## **A PROJECT REPORT**

## on

**“CREDIT CARD FRAUD DETECTION SYSTEM”**

**Submitted to**

# KIIT Deemed to be University

# 

### In Partial Fulfilment of the Requirement for the Award of BACHELOR’S DEGREE IN COMPUTER SCIENCE AND SYSTEM ENGINEERING

**BY**

**SWATI THAKUR 1828214**

**UTKARSH JAISWAL 1828217**

**VIKASH KUMAR 1828218**

**UNDER THE GUIDANCE OF**

## Prof. Lalit Vashishtha



**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA - 751024**

**April 2022**

**KIIT Deemed to be University**

**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA - 751024**



## CERTIFICATE

**This is certify that the project entitled**

**“CREDIT CARD FRAUD DETECTION SYSTEM”**

**submitted by**

**SWATI THAKUR 1828214**

**UTKARSH JAISWAL 1828217**

**VIKASH KUMAR 1828218**

**is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Sci-ence & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during the year 2022-2023, under our guidance.**

**Date: / /**

**Prof. Lalit Vashishtha**

## **Acknowledgements**

**First of all, we would like to sincerely thank our supervisor Prof. Lalit Vashishtha for sharing his knowledge and expertise in this study. We would not be able to thank him enough for making himself available to even clear our silliest doubts and also for teaching and guiding us throughout the process of research.**

**We would further be thankful to the Dean of Computer Communication and Computer System, Dr. Amulya Ratan Swain, and all faculty members of School of Computer Engineering and technical assistants of School of Computer Engineering for their constant support.**

**SWATI THAKUR 1828214**

**UTKARSH JAISWAL 1828217**

**VIKASH KUMAR 1828218**

## **ABSTRACT**

**As an engineer, we always look forward to contributing our time and working towards the betterment of society so any project idea which saves our time is worth making. As today’s world is more dependent on machines.**

**Digitalization is making our daily life work more easier but here comes the disadvantage of this concept which is becoming a major threat to us as human beings.**

**Due to the pandemic's drive for contactless payments, a lack of understanding and weaknesses in sensitive card credentials are leading to an increase in digital fraud.**

**So here we came up with this project “ONLINE CREDIT CARD FRAUD DETECTION SYSTEM”.**

**Keywords:**

* ONLINE FRAUD DETECTION
* IMBALANCE DATASET
* UNDERSAMPLING TECHNIQUES
* OVERSAMPLING TECHNIQUES(SMOTE)
* REAL TIME MODEL

**Contents**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | **Introduction** | | | **1** |
| **2** | **Basic Concepts/ Literature Review** | | | **2** |
|  | **2.1** | **Sub Section Name...........................** | | **2** |
| **3** | **Problem Statement / Requirement Specifications (SRS)** | | | **3** |
|  | **3.1** | **Project Planning...........................** | | **3** |
|  | **3.2** | **Project Analysis...........................** | | **4** |
|  | **3.3** | **System Design …………………..** | | **4** |
|  |  | **3.3.1** | **Design Constraints ……** | **4** |
|  |  | **3.3.2** | **System Architecture (UML) / Block Diagram …** | **4** |
| **4** | **Implementation** | | | **5** |
|  | **4.1** | **Methodology / Proposal ...........................** | | **5** |
|  | **4.2** | **Testing / Verification Plan …………….** | | **6** |
|  | **4.3** | **Result Analysis / Screenshots ………….** | | **7** |
|  | **4.4** | **Quality Assurance ……………………..** | | **7** |
| **6** | **Conclusion and Future Scope** | | | **8** |
|  | **6.1** | **Conclusion ………………………..** | | **8** |
|  | **6.2** | **Future Scope ……………………….** | | **8** |
| **References** | | | | **9** |
| **Individual Contribution** | | | | **10-12** |
| **Plagiarism Report** | | | | **13** |

*CREDIT CARD FRAUD DETECTION SYSTEM*

**Chapter 1**

Introduction

**Fraud can be committed in many ways and in many areas. Most detection methods combine a variety of fraud detection data sets to form a connected view of valid and invalid payment data to make a decision. According to recent reports in step with data from the Reserve bank of India (RBI), the best variety of such frauds inside the said period have been reported within the ATM/debit card class which noticed eleven,058 cases concerning Rs 94.5 crores. That is observed through 6,117 credit score card frauds concerning Rs 19.7 crore, and 3,866 net banking frauds concerning Rs 13.6 crores.**

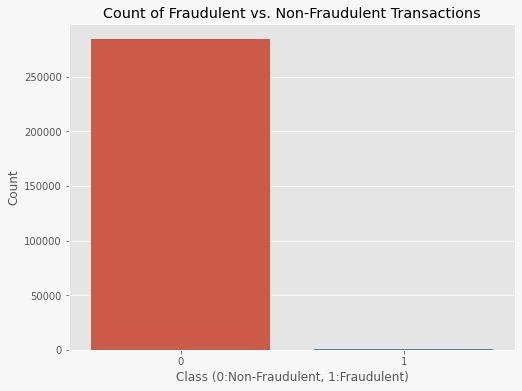


Figure 1.1 FRAUD VS NON-FRAUD

*School of Computer Engineering, KIIT, BBSR* 1

*CREDIT CARD FRAUD DETECTION SYSTEM*

**Chapter 2**

Basic Concepts/ Literature Review

**HOW DOES THIS WORK?**

The fraud detection system examines the input data and looks for patterns in prior financial records before classifying the data as fraud or non-fraud. On the dataset, we used multiple categorization algorithms and compared their results. LOGISTIC REGRESSION, RANDOM FOREST CLASSIFIER , AND SUPPORT VECTOR REGRESSION were implemented.

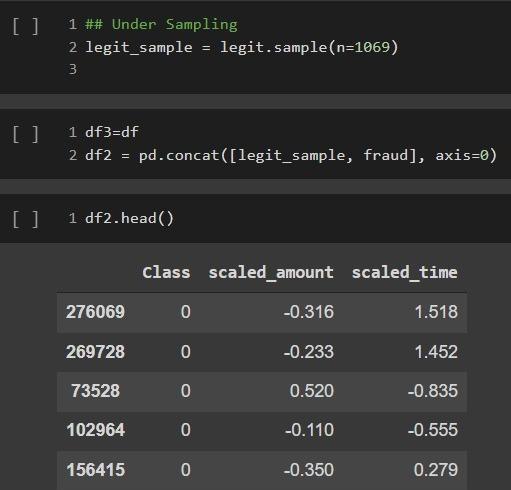
We have faced some challenges in this and few are mentioned below:

strong class imbalance.

Inclusion of the labelled and unlabeled data.

As the number of non-fraud transactions were more as compared to the fraud ones so what we have done is :

We have implemented an algorithm in which we have extracted the non-fraud transactions and then under-sampled them. In which to extract some data from non fraud entries and then mix it with the fraud transaction data. This is how under-sampling is performed on our dataset.

****

UNDER-SAMPLING

*School of Computer Engineering, KIIT, BBSR* 2

*CREDIT CARD FRAUD DETECTION SYSTEM*

**Chapter 3**

Problem Statement / Requirement Specifications

**3.1 Project Planning**

**Problem set identify**

**At first, we identified the problem statement. Due to post covid times we have observed and collected the information that cybercrimes were increasing at a very high pace so we observed some of the figures from past years and summarised them as cybercriminals have done a fraud of approximately Rs 128 crore in terms related to debit and credit cards and every day it is rising in numbers.**

**3.2 Project Analysis**

**Post covid time has pushed people to go with more of cashless transactions**

**three main challenges with card frauds analogous to dataset i.e. strong class imbalance of the data, the inclusion of labeled and unlabeled data, and to increase the ability to process a large number of transactions.**

**We have observed our dataset and then due to data imbalance, we have performed**

**two different operations on our datasets**

* **UnderSampling-Took randomized sample from majority biased class to avoid data imbalance**
* **OverSampling-Resampled the minority biased class to avoid data imbalance**

**If we are not doing oversampling and undersampling then the amount of**

**false positive was relatively high while using these two methods we have observed that when we have done oversampling then the accuracy of our model collapsed and generally we have observed that this leads to overtraining of the model which also results in unstable model.**

*School of Computer Engineering, KIIT, BBSR*  ***3***

|  |
| --- |
| **On the other hand, we have observed that when we have trained our dataset by doing undersampling on the fraud data the accuracy was about 78 percent this is how we have avoided the overtraining of our data and hence we are not using some of the time taking mechanisms hence solving one of the major problems that are results in real-time. We have observed that the number of fraud cases was more during day time and more frauds were observed under the transaction of Rs 25000.**  **We have done bucketization then we have plotted the amount then observed the amount of cluster forms then we have used the cluster to identify the amount group to which it belongs.**  **Problems faced in this statement**  **There are majorly 3 main challenges:**   * **Strong Imbalance Dataset** * **Non-anonymized features** * **Class biased feature**     Non-Anonymized features  **3.3 System Design**  **3.3.1 Design Constraints**  **We have used-:**   * **Jupyter notebook** * **Google colaboratory** * **Python libraries**  1. **numpy** 2. **pandas** 3. **seaborn** 4. **matplotlib** 5. **sklearn** 6. **scipy**   **for designing this model.**  *School of Computer Engineering, KIIT, BBSR*  ***4*** |

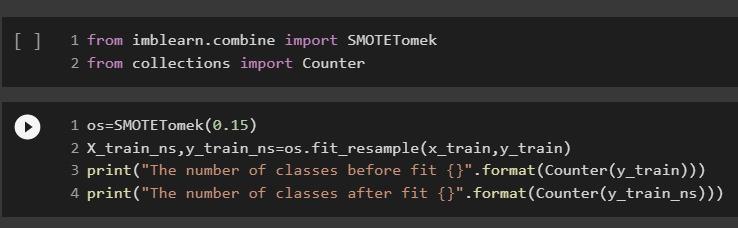
*CREDIT CARD FRAUD DETECTION SYSTEM*

**Chapter 4**

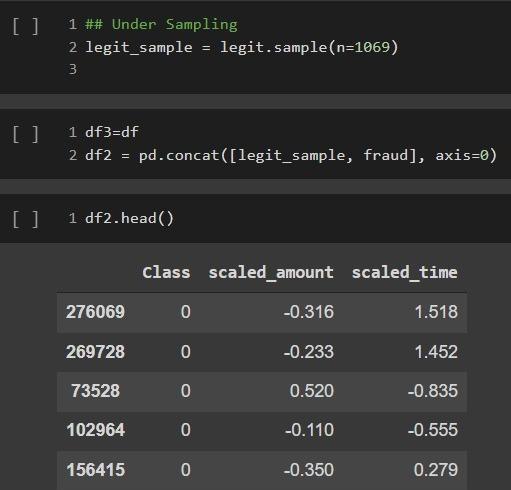
**Implementation**

**4.1 Methodology OR Proposal**

**In order to avoid the biases in the dataset, we have performed undersampling and oversampling.**



Over-Sampling(SMOTE)

****

Under-Sampling

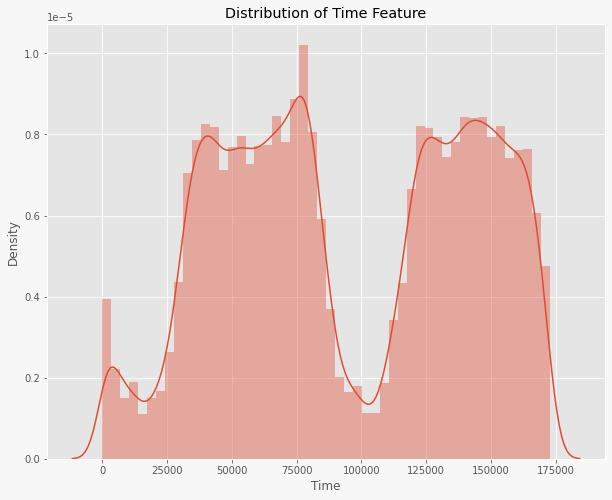
*School of Computer Engineering, KIIT, BBSR* 5

**4.2 Testing OR Verification Plan**

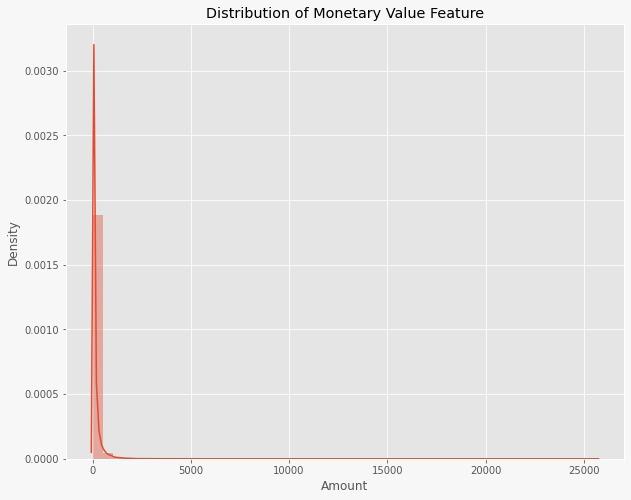
We were provided the data of two days which includes the time column also.

We have used this time as a feature for training our model but by doing some modifications in it.

Actually, the timestamp was given in seconds and we changed that two days timestamp to one day timestamp and then included it in our feature.

****

Before modification

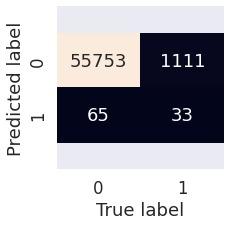
****

After modification

We tried two methods undersampling and oversampling(SMOTE) for training and as undersampling was giving us more accuracy.In undersampling, we undersample the majority biased data to reduce the data imbalancement.***School of Computer Engineering, KIIT, BBSR* 6**

**4.3 Result Analysis OR Screenshots**

**In this subsection, the output of the experiment or study in terms of some graphs, plots must be presented. Also, if some implementation is done then its screenshots can be presented here, so as to showcase the proof of the output.**

****

**OVERSAMPLING**

**UNDERSAMPLING**

**4.4 Quality Assurance**

**In the working organization, if some department is there to verify the quality of your work, they can produce a certificate or guidelines followed.**

***School of Computer Engineering, KIIT, BBSR* 7**

***CREDIT CARD FRAUD DETECTION SYSTEM***

**Chapter 5**

**Conclusion and Future Scope**

**6.1 Conclusion**

Credit card fraud is the most common form of fraud occurring these days. It’s a complex issue that needs perfect planning to deploy and detect it, and then our machine learning The model comes into the game.

Many algorithms like Random Forest,Logistic Regression,XGB Boost,Support Vector Regressor were used in building this model, where Random forest gives the best accuracy without overfitting and overtraining of our model.

**6.2 Future Scope**

For detecting fraudulent transactions with non-anonymized features would make this captivating as outputting the feature would enable us to see what fixed factors are most important for detection.

It’s a model of machine learning for the good, which makes sure that the customer’s money is safe and not easily interfered with.

***School of Computer Engineering, KIIT, BBSR*** 8

*CREDIT CARD FRAUD DETECTION SYSTEM*

Reference

Schaffer, C. 1993. Overfitting avoidance as bias. Machine Learning 10:153-178.

Japkowicz, N. 2000. The class imbalance problem: Significance and strategies. In Proceedings of the 2000 International Conference on Artificial Intelligence (ICAI 2000).

Fawcett, T. E., and Provost, F. 1997. Adaptive fraud detection. Data Mining and Knowledge Discovery 1(3):291-316.

Jon TS Quah and M Sriganesh. Real-time credit card fraud detection using computational Intelligence. Expert Systems with Applications, 35(4):1721-1732, 2008.

Tej Paul Bhatla, Vikram Prabhu, and Amit Dua. Understanding credit card frauds. Cards business review, 1(6), 2003.

Sahin, Y., & Duman, E. (2011). Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. Hong Kong, China: The International MultiConference of Engineers and Computer Scientists.

Huang, S. (2013). Fraud Detection Model by Using Support Vector Machine Techniques. Chiayi, Taiwan: International Journal of Digital Content Technology & its Applications.

L.J.P. van der Maaten and G.E. Hinton, Visualizing High-Dimensional Data Using t-SNE (2014), Journal of Machine Learning Research.

Nathalie Japkowicz, Learning from Imbalanced Data Sets: A Comparison of Various Strategies (2000), AAAI Technical Report WS-00–05.

FW-SMOTE: A feature-weighted oversampling approach for imbalanced classification(2022)

Sebastián Maldonadoad, Carla Vairettibd Alberto, Fernandezc Francisco Herrerac.

Machine Learning based Prediction of Dropout Students from the Education University using SMOTE(2022) IEEE - M. Revathy; S. Kamalakkannan; P. Kavitha

# Machine Learning-based Approach on Dealing with Binary Classification Problem in Imbalanced Financial Data(2022), IEEE - Dennis; Ivan Reginald Budianto; Russell Kenael Azaria; Alexander A.S Gunawan

*School of Computer Engineering, KIIT, BBSR* 9

*CREDIT CARD FRAUD DETECTION SYSTEM*

**INDIVIDUAL CONTRIBUTION REPORT 1:**

**CREDIT CARD FRAUD DETECTION SYSTEM**

SWATI THAKUR

1828214

**Abstract:** As an engineer, we always look forward to contributing our time and working towards the betterment of society so any project idea which saves our time is worth making. As today’s world is more dependent on machines.

Digitalization is making our daily life work more easier but here comes the disadvantage of this concept which is becoming a major threat to us as human beings.

Due to the pandemic's drive for contactless payments, a lack of understanding and weaknesses in sensitive card credentials are leading to an increase in digital fraud.

So here we came up with this project “ONLINE CREDIT CARD FRAUD DETECTION SYSTEM”.

**Individual contribution and findings:** I embolden my subordinate team members to brainstorm ideas and input for our project. I have deployed significant time to guide my team members on various tasks all along the project. I did all the research part of the project, which is my expertise which helped us greatly throughout the project cycle.

I have manifested strong leadership skills throughout the project by assigning, managing and assisting in various aspects of the project throughout its life cycle. I made sure that my subordinates worked seamlessly and I have made this report which comprises all the work which me and my team have done throughout this project.

**Individual contribution to project report preparation:** My role in the project report preparation was to create the “standard adoption” and “conclusion and future scope’’ part of the report.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

SWATI THAKUR

*School of Computer Engineering, KIIT, BBSR* 10

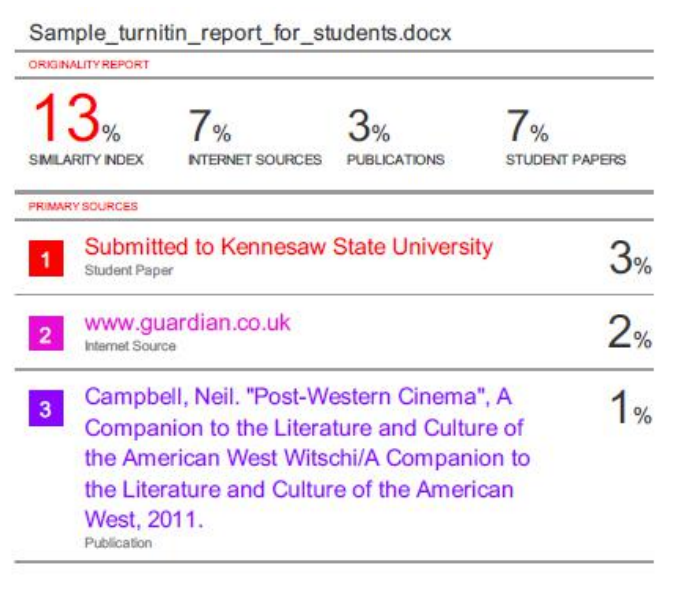
|  |
| --- |
| *CREDIT CARD FRAUD DETECTION SYSTEM*  **INDIVIDUAL CONTRIBUTION REPORT 2:**  **CREDIT CARD FRAUD DETECTION SYSTEM**  UTKARSH JAISWAL  1828217  **Abstract:** As an engineer, we always look forward to contributing our time and working towards the betterment of society so any project idea which saves our time is worth making. As today’s world is more dependent on machines.  Digitalization is making our daily life work more easier but here comes the disadvantage of this concept which is becoming a major threat to us as human beings.  Due to the pandemic's drive for contactless payments, a lack of understanding and weaknesses in sensitive card credentials are leading to an increase in digital fraud.  So here we came up with this project, “ONLINE CREDIT CARD FRAUD DETECTION SYSTEM”.  **Individual contribution and findings:**I took over the coding part of the problem, where we first read various identified problem statements which we had identified, then we read various research papers corresponding to the problem statement.  I then gathered all the information which we received from the research papers and drew out a solution to those problem statements and implemented them through coding, where we observed that our findings were giving us fruitful results which worked.    **Individual contribution to project report preparation:** My role in the project report preparation was to create the “Implementation” and “References’’ part of the report.    Full Signature of Supervisor: Full signature of the student:  ……………………………. ……………………………..  UTKARSH JAISWAL  *School of Computer Engineering, KIIT, BBSR* 11 |

|  |
| --- |
| *CREDIT CARD FRAUD DETECTION SYSTEM*  **INDIVIDUAL CONTRIBUTION REPORT 2:**  **CREDIT CARD FRAUD DETECTION SYSTEM**  VIKASH KUMAR  1828218  **Abstract:** As an engineer, we always look forward to contributing our time and working towards the betterment of society so any project idea which saves our time is worth making. As today’s world is more dependent on machines.  Digitalization is making our daily life work more easier but here comes the disadvantage of this concept which is becoming a major threat to us as human beings.  Due to the pandemic's drive for contactless payments, a lack of understanding and weaknesses in sensitive card credentials are leading to an increase in digital fraud.  So here we came up with this project, “ONLINE CREDIT CARD FRAUD DETECTION SYSTEM”.  **Individual contribution and findings:**I took over the coding part of the problem, where we first read various identified problem statements which we had identified, then we read various research papers corresponding to the problem statement.  I then gathered all the information which we received from the research papers and drew out a solution to those problem statements and implemented them through coding, where we observed that our findings were giving us fruitful results which worked.    **Individual contribution to project report preparation:** My role in the project report preparation was to create the “Implementation” and “References’’ part of the report.    Full Signature of Supervisor: Full signature of the student:  ……………………………. ……………………………..  VIKASH KUMAR  *School of Computer Engineering, KIIT, BBSR* 12 |

TURNITIN PLAGIARISM REPORT

**(This report is mandatory for all the projects and plagiarism**

**must be below 25%)**



*School of Computer Engineering, KIIT, BBSR* 13