**Credit Card Fraud Detection Project Report**

**ABSTRACT :**

Credit card plays a very important rule in today's economy. It becomes an unavoidable part of household, business and global activities. Although using credit cards provides enormous benefits when used carefully and responsibly, significant credit and financial damages may be caused by fraudulent activities. Many techniques have been proposed to confront therewith credit card fraud. However, all of these techniques have the same goal of avoiding the credit card fraud; each one has its own drawbacks, advantages and characteristics. In this paper, after investigating difficulties of credit card fraud detection, we seek to review the state of the art in credit card fraud detection techniques, datasets and evaluation criteria. The advantages and disadvantages of fraud detection methods are enumerated and compared. Furthermore, a classification of mentioned techniques into two main fraud detection approaches, namely, misuses (supervised) and anomaly detection (unsupervised) is presented. Again, a classification of techniques is proposed based on capability to process the numerical and categorical datasets. Different datasets used in literature are then described and grouped into real and synthesized data and the effective and common attributes are extracted for further usage. Moreover, evaluation employed criterions in literature are collected and discussed. Consequently, open issues for credit card fraud detection are explained as guidelines for new researchers.

**2.LITERATURE SURVEY :**

Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on neural networks, data mining and distributed data mining have been suggested. Ghosh and Reilly have proposed credit card fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labelled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and non-received issue (NRI) fraud. Recently, Syeda et al. have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in credit card fraud detection. A complete system has been implemented for this purpose. Stolfo et al. suggest a credit card fraud detection system (FDS) using Meta learning techniques to learn models of fraudulent credit card transactions. Meta learning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A Meta classifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection. They use Java agents for Meta learning (JAM), which is a distributed data mining system for credit card fraud detection A number of important performance metrics like True Positive— False Positive (TP-FP) spread and accuracy have been defined by them. Alekerov et al. present CARDWATCH, a database mining system used for credit card fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim have identified skewed distribution of data and mix of legitimate and fraudulent transactions as the two main reasons for the complexity of credit card fraud detection. Based on this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of misdetections.

**3). Value Proposition :**

Provide financial institutions(banks, credit card companies) with an advanced fraud detection system that significantly reduces fraudulent transactions, minimizing losses and enhancing customer trust.

Enable real-time detection and prevention of fraudulent activities, allowing for swift action and mitigation.

Improve the overall security of financial transactions and protect the financial well being of customers.

**4).Customer Segments:**

Financial institutions: Banks, credit card companies, and other payment processors are the primary customers. These organizations are motivated to minimize fraud losses and maintain their reputation for secure financial transactions.

E-Commerce Businesses: Online retailers and merchants can also benefit from fraud detection systems to prevent unauthorized transactions.

**5). Revenue Streams:**

Subscription or licensing model: charge financial institutions a recurring subscription fee or a licensing fee based on the volume of transactions they process through the fraud detection system.

Customization and Consultation: Offer customization of the fraud detection models based on the specific needs of each institution. Provide consultation services for optimizing and integrating the system into their existing infrastructure.

6. **Emerging Technologies:**

Keep an eye on emerging technologies that could influence fraud patterns.

For instance, as AI and machine learning techniques become more accessible, fraudsters might adapt by using more sophisticated attacks.

**7. Collaboration and Industry Trends:**

Engage with industry association attend conferences, and stay informed about the least developments and trends in fraud detection. Collaborate with other experts and professionals to gain insights into emerging fraud techniques.

Market Forecasting in credit and debit card fraud detection is crucial for staying proactive in identifying and addressing emerging fraud threats. By continuously analyzing trends, adapting strategies, and leveraging advanced analytics, business can better protect their customer and assets from evolving fraud schemes.

**Project Pipeline**

The project pipeline can be briefly summarized in the following four steps:

* **Data Understanding:** Here, we need to load the data and understand the features present in it. This would help us choose the features that we will need for your final model.
* **Exploratory data analytics (EDA):** Normally, in this step, we need to perform univariate and bivariate analyses of the data, followed by feature transformations, if necessary. For the current data set, because Gaussian variables are used, we do not need to perform Z-scaling. However, you can check if there is any skewness in the data and try to mitigate it, as it might cause problems during the model-building phase.
* **Train/Test Split:** Now we are familiar with the train/test split, which we can perform in order to check the performance of our models with unseen data. Here, for validation, we can use the k-fold cross-validation method. We need to choose an appropriate k value so that the minority class is correctly represented in the test folds.
* **Model-Building/Hyperparameter Tuning:** This is the final step at which we can try different models and fine-tune their hyperparameters until we get the desired level of performance on the given dataset. We should try and see if we get a better model by the various sampling techniques.
* **Model Evaluation:** We need to evaluate the models using appropriate evaluation metrics. Note that since the data is imbalanced it is is more important to identify which are fraudulent transactions accurately than the non-fraudulent. We need to choose an appropriate evaluation metric which reflects this business goal.

**Conclusion:**

In conclusion, this project aimed to develop a machine learning-based credit card fraud detection system to address the increasing concern of fraud in the financial sector. Through the use of various techniques such as Undersampling, Oversampling, SMOTE, and Adasyn, we were able to balance the data and build models such as Logistic Regression and XGBoost to detect fraudulent transactions. The best model was found to be the Logistic Regression model with SMOTE, as it had a high ROC score, good interpretability, and lower resource requirements compared to other models. Additionally, financial modeling was used to evaluate the feasibility and potential profitability of the project. We have also discussed various business models such as pay per use and subscriptionbased model, that could be used to monetize the developed solution.