K. S. INSTITUTE OF TECHNOLOGY, BANGALORE-560109
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
PROJECT PHASE 1 + SEMINAR (18CSP77)

PROJECT ABSTRACT SUBMISSION 7TH SEM A & B SEC 2021-22 (Odd Sem)

PROJECT TITLE: Dataset Shift Quantification for Credit Card fraud Detection using Machine learning

BatchNo.: 2021_CSE_14	Guide Name: Mr Roopesh Kumar BN	
Sl. No.	**NAME**	**USN**
1	BHAGYASHREE V	1KS18CS128
2	BI BI AYESHA	1KS18CS129
3	LIKITHA S	1KS18CS130
4	SHALINI KP	1KS18CS131

ABSTRACT

One of the greatest concerns of many business owners is how to protect their company from fraudulent activity. This concern motivated large companies to save data relative to their past frauds, however, whoever performs a fraud aims not to be caught then this kind of data usually is unlabeled or partially labeled.

we will talk about how to discover frauds on a credit card transaction dataset, unlike most fraud datasets this dataset is completely labeled however, we won't use the label to discover frauds. Credit card fraud is when someone uses another person's credit card or account information to make unauthorized purchases or access funds through cash advances. Credit card fraud doesn't just happen online; it happens in brick-and-mortar stores, too. As a business owner, you can avoid serious headaches — and unwanted publicity — by recognizing potentially fraudulent use of credit cards in your payment environment.

One of the most common approach to find fraudulent transactions was randomly select some transactions and ask and auditor to audit it. This approach was quite unaccurate since the relation between the number of fraudulent transactions and normal transactions is close to 0.1%.
Then, we aim to leverage machine learning to detect and prevent frauds and make fraud fighters more efficient and effective. Commonly, there are the supervised and the unsupervised approach:
Also, these models can then be deployed to automatically identify new instances/cases of known fraud patterns/types in the future. Ideally the validation of this type of machine learning algorith sometimes need to be a temporal validation since fraud patterns can change over time, however to simplify this article, the validation will be simplified.

The project uses a dataset of credit card transactions which have been taken from Kaggle.The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, the original features and more background information about the data are not

K. S. INSTITUTE OF TECHNOLOGY, BANGALORE-560109
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
PROJECT PHASE 1 + SEMINAR (18CSP77)

PROJECT ABSTRACT SUBMISSION 7TH SEM A & B SEC 2021-22 (Odd Sem)

provided. Features V1, V2, …, V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are "Time" and "Amount", and there are no null values

This project proposes an unsupervised approach to detect frauds, the only place the labels are used is to evaluate the algorithm. One of the biggest challenge of this problem is that the target is highly imbalanced as only 0.17% cases are fraudulent transactions. But the advantage of the representation learning approach is that it is still able to handle such imbalance nature of the problems.

The objective of the approach was fulfilled, making possible to detect frauds with an 100% unsupervised approach. Nevertheless, there are several ways to make this approach work better.

System Requirements (H/W and S/W)

Proccesor : above i5 procceser
RAM : 8GB
Hard Disk : 500 GB

Software Requirements:
Operating System: Windows v10
Virtual Box:Oracle VM Virtual Box
Python Version:Python v3.7.2
Python IDE: PyCharm

Base Paper Submitted: (Yes/No)

Provide Base Paper Details in the following format:

SampleReference:
https://towardsdatascience.com/anomaly-detection-with-practical-example-d06b90f89caf   1. https://towardsdatascience.com/fraud-detection-unsupervised-anomalydetection-df43d81fce67

 2. Dataset collected from https://www.kaggle.com/datasets

 3. Taha, S. J. Malebary, "Intelligent Approach to Credit Card Fraud Detect ion Using an OLightGBM", IEEE Access (2020), pp.25579-25587

 4. S. N. Kalid, K. H NG, G. K Tong, K. C Khore., "A Mult iple Classifiers System for Anomaly Detect ion in Credit Card Data With Unbalanced and Overlapped Classes", IEEE Access (2020), Vol. 8, pp. 28210-28221

 5. S. Makki, Z. A Assaghir, Y. Taher, R. Haque, M. S Hacid, H. Zeineddine, "An Experimental Study With Imbalanced Classificat ion Approaches for Credit Card Fraud Det ect ion", Special Sect ion On Advanced

PROJECT ABSTRACT SUBMISSION 7TH SEM A & B SEC 2021-22 (Odd Sem)

Software And Data Engineering For Secure Societ ies, IEEE Access (2019), Vol 7, pp. 93010-93022 6.
https://www.avenga.com/magazine/anomaly-detection/
Note: Not for Student Use

ACCEPTED	REJECTED	RE SUBMIT
Reason for Rejection:		
Reason for Re Submit:		

Project Coordinator   **HOD**