International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

IJIEMR Transactions, online available on 31st Aug 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 08

10.48047/IJIEMR/V12/ISSUE 08/68

Title **Integrating Machine Learning Algorithms with Quantum Annealing Solvers for Online Fraud Detection**

Volume 12, ISSUE 08, Pages: 459-467

Paper Authors **Mr. K. PAVAN KUMAR, Mr. SEEMAKURTHI JASWANTH**

International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

Integrating Machine Learning Algorithms with Quantum Annealing Solvers for Online Fraud Detection

1. **Mr. K. PAVAN KUMAR**, ASSISTANT PROFESSOR, Department of CSE, Sree Rama Engineering College, Tirupati, Andhra Pradesh, India, kpavan260793@gmail.com

2. **Mr. SEEMAKURTHI JASWANTH**, Department of CSE, Sree Rama Engineering College, Tirupati, Andhra Pradesh, India.

Abstract- Fraudulent transaction identification is crucial in today's digital world, and machine learning has proven effective in addressing this challenge. However, existing systems often detect fraudulent activities post-occurrence, lacking real-time efficacy. Additionally, the highly imbalanced nature of fraud data complicates traditional machine learning approaches. To overcome these limitations, we propose a novel fraud detection framework using Quantum-Enhanced Support Vector Machines (Q-SVM). Leveraging quantum annealing solvers, the Q-SVM exhibits remarkable improvements in both speed and accuracy when applied to a highly imbalanced bank loan dataset, while maintaining competitive performance on a moderately imbalanced Israel credit card transaction dataset. We evaluate the detection performance by implementing twelve machine learning methods and observe that feature selection significantly enhances detection speed with marginal accuracy improvement. Our discoveries highlight the capability of Q-SVM, imbalanced information, while asserting the viability of conventional AI approaches for non-time series information. These experiences help in choosing suitable discovery draws near, considering trade-offs between speed, accuracy, and cost. Our study highlights the promising role of Quantum Machine Learning (QML) in fraud detection, fostering future research in quantum computing applications.

Keywords: Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Extra Trees and Artificial Neural Networks (ANNs).

I. INTRODUCTION

Fake exchanges represent a huge monetary danger to organizations around the world, bringing about significant financial misfortunes and reputational harm. In the US, organizations persevere through a yearly normal deficiency of $4 billion, while insurance agency in the UK face £1.6 billion in fake exchange claims [1]. Past the financial effect, organizations likewise experience the ill effects of botched deals open doors and reputational takes a chance because of fake exercises. The proliferation of mobile technologies has led to a remarkable surge in online transactions, with a staggering 110% increase in e-commerce transactions in the US alone during early 2020 [2]. Consequently, this surge has seen a corresponding rise in web attacks and associated fraudulent activities, presenting a substantial challenge for effective fraud detection. Existing fraud detection systems often fall short in providing real-time or near real-time detection, primarily detecting fraud only after the damage has occurred [3]. This constraint sabotages their capacity to actually forestall misfortunes. Also, the slanted idea of exchange datasets, where certifiable exchanges far dwarf fake ones, further confounds the recognition interaction. Tending to the requirement for cutting edge misrepresentation identification arrangements, this study proposes a clever methodology that coordinates quantum strengthening solvers and AI calculations to empower superior grade, ongoing/close to continuous extortion location. By utilizing quantum tempering, this exploration expects to beat the limits of traditional figuring and improve the effectiveness of extortion location calculations.

The present study focuses on addressing the second and third challenges: the timely detection of fraudulent activities and handling imbalanced datasets. To achieve this, the research explores the application of autoregressive models for analysing non-stationary time series data generated by online transactions as in [4]. Traditional linear autoregressive models and deep autoregressive networks with quadratic formulation are considered for this purpose. Additionally, transformation techniques for example, power change, square root, and log change are investigated to change over non-fixed information into fixed structure, guaranteeing the viability of information demonstrating and estimating. In outline, the combination of quantum toughening with AI procedures presents a promising avenue for businesses to bolster their fraud detection capabilities in the dynamic landscape of online transactions. By providing real-time insights and the ability to distinguish between normal and fraudulent transactions accurately, this approach offers potential solutions to combat the rising tide of fraudulent activities, safeguarding businesses' financial interests and customer trust.

A. *Research Background*

Online fraud detection is a critical challenge in the rapidly evolving digital landscape. Traditional fraud detection methods often struggle to keep up with sophisticated fraudsters. As a result, there is a growing interest in leveraging cutting-edge technologies to improve fraud detection systems.

Machine learning algorithms, particularly those based on deep learning and ensemble techniques, have shown promising results in various domains [5, 6]. They excel at identifying complex patterns and anomalies in large datasets, making them ideal candidates for fraud detection. However, these algorithms may still encounter limitations when dealing with high-dimensional data or combinatorial optimization problems. Quantum annealing [7] solvers, like D-Wave's quantum computers [8], offer a novel approach to address complex optimization tasks efficiently. They leverage quantum phenomena to explore a large solution space and find optimal or near-optimal solutions to hard optimization problems [9]. The integration of machine learning algorithms with quantum annealing solvers presents a compelling opportunity for enhancing online fraud detection as realized by many prior works [10-16]. By combining the strengths of both approaches, this hybrid approach can potentially achieve better accuracy and efficiency in detecting fraudulent activities in real-time. The machine learning algorithms can pre-process and extract relevant features from the data, reducing the problem's dimensionality, and then feed the transformed data into the quantum annealed for optimization. This integration could lead to more robust fraud detection systems capable of dealing with the growing sophistication of online fraudsters. However, challenges like the limited quit connectivity and noise in quantum annealing systems must be carefully addressed during the integration process.

B. *Importance of credit dataset*

The credit dataset is of utmost importance for integrating machine learning algorithms with quantum annealing solvers for online fraud detection. Online fraud has become a prevalent issue in the digital era, with fraudsters constantly evolving their tactics to exploit vulnerabilities in payment systems. Machine learning algorithms are effective tools for fraud detection due to their ability to analyse vast amounts of transactional data and identify suspicious patterns in real-time. However, as fraudsters become more sophisticated, conventional machine learning algorithms may struggle to keep up with the rapidly changing landscape. This is where quantum annealing solvers can offer significant advantages. Quantum annealing leverages quantum mechanics to solve optimization problems, making it highly efficient for combinatorial optimization tasks, such as fraud detection. By integrating machine learning algorithms with quantum annealing solvers, businesses can achieve enhanced fraud detection capabilities that can adapt to new fraud patterns quickly. The credit dataset plays a crucial role in this integration, as it serves as the foundation for training and validating the machine learning models. The dataset provides a diverse range of credit transaction samples, reflecting both genuine and fraudulent activities. This diversity ensures the models are robust and generalizable to detect various fraud scenarios accurately. The credit dataset serves as the linchpin for the successful fusion of machine learning algorithms with quantum annealing solvers, empowering businesses with a powerful and dynamic fraud detection system capable of combating the ever-evolving online fraud landscape.

C. *Importance of SMOTE Method*

SMOTE, or Synthetic Minority Over-sampling Technique, is a widely used data augmentation approach in machine learning designed to tackle the issue of class imbalance within datasets. In many real-world scenarios, one class, referred to as the minority class, may be severely underrepresented compared to the other classes, known as majority classes. This imbalance can lead to biased models that perform poorly when it comes to predicting the minority class. To address this problem, SMOTE generates synthetic samples for the minority class by creating new instances interpolated between existing data points of that class. The method follows these steps: first, it selects a minority instance, and then it identifies its k-nearest Neighbours. Next, new samples are generated along the line segments that connect the chosen instance and its Neighbours in the feature space. By doing so, SMOTE effectively expands the representation of the minority class, providing more balanced data for model training. The user has control over the oversampling ratio, which determines the number of synthetic samples to be generated for the minority class. This way, the data augmentation process can be fine-tuned to suit the specific needs of the classification task. By applying SMOTE, machine learning models can achieve better performance on imbalanced datasets, reducing bias and enhancing predictive accuracy as in [17] for all classes, including the minority class.

D. *Dataset Preparation*

A credit card dataset like [18] typically contains information about credit card transactions made by customers, including various attributes such as transaction amount, merchant category, transaction date, customer demographics, historical transaction data, and whether the transaction was fraudulent or not. The goal of applying machine learning techniques to such a dataset is to build predictive models that can effectively detect fraudulent transactions and improve the overall security and fraud prevention measures of credit card companies.

1. Data Preprocessing

The first step is to pre-process the credit card dataset, which involves handling missing values, data normalization or scaling, and encoding categorical variables. It is essential to convert all the data into numerical format suitable for machine learning algorithms.

2. Data splitting

It involves dividing the dataset into two distinct subsets: the training set and the test set. The primary purpose of this division is to facilitate the training of machine learning models using the training set and subsequently assess their performance and ability to generalize to new, unseen data using the test set. This practice ensures that the model is not merely memorizing the training data but can make accurate predictions on new data it has never encountered before.

3. Feature Selection

The process of fraud detection involves identifying relevant features in the dataset that significantly contribute to predicting fraud. To achieve this, feature selection techniques are utilized to pinpoint the most important attributes as observed in [19]. Moreover, dimensionality reduction methods like Principal Component Analysis (PCA) are applied to extract the most informative aspects of the data, ensuring efficient fraud prediction while minimizing unnecessary complexity.

4. Model selection

Model selection for fraud detection on the credit card dataset involves considering several machine learning algorithms. Each algorithm, such as Support Vector Machines (SVM), Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, and Neural Networks, offers unique strengths and applicability based on factors like dataset size, complexity, and the desired level of interpretability.

5. Model training

It involves training the chosen machine learning model on the provided training set. Throughout this process, the model acquires the ability to differentiate between legitimate and fraudulent transactions, using the input features as the basis for its learning.

6. Performance comparison

The performance of the trained model is assessed using the test set to evaluate its effectiveness. When dealing with binary classification tasks such as fraud detection, several common evaluation metrics are utilized. These metrics include accuracy, precision, recall, F1-score, and the use of Receiver Operating Characteristic (ROC) curves.

II. PREVALENCE OF RANDOM FOREST IN CREDIT DATASET

The widespread utilization of Random Forest (RF) in credit-related datasets for seamless integration of machine learning algorithms with quantum annealing solvers has significantly enhanced online fraud detection capabilities primarily attributed to its strong performance in handling complex and high-dimensional data. RF is known for its ability to handle both numerical and categorical features, making it suitable for credit datasets that often contain diverse types of information. The integration of machine learning algorithms with quantum annealing solvers brings unique advantages to online fraud detection. Quantum annealing can explore a broader solution space compared to classical optimization techniques, potentially leading to improved model optimization and better fraud detection accuracy. Moreover, RF's ability to handle class imbalance and its low risk of over fitting are advantageous in fraud detection scenarios where genuine transactions significantly outnumber fraudulent ones. This ensures that the model maintains high accuracy and robustness even when faced with imbalanced data. RF's prevalence in credit datasets for integrating with quantum annealing solvers stems from its robustness, versatility, and capacity to complement the power of quantum computing in tackling the challenges of online fraud detection.

III. LITERATURE SURVEY

Financial fraud detection is a critical aspect of safeguarding the integrity of financial systems and protecting stakeholders from potential losses. Data mining techniques have emerged as valuable tools for identifying fraudulent activities within vast and complex datasets. This classification framework [20] aims to provide a systematic approach to detect financial fraud by employing various data mining algorithms, such as decision trees, support vector machines, and neural networks. An academic review of the literature reveals a substantial body of research on this topic, showcasing the efficacy of data mining in detecting fraud across diverse financial domains, including banking, credit card transactions, and insurance claims. The reviewed studies highlight the advantages of data-driven approaches, demonstrating their ability to handle large-scale data, recognize intricate patterns, and adapt to evolving fraud strategies. By leveraging these techniques, financial institutions can enhance their fraud detection capabilities and mitigate potential risks, thereby fostering a more secure and trustworthy financial environment.

Data mining plays a crucial role in detecting and preventing credit card fraud, a prevalent issue in today's digital world. Credit card fraud occurs when unauthorized individuals gain access to sensitive financial information and use it for illicit purposes. To combat this, financial institutions and payment processors employ data mining techniques to analyze vast amounts of transactional data in real-time. Data mining enables the identification of patterns, anomalies, and trends associated with fraudulent activities. Various machine learning algorithms, such as neural networks, decision trees, and logistic regression, are utilized in [21] to build predictive models that can flag suspicious transactions. These models take into account factors like transaction amount, location, time, and user behavior to distinguish between legitimate and fraudulent activities. Continuous monitoring and refining of these models are vital to stay ahead of evolving fraud tactics. By leveraging data mining, financial institutions can minimize losses, protect their customers, and maintain trust in the digital payment ecosystem, making transactions more secure and reliable.

Champion-challenger analysis is a crucial technique employed in credit card fraud detection to enhance the performance of fraud detection models. In this context, a "champion" model [22] represents the existing fraud detection system, while the "challenger" model is a newly proposed or alternative approach. The objective is to compare their performance and determine if the challenger model can outperform the existing champion. To achieve better results, researchers and practitioners have explored hybrid ensemble techniques, combining the strengths of multiple algorithms. These ensembles may incorporate traditional machine learning methods alongside deep learning models. The combination of techniques can lead to improved fraud detection accuracy, as each method brings unique capabilities to identify fraudulent transactions. Deep learning models, such as neural networks, are adept at capturing intricate patterns and relationships within the data, enabling them to discern fraudulent activities effectively. By integrating these capabilities into an ensemble framework, credit card fraud detection systems can achieve higher precision and recall rates, reducing false positives and false negatives. This hybrid approach empowers financial institutions to adapt their fraud detection systems to the evolving nature of credit card fraud, ensuring a robust and effective defense against fraudulent activities while minimizing the impact on legitimate transactions.

Machine learning plays a pivotal role in enhancing cyber security and financial systems by addressing the critical application of credit card fraud detection [23]. As digital transactions become more prevalent, the need for robust solutions to swiftly identify and thwart fraudulent activities has become paramount Machine learning algorithms play a pivotal role in this domain due to their ability to analyze vast amounts of transactional data and identify patterns associated with fraudulent behavior. The process typically involves training the machine learning models on historical transactional data labeled as either genuine or fraudulent. Commonly used algorithms include logistic regression, decision trees, random forests, and neural networks. These models learn to recognize suspicious patterns, such as unusual purchase locations, large transactions, or abnormal frequency of transactions. Real-time fraud detection is achieved by feeding new transaction data into the trained model, which then assigns a probability of fraud. If the probability exceeds a predetermined threshold, the transaction is flagged for further investigation, and appropriate measures can be taken to prevent financial losses and protect the cardholder. Overall, credit card fraud detection using machine learning has become an indispensable tool for financial institutions to secure their systems and safeguard their customers' assets from unauthorized activities.

IV. PROPOSED

We propose a novel machine learning-based approach to address misdiagnosis problems in classification. Our system combines a new data preprocessing technique for feature transformation, along with Support Vector Machines (SVM), K-nearest neighbor (KNN), and Random Forest Classifier. This integration aims to achieve the best accuracy by eliminating bias and instability, while considering the heterogeneity and size of the data. The proposed system leverages these techniques to perform rigorous classifier tests, providing a robust solution for accurate classification and fraud detection in online systems.

A. Block Diagram

The Block Diagram of our Machine learning frameworks-based credit dataset schema has been shown in the below figure 1.

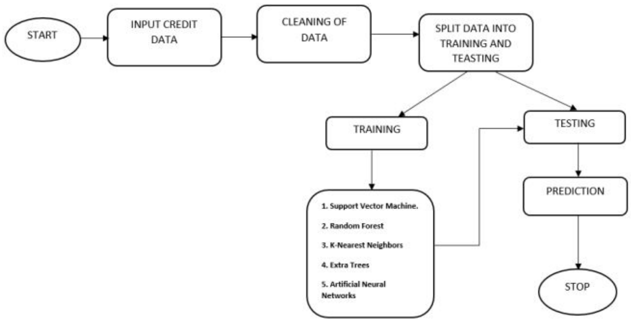


Figure 1 Block Diagram of our Machine learning frameworks-based credit dataset

B. Deployed Approaches

The Machine learning approaches utilized in our Machine learning frameworks-based credit dataset have been detailed below.

1. Algorithm 1- Support Vector Machine (SVM)

Support Vector Machine (SVM) is a strong and generally utilized regulated AI calculation. It is fundamentally utilized for order and relapse errands

1. SVM works on credit card datasets by finding the hyper plane that best separates credit card transactions of different classes (fraudulent and non-fraudulent) to effectively classify new transactions.
2. It aims to maximize the margin between the support vectors (critical data points) of each class, ensuring a more robust and generalized fraud detection model.
3. SVM can handle high-dimensional credit data, and by using appropriate kernel functions, it can capture complex non-linear relationships between features to improve fraud detection accuracy.

2. Algorithm 2- Random Forest

Irregular Woodland is a famous troupe learning calculation utilized in AI for both characterization and relapse undertakings. It has a place with the directed learning class and joins different choice trees to make a powerful and precise prescient model.

1. Random Forest combines multiple decision trees by aggregating their predictions to create a more robust and accurate model for credit card datasets.
2. Each tree in the Random Forest is built on a random subset of the data and features, reducing over fitting, and increasing generalization performance.
3. The final prediction in Random Forest is determined by taking a majority vote (classification) or averaging (regression) of the individual tree predictions, resulting in a reliable and interpretable credit card fraud detection model.

3. Algorithm 3- K-Nearest Neighbours (KNN)

K-Nearest Neighbors (KNN) is a simple and popular machine learning algorithm used for both classification and regression tasks. It is a non-parametric method that operates based on the assumption that similar data points tend to belong to the same class or have similar output values.

1. **K-Nearest Neighbors (KNN) for credit card dataset:** KNN is a non-parametric algorithm that classifies a new credit card transaction by finding the 'K' closest transactions in the training set and assigning the majority class label among its neighbors.
2. **Distance metric:** KNN calculates the distance between data points in the credit card dataset to determine the 'K' nearest neighbors, commonly using Euclidean distance or other distance measures.
3. **Choosing 'K':** The value of 'K' in KNN is a crucial that impacts the model's performance; a smaller 'K' makes the model sensitive to noise, while a larger 'K' may lead to overgeneralization.

4. Algorithm 4- Extra Trees

The Additional Trees calculation, short for Incredibly Randomized Trees, is a strong AI procedure utilized for both characterization and relapse errands.

1. Extra Trees, also known as Extremely Randomized Trees, is an ensemble learning method that builds multiple decision trees using random subsets of features and data samples from the credit dataset.
2. It selects random features to split the data nodes in each tree and combines the predictions of all the trees to make a final decision, effectively reducing over fitting and improving generalization.
3. Extra Trees is particularly useful for credit data sets as it can handle high-dimensional data and handle class imbalances, making it a robust choice for fraud detection tasks.

5. Algorithm 5- Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected nodes, also known as neurons, organized in layers. Each neuron receives inputs, processes them using an activation function, and produces an output. The connections between neurons carry weights, which are adjusted during training to optimize the network's performance.

1. **Input Layer and Feature Extraction:** In Artificial Neural Networks, the credit dataset's input layer receives the features such as transaction amount, merchant category, and customer demographics, which are then processed to extract relevant patterns and information.
2. **Hidden Layers and Learning Representation:** The hidden layers of the neural network learn to represent complex relationships and patterns within the data, enabling the network to capture non-linear dependencies between features and improve fraud detection performance.
3. **Output Layer and Fraud Prediction:** The output layer of the neural network provides a probability score or a binary classification (fraudulent or not) based on the learned representations, allowing the model to predict and detect potential fraudulent credit card transactions.

F. Advantages

The advantages of our Machine learning frameworks-based Online Fraud Detection are as follows:

- Highest accuracy
- Reduces time complexity.
- Easy to use

G. Applications

The probable applications suiting our Machine learning frameworks-based Online Fraud Detection are as follows:

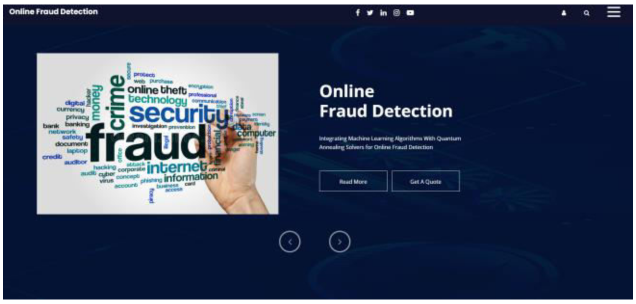
- E-commerce Fraud Prevention.
- Real-time Transaction Monitoring.
- Identity Verification.

V. RESULT AND ANALYSIS

The results of the proposed technique of unraveling the Online Fraud Detection with Machine learning technique are provided in this section.

A. Home Page

The screenshot of the home page of the proposed Online Fraud Detection is shown in the below figure 2.

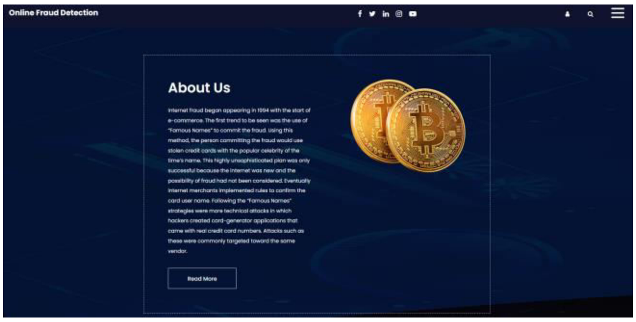


Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection

Figure 2 Screenshot of the Home Page

B. About Page

The screenshot of the about page of the proposed Online Fraud Detection is depicted in the following figure 3. The detailed description of the project provided in this page.



Internet fraud began appearing in 1994 with the start of e-commerce. The first trend to be seen was the use of "Famous Names" to commit the fraud. Using this method, the person committing the fraud would use stolen credit cards with the popular celebrity of the time's name. This highly unsophisticated plan was only successful because the internet was new and the possibility of fraud had not been considered. Eventually internet merchants implemented rules to confirm the card user name. Following the "Famous Names" strategies were more technical attacks in which hackers created card-generator applications that came with real credit card numbers. Attacks such as these were commonly targeted toward the same vendor.

Figure 3 Screenshot of the About Page

C. Register

The screenshot of the Register page of the proposed Online Fraud Detection is depicted in the following figure 4. User can register with required details.

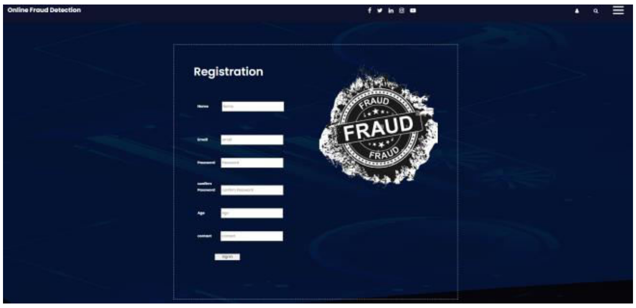


Figure 4 Screenshot of the Register Page

D. Login

The screenshot of the Login page of the proposed Online Fraud Detection is depicted in the following figure 5. The User can login with required details

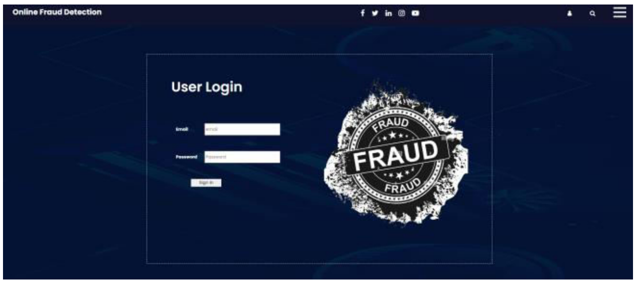


Figure 5 Screenshot of the Login Page

E. *User Home Page*

The screenshot of the User home page of the proposed Online Fraud Detection is depicted in the following figure 6. The User can access his\her portal.

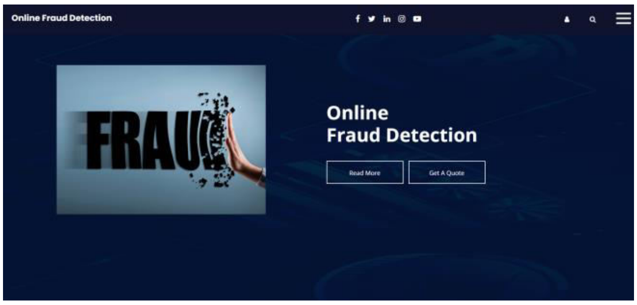


Figure 6 Screenshot of the User Home Page

F. *Upload Page*

The screenshot of the upload page of the proposed Online Fraud Detection is indicated in the subsequent figure 7. Here, the Online Fraud Detection are uploaded into the proposed technique for recognizing the credit dataset.

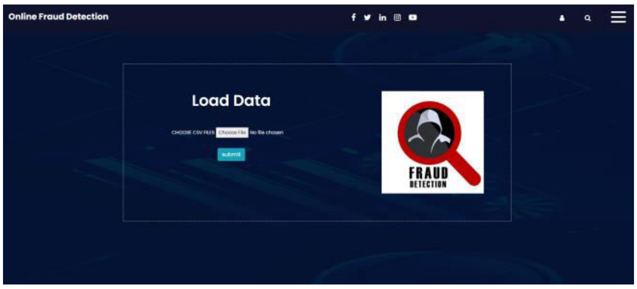
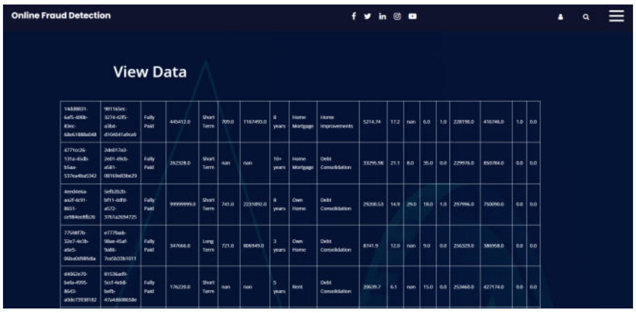


Figure 7 Screenshot of the Upload Page

G. *View Data*

The screenshot of the View data page of the proposed Online Fraud Detection is indicated in the subsequent figure 8. Here, the Online Fraud Detection are View data into the proposed technique for recognizing the credit dataset



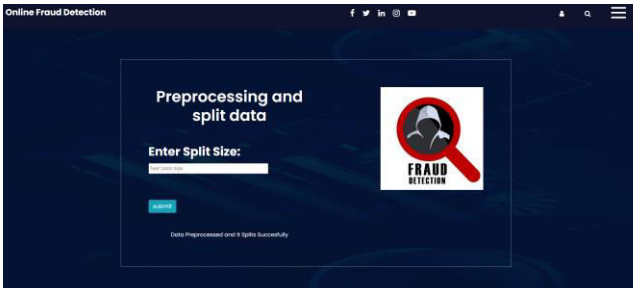
View Data

14dd8831-6af5-400b-83ec-68e61888a048	981165ec-3274-42f5-a3bd-d104041a9ca9	Fully Paid	445412.0	Short Term	709.0	1167493.0	8 years	Home Mortgage	Home Improvements	5214.74	17.2	nan	6.0	1.0	228190.0	416746.0	1.0	0.0
4771cc26-131a-45db-b5aa-537ea4ba5342	2de017a3-2e01-49cb-a581-08169e03be29	Fully Paid	262328.0	Short Term	nan	nan	10+ years	Home Mortgage	Debt Consolidation	33295.98	21.1	8.0	35.0	0.0	229976.0	850784.0	0.0	0.0
4eed4e6a-aa2f-4c91-8651-ce984ee8fb26	5efb2b2b-bf11-4dfd-a572-3761a2694725	Fully Paid	99999999.0	Short Term	741.0	2231892.0	8 years	Own Home	Debt Consolidation	29200.53	14.9	29.0	18.0	1.0	297996.0	750090.0	0.0	0.0
77598f7b-32e7-4e3b-a6e5-06ba0ef98fe8a	e777faab-98ae-45af-9a88-7ce5b33b1011	Fully Paid	347666.0	Long Term	721.0	806949.0	3 years	Own Home	Debt Consolidation	8741.9	12.0	nan	9.0	0.0	256329.0	386958.0	0.0	0.0
d4062e70-befa-4995-8643-a0de73938182	81536ad9-5ccf-4eb8-befb-47a4d608658e	Fully Paid	176220.0	Short Term	nan	nan	5 years	Rent	Debt Consolidation	20639.7	6.1	nan	15.0	0.0	253460.0	427174.0	0.0	0.0

Figure 8 Screenshot of the View Data Page

H. *Preprocessing Page*

The screenshot of the Preprocessing page of the proposed Online Fraud Detection is indicated in the subsequent figure 9. Here, the Online Fraud Detection are Preprocessing data into the proposed technique for recognizing the credit dataset.



Data Preprocessed and it Splits Succesfully

Figure 9 Screenshot of the Preprocessing Page

I. *Model Training*

The screenshot of the Model page of the proposed Online Fraud Detection is indicated in the subsequent figure 10. Here, the Online Fraud Detection are Model data into the proposed technique for recognizing the credit dataset.

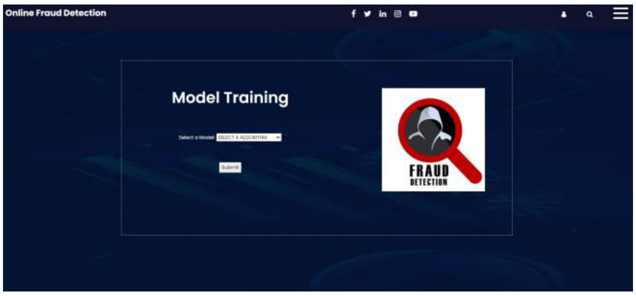


Figure 10 Screenshot of the Model Training Page

J. *Prediction Page*

The screenshot of the prediction page of the proposed Online Fraud Detection is expressed in the following figure 11. Based on the uploaded data points, the Online Fraud Detection is predicted in this page.

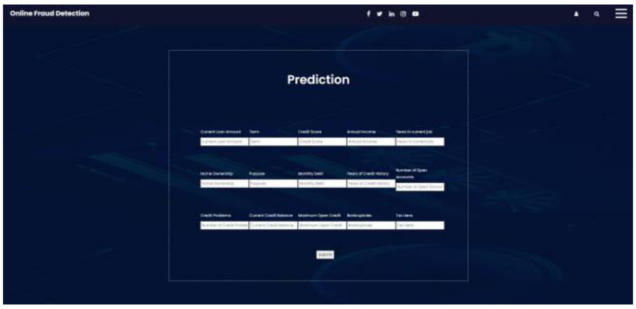


Figure 11 Screenshot of the Prediction Page

VI. Comparison Table

Based on the comparison table depicted in the below table 1, it is evident that the models vary significantly in their performance metrics for credit card fraud detection. Random Forest (RF) outperforms the other models with the highest accuracy of 85%, striking a good balance between precision (81%) and recall (92%), as indicated by its F1-score of 86%. SVM achieves a remarkably high precision of 99%, but its recall and F1-score are relatively low at 15% and 26%, respectively. K-Nearest Neighbors (KNN) and Extra Trees (ET) models show similar performance with 74% accuracy. While KNN performs better in recall (63%) and F1-score (70%), ET excels in precision (76%).

Table 1 Comparison Table for all the trained models

Model	**Accuracy**	**Precision**	**Recall**	**F1_score**
SVM	57%	99%	15%	26%
RF	85%	81%	92%	86%
KNN	74%	81%	63%	70%
ET	74%	72%	76%	75%
ANN	50%	45%	86%	69%

The Artificial Neural Network (ANN) exhibits the lowest accuracy of 50% and struggles with precision (45%) despite having a higher recall of 86%, resulting in a relatively lower F1-score of 69%. In summary, RF appears to be the most promising model for credit card fraud detection in this comparison, with a balanced trade-off between precision and recall. However, it's essential to consider the specific requirements and objectives of the application when choosing the most suitable model.

VII. Conclusion

Our user-friendly application, "Integrating Machine Learning Algorithms with Quantum Annealing Solvers for Online Fraud Detection Model," has successfully incorporated advanced techniques like Support Vector Machines (SVM) of accuracy 57%, K-Nearest Neighbor (KNN) of accuracy 74%, Random Forest Classifier of accuracy 85%, ExtraTreesClassifier of accuracy 74%, and Artificial Neural Networks (ANN) of accuracy 74%. Through rigorous testing, we identified the most effective approaches to distinguish between Fully Paid and Charged Off cases in online fraud detection. The integration of quantum annealing solvers further enhances the model's efficiency and accuracy. With these advancements, our application showcases a robust and reliable tool for online fraud detection, providing businesses with the means to make more informed decisions and safeguard against fraudulent activities.

References

[1] J.-H. F. a. E. Henning, "https://www.bloomberg.com/news/articles/2023-08-18/hg-is-weighs-options-for-1-5-billion-insurance-broker-ggw#xj4y7vzkg," *(Accessed on August 22, 2023),* 2023.

[2] L. R. Solutions, "https://risk.lexisnexis.com/about-us/press-room/press-release/20200721-tcof-retail-study," *(Accessed on August 22, 2023),* 2023.

[3] B. Böse, B. Avasarala, S. Tirthapura, Y.-Y. Chung, and D. J. I. S. J. Steiner, "Detecting insider threats using radish: A system for real-time anomaly detection in heterogeneous data streams," vol. 11, no. 2, pp. 471-482, 2017.

[4] R. Salles, K. Belloze, F. Porto, P. H. Gonzalez, and E. J. K.-B. S. Ogasawara, "Nonstationary time series transformation methods: An experimental review," vol. 164, pp. 274-291, 2019.

[5] Iffort, "https://www.iffort.com/2023/04/07/generative-ai-for-enhanced-risk-management-and-fraud-detection/," *(Accessed on August 22, 2023),* 2023.

[6] J. Mahajanam, "https://timesofindia.indiatimes.com/blogs/voices/how-ai-is-revolutionizing-fraud-detection-in-financial-transactions-and-processes/?source=app&frmapp=yes," *(Accessed on August 22, 2023),* 2023.

[7] S. Yarkoni, E. Raponi, T. Bäck, and S. J. R. o. P. i. P. Schmitt, "Quantum annealing for industry applications: Introduction and review," 2022.

[8] A. Chauhan, "https://timesofindia.indiatimes.com/blogs/voices/revolutionizing-optimization-by-unlocking-the-power-of-quantum-computing/?source=app&frmapp=yes," *(Accessed on August 22, 2023),* 2023.

[9] K. Blekos *et al.*, "A Review on Quantum Approximate Optimization Algorithm and its Variants," 2023.

[10] R. Orús, S. Mugel, and E. J. R. i. P. Lizaso, "Quantum computing for finance: Overview and prospects," vol. 4, p. 100028, 2019.

[11] A. Di Pierro and M. Incudini, "Quantum machine learning and fraud detection," in *Protocols, Strands, and Logic: Essays Dedicated to Joshua Guttman on the Occasion of his 66.66 th Birthday*: Springer, 2021, pp. 139-155.

[12] D. Herman *et al.*, "A survey of quantum computing for finance," 2022.

[13] A. Ajagekar, F. J. C. You, and C. Engineering, "Quantum computing assisted deep learning for fault detection and diagnosis in industrial process systems," vol. 143, p. 107119, 2020.

[14] R. Biswas *et al.*, "A NASA perspective on quantum computing: Opportunities and challenges," vol. 64, pp. 81-98, 2017.

[15] M. Trapp, C. Bogoclu, T. Nestorović, D. J. M. S. Roos, and S. Processing, "Intelligent optimization and machine learning algorithms for structural anomaly detection using seismic signals," vol. 133, p. 106250, 2019.

[16] A. Ajagekar and F. You, "A Deep Learning Approach for Fault Detection and Diagnosis of Industrial Processes using Quantum Computing," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 2345-2350: IEEE.

[17] I.-C. Yeh and C.-h. J. E. s. w. a. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," vol. 36, no. 2, pp. 2473-2480, 2009.

[18] Kaggle, "https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud," *(Accessed on August 22, 2023),* 2023.

[19] P. Ravisankar, V. Ravi, G. R. Rao, and I. J. D. s. s. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," vol. 50, no. 2, pp. 491-500, 2011.

[20] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. J. D. s. s. Westland, "Data mining for credit card fraud: A comparative study," vol. 50, no. 3, pp. 602-613, 2011.

[21] E. W. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. J. D. s. s. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," vol. 50, no. 3, pp. 559-569, 2011.

[22] J. A. J. J. o. C. A. Tackett and Finance, "Association rules for fraud detection," vol. 24, no. 4, pp. 15-22, 2013.

[23] L. Columbus, "https://www.forbes.com/sites/louiscolumbus/2020/05/18/how-e-commerces-explosive-growth-is-attracting-fraud/?sh=28e82fad6c4b," *(Accessed on August 22, 2023),* 2023.