Comparative Analysis of Machine Learning Models for Credit Card Fraud Detection

S.G. Santhiya[1#] A.Merry Ida [2#] S.Angel Nithya [3#] M.Antro Monica Sanjas [4#] P.Anitha [5#]

[1]*Assistant Professor,* Dept. of CSE(AIML), *Loyola Institute of Technology and Science, Nagercoil, India.*
[#]*santhiyasujin96@gmail.com*
[2]*Assistant Professor,* Dept. of CSE, *Loyola Institute of Technology and Science, Nagercoil, India*
[3]*Assistant Professor,* Dept. of CSE(AIML), *Loyola Institute of Technology and Science, Nagercoil, India*
[4]*Assistant Professor,* Dept. of CSE(AIML), *Loyola Institute of Technology and Science, Nagercoil, India*
[5]*Assistant Professor,* Dept. of CSE(AIML), *Loyola Institute of Technology and Science, Nagercoil, India*

ABSTRACT: - The rapid growth of online transactions has intensified the risk of credit card fraud, posing major challenges to consumers and financial institutions. This study explores various machine learning approaches to accurately distinguish fraudulent transactions from legitimate ones. Using a publicly available dataset, multiple algorithms including Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Naïve Bayes, and Artificial Neural Networks (ANN) were implemented and compared. The models were evaluated on accuracy, precision, recall, and misclassification rates to identify the most effective technique. Experimental findings indicate that SVM achieved the highest accuracy among the tested models, demonstrating strong potential for real-world fraud prevention applications. The results emphasize the importance of data preprocessing, model selection, and continuous improvement to combat the evolving tactics of fraudsters.

INTRODUCTION

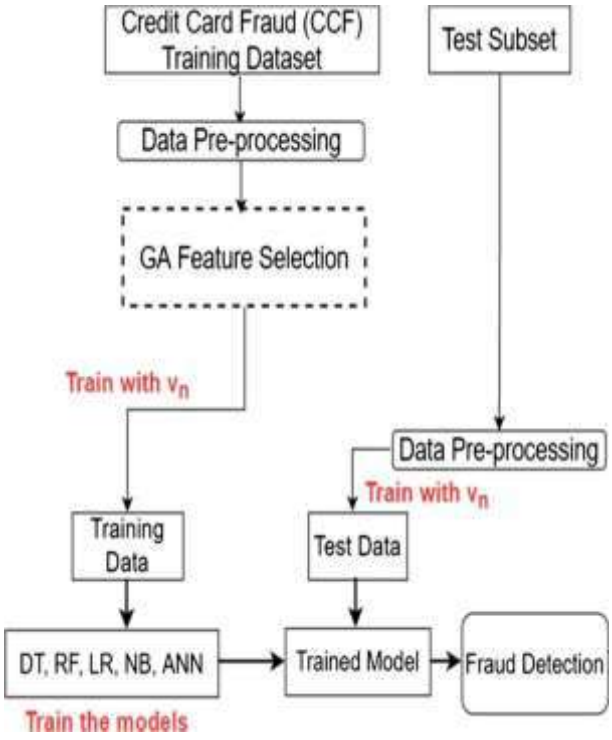
As more individuals rely on credit cards for everyday purchases, financial providers must prioritize robust protection measures for their clients. Global credit card usage reached 2.8 billion people in 2019, with about 70% holding at least one card, according to 2021 statistics. In the United States alone, fraud reports jumped 44.7% from 271,927 cases in 2019 to 393,207 in 2020. Fraud typically falls into two categories: creating fake accounts using stolen identities (up 48% from 2019 to 2020) or hijacking existing accounts via pilfered details (up 9% in the same period) (Daly, 2021). These alarming trends motivated this work to apply machine learning techniques for identifying fraudulent transactions within large volumes of data, aiming to enhance security and reduce risks.

LITERATURE REVIEW

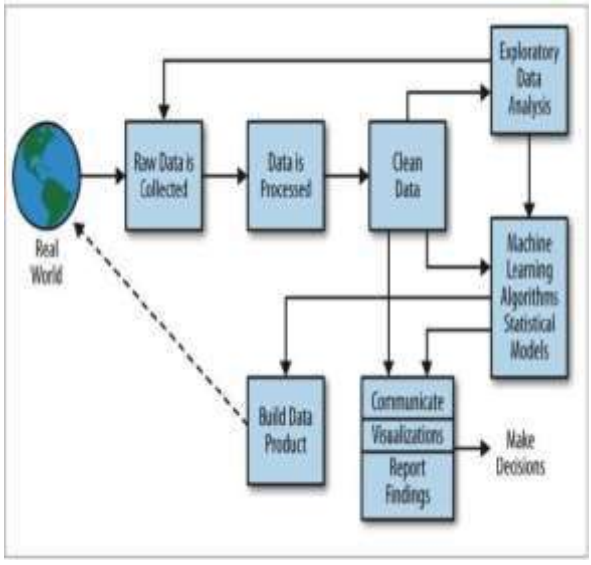
Researchers have applied various machine learning (ML) techniques to detect credit card fraud, focusing on accuracy, speed, and cost. Zareapoor et al. (2012) compared Neural Networks, Bayesian Networks, SVM, and KNN, finding Bayesian Networks the fastest and most accurate, though costly. Alenzi and Aljehane (2020) used Logistic Regression, achieving 97.2% accuracy, outperforming Voting Classifier and KNN. Maniraj et al. (2019) achieved 99.7% detection using a model designed to minimize misclassifications. Dheepa and Dhanapal (2012) used SVM with behavior-based features, reaching over 80% accuracy. Malini and Pushpa (2017) showed KNN outperformed Outlier Detection, proving effective with memory limits. Maes et al. (2002) found Bayesian 8% more effective than ANN, with faster training times. Awoyemi et al. (2017) tested KNN, Naïve Bayes, and Logistic Regression, where NB achieved 97.5% accuracy in imbalanced data. Jain et al. (2019) reported ANN achieving 99.71% accuracy, outperforming SVM and KNN. Gupta et al. (2021) highlighted Naïve Bayes with 80.4% accuracy. Other studies showed high performance of Random Forest (Varmedja et al., 2019; Sailusha et al., 2020), hybrid models like NB-KNN (Kiran et al., 2018), and advanced

approaches like BiLSTM and BiGRU (Najadat et al., 2020). Overall, ensemble, hybrid, and deep learning models consistently outperformed single classifiers in fraud detection.

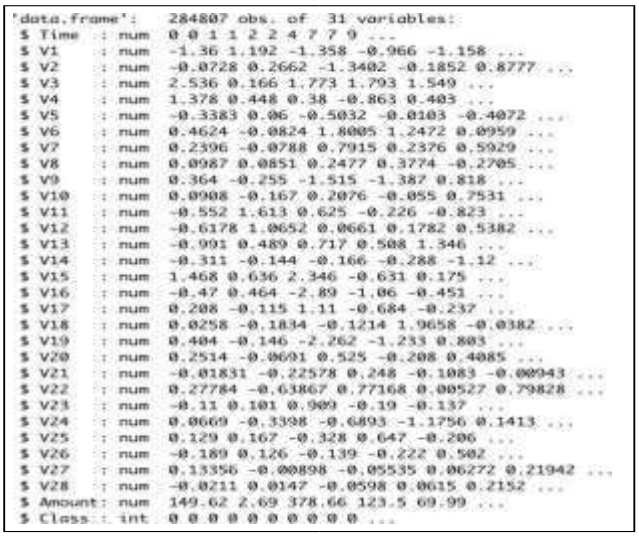
SYSTEM ARCHITECTURE:



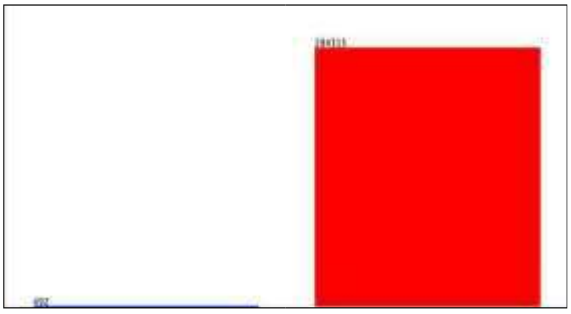
DATA FLOW:

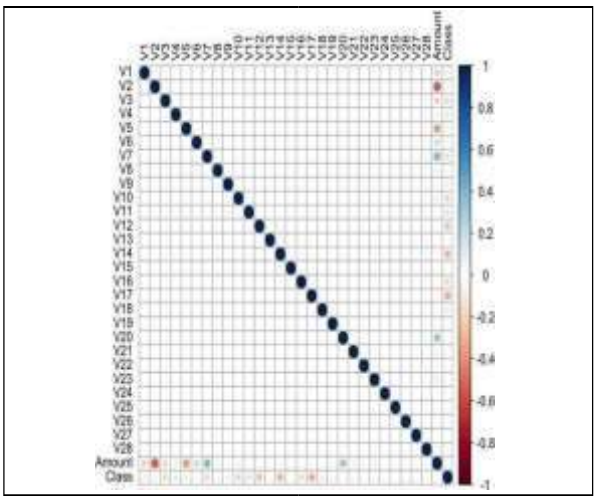


DATA PREPARATION:

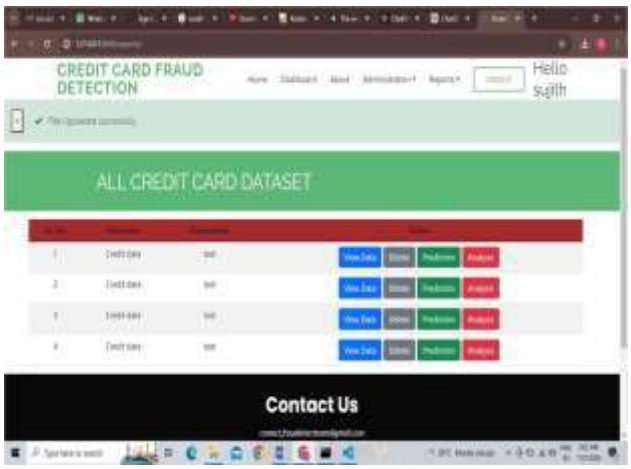


```
'data.frame':   284807 obs. of  31 variables:
 $ Time  : num  0 0 1 1 2 2 4 7 7 9 ...
 $ V1    : num  -1.36 1.192 -1.358 -0.966 -1.158 ...
 $ V2    : num  -0.0728 0.2662 -1.3402 -0.1852 0.8777 ...
 $ V3    : num  2.536 0.166 1.773 1.793 1.549 ...
 $ V4    : num  1.378 0.448 0.38 -0.863 0.403 ...
 $ V5    : num  -0.3383 0.06 -0.5032 -0.0103 -0.4072 ...
 $ V6    : num  0.4624 -0.0824 1.8005 1.2472 0.0959 ...
 $ V7    : num  0.2396 -0.0788 0.7915 0.2376 0.5929 ...
 $ V8    : num  0.0987 0.0851 0.2477 0.3774 -0.2705 ...
 $ V9    : num  0.364 -0.255 -1.515 -1.387 0.818 ...
 $ V10   : num  0.0908 -0.167 0.2076 -0.055 0.7531 ...
 $ V11   : num  -0.552 1.613 0.625 -0.226 -0.823 ...
 $ V12   : num  -0.6178 1.0652 0.0661 0.1782 0.5382 ...
 $ V13   : num  -0.991 0.489 0.717 0.508 1.346 ...
 $ V14   : num  -0.311 -0.144 -0.166 -0.288 -1.12 ...
 $ V15   : num  1.468 0.636 2.346 -0.631 0.175 ...
 $ V16   : num  -0.47 0.464 -2.89 -1.06 -0.451 ...
 $ V17   : num  0.208 -0.115 1.11 -0.684 -0.237 ...
 $ V18   : num  0.0258 -0.1834 -0.1214 1.9658 -0.0382 ...
 $ V19   : num  0.404 -0.146 -2.262 -1.233 0.803 ...
 $ V20   : num  0.2514 -0.0691 0.525 -0.208 0.4085 ...
 $ V21   : num  -0.01831 -0.22578 0.248 -0.1083 -0.00943 ...
 $ V22   : num  0.27784 -0.63867 0.77168 0.00527 0.79828 ...
 $ V23   : num  -0.11 0.101 0.909 -0.19 -0.137 ...
 $ V24   : num  0.0669 -0.3398 -0.6893 -1.1756 0.1413 ...
 $ V25   : num  0.129 0.167 -0.328 0.647 -0.206 ...
 $ V26   : num  -0.189 0.126 -0.139 -0.222 0.502 ...
 $ V27   : num  0.13356 -0.00898 -0.05535 0.06272 0.21942 ...
 $ V28   : num  -0.0211 0.0147 -0.0598 0.0615 0.2152 ...
 $ Amount: num  149.62 2.69 378.66 123.5 69.99 ...
 $ Class : int  0 0 0 0 0 0 0 0 0 0 ...
```





Django



CONCLUSION:

This project's core goal was to determine the optimal machine learning model for credit card fraud detection among the selected techniques. By constructing and testing the models, we found Support Vector Machine to be the leader with 99.94% accuracy and just 51 misclassifications. Implementing such a model could significantly curb fraud incidents, boosting customer trust through enhanced security and smoother experiences.

REFERENCE

[1] Adepoju, O., Wosowei, J., Lawte, S., & Jaiman, H. (2019). Comparative evaluation of credit card fraud detection using machine learning techniques. *2019 Global Conference for Advancement in Technology (GCAT).* https://doi.org/10.1109/GCAT47503.2019.8978372

[2] Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. *International Journal of Advanced Computer Science and Applications, 11*(12). https://doi.org/10.14569/IJACSA.2020.0111265

[3] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *2017 International Conference on Computing Networking and Informatics (ICCNI).* https://doi.org/10.1109/ICCNI.2017.8123782

[4] Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., & Rohith, R. (2020). Credit card fraud detection using machine learning algorithms. *Journal of Research in Humanities and Social Science, 8*(2), 04–11.

[5] Dheepa, V., & Dhanapal, R. (2012). Behavior-based credit card fraud detection using support vector machines. *ICTACT Journal on Soft Computing, 2*(4), 391–397. https://doi.org/10.21917/ijsc.2012.0061

[6] Dighe, D., Patil, S., & Kokate, S. (2018). Detection of credit card fraud transactions using machine learning algorithms and neural networks: A comparative study. *2018 Fourth International Conference on Computing, Communication, Control and Automation (ICCUBEA).* https://doi.org/10.1109/ICCUBEA.2018.8697799

[7] Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naïve Bayes algorithm in highly imbalanced dataset. *Journal of Discrete Mathematical Sciences and Cryptography, 24*(5), 1559–1572. https://doi.org/10.1080/09720529.2021.1969733

[8] Itoo, F., Meenakshi, & Singh, S. (2020). Comparison and analysis of logistic regression, naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology, 13*(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y

[9] Jain, Y., Tiwari, N., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. *International Journal of Recent Technology and Engineering, 7*(5S2), 402–407.

[10] Kiran, S., Guru, J., Kumar, R., Kumar, N., Katariya, D., & Sharma, M. (2018). Credit card fraud detection using naïve Bayes model based and KNN classifier. *International Journal of Advance Research, Ideas and Innovations in Technology, 4*(3).