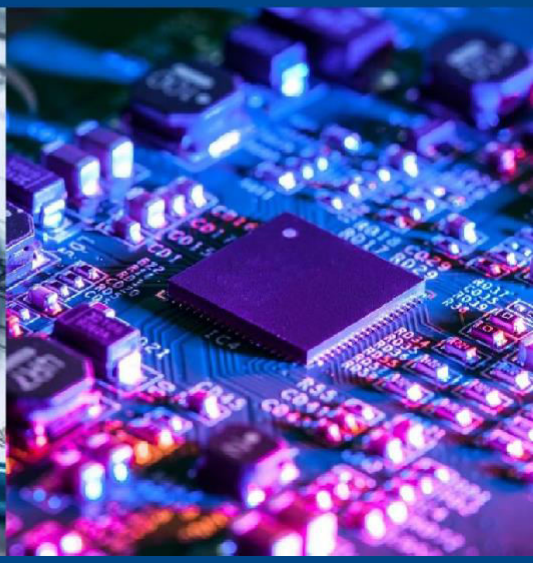
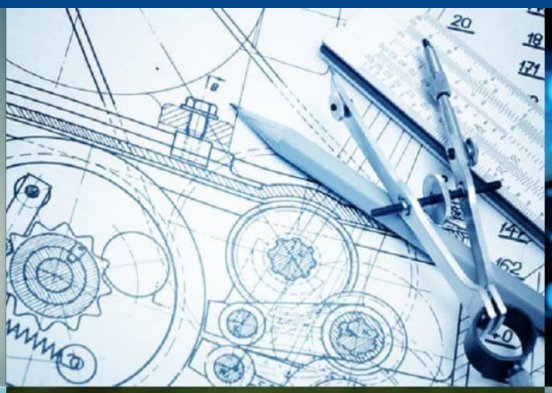
ISSN: 2582-7219

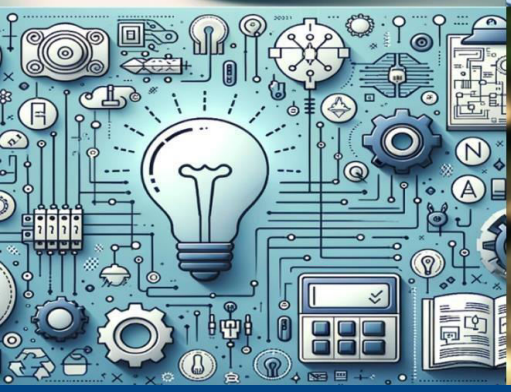




International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 8, Issue 12, December 2025

© 2025 IJMRSET | Volume 8, Issue 12, December 2025| DOI:10.15680/IJMRSET.2025.0812003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

SecureShield: A Machine Learning Based Framework for Credit Card Fraud Detection

Prerana Sanjay Patel[1], Parth Ashuthosh Kulkarni[1], Prerna Vijay Nagare[1], Shrutika Thakare[1], Prof.M.S.Shelar[2], Dr.Sharmila.P.Zhope[2]

Final Year Students, Department of Computer Science and Engineering, Jawahar Education Society Institute of Management Technology, Nashik, India[1]

Department of Computer Science and Engineering, Jawahar Education Society Institute of Management Technology, Nashik, India[2]

ABSTRACT: Credit card has become popular mode of payment for both online and offline purchase, which leads to increasing daily fraud transactions. An Efficient fraud detection methodology is therefore essential to maintain the reliability of the payment system. In this study, we perform a comparison study of credit card fraud detection by using various supervised and unsupervised approaches. Specifically, 6 supervised classification models, i.e., Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), as well as 4 unsupervised anomaly detection models, i.e., One-Class SVM (OCSVM), Auto-Encoder (AE), Restricted Boltzmann Machine (RBM), and Generative Adversarial Networks (GAN), are explored in this study. We train all these models on a public credit card transaction dataset from Kaggle website, which contains 492 frauds out of 284,807 transactions. The labels of the transactions are used for supervised learning models only. The performance of each model is evaluated through 5-fold cross validation in terms of Area Under the Receiver Operating Curves (AUROC). Within supervised approaches, XGB and RF obtain the best performance with AUROC = 0.989 and AUROC= 0.988, respectively. While for unsupervised approaches, RBM achieves the best performance with AUROC = 0.961, followed by GAN with AUROC = 0.954. The experimental results show that supervised models perform slightly better than unsupervised models in this study. Anyway, unsupervised approaches are still promising for credit card fraud transaction detection due to the insufficient annotation and the data imbalance issue in real-world applications.

I. INTRODUCTION

Credit card fraud detection has recently become an active research topic with the exploting growth of big data and AI techniques. Also, it plays an important role in banks as it would help to reduce loss caused by fraudulent transactions. Although many proposed methods (Zareapoor and Shamsolmoali 2015; Randhawa et al. 2018) have achieved promising results, it is still very challenging to accurately and promptly detect credit card fraudulent transactions due to dramatic data imbalance and large variations of fraud transactions.Both supervised and unsupervised learning have been investigated in credit card fraud detection. For example, a combination of multiple learned fraud detectors (Chan et al. 1999) is proposed under a so-called “cost model” to solve the problem of skewed distribution for training data. In contrast, an unsupervised method (Bolton, Hand, and others 2001) is proposed to detect changes in behavior of usual credit card transactions rather than relying on labels of fraudulent historical transaction data. Also, some surveys have comprehensively studied machine learning techniques applied to credit card fraud detection. For example, the survey (Zojaji et al. 2016) reviews the techniques, datasets and evaluation criteria in credit card fraud detection. However, no one has evaluated machine learning models and compared credit card fraud detection performance in a supervised vs unsupervised manner.

In this paper, we evaluate 5 supervised learning models and 4 unsupervised learning models on a Kaggle credit card transaction dataset. The supervised learning models include Support Vector Machines (SVM) (Cortes and Vapnik 1995), K-Nearest Neighbors (KNN) (Altman 1992), Extreme Gradient Boosting (XGB) (Chen et al. 2015), Logistic Regression (LR) (Neter et al. 1996), Decision Tree (DT) (Quinlan 1986) and Random Forest (RF) (Breiman 2001), while the unsupervised learning methods containOne-Class SVM (OCSVM) (Scholkopet al. 2000), Auto-Encoder (AE) (Deng et al. 2010), Restricted Boltzmann Machine (RBM) (Sutskever, Hinton, and Taylor 2009), and Generative

Adversarial Networks (GAN) (Goodfellow et al. 2014). The supervised learning models leverage transaction labels to train classifiers that are able to distinguish between normal and abnormal transactions. In contrast, the unsupervised learning models use unlabeled data for training to capture normal data distribution and then determine whether an unknown test sample is normal or abnormal. As labeling data is time-consuming and labor intensive, labeled data is very expensive, especially when abnormal samples are much smaller than normal one. In this case, the unsupervised learning models would be more useful than the supervised one.

The main contribution of this paper is that we comprehensively studied both supervised and unsupervised learn-ing models for credit card fraud detection and evaluate these machine learning algorithms on a Kaggle credit card transaction dataset in a supervised vs unsupervised way. According to our best knowledge, we are the first to conduct this sort of comparison study between supervised and unsupervised learning on credit card fraud detection.

II. RELATED WORKS

Traditional Machine Learning Methods

It is very time-consuming for people to check credit card transactions one-by-one as transaction amount is tremendously large. Hence, an automated method is desired for credit card fraud detection. In decades, many machine learning methods have been used to solve this problem. Next, we will review some of them to have a big picture of this research area. The traditional neural networks (compared to the current deep neural networks) have already been used for credit card fraud detection in (Dorronsoro et al. 1997). Hidden Markov Model (HMM) (Srivastava et al. 2008) is utilized to model the sequence of operations in credit card transaction processing and detect frauds. In (Bhattacharyya et al. 2011), Support Vector Machine (SVM) and Random Forest (RF) are investigated together with Logistic Regression (LR) based on real-life data from international credit card transactions. Also, a cost-sensitive decision tree based method (Sahin, Bulkan, and Duman 2013) is proposed for credit card fraud detection and evaluated on a real world dataset. In another work (Mahmoudi and Duman 2015), a modified Fisher discriminant function is proposed for credit card fraud detection to be more sensitive to important instances. Besides using machine learning methods, a framework for transaction aggregation (Whitrow et al. 2009) is proposed to solve the problem of preprocessing credit card transaction data for supervised fraud classification. Also, a novel learning strategy (Dal Pozzolo et al. 2018) is proposed to solve three issues of class imbalance, concept drift and verification latency in credit card fraud detection.

Advanced Deep Learning Methods

Recently, deep learning algorithms have achieved promising results in many areas such as image processing (Wang et al. 2015). Therefore, we will review several deep learning based works for credit card fraud detection as follows. Long Short-Term Memory (LSTM) is utilized in (Jurgovsky et al. 2018) to formulate the credit card fraud detection as a sequence classification problem belonging to supervised learning. Also, an unsupervised model (Pumsirirat and Yan 2018) of deep Auto-Encoder (AE) and Restricted Boltzmann Machine (RBM) is proposed to reconstruct credit card normal transactions and detect anomalies. Specifically, a framework tuning parameters of deep learning topologies is proposed for credit card fraud detection in (Roy et al. 2018). It is necessary to mention that Generative Adversarial Network (GAN) is a remarkable model in unsupervised and semisupervised learning. Not only it is employed to detect activity fraud and malicious users in online social networks (Zheng et al. 2018), but also it has been used in credit card fraud detection (Fiore et al. 2017) to augment minority class examples for the classification between fraudulent and nonfraudulent samples. In this paper, the GAN model will also be studied and evaluated as one of unsupervised learning methods.

III. SUPERVISED LEARNING METHODS

Some machine learning methods treat fraud transaction as a supervised classification problem. In this way, we can train a classifier based on training data together with annotations, then classify test transaction data into normal and abnormal categories. In this Section, we briefly discuss 6 widelyused supervised machine learning approaches for credit card fraud detection.

Logistic Regression

Logistic regression was developed by statistician David Cox in 1958 and is a regression model where the response variable Y is categorical. Logistic regression allows us to estimate the probability of a categorical response based on one or more predictor variables x. It allows one to say that the presence of a predictor increases (or decreases) the

© 2025 IJMRSET | Volume 8, Issue 12, December 2025| DOI:10.15680/IJMRSET.2025.0812003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

probability of a given outcome by a specific percentage. Mathematically, logistic regression estimates a multiple linear regression function defined as:

$$Yi = \beta 0 + \beta 1xi,1 + \beta 2xi,2 + \ldots + \beta pxi,p \qquad (1)$$

where $x_{i,j}$ refers to the j^{th} predictor variable for the i^{th} observation, Y_i is the output of i^{th} observation.

K-Nearest Neighbors

In the classification setting, the KNN algorithm essentially boils down to forming a majority vote between the K most similar instances to a given unseen observation. Similarity is defined according to a distance metric between two data points x and x^0. A popular choice is the Euclidean distance given by

$$d(x,x^0) = \sqrt{(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2 + \ldots + (x_n - x_n^0)^2} \qquad (2)$$

But other measures can be more suitable for a given setting and include the Manhattan, Chebyshev and Hamming distance. More formally, given a positive integer K, an unseen observation x and a similarity metric d, KNN classifier performs the following two steps: It runs through the whole dataset computing d between x and each training observation. Suppose the K points in the training data that are closest to x are denoted as set A. It then estimates the conditional probability for each class, that is, the fraction of points in A with that given class label.

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in A} I(y^i = j) \qquad (3)$$

where $I(x)$ is the indicator function which evaluates to 1 when the argument x is true and 0 otherwise Finally, the input x is assigned to the class with the largest probability.

Support Vector Machine

SVM was first introduced by Vapnik in 1995 to solve the classification and regression problems. The basic idea of SVM is to derive an optimal hyperplane that maximizes the margin between two classes. A nice property of SVMs is that it can find a non-linear decision boundary by projecting the data through a nonlinear function φ to a space with a higher dimension. This means that data points which cannot be separated by a straight line in their original input space are lifted to a feature space F where there can be a linear hyperplane separating the data points of one class from an other. When that hyperplane would be projected back to the input space I, it would have the form of a non-linear curve. Mathematically, given n training data samples

$$\{(x_i, y_i)\}_{i=1}^n, \quad x_i \in R^N, y_i \in \{-1, 1\}$$

SVM is formulated by the following optimization problem:

$$Minimize \quad \Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \qquad (4)$$

where the kernel function φ maps training points x_i from input space into a higher dimensional feature space. The regularization parameter C controls the trade-off between achieving a low error on the training data and minimising the norm of the weights.

Decision Tree

Decision trees are simple but intuitive models that utilize a top-down approach in which the root node creates binary splits until a certain criteria is met. This binary splitting of nodes provides a predicted value based on the interior nodes leading to the terminal (final) nodes. In a classification context, a decision tree will output a predicted target class for each terminal node produced.

Decision trees tend to have high variance when they utilize different training and test sets of the same data, since they tend to overfit on training data. This leads to poor performance on unseen data. Unfortunately, this limits the usage of

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

decision trees in predictive modeling. However, using ensemble methods, we can create models that utilize underlying decision trees as a foundation for producing powerful results.

Random Forest

The random forest algorithm, proposed by L. Breiman in 2001, has been successful as a general-purpose classification and regression method. The approach, which combines several randomized decision trees and aggregates their predictions by averaging, has shown excellent performance in the setting where the number of variables is much larger than the number of observations. Moreover, it is versatile enough to be applied to large-scale problems, is easily adapted to various ad-hoc learning tasks, and returns measures of variable importance.

In the classification context, the random forest classifier m is obtained via a majority vote among K classification trees with input x, that is,

$$m(x:\Theta_1,...,\Theta_K) = \begin{cases} 1 \; if \frac{1}{K}\sum_{j=1}^{K} m(x;\Theta_j) > \frac{1}{2} \\ 0 \; otherwise \end{cases} \quad (6)$$

Extreme Gradient Boosting

Gradient boosting is a powerful machine learning technique for regression, classification and ranking problems, which produces a prediction model in the form of an ensemble of weak prediction models like decision trees. The model is built in a stage-wise manner. In each stage, it introduces a new weak learner to compensate the shortcomings of the existing weak learners. XGB stands for eXtreme Gradient Boosting, one of the implementations of gradient boosting concept. The unique of XGB is that it uses a more regularized model formalization to control over-fitting and achieves better performance.

Gradient boosting relies on regression trees, where the optimization step works to reduce mean square error, and for binary classification the standard log loss is used. For a multi-class classification problem, the objective function is to optimize the cross entropy loss. Combining the loss function with a regularization term arrives at the objective function. The regularization term controls the complexity and reduces the risk of over-fitting. XGB uses gradient descent for optimization to improve the predictive accuracy at each optimization step by following the negative of the gradient as we are trying to find the sink in a n-dimensional plane.To learn the set of functions used in the model, XGB minimizes the following regularized objective

$$L(\Theta) = {}^{X}l(y_i, \hat{y}_i) + \Omega(\Theta) \quad (7)$$

where Θ is the learned parameter set, l is a differentiable convex loss function that measures the difference between the predictions $\hat{y}_i$ and the target y_i, and Ω is the regularization term.

Unsupervised Learning Methods

There is a recent surge of interest in developing unsupervised generative models for anomaly detection. Generative models are trained to model the distribution of the normal transaction data (without annotations) distribution. Any transaction that does not follow the distribution is considered to be anomalous. In such a way, the fraud transaction can be detected in an unsupervised manner. In this Section, we briefly discuss 4 unsupervised machine learning approaches for credit card fraud detection.

One-Class Support Vector Machine

One-Class SVM (OCSVM) was proposed by scholkopf to identify novelty / anomaly in an unsupervised manner without labeled training data. The algorithm learns a soft boundary in order to embrace the normal data instances using the training set, and then, using the testing instance, it tunes itself to identify the abnormalities that fall outside the learned region.

Mathematically, OCSVM is formulated by the following optimization problem :

$$Minimize \;\; \Phi(w) = \frac{1}{2}w^T w + \frac{1}{\upsilon n}\sum_{i=1}^{n}\xi_i - \rho \quad (8)$$

© 2025 IJMRSET | Volume 8, Issue 12, December 2025| DOI:10.15680/IJMRSET.2025.0812003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

The parameter υ sets an upper bound on the fraction of outliers and a lower bound on the number of training examples used as support vectors.

Restricted Boltzmann Machine

A RBM model consists of visible and hidden layers, which are connected through symmetric weights. The inputs x correspond to the neurons in the visible layer. The response of the neurons h in the hidden layer model the probability distribution of the inputs. The probability distribution is derived by learning the symmetrical connecting weights between the visible and the hidden layers. The neurons in the same layer are not connected. The conditional probability of a configuration of the hidden neurons (h), given a configuration of the visible neurons associated with inputs (x), is:

$$p(h|x) = \prod_i p(h_i|x) \qquad (10)$$

The objective of the generative training in RBM is to learn the unknown (h) iteratively using the input (x).The generative training phase iterates until the reconstructed samples most closely approximates x. It is performed using the maximum likelihood criterion, and implemented by minimizing the negative log probability of the training data:

$$L_{gen} = -^{X} logP(x|(w_{ij}, b_i, c_j) \quad (11)$$

where b_i and c_j are the bias in the input and hidden layers, respectively. w_{ij} denotes the weights between the inputs and hidden layers.

Auto-Encoder

An auto-encoder (AE) learns to map from input to output through a pair of encoding and decoding phases. The encoder maps from the input to hidden layer, the decoder maps from the hidden layers to the output layer to reconstruct the inputs. The hidden layers of the auto-encoder are lowdimensional and nonlinear representation of the input data. The AE is formulated as follows,

$$X^{\wedge} = D(E(X)) \qquad (12)$$

where X is the input data, E is an encoding map, D is a decoding map, and $^{X^{\wedge}}$ is the reconstructed input data. The objective of the auto-encoder is to approximate the distribution of X as accurately as possible. In particular, an autoencoder can be viewed as a solution to the following optimization problems:

$$min_{D,E} \mathrm{k} X - D(E(X)) \mathrm{k} \qquad (13)$$

where k · k is usually 2-norm. Complex distributions of X can be modelled using a deep auto-encoder with multiple layers, which refers to multiple pairs of encoders and decoders.

Generative Adversarial Networks

GAN is a generative model designed by Goodfellow in 2014. In a GAN setup, two differentiable functions (generator G and discriminator D), represented by neural networks, are competing and trained simultaneously, which eventually drive the generated samples to be indistinguishable from real data.The GAN model in this study is based on AnoGAN (Schlegl et al. 2017) recently developed for anomaly detection by T. Schlegl etc. We modified the original AnoGAN by simultaneously learn an encoder E that maps input samples x to a latent representation z, along with a generator G and discriminator D during training. This enables us to avoid the computationally expensive SGD step for recovering a latent representation at test time.

After we train the model on the normal data to yield G, D and E for inference, we also define a score function $A(x)$ that measures how anomalous an example x is, based on a convex combination of a reconstruction loss L_G and a discriminator-based loss L_D:

$$A(x) = \alpha * L_G(x) + (1 - \alpha) * L_D(x) \qquad (14)$$

© 2025 IJMRSET | Volume 8, Issue 12, December 2025| DOI:10.15680/IJMRSET.2025.0812003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

where α is a weighting parameter ranged in (0,1), σ is the cross-entropy loss from the discriminator of x being a real example (class 1). The definition of $L_G(x)$ indicates how well the trained encoder and generator can reconstruct an input example x. The definition of $L_D(x)$ captures the discriminator confidence that a sample is derived from the real data distribution.

IV. EXPERIMENTAL RESULTS

Data Set and Preprocessing

This public dataset contains credit card transactions made in September 2013 by European cardholders. The transactions occurred in two days include 492 fraud records out of 284,807 transactions. It is obvious that the dataset is highly unbalanced (Fig.1). The fraudulent class only accounts for 0.172% of all transactions.

The dataset contains numerical input variables which are from a PCA transformation due to confidentiality issue. For the non-numerical features of "Time" and "Amount", we normalize them by using RobustScaler which scales the data according to the quantile range. Specifically for the supervised learning models, to tackle the heavily unbalanced problem, random downsampling is used to avoid the bias results toward the non-fraudulent class. Through random downsampling, non-fraud transactions (Class = 0) are randomly reduced to the same amount as fraud transactions (Class = 1), which is equivalent to 492 cases of frauds and 492 cases of non-fraud transactions.To avoid overfitting issues, in this study, k-fold crossvalidation technique is used to estimate fraud detection performance. In one round of k-fold cross-validation, the data set is first randomly divided into k subsets (or folds), which are of approximately equal size and are mutually exclusive. A machine learning model is then trained and tested k times, where in each time

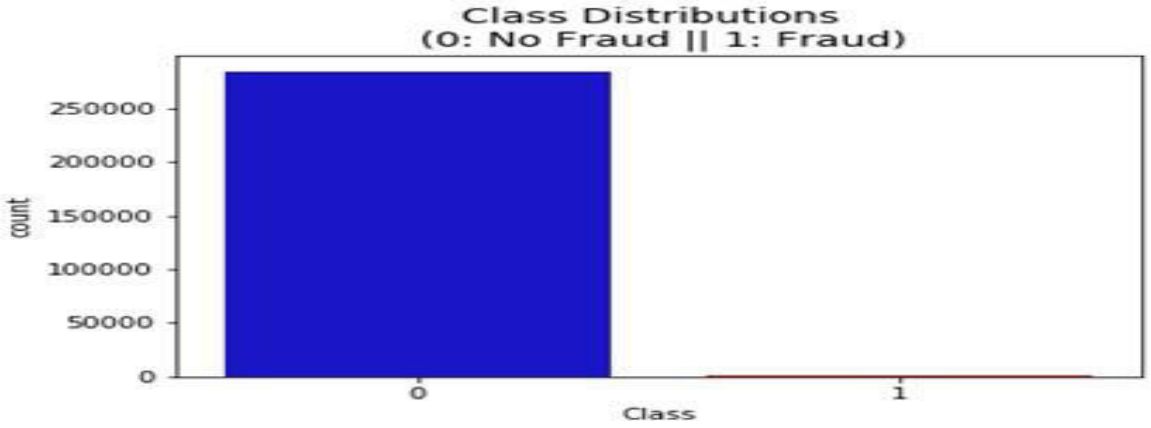


Figure 1: Number of Different Classes

Evaluation Metrics

As mentioned above, the studied data set is highly imbalanced with 492 fraud records out of 284,807 transactions. Even all the samples are classified into non-fraud category, the classification accuracy is still extremely high, that means traditional evaluation metrics like accuracy is not suitable for this study. Instead, we report the Area Under the Receiver Operating Curves (AUROC) () in our experimental study. AUROC combines the false positive rate (FPR) and the true positive rate (TPR) into one single metric. With the assumption that fraud class is "positive" and non-fraud class is "negative, the definition of FPR and TPR are as follows:

$$TPR = TP/P$$

and

$$FPR = FP/N$$

where P and N are the number of samples from positive and negative classes, respectively. TP (True Positive) represents the number of samples predicted to be positive while they are actually positive, and FP (False Positive) the number of samples predicted to be positive while they are actually negative.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

To avoid overfitting issues, in this study, *k*-fold crossvalidation technique is used to estimate fraud detection performance. In one round of k-fold cross-validation, the data set is first randomly divided into *k* subsets (or folds), which are of approximately equal size and are mutually exclusive. A machine learning model is then trained and tested *k* times, where in each time, one of the subsets is set aside as the testing data and the remaining *k*1 subsets are used as training data. The final testing results are predicted from *k* trained sub-models. In our experimental studies, 5 cross validations (i.e., $k = 5$) are used as the validation method.

Parameter Settings

The key parameters of most studied models are determined by grid-search through cross validation, which are listed below:

- LR: 'C': 0.1, 'penalty': 'l1'
- KNN: 'algorithm': 'auto', 'n neighbors': 4
- SVM: 'C': 0.5, 'kernel': 'linear'
- DT: 'criterion':'entropy','max depth': 3,'min
- samples leaf': 6
- RF: 'n estimators': 30, 'oob score': 'True'
- XGB: 'learning rate': 0.4, 'max depth': 4
- OCSVM: 'nu': 0.1, 'gamma': 0.001
- RBM: 'learning rate': 0.0005 'num hidden': 10

While the neural network architectures for Auto-encoder and Generative Adversarial Networks are shown below:

- AE: The encoder has two dense layers with 16 and 32 Relu units, each. The decoder has two dense layers of 32 and 16 Relu units, respectively.
- GAN: The encoder has two dense layers with 32 leaky ReLu and 32 linear units, each. The generator has three dense layers of 32 ReLu, 64 ReLu and 28 linear units, respectively. And the discriminator has one dense layer of 32 leaky ReLu units followed one linear layer with single unit.

Results

The AUROC values of the 6 supervised models on the studied credit card transaction dataset are shown in Fig.2. It can be seen that all the models perform well on this data set, with XGB achieves the best performance with AUROC=0.99, while DT obtains the lowest AUROC value of 0.95. It is expected that the ensemble methods like XGB and RF perform better than the basic methods like DT. Fig.3 shows the AUROC values obtained by unsupervised models, with the RBM, GAN and AE obtain AUROC values above 0.95, while the OC-SVM performs not very well with AUROC = 0.90. Overall, it can be observed that supervised models perform slightly better than unsupervised models, at the expense of additional preprocessing procedures like outliers remove.

V. DISCUSSIONS

In credit card fraud detection, supervised learning aims to train a binary classification model to distinguish between fraudulent and non-fraudulent instances by feeding labeled data, while unsupervised learning is intended to model data distribution of one class and determine whether a test sample belongs to this class or not. In this section, we will discuss the pros and cons of both supervised and unsupervised learning.

Assuming there are sufficient labeled data, supervised learning models, especially for deep neural networks, are able to achieve very promising classification performance. For example, AlexNet (Krizhevsky, Sutskever, and Hinton 2012) significantly reduce error rates for image classification on a large-scale image dataset with more than 1 million labeled images. However, in credit card fraud detection, the training data in two classes are dramatically imbalanced. The fraudulent transactions are much less than the non-fraudulent ones. As a result, the trained classifier will be biased by the majority class whereas it should pay more attention to the minority one. Another issue for supervised

.

Although unsupervised learning is not so attractive as the supervised one, it is suitable for credit card fraud detection as it does not require balanced label data. For example, the AnoGAN model (Schlegl et al. 2017) is able to learn the normal data distribution and indicate whether an unknown test data is normal or abnormal by using their proposed anomaly scoring scheme. This sort of unsupervised learning model would be more prominent if label data is insufficient and data imbalance is severe. Another advantage for unsupervised learning is that a fraudulent credit card use could be detected promptly because the unsupervised model can be updated in low latency by using online

© 2025 IJMRSET | Volume 8, Issue 12, December 2025| DOI:10.15680/IJMRSET.2025.0812003

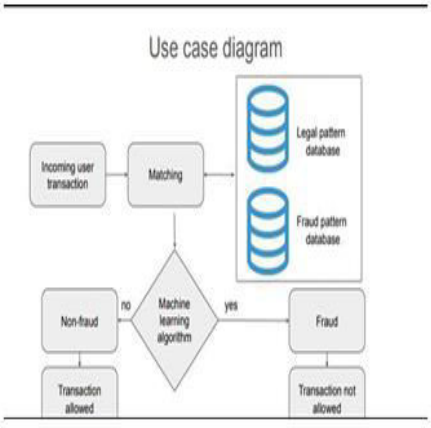
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

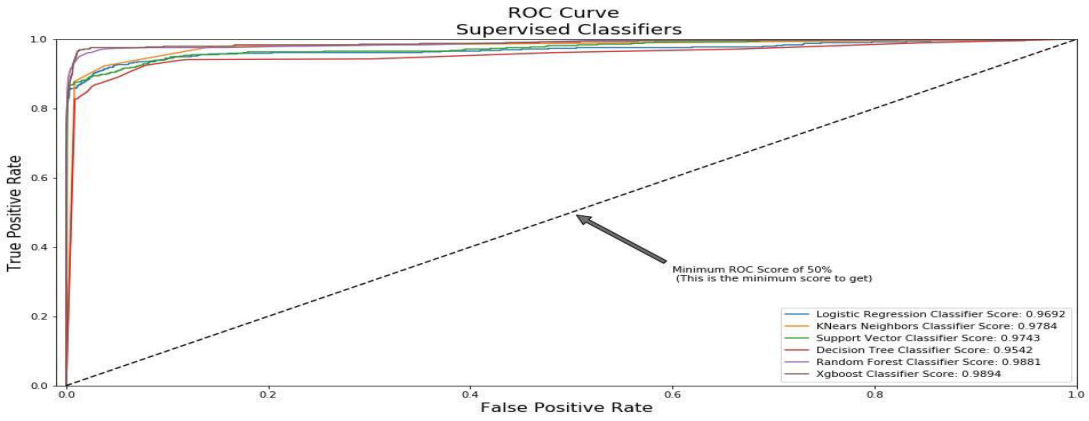
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

unlabeled data in banks and financial institutes. For example, one of unsupervised learning models, Self-Organizing Map (SOM) (Zaslavsky and Strizhak 2006), is used to build a framework for unsupervised credit card fraud detection. The proposed automated system is able to continuously modify the model by using new added transactions because the SOM model does not require priori information, e.g., whether a transaction is done by the cardholder or not. In sum, the advantage of unsupervised learning methods are quite obvious for credit card fraud detection, while the disadvantage may be the difficulty of making some unsupervised model (e.g., GAN) converge

V. USE CASE DIAGRAM





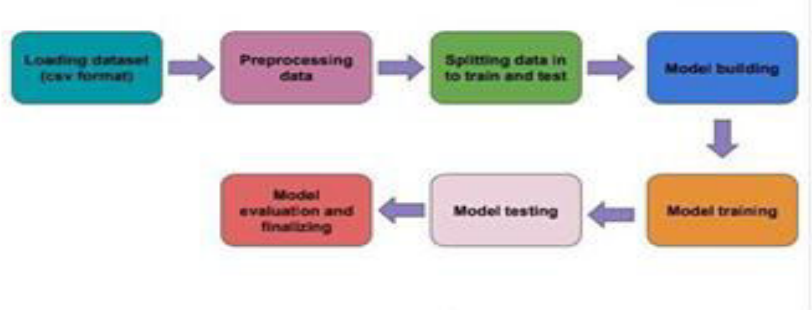


Fig : Proposed system

Figure 2: Plot of AUROC by supervised approaches

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

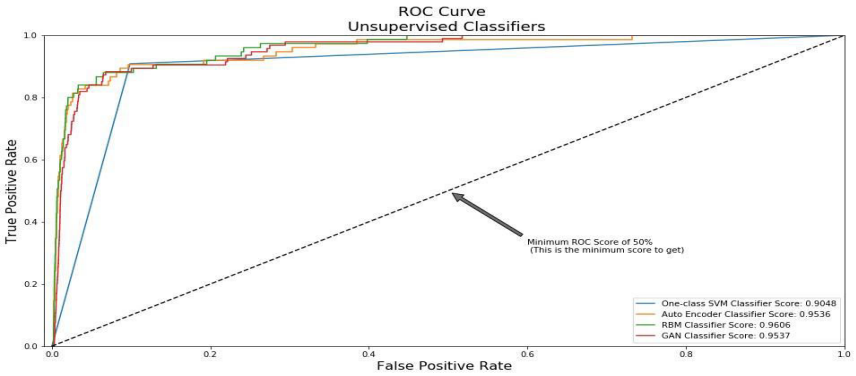


Figure 3: Plot of AUROC by unsupervised approach

V. CONCLUSION

In this paper, we presented **SecureShield**, a machine learning–based framework for detecting fraudulent credit card transactions. By leveraging multiple By leveraging multiple classifiers such as K-Nearest Neighbors (KNN) and Random Forest, and analyzing anonymized transaction features (V1–V28, Amount, Time), the system effectively identifies fraudulent activities in highly imbalanced datasets. The framework also allows flexibility in model selection and evaluation, making it suitable for real-time deployment in online transaction systems. Future work can focus on integrating deep learning models, improving real-time performance, and enhancing the system with explainability techniques such as SHAP or LIME to provide insights into model predictions.

REFERENCES

1. [Altman 1992] Altman, N. S. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46(3):175–185.
2. [Bhattacharyya et al. 2011] Bhattacharyya, S.; Jha, S.; Tharakunnel, K.; and Westland, J. C. 2011. Data mining for credit card fraud: A comparative study. *Decision Support Systems* 50(3):602–613.
3. [Bolton, Hand, and others 2001] Bolton, R. J.; Hand, D. J.; et al. 2001. Unsupervised profiling methods for fraud detection. *Credit Scoring and Credit Control VII* 235–255.
4. [Breiman 2001] Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.
5. [Chan et al. 1999] Chan, P. K.; Fan, W.; Prodromidis, A. L.; and Stolfo, S. J. 1999. Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications* 14(6):67–74.
6. [Chen et al. 2015] Chen, T.; He, T.; Benesty, M.; et al. 2015. Xgboost: extreme gradient boosting. *R package version 0.42* 1–4.
[Cortes and Vapnik 1995] Cortes, C., and Vapnik, V. 1995. Support-vector networks. *Machine learning* 20(3):273–297.
6.[Dal Pozzolo et al. 2018] Dal Pozzolo, A.; Boracchi, G.; Caelen, O.; Alippi, C.; and Bontempi, G. 2018. Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE transactions on neural networks and learning systems* 29(8).
7. [Deng et al. 2010] Deng, L.; Seltzer, M. L.; Yu, D.; Acero, A.; Mohamed, A.-r.; and Hinton, G. 2010. Binary coding of speech spectrograms using a deep auto-encoder. In *Eleventh Annual Conference of the International Speech Communication Association*.
8. [Dorronsoro et al. 1997] Dorronsoro, J. R.; Ginel, F.; Sanchez,´ C. R.; and Santa Cruz, C. 1997. Neural fraud detection in credit card operations. *IEEE transactions on neural networks*.[Fiore et al. 2017] Fiore, U.; De Santis, A.; Perla, F.; Zanetti, P.; and Palmieri, F. 2017. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*.
9. [Goodfellow et al. 2014] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville,
A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

10. [Jurgovsky et al. 2018] Jurgovsky, J.; Granitzer, M.; Ziegler, K.; Calabretto, S.; Portier, P.; He-Guelton, L.; and Caelen, O. 2018. Sequence classification for credit-card fraud detection. *Expert Syst. Appl.* 100:234–245.
11. [Krivko 2010] Krivko, M. 2010. A hybrid model for plastic card fraud detection systems. *Expert Systems with Applications* 37(8):6070–6076.
12. [Krizhevsky, Sutskever, and Hinton 2012] Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105. [Mahmoudi and Duman 2015] Mahmoudi, N., and Duman, E. 2015. Detecting credit card fraud by modified fisher discriminant analysis. *Expert Systems with Applications* 42(5):2510–2516.
13. [Neter et al. 1996] Neter, J.; Kutner, M. H.; Nachtsheim, C. J.; and Wasserman, W. 1996. *Applied linear statistical models*, volume 4. Irwin Chicago.
14. [Pumsirirat and Yan 2018] Pumsirirat, A., and Yan, L. 2018. Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. *International Journal of Advanced Computer Science and Applications* 9(1).
15. [Quinlan 1986] Quinlan, J. R. 1986. Induction of decision
trees. *Machine learning* 1(1):81–106.
16. [Randhawa et al. 2018] Randhawa, K.; Loo, C. K.; Seera, M.; Lim, C. P.; and Nandi, A. K. 2018. Credit card fraud detection using adaboost and majority voting. *IEEE ACCESS* 6:14277–14284.
17. [Roy et al. 2018] Roy, A.; Sun, J.; Mahoney, R.; Alonzi, L.; Adams, S.; and Beling, P. 2018. Deep learning detecting fraud in credit card transactions. In *Systems and Information Engineering Design Symposium (SIEDS), 2018*, 129– 134. IEEE.
18. [Sahin, Bulkan, and Duman 2013] Sahin, Y.; Bulkan, S.; and Duman, E. 2013. A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications* 40(15):5916–5923.
19. [Schlegl et al. 2017] Schlegl, T.; Seebock,¨ P.; Waldstein, S. M.; Schmidt-Erfurth, U.; and Langs, G. 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*, 146– 157. Springer.
20. [Scholkopf¨ et al. 2000] Scholkopf,¨ B.; Williamson, R. C.; Smola, A. J.; Shawe-Taylor, J.; and Platt, J. C. 2000. Support vector method for novelty detection. In *Advances in neural information processing systems*, 582–588.
21. [Srivastava et al. 2008] Srivastava, A.; Kundu, A.; Sural, S.; and Majumdar, A. 2008. Credit card fraud detection using hidden markov model. *IEEE Transactions on dependable and secure computing* 5(1):37–48.
22. [Sutskever, Hinton, and Taylor 2009] Sutskever, I.; Hinton, G. E.; and Taylor, G. W. 2009. The recurrent temporal restricted boltzmann machine. In *Advances in neural information processing systems*, 1601–1608.
23. [Wang et al. 2015] Wang, L.; Liu, T.; Wang, G.; Chan, K. L.; and Yang, Q. 2015. Video tracking using learned hierarchical features. *IEEE Transactions on Image Processing* 24(4):1424–1435.
24. [Whitrow et al. 2009] Whitrow, C.; Hand, D. J.; Juszczak, P.; Weston, D.; and Adams, N. M. 2009. Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery* 18(1):30–55.
25, [Zareapoor and Shamsolmoali 2015] Zareapoor,M., and
Shamsolmoali, P.2015. Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia Computer Science* 48:679–685.
26. [Zaslavsky and Strizhak 2006] Zaslavsky, V., and Strizhak, A. 2006. Credit card fraud detection using self-organizing maps. *Information and Security* 18:48.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com