Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 11, Issue. 2 : 2020
ISSN : **1906-9685**

UNCOVERING FRAUD: HARNESSING THE POWER OF DEEP LEARNING AND MACHINE LEARNING

[#1]**Mrs.SHAGUFTHA BASHEER,** *Assistant Professor*
[#2]**Mr.N SANTHOSH KUMAR,** *Assistant Professor*
Department of Computer Science and Engineering,
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

Abstract— Forgeries are difficult to detect due to their fluidity and lack of distinguishable patterns. Fraudsters are making use of new technologies. They employ these tactics to get around security measures at their own risk. Data mining can reveal unusual patterns of activity, enabling the discovery and investigation of potentially fraudulent transactions. In a monetary transaction, two or more parties exchange money or other financial assets. K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVM) are just a few of the machine learning and deep learning methods reviewed in this study, which also includes Autoencoders, CNNs, RBMs, and DBNs. Databases from the European Union (EU), Australia, and Germany will be included. Area under the curve (AUC), Matthews correlation coefficient (MCC), and failure cost are common performance indicators.

Keywords—credit card, fraud detection, machine learning, deep learning, random forest, k nearest neighbor, support vector machine, autoencoder, restricted boltzmann machine, deep belief networks, convolutional neural networks

1. INTRODUCTION

Dishonest crooks can simply target naive clients to obtain their credit card information and commit crimes thanks to credit cards and internet payment options. Illegal purchases are made on a daily basis. To combat financial fraud, banks and e-commerce enterprises work together. Pre-transaction fraud is avoided using machine and deep learning techniques.

Machine learning, a sort of AI, is becoming more prevalent. Machine learning is being used by more businesses to improve their products and services. Machine learning is powered by algorithms and statistical models. The model may learn from its experiences, forecast, and respond using "training data." Deep learning based on ANNs is included in machine learning. There are deep belief, auto-encoder, recurrent, and limited Boltzmann machines. A trained neural network can detect patterns in big datasets.

Fraud or unauthorized credit card access defrauds account holders or businesses. SEPA credit card theft is expected to cost €1.8 billion in 2016, a 0.4% decrease over 2015. Divide 4.38 trillion Euros in transactions by 1,000 to get this. Nelson projects global credit card losses of $21.84 billion in 2015, rising to $32 billion by 2020.

The research will look at three different datasets. This study looked at datasets from Europe, Australia, and Germany. Deep learning and machine learning are compared. Ensembles are created by combining the top three models from each dataset. The conclusions were reached through comparisons of ML and deep neural network-trained models.

2. RELATED WORK

According to Tuyls et al. (year), identifying fraud presents a number of significant obstacles. The datasets used for this application

are highly skewed and have a low fraud occurrence, making it challenging to train successful models. The inclusion of unstructured data and the creation of overlapping patterns may also cause significant challenges. An essential aspect is the ability of classification algorithms to successfully adapt to the ever-changing dynamics of fraudulent behavior. In this lecture, we will look at a few typical projects that used machine learning and deep learning to detect fraud.

A Comparative study on KNN and SVM

Zareapoor and Shamsolmoali (year) used Naive Bayes, KNN, SVM, and Bagging Ensemble Classifier to detect fraud. The essay investigates the lack of relevant empirical data as a result of banks' and other financial institutions' reluctance to provide sensitive information. If 2% of transactions are fraudulent and 98% are genuine, data is suspect. This is true for large amounts of data as well as time restrictions. Numerous studies show that fraud's constant evolution is tough. For everyone, no single act or method constitutes fraud. To detect and respond to illegal activities, machine learning algorithms must be frequently updated.

Researchers investigated 100,000 online merchant credit card transactions using 20 criteria for the UCSD-FICO competition. With 2,293 fraudulent transactions, the ratio is 100:3. Four datasets had skew of 20%, 15%, 10%, or 3%. The accuracy and error rate were determined to be low. Researchers used the True Positive Rate, True Negative Rate, Erroneous Positive Rate, and Erroneous Negative Rate to account for erroneous positives and negatives. For efficiency evaluation, these four parameters are more significant than accuracy and error rate. Experiments were subjected to ten cross-validations. On all four subsets, KNN outperformed SVM and Naive Bayes Classifier in detecting false alarms and fraud.

Random Forest in Fraud Detection

Randhawa et al. use Naive Bayes, Random Forest, and Gradient Boosted Tree to investigate AdaBoost and majority voting to avoid credit card fraud. Use "Majority Voting" to combine algorithms. According to the findings, outliers and anomalies can confuse AdaBoost ensemble models. RapidMiner searches credit cards in Southeast Asia. Class bias is evident in a dataset with fewer than 1% fraud. To eliminate bias, all classifiers were cross-validated ten times. Classifiers were evaluated using the Matthews Correlation Coefficient (MCC).

The MCC, which looks at two-class problems, employs TP, TN, FP, and FN rates. Random Forest surpassed SVM, gradient-boosted trees, and other complex approaches with a score of 0.990. With 100% accuracy and MCC 1, AdaBoost outperformed Random Forest. The model's generalizability to fresh data is tested in this study.

According to the study, hybrid classifiers are more accurate.

Detecting Fraud using AutoEncoders based on Reconstruction Error

Tom Sweers' undergraduate thesis is on autoencoders, which are neural networks that encode and decode data quickly and consistently. For autoencoder training, standard data points are used. Reconstruction error after autoencoder training labels anomalous points as "fraud" or "no fraud." For untrained anomalies, expect significant reconstruction mistakes. Any quantity that exceeds a certain threshold is considered abnormal. This method was used in the autoencoder-based network anomaly identification model developed by Z. Chen et al.

Chen discovered that many stacked AutoEncoders outperform a single hidden layer for anomaly identification. The AutoEncoder network was used in the study, with four input/output settings: 30-2-30, 30-10-30, 30-20-10-2-10-20-30, and 30-25-20-10-20-25-30. Two layouts concealed one layer, while two others concealed five. Python and Tensorflow were used to create each framework component. The neural network was trained over 100 iterations at 0.01 learning rate. Recall and precision-at-k (k times) were calculated.

At k=1000, the single-hidden-layer AutoEncoder outperformed the stacked multilayered one. As k grew, the layered model outperformed the single layer model.

Using Restricted Boltzmann for Fraud Detections

Restricted Boltzmann Machines (RBMs) provide data reconstruction in the context of

unsupervised learning [9]. The Keras framework was used to create the high-level neural network architecture described in Pumsirirat and Yan's (2019) publication [9]. The H2O platform was used to compute Mean Squared Error, Root Mean Square Error, and Variable Significance for the attributes of each dataset. Keras computed the AUC as well as the confusion matrix for each scenario.

Area Under the Curve (AUC) and Accuracy assessments demonstrate that European datasets outperform German and Australian datasets for the Restricted Boltzmann Machine (RBM). Smaller datasets may diminish the efficacy of the detection process; thus, limited data availability may complicate the task of fraud detection. When using large datasets, it is simpler to spot occurrences categorized as "not-fraud" because there is more data accessible for learning and training purposes.

Using CNN for for detecting Suspicious Activity

Chouiekha and El Haj employed convolutional neural networks (CNNs) to detect dishonesty in their investigation. A large database of 18,000 digitally manipulated images documenting the daily activities of 300 people over the course of 60 days was developed. Long discussions and odd coupon behavior were detected using Customer Details Records. Images are evaluated using CNN (Convolutional Neural Networks) to detect instances of fraud. Half of the data was used for training, a quarter for model validation, and the remaining quarter for testing. To boost classifier performance, image rescaling was used. The Deep Convolutional Neural Network (DCNN) included seven layers in total, including three convolutional layers, two pooling levels, one fully connected layer, and one SoftMax regression layer.

The findings' dependability was evaluated. We will compare Deep Convolutional Neural Networks (CNN) to other popular models such as SVMs, Random Forests, and Gradient Boosting Classifiers (GBCs). When compared to SVM, Random Forest, and GBC, DCNN outperforms them all by 5%. Deep convolutional neural networks (CNNs) take around half the time to train as previous methods.

Neural Networks vs Bayesian Belief Network

The goal of this study is to look at the similarities and differences between BBNs and ANNs. BBN detectors are more effective and require less training when it comes to detecting incidences of fraud. When employed in real-time scenarios, artificial neural networks (ANN) found to be more effective in generating timely predictions.

Summary and Motivation

SVMs, KNN, K-Means, Random Forest, and Naive Bayes are all used to detect fraud. Many research make use of skewed data that has few flaws. Fraud is shown by the True Positive Rate, False Negative Rate, and Matthews Correlation Coefficient. Several research have shown that neural networks are capable of correcting skewed data.

Tom Sweers utilizes autoencoders and regular data to detect thesis fraud using reconstruction error. Pumsirirat et al. use AUC and the Confusion Matrix to assess model accuracy. According to Australian and German studies, deep learning fails with less data. Small datasets have a negative impact on forecasts. Deep learning may be hampered by small datasets. Chouiekha et al. discovered that DCNNs outperformed SVMs, Random Forest, and Gradient Boosted Classifiers. Tuyls et al. contend that ANNs detect fraud better than Bayesian Belief Networks.

Machine learning and deep learning models were first empirically compared. We want to compare models with varied dataset sizes, complexity, and properties. This research is looking for the best fraud detection model.

SVM, KNN, and Random Forest are compared to advanced deep learning algorithms such as Autoencoders, RBM, DBN, and CNN in this study. Test these machine learning methods on three different datasets. This work improves classifier performance by modifying hyperparameters, reducing features, and purifying PCA data.

The top three models are combined based on majority voting. Following are the foundational learning models, data, and experimental plan.

IMPLEMENTATION AND ANALYSIS

Data Sets

The goal of this study is to compare and contrast three alternative ways of data collection. Purchases conducted on two separate days in September 2013 are included

in the European credit card transaction dataset. All other variables, with the exception of the temporal and quantitative dimensions, have been converted using principal component analysis (PCA). So yet, only 492 out of a potential 284,807 have been validated as bogus. Australian and German datasets are available in the UCI Machine Learning repository [4]. Anonymization techniques have removed identifying information from the databases.

According to the Australian, there are 383 real cases and 307 bogus cases. The German dataset contains 1000 observations, with 700 representing "normal" data and 300 representing "fake" data. The size of European databases differs significantly from those of Australia and Germany.

The goal of this study was to look at how machine learning and deep learning models performed on a variety of datasets ranging in size and complexity.

Experimental Setup

Python, NumPy, Pandas, Keras, Scikit-Learn, and Tensorflow were all employed as part of the stack. For the data cleansing phase, Rstudio was used.

Cross-validation on the training dataset is used to discover the ideal value for K-nearest neighbor for each dataset. The best potential value of K is used when examining all of the data sets together.

To discover the best parameter, both Support Vector Machines (SVMs) and Random Forests employ a grid-based search technique. Figures 1 and 2 depict the Python GridSearchCV function's support vector machine (SVM) and random forest settings.

Fig. 1. Support Vector Machine Variables

The models are evaluated using all available data, and the best parameters are established.

Fig. 2.Parameters for Random Forest

Autoencoders are designed with one goal in mind: to reconstruct the original input. When training autoencoders for fraud detection, only normal financial transactions are employed. When we executed the experiment with test data, we encountered numerous reconstruction mistakes. It is reasonable to expect a lower rate of reconstruction errors in honest transactions vs dishonest ones. If the rate of reconstruction mistakes in a given instance or transaction exceeds a certain threshold, the instance or transaction is most likely fraudulent. The transaction will be judged depending on whether or not the requirement is met. The threshold value experiments are carried out, and the results are shown.

The restricted Boltzmann machine (RBM), like Autoencoders, creates free energy that can be compared to a threshold to identify possibly fraudulent financial transactions. Weiman Wang created the RBM model to detect fraudulent activity.

An improved AlbertUP model has been implemented in the Tensorflow framework for both supervised and unsupervised pattern identification in deep belief networks.

The dataset is altered by the convolutional neural network (CNN), which generates a two-dimensional array. Following the convolutional and max-pooling layers, a flattening layer is utilized. The SoftMax layer is in charge of data classification. Figure 3 displays our CNN's general design.

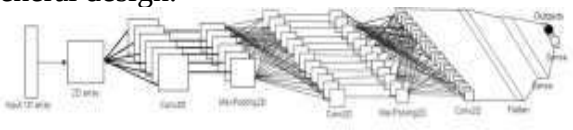

Fig. 3.CNN Architecture

The models are cross-validated, and the three best models are picked by consensus and combined. The model's basic architecture is represented in the graphic below.

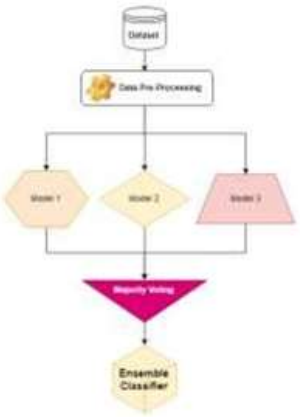

Fig. 4. The final model is chosen through consensus vote

Evalulation Metrics

The key study quality indicators are listed below.

Matthews Correlation Coefficient assesses binary classifiers that distinguish between two classes. It was proposed by Brain W. Matthews in 1975. Perfect forecasts have a coefficient of +1, while random guesses have a coefficient of 0. phi is expressed via Matthews Correlation. Davide Chicco prefers MCC above accuracy and F1 score since they omit all four regions of the confusion matrix.

Receiver operating characteristic (ROC) curves depict the performance of a binary classification model. This method facilitates model correctness evaluation in unbalanced data sets. On the x and y axes, ROC curves compare TPR and FPR. False positives cost money when the AUCs of two ROC curves are identical.

The organization rewards $1,000 for a "False Negative," or fraud misinterpreted as normal. Each False Positive (legal behavior misclassified as fraud) costs $100. We utilize this strategy to analyze the top three models since it compares MCC and AUC values. Costs for ensemble classifiers are also calculated.

3. RESULTS

The results of the evaluations of the European, Australian, and German datasets are as follows. The debate revolves upon statistics connected to Europe.

European Dataset

TABLE 1. THE FINDINGS OF DATA ANALYSIS FROM EUROPE.

Method	MCC	AUC	Cost of Failure
RBM	0.176	0.9109	**227360**
Autoencoders	0.2315	0.8943	**127220**
Random Forest	0.7947	0.8507	**30340**
CNN	0.8096	0.8764	**25700**
SVM	0.8145	0.9004	**21220**
KNN	0.8354	0.8887	**22660**
Ensemble (KNN, SVM and CNN)	**0.8226**	**0.8964**	**21740**

The results of the European Dataset are shown in Table 1. The table shows the AUC and MCC for a variety of machine learning models.

RBMs and AEs with a high number of false alarms reduce the Matthews Correlation Coefficient (MCC) and increase expenses. MCC and AUC are improved by using Random Forest. MCC and AUC are best classified by CNN, SVM, and SVM. Autoencoders and RBM have the highest failure cost, whereas SVM has the lowest. Random forest is a computationally difficult but useful algorithm. SVM, KNN, and CNN algorithms perform admirably on this data.

The best three models are combined in a majority vote classifier. Ensemble outperforms SVM and CNN without any additional costs. AUCs for SVMs are greater. Support Vector Machines (SVM) are less expensive to teach and test than ensemble approaches.

Australian Dataset

TABLE 2. THE DATA COLLECTION IN AUSTRALIA, RESULTS DERIVED FROM AUSTRALIAN DATA SETS

Method	MCC	AUC	Cost of Failure
RBM	0.15	0.5546	**24600**
Autoencoders	0.2318	0.6174	**12220**
CNN	0.6408	0.8227	**6430**
Random Forest	0.684	0.8416	**4700**
KNN	0.6905	0.8425	**6460**
DBN	0.6999	0.8441	**6790**
SVM	0.7085	0.8551	**3380**
Ensemble1 (KNN, SVM, DBN)	**0.7144**	**0.8573**	**5290**
Ensemble2 (KNN,SVM, Random Forest)	**0.7281**	**0.8655**	**3470**

The results of the Australian Dataset are shown

in Table 2. Testing of RBMs and autoencoders fails. SVM, DBN, and KNN benefit from AUC and MCC. Random Forest and CNN are the best models. Two ensemble models were investigated. DBNs, SVMs, and K-Nearest Neighbors are used in Ensemble 1. Top Ensemble 2 models include KNN, SVM, and Random Forest. False positives make RBM and AE vulnerable to costly failures.

In MCC and AUC, Ensemble 1, which consists of KNN, SVM, and DBN, outperforms solo SVM and other techniques, according to Table 2. This approach is more expensive than random forest and SVM. The high failure rates of KNN and DBN may decrease classification quality. In Ensemble 2, KNN, SVM, and Random Forest classifiers reduce failure cost while maximizing MCC, AUC, and total cost. Because Ensemble 2 has the highest MCC, AUC, and price, using many techniques improves performance.

German Dataset

TABLE 3. THE RESULTS FROM THE GERMAN DATA SET ARE AS FOLLOWS.

Method	MCC	AUC	Cost of Failure
RBM	0.0984	0.5524	**14160**
Autoencoders	0.139	0.5614	**22640**
KNN	0.2487	0.6047	**21100**
DBN	0.2725	0.5873	**23640**
Random Forest	0.2912	0.6437	**16970**
SVM	0.4038	0.6857	**16400**
CNN	0.4291	0.7056	**14220**
Ensemble (SVM, CNN, Random Forest)	**0.4439**	**0.7011**	**15620**

Table 3 displays the results of the Germany dataset. According to AUC and MCC statistics, SVM, Random Forest, and CNN models perform well. Other approaches, like as random forest, CNN, and SVM, have higher failure costs. These three models are used to build a majority-voting classifier.

In all three datasets, better models outperform individual models, as seen in Tables 1 and 3. Smaller data sets, such as those from Germany and Australia, benefit from ensemble improvement. On the European dataset, SVM outperforms the model.

Random Forest prefers lower sample sizes. CNN deep learning fared best on datasets from Europe and Germany. CNNs were placed fourth in Australian datasets. CNN failure costs were comparable to K-Nearest Neighbors. Because of its low cost, German data was utilised.

Table 4 displays the frequency with which each model scores in the top three across datasets. SVM performed admirably overall. K-Nearest Neighbors (KNN) works well with big data.

TABLE 4. THOSE MODELS THAT HAVE THE BEST TRACK RECORDS

Method	Number of times in Top 3
Support Vector Machines	3 Times
K-Nearest Neighbors	2 Times
Convolutional Neural Networks	2 Times
Random Forest	2 Times
Deep Belief Network	1 Time

4. CONCLUSION

For nearly two decades, researchers in fraud detection have relied on human inspection and consumer end authentication. In this case, ML models surpass humans. Deep learning models are used in many applications because of their high processing capacity and inexpensive cost. Across datasets, the empirical study compares machine learning and deep learning fraud detection algorithms. This study focuses on the best techniques for various datasets. Because more firms are implementing new business improvement approaches, this research assists practitioners and businesses in understanding the operational dynamics of various tactics on diverse datasets.

In large datasets, we discovered that SVMs identify fraud better than CNNs. SVM, Random Forest, and K-Nearest Neighbors are useful for smaller datasets. CNNs outperformed Autoencoders, RBMs, and DBNs.

Unfortunately, supervised learning is used in the majority of fraud detection systems. In dynamic scenarios, supervised learning algorithms such as CNNs, KNNs, and Random Forest have limitations. The mechanisms of fraud change with time, making identification difficult. Machine learning models must be retrained, and new data must be collected.

Autoencoders save time while teaching students about legal traffic. Inconsistencies imply fraud. Trainable autoencoders label datasets, but they are costly. Data with labels is

used to retrain and monitor models.

REFERENCES

[1] European Central Bank, "Fifth report on card fraud, September 2018," 26 September 2018. [Online], Available: https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport20 1809.en.html.

[2] Credit card fraud detection anonymized credit card transaction labeled as fraudulent or genuine [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud.

[3] Dheeru Dua and Casey Graff. UCI Machine Learning Repository. 2017 [Online]. Available: http://archive.ics.uci.edu/ml/datasets/

[4] Masoumeh Zareapoora, Pourya Shamsolmoalia. "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier" International Conference on Intelligent Computing, Communication & Convergence (ICCC-2015). Procedia Computer Science 48 pp 679686. 2015.

[5] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim, Asoke K. Nandi. "Credit card fraud detection using AdaBoost and majority voting". IEEE Access (Volume: 6), pp 14277 – 14284. 2018.

[6] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau. "Autoencoder based network anomaly detection." Wireless Telecommunications Symposium, pp 1-5. 2018.

[7] Apapan Pumsirirat, Liu Yan. "Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine". International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2, pp 18-25. 2018.

[8] Alae Chouiekha, EL Hassane Ibn EL Haj. "ConvNets for Fraud Detection analysis". Procedia Computer Science 127, pp.133–138. 2018.

[9] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick. "Credit Card Fraud Detection Using Bayesian and Neural Networks". 2002. [Online]. Available:

[10] Albertbup, "A Python implementation of Deep Belief Networks built upon NumPy and TensorFlow with scikit-learn compatibility". GitHub. October 2018. [Online] Available: https://github.com/albertbup/deep-belief-network.