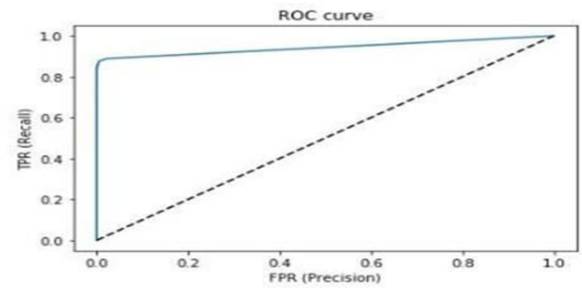
The evaluation criteria are explained[Figure.4] and the precision, recall, F1-score are the same for that of the non-fraud cases and differ for that of the fraud cases.

```
Confusion Matrix on train data
[[190490      0]
 [     0    330]]

Confusion Matrix on test data
[[93818    37]
 [    7   125]]
```

Figure.5 Confusion Matrix for Random Forest

The confusion matrix[Figure.5] shows us that for the train data the true positives are 190490 and false positives are 0, the true negatives are 0 and the false negatives are 330. For the test data, the true positives are 93818 and false positives are 37, the true negatives are 7 and the false negatives are 125



Area under curve (AUC): 0.9429434888303349

Figure.6 ROC curve for Random Forest

IJIRSET
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405398|

Now the dataset is applied for the Adaboost algorithm. The results are obtained similar to that of the Random Forest Algorithm.

```
Accuracy = 0.9990743400683073
              precision    recall  f1-score   support

           0  0.99938202 0.99969091 0.99953644     93825
           1  0.78195489 0.64197531 0.70508475       162
```

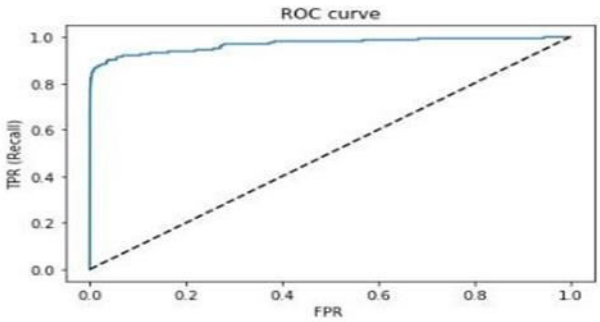
Figure.7 Output for Adaboost

The evaluation criteria[Figure.7] shows us that the evaluation criteria like the precision, recall, and F1-score differ less in the case of the non-fraud cases and differ greatly in those of the fraud cases.

```
Confusion Matrix on train data
[[190464    120]
 [    26    210]]
Confusion Matrix on test data
[[93811    65]
 [   14    97]]
```

Figure.8 Confusion Matrix for Adaboost

The confusion matrix [Figure.8] shows us that for the train data the true positives are 190464 and false positives are 120, the true negatives are 26 and false negatives are 201. For the test data, the true positives are 93811 and false positives are 65, the true negatives are 14 and false negatives are 97.



Area under curve (AUC): 0.968858394765669

Figure.9 ROC curve for Adaboost

Now the comparison of the random forest and the Adaboost algorithms is shown [Figure.9]. The two algorithms have the same accuracy but the precision, recall, and the F1-score of the two algorithms differ. The random forest algorithms have the highest precision, recall, and F1-score.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405398|

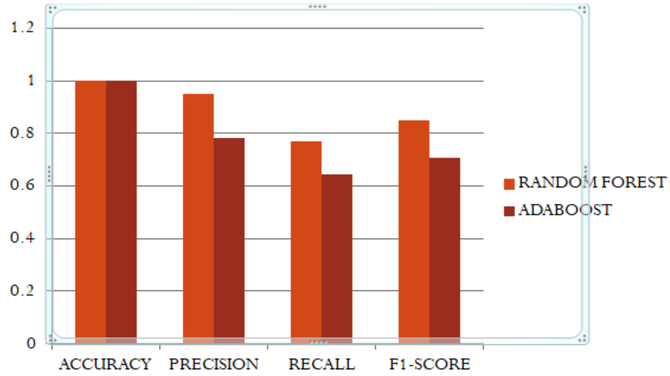


Figure.10 Comparision of Algorithms

V. CONCLUSION AND FUTURE WORK

Even though there are many fraud detection techniques we can't say that this particular algorithm detects the fraud completely. From our analysis, we can conclude that the accuracy is the same for both the Random Forest and the Adaboost algorithms. When we consider the precision, recall, and the F1-score the Random Forest algorithm has the highest value than the Adaboost algorithm. Hence we conclude that the Random Forest Algorithm works best than the Adaboost algorithm to detect credit card fraud. This study aimed to evaluate the effectiveness of four machine learning algorithms in detecting credit card fraud. The results show that all four algorithms, namely Stacking classifier , Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), Decision Trees, and Random Forests, demonstrated promising performance in identifying fraudulent transactions. Stacking classifier achieved an accuracy of 99.6% in detecting fraud CNNs achieved an accuracy of 88.6% in detecting fraud, outperforming the other algorithms. This suggests that CNNs are effective in capturing intricate patterns in transaction data.

LSTMs showed an accuracy of 92.2%, indicating their ability to model sequential dependencies and temporal dynamics in transaction data.

From the above analysis, it is clear that many machine learning algorithms are used to detect the fraud but we can observe that the results are not satisfactory. So, we would like to implement deep learning algorithms to detect credit card fraud accurately. While this study makes significant strides in enhancing the accuracy and efficiency of credit card fraud detection through the implementation of various machine learning algorithms, there remain numerous avenues for further research and exploration to continue improving the robustness and adaptability of fraud detection systems.
The inclusion of additional features and more sophisticated feature engineering techniques could improve model performance. This might include leveraging external data sources such as social media activity or geolocation data, which could provide additional context to transactions. Additionally, advanced feature selection methods can help in identifying the most relevant features, thereby reducing noise and improving prediction accuracy. Implementing real-time fraud detection systems that can instantly flag suspicious transactions is crucial. This requires not only fast and efficient algorithms but also an infrastructure that supports rapid data processing and immediate action. Exploring the use of cloud computing and edge computing could be beneficial in this context.

REFERENCES

1. Adi Saputra1, Suharjito2L: Fraud Detection using Machine Learning in e-Commerce, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.
2. Dart Consulting, Growth of Internet Users in India and Impact on Country's Economy: https://www.dartconsulting.co.in/marketnews/growth-of-internet-users-in-india-and-impact-on-countryseconomy/
3. Ganga Rama Koteswara Rao and R.Satya Prasad, " -Shielding The Networks Depending On Linux Servers Against Arp Spoofing, International Journal of Engineering and Technology(UAE),Vol. 7, PP.75-79, May 2018, ISSN No: 2227-524X, DOI - 10.14419/ijet.v7i2.32.13531.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 5, May 2025

|DOI: 10.15680/IJIRSET.2025.1405398|

4. Heta Naik , Prashasti Kanikar: Detctionsbased on Machine Learning Algorithms,International Journal of Computer Applications (0975 – 8887) Volume 182 – No. 44, March 2019.
5. Navanshu Khare ,Saad Yunus Sait: DetctionsUsing Machine Learning Models and Collating Machine Learning Models, International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838 ISSN: 1314-3395.
6. Randula Koralage, , Faculty of Information Technology, University of Moratuwa,Data Mining Techniques for Credit Card Fraud Detection.
7. Roy, Abhimanyu, et al:Deep learning detecting fraud in credit card transactions, 2018 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2018.
8.Sahayasakila.V, D. Kavya Monisha, Aishwarya, Sikhakolli VenkatavisalakshiseshsaiYasaswi: DetctionsSystem using Smote Technique and Whale Optimization Algorithm,International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019.