

## **CREDIT CARD FRAUD DETECTION**

### **A COURSE PROJECT REPORT**

*Submitted by*

**SANDEEP REDDY (RA2011027010165)**

**SHASHI AKSHITH (RA2011027010172)**

**SIVANESH G (RA2011027010177)**

*Under the guidance of*

**Dr. A. Shanthini**

*In partial fulfilment for the Course*

*of*

**Data science (18CSE396T)**

*in*

**DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS**



**SRM**

INSTITUTE OF SCIENCE & TECHNOLOGY

*Deemed to be University u/s 3 of UGC Act, 1956*

**SCHOOL OF COMPUTING**

**COLLEGE OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

*(Deemed to be University u/s 3 of UGC Act, 1956)*

**KATTANKULATHUR - 603 203**

**November, 2022**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

*(Under Section 3 of UGC Act, 1956)*

**BONAFIDE CERTIFICATE**

Certified that this mini project titled “**CREDIT CARD FRAUD DETECTION**” is the bonafide work of **SANDEEP REDDY (RA2011027010165), SHASHI AKSHITH (RA2011027010172), SIVANESH G(RA2011027010177)**, who carried out the project work under my supervision.

**SUPERVISOR**

Dr. A. Shanthini  
Associate Professor  
Department of Data Science and Business Systems  
SRM Institute of Science and Technology  
Kattankulathur – 603 203

**HEAD OF THE DEPARTMENT**

Dr. M. LAKSHMI  
Professor & Head  
Department of Data Science and Business Systems  
SRM Institute of Science and Technology  
Kattankulathur – 603 203

## **ABSTRACT**

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyze frauds in online transactions. The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyze the past transaction details of the customers and extract the behavioral patterns. Where cardholders are clustered into different groups based on their transaction amount. Then using sliding window strategy, to aggregate the transaction made by the cardholders from different groups so that the behavioral pattern of the groups can be extracted respectively. Later different classifiers are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift . In this paper, we worked with European credit card fraud dataset. Due to cashless Transaction every people use ATM card and credit card for transaction, so fraud can also be increase. Billions of dollars of loss are caused every year by fraudulent credit card transactions. The design of efficient fraud detection algorithms is key for reducing these losses, and more and more algorithms rely on advanced machine learning techniques to assist fraud investigators.

## **ACKNOWLEDGEMENT**

We express our heartfelt thanks to our honorable **Vice Chancellor Dr. C. MUTHAMIZHCHELVAN**, for being the beacon in all our endeavors.

We would like to express my warmth of gratitude to our **Registrar Dr. S. Ponnusamy**, for his encouragement

We express our profound gratitude to our **Dean (College of Engineering and Technology) Dr. T. V.Gopal**, for bringing out novelty in all executions.

We would like to express my heartfelt thanks to Chairperson, School of Computing **Dr. Revathi Venkataraman**, for imparting confidence to complete my course project

We wish to express my sincere thanks to **Course Audit Professor and Course Coordinator** for their constant encouragement and support.

We are highly thankful to my Course project Faculty **Dr. A. Shanthini , Associate Professor, Department of Data Science and Business Systems**, for his assistance, timely suggestion and guidance throughout the duration of this course project.

We extend our gratitude to our **HoD, Dr. M. Lakshmi, Professor, Department of Data Science and Business Systems**, and my Departmental colleagues for their Support.

Finally, we thank our parents and friends near and dear ones who directly and indirectly contributed to the successful completion of our project. Above all, I thank the almighty for showering his blessings on me to complete my Course project.

## **TABLE OF CONTENTS**

<b>CHAPTERS</b>	<b>CONTENTS</b>	<b>PAGE NO.</b>
1.	<b>ABSTRACT</b>	
2.	<b>INTRODUCTION</b>	
3.	<b>LITERATURE SURVEY</b>	
4.	<b>REQUIREMENT ANALYSIS</b>	
5.	<b>ARCHITECTURE &amp; DESIGN</b>	
6.	<b>DATA SET DESCRIPTION</b>	
7.	<b>IMPLEMENTATION</b>	
8.	<b>METHOD/ALGORITHM/MODEL USED</b>	
9.	<b>EXPERIMENT RESULTS &amp; ANALYSIS</b>	
	9.1. <b>RESULTS</b>	
	9.2. <b>RESULT ANALYSIS</b>	
10.	<b>CONCLUSION &amp; FUTURE ENHANCEMENT</b>	
11.	<b>REFERENCES</b>	

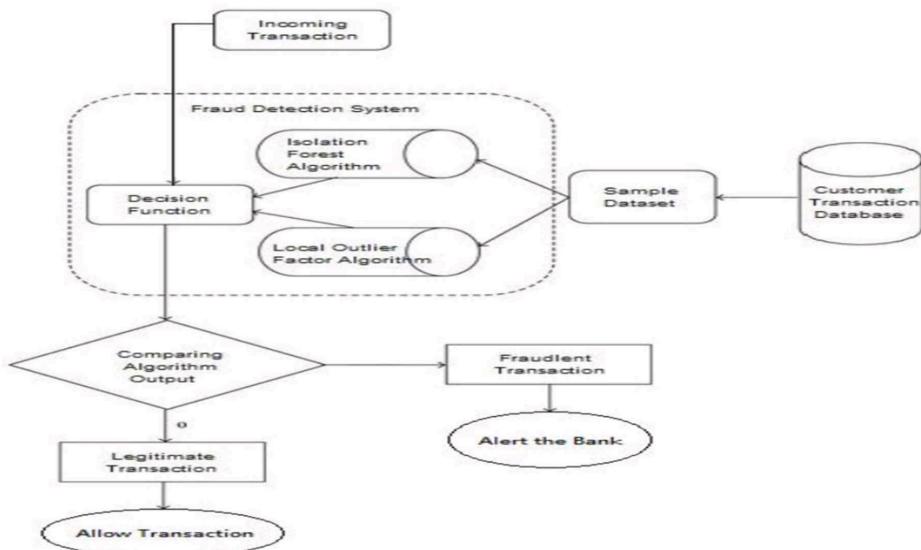
## 2. INTRODUCTION

### 1.1 Scenario Description

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behavior of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future. In other words, Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting. This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time.

Fraud detection methods are continuously developed to defend criminals in adapting to their fraudulent strategies. These frauds are classified as:

- Credit Card Frauds: Online and Offline
- Card Theft • Account Bankruptcy
- Device Intrusion
- Application Fraud
- Counterfeit Card
- Telecommunication Fraud



### **3. LITERATURE SURVEY**

- Fraud act as the unlawful or criminal deception intended to result in financial or personal benefit. It is a deliberate act that is against the law, rule or policy with an aim to attain unauthorized financial benefit. Numerous literatures pertaining to anomaly or fraud detection in this domain have been published already and are available for public usage.
- A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. In another paper, Suman, Research Scholar, GJUS&T at Hisar HCE presented techniques like Supervised and Unsupervised Learning for credit card fraud detection. Even though these methods and algorithms fetched an unexpected success in some areas, they failed to provide a permanent and consistent solution to fraud detection.
- A similar research domain was presented by Wen-Fang YU and Na Wang where they used Outlier mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment of credit card transaction data set of one certain commercial bank.
- Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that aren't genuine. They have taken attributes of customer's behaviour and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value.
- Unconventional techniques such as hybrid data mining/complex network classification algorithm is able to perceive illegal instances in an actual card transaction data set, based on network reconstruction algorithm that allows creating representations of the deviation of one instance from a reference group have proved efficient typically on medium sized online transaction. There have also been efforts to progress from a completely new aspect. Attempts have been made to improve the alert feedback interaction in case of fraudulent transaction.
- In case of fraudulent transaction, the authorized system would be alerted and a feedback would be sent to deny the ongoing transaction. Artificial Genetic Algorithm, one of the approaches that shed new light in this domain, countered fraud from a different direction. It proved accurate in finding out the fraudulent transactions and minimizing the number of false alerts.

## **4. REQUIREMENTS**

### **4.1 Requirement Analysis**

- Dataset of Credit Card and their Hex Code values
- Python Libraries (OpenCV, Pandas)
- LOGISTIC REGRESSION

### **4.2 Hardware Requirement**

RAM: 4GB and Higher

Processor: Intel i3 and above

Hard Disk: 500GB: Minimum

### **4.3 Software Requirement**

OS: Windows or Linux

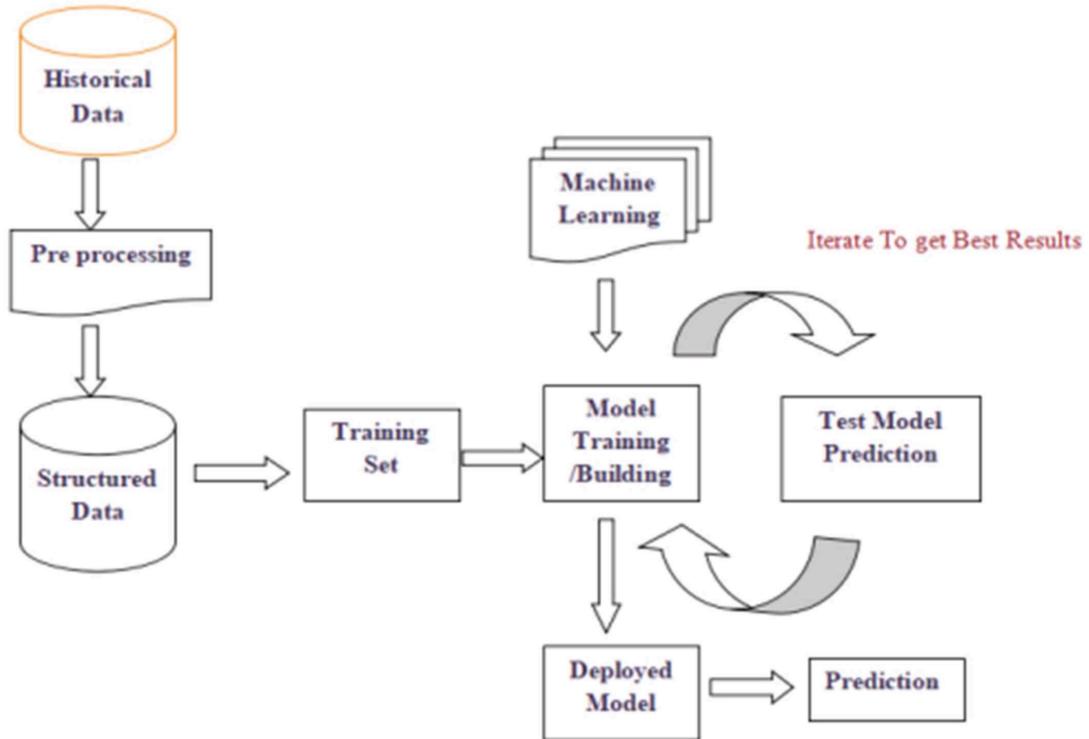
Python IDE: python 2.7.x and above

Jupyter Notebook

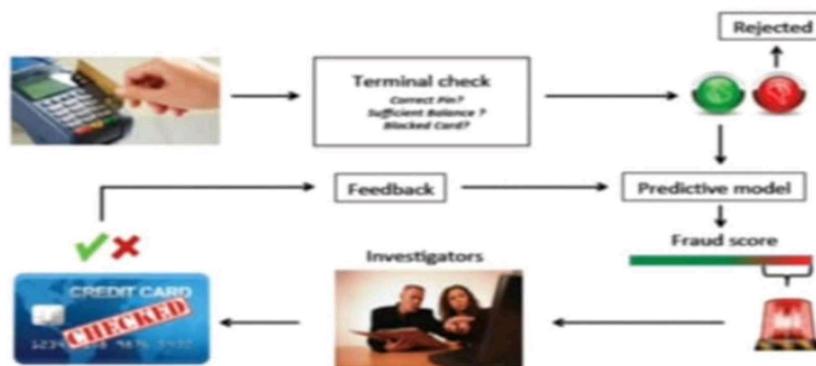
Setup tools and pip to be installed for 3.6 and above

Language : Python

## 5. ARCHITECTURE & DESIGN



### Fraud detection process



## 6. DATA SET DESCRIPTION

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. 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, we cannot provide the original features and more background information about the data.

<https://www.kaggle.com/code/renjithmadhavan/credit-card-fraud-detection-using-python>

## using-python

Time	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	Class		
2	0	-1.35881	-0.07278	1.538437	1.781515	-0.38832	0.462388	0.239599	0.098686	0.583787	0.091704	-0.5166	-0.6178	0.91939	-0.1137	1.448177	-0.4704	0.207971	0.25791	0.408989	0.251421	0.20183	0.277883	-0.11047	0.066938	0.123539	-0.18911	0.133558	-0.02105	149.82	0						
3	0	1.09187	0.26511	0.16468	0.44584	0.00168	-0.0788	0.05102	0.05453	-0.16697	0.162127	0.160525	0.04905	-0.14377	0.144837	-0.1148	-0.1836	-0.13548	-0.05690	0.25578	-0.061288	0.13579	0.153895	-0.03889	0.04742	0.13579	0.153895	-0.03889	0.04742	0.13579	0.153895	-0.03889	0.04742	0.13579	0.153895	0.03889	0.04742
4	1	-1.35835	-0.14400	1.73209	0.37978	-0.5021	0.18049	0.719481	0.247637	-0.15146	0.207643	0.62450	0.06064	0.71293	-0.28908	0.11996	-0.21236	-0.26186	0.2494	0.247998	0.771679	0.071679	0.68628	-0.32764	-0.1391	-0.05535	-0.05975	378.66	0								
5	1	-0.98662	-0.18523	1.79393	-0.8638	-0.10291	0.147203	0.27369	0.37746	-0.3872	0.05495	0.22649	0.17828	0.0577	-0.2872	0.03451	-0.13645	-0.08495	0.1689	0.167575	0.12302	0.2084	0.1083	0.0502	0.19032	-0.17558	0.647373	0.22193	0.62713	0.06458	123.5	0					
6	1	-1.35823	-0.07789	1.54781	0.40303	-0.0477	0.095921	0.592541	0.20730	0.01777	0.53074	0.28824	0.13848	0.134852	-0.134852	0.03451	-0.13645	-0.08495	0.03451	0.084547	0.0943	0.0943	0.084547	0.0943	0.084547	0.0943	0.084547	0.0943	0.084547	0.0943	0.084547	0.0943	0.084547	0.0943	0.084547		
7	2	-0.42597	0.06524	0.141109	0.16825	0.02497	-0.02793	0.07460	0.265041	-0.56867	0.13471	0.14262	0.03598	0.35989	-0.1713	0.05173	0.04671	-0.02767	0.05173	0.04671	0.05173	0.04671	0.05173	0.04671	0.05173	0.04671	0.05173	0.04671	0.05173	0.04671	0.05173	0.04671	0.05173	0.04671			
8	4	1.22958	0.10004	0.04571	0.20215	0.18818	0.27720	-0.05516	0.18313	0.46496	-0.09923	0.14911	-0.1538	0.07707	0.05149	0.44359	-0.0281	0.1619	-0.04558	0.18196	-0.16772	0.05149	-0.2734	0.034507	0.051588	4.99	0										
9	7	-0.64427	0.14794	1.70478	-0.4922	0.08494	0.28118	0.11631	0.80786	0.615357	0.19374	0.16397	0.29147	0.157764	-0.13287	0.08613	0.07163	-0.12223	0.08613	0.07163	0.08613	0.07163	0.08613	0.07163	0.08613	0.07163	0.08613	0.07163	0.08613	0.07163	0.08613	0.07163	0.08613				
10	7	-0.89429	0.26515	-0.13199	-0.27113	0.46699	0.72188	0.37045	0.05182	-0.39025	-0.10403	-0.7052	-0.11045	0.28652	0.04755	-0.32878	-0.21008	-0.49977	0.118765	0.70328	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	0.073205	
11	9	-0.38362	1.11598	0.144367	-0.22211	0.04939	-0.24676	0.15580	0.05659	0.05659	0.73673	0.16688	0.17616	0.38684	0.04452	0.04529	0.05193	0.05193	0.04452	0.04529	0.04677	0.04677	0.04677	0.04677	0.04677	0.04677	0.04677	0.04677	0.04677	0.04677	0.04677						
12	10	0.44904	-1.17653	0.191386	-0.17357	-0.17194	-0.62519	-0.42324	0.04458	-0.17204	0.162659	0.19944	-0.67144	-0.31959	0.05939	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197	0.03197						
13	10	0.38478	0.61630	0.09461	0.29464	0.12584	0.31707	0.04705	0.53847	0.13082	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847	0.05384	0.53847						
14	12	0.24999	-1.21264	0.38393	-0.1249	-0.5373	-0.6893	0.2479	-0.09401	0.12373	0.27666	-0.22766	-0.12047	0.2181	-0.072567	-0.08819	0.04779	-0.08819	0.04779	-0.08819	0.04779	-0.08819	0.04779	0.04779	0.04779	0.04779	0.04779	0.04779	0.04779	0.04779	0.04779	0.04779					
15	11	0.16953	0.37822	0.17523	0.17524	0.17525	0.17526	0.17527	0.17528	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529	0.17529							
16	12	-0.27195	-0.32778	1.64175	1.76747	-0.10369	0.070598	-0.4228	0.07101	0.75774	0.15304	0.16343	0.05355	0.19348	-0.0281	0.04056	-0.30308	-0.15858	0.07285	0.221868	0.15136	0.15136	0.082317	0.15136	-0.15558	-0.15558	-0.15558	-0.15558	-0.15558	-0.15558	-0.15558	-0.15558	-0.15558	-0.15558			
17	13	-0.17053	0.34864	0.18891	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991	0.08991	-0.08991							
18	13	-0.05743	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895	0.18895							
19	13	-0.43691	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896	0.18896							
20	14	-0.54012	0.45010	1.18035	1.73869	0.24109	0.17641	-0.15974	-0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974	0.15974						
21	21	0.42936	-0.12307	0.24598	-0.45795	-0.13803	-0.55496	-0.17394	-0.08048	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310	0.18310						
22	16	0.56485	-0.18562	0.10223	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519	0.14519							
23	17	0.16749	0.33846	-0.17148	0.21023	0.11566	0.19968	0.06071	-0.08368	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847	0.17847						
24	17	0.24789	0.17766	0.11847	-0.0926	-0.1348	-0.1502	-0.04636	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474						
25	26	-0.19455	0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557	-0.04557							
26	23	1.17352	0.33489	0.28905	0.13555	-0.17258	0.91605	0.38603	0.37275	-0.24648	0.04544	-0.15975	0.79735	0.24104	0.78148	-0.05306	-0.35906	-0.27087	0.07603	0.27182	-0.15049	0.43550	0.724825	-0.33708	0.03842	0.30041	0.4188	0.04563	0.22133	0.04563	0.04563	0.04563					
27	22	-0.33207	0.17347	0.45457	0.45763	-0.08376	-0.17428	0.02033	-0.24648	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044	0.18044						
28	30	-0.31402	0.09457	0.27443	0.17474	0.07441	0.20033	0.07422	0.02923	-0.05393	0.04516	-0.02409	0.18769	0.07417	-0.15587	0.12492	-0.5228	-0.24848	-0.32385	-0.03771	0.04493	0.23108	-0.15587	0.03885	0.442322	0.273643	0.03885	0.442322	0.273643	0.03885	0.442322	0.273643	0.03885	0.442322			
29	31	0.15987	0.17562	1.76131	1.86111	-0.178	0.7863	0.45763	-0.07670	0.04094	-0.0674	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474	0.15474							
30	24	1.23749	0.16040	0.3																																	

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI
2723	-0.2681	1.357871	0.754686	-0.142666	0.3501	0.178405	-0.0034	0.86766	-0.41293	-0.39683	1.570941	0.199014	-1.7859	0.356648	0.836346	1.61035	0.510176	0.31546	-0.55112	0.10571	-0.16199	0.49882	0.159564	-0.3805	-0.314	1.12504	0.259988	0.04192	9.89	0		
2742	-0.2667	-0.32146	1.48580	0.40488	1.3184	-0.4824	-0.75534	0.490204	-0.5425	0.48701	-0.4947	-0.1719	1.84598	-0.58192	0.34767	-0.2393	0.526886	-0.15126	-0.15623	0.01551	0.05021	0.04467	-0.0007	0.0556	0.16349	0.41212	0.16593	0.02844	66	0		
2743	-0.2659	0.29340	1.19487	0.14084	0.582526	-0.2911	-0.58537	0.0995	0.06105	-0.0287	-0.0284	1.09245	0.099623	-0.2861	0.05681	0.57658	0.75228	-0.0747	0.21218	-0.10827	-0.26515	0.87984	0.075447	-0.278	0.17554	0.09886	-0.037	0.1792	19.99	0		
2747	-0.2673	0.281886	0.253826	0.755282	0.3114	0.17492	0.246627	-0.3785	-0.4758	0.18383	1.19947	0.74684	0.797010	-0.3579	0.92938	-0.32573	0.3873	0.08848	0.04248	-0.00313	0.18752	0.03042	-0.04241	0.30627	0.22028	0.561473	-0.3924	-0.475	57.75	0		
2748	-0.2674	0.047742	0.11877	-0.35653	0.21518	-0.20518	-0.00588	0.05704	0.07065	0.24115	-0.30732	-0.2517	0.28671	-0.0273	0.0517	0.0273	0.04467	0.07773	0.28709	0.07519	-0.10841	0.08649	-0.19167	0.1135538	0.11887	0.031988	9.98	0				
12044	-0.2684	-0.30893	0.885365	0.151249	-0.40463	0.59084	0.358943	0.141562	-0.3959	-0.21877	0.02874	-0.02701	-0.59878	-0.1584	-0.8558	0.15254	-0.96038	0.04704	-0.02607	-0.47039	0.6946	-0.02441	0.45887	0.080607	0.02175	0.278	0					
2723	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2719	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2741	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2742	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2743	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2744	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2745	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2746	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2747	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2748	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2749	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2750	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2751	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2752	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2753	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2754	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2755	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2756	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2757	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2758	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2759	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2760	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2761	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2762	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2763	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2764	-0.2671	0.2119	1.151298	0.698043	0.18804	0.1951	0.11881	1.1367	0.81303	0.677598	-0.4868	0.47658	0.352616	-0.8752	0.15709	0.12547	-0.10533	0.19939	-0.58984	0.14007	0.038158	-0.375	-0.316	0.06784	-0.17842	0.17868	0.3029	0.29193	0.02078	21.97	0	
2765	-0.2671	0.2119																														

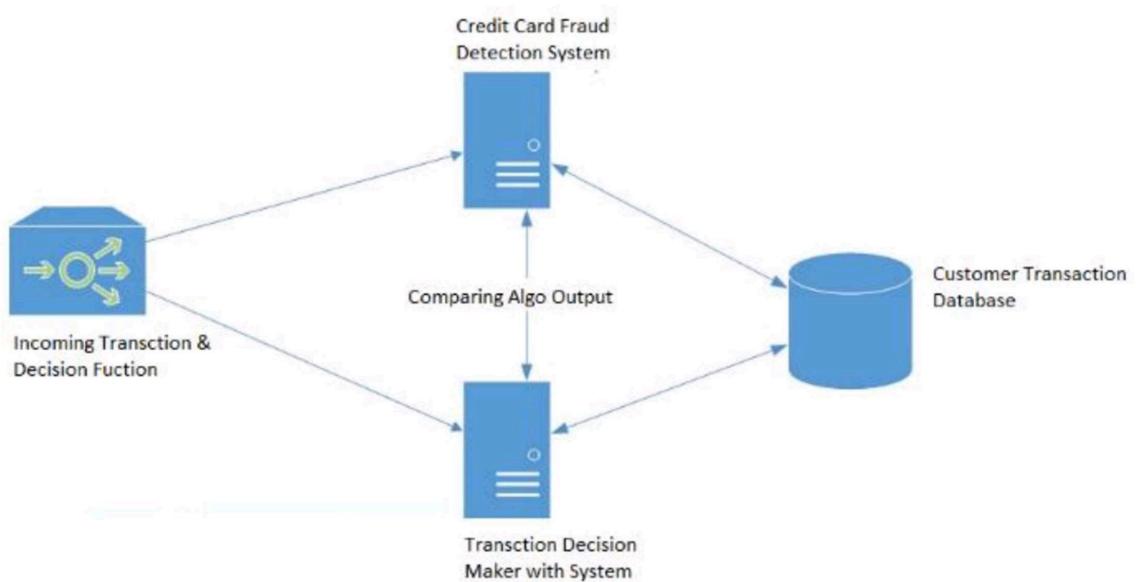
We are using this data set to detect the credit card frauds happening in the whole system.

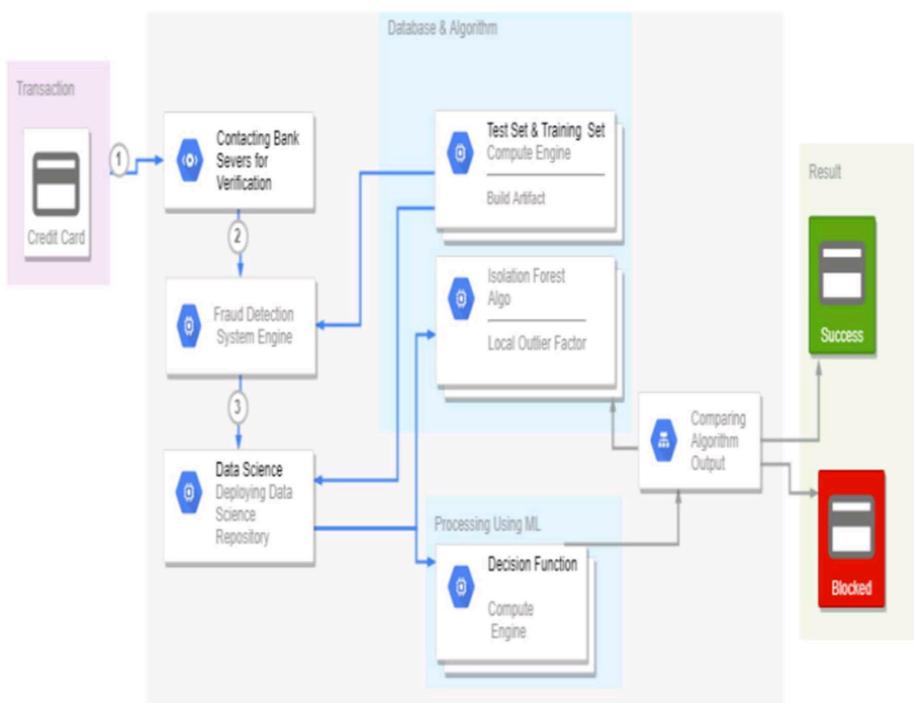
## 7. METHOD/ALGORITHM/MODEL USED

Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set.

A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables. For example, a logistic regression could be used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted or not to a particular college. These binary outcomes allow straightforward decisions between two alternatives.

A logistic regression model can take into consideration multiple input criteria. In the case of college acceptance, the logistic function could consider factors such as the student's grade point average, SAT score and number of extracurricular activities. Based on historical data about earlier outcomes involving the same input criteria, it then scores new cases on their probability of falling into one of two outcome categories.





The approach that this paper proposes, uses the latest machine learning algorithms to detect anomalous activities, called outliers. First of all, we obtained our dataset from Kaggle, a data analysis website which provides datasets. Inside this dataset, there are 31 columns out of which 28 are named as v1-v28 to protect sensitive data. The other columns represent Time, Amount and Class.

## **8.IMPLEMENTATION**

### **Define project goals, measurement metrics and assign resources**

- What are the fraud cases that we wanted to identify?
- What kind of analytics techniques have already been implemented to combat fraud?
- What are the key measurement metrics that we wanted to focus on when assessing the effectiveness of our fraud detection system?

### **Identify proper data sources**

The common data sources for detecting fraud includes:

- client profile
- risk profile
- product usage
- billing data

### **Develop the data engineering, transformation and modelling pipelines**

- For the data engineering pipeline, we need to ingest and merge the data from different sources, aggregate the data based on business metrics, and set up batch processes.
- For the data transformation pipeline, the main goal is to improve the data quality, deal with data issues such as missing & incorrect data and convert the data so that it could be fed into machine learning models.
- For the machine learning model pipeline, we focus on building and comparing diversified ML models based on key business metrics. A module for automated model accuracy testing and re-training is a necessity in the production environment to avoid model drifting issue.

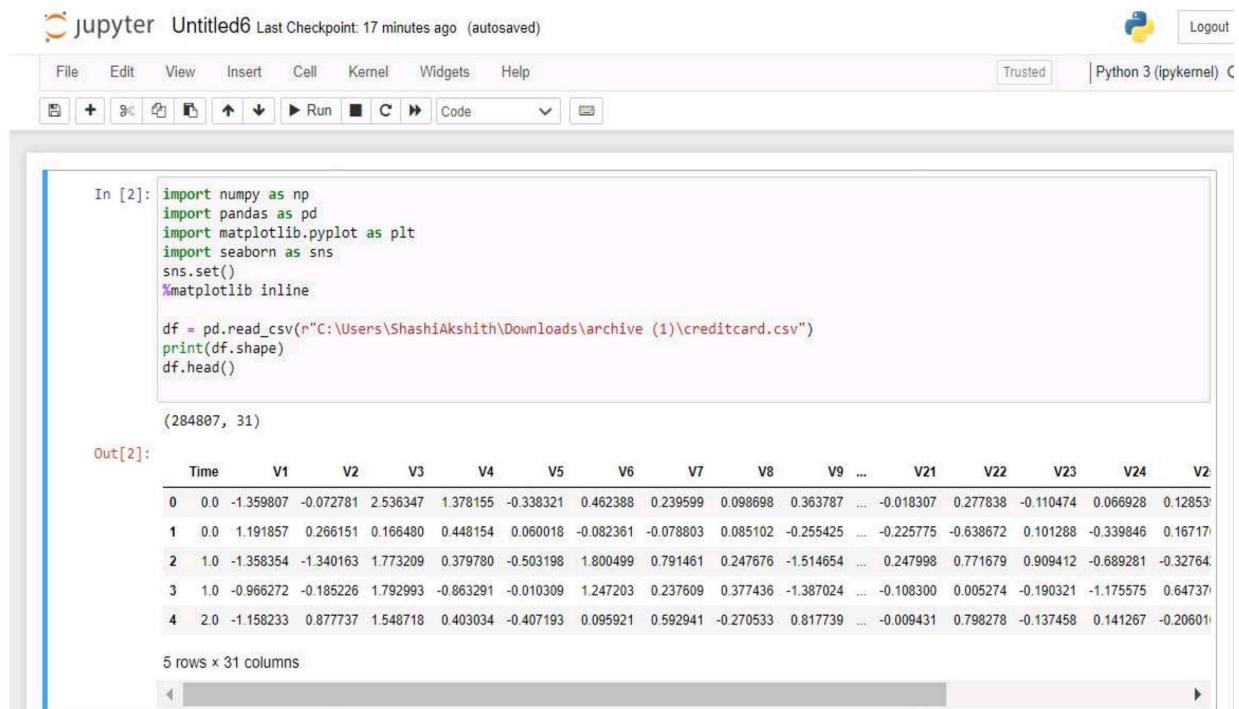
### **Integrate the model into the case management system**

The final step is to incorporate our best performing ML model into the case management system. We can rank the risk level of individual case based on the risk score that we generated. Then, a list of highly suspicious cases will be sent and assigned to relationship managers for further review through the case management system.

## 9. RESULTS AND DISCUSSION

The code prints out the number of false positives it detected and compares it with the actual values. This is used to calculate the accuracy score and precision of the algorithms. The fraction of data we used for faster testing is 10% of the entire dataset. The complete dataset is also used at the end and both the results are printed. These results along with the classification report for each algorithm is given in the output as follows, where class 0 means the transaction was determined to be valid and 1 means it was determined as a fraud transaction. This result matched against the class values to check for false positives.

Before proceeding with the Random Under Sampling technique we have to separate the original data frame. Why? for testing purposes, remember although we are splitting the data when implementing Random Under Sampling or Oversampling techniques, we want to test our models on the original testing set not on the testing set created by either of these techniques. The main goal is to fit the model either with the that were under sample and oversample (in order for our models to detect the patterns), and test it on the original testing set. Random Under Sampling is basically consists of removing data in order to have a more balanced dataset and thus avoiding our models to overfitting.



The screenshot shows a Jupyter Notebook interface with the following details:

- Header:** jupyter Untitled6 Last Checkpoint: 17 minutes ago (autosaved)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, Python 3 (ipykernel) C, Logout
- In [2]:** Contains Python code for importing numpy, pandas, matplotlib.pyplot, and seaborn, setting sns.set(), and reading a CSV file named creditcard.csv. It also prints the shape of the DataFrame and its head.
- Out [2]:** Displays the first five rows of the DataFrame. The columns are labeled Time, V1, V2, V3, V4, V5, V6, V7, V8, V9, ..., V21, V22, V23, V24, and V25. The data shows numerical values for each row, with the first row being (0.0, -1.359807, -0.072781, 2.536347, 1.378155, -0.338321, 0.462388, 0.239599, 0.098698, 0.363787, ..., -0.018307, 0.277838, -0.110474, 0.066928, 0.12853).
- Bottom:** A message indicates 5 rows x 31 columns.

```
In [5]: class_names = {0:'Not Fraud', 1:'Fraud'}
print(df.Class.value_counts().rename(index = class_names))

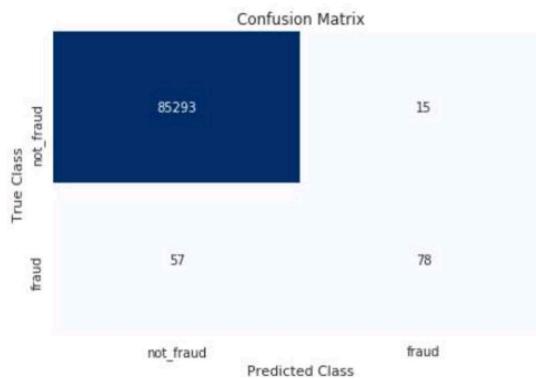
Not Fraud    284315
Fraud        492
Name: Class, dtype: int64
```

```
In [6]: fig = plt.figure(figsize = (15, 12))

plt.subplot(5, 6, 1) ; plt.plot(df.V1) ; plt.subplot(5, 6, 15) ; plt.plot(df.V15)
plt.subplot(5, 6, 2) ; plt.plot(df.V2) ; plt.subplot(5, 6, 16) ; plt.plot(df.V16)
plt.subplot(5, 6, 3) ; plt.plot(df.V3) ; plt.subplot(5, 6, 17) ; plt.plot(df.V17)
plt.subplot(5, 6, 4) ; plt.plot(df.V4) ; plt.subplot(5, 6, 18) ; plt.plot(df.V18)
plt.subplot(5, 6, 5) ; plt.plot(df.V5) ; plt.subplot(5, 6, 19) ; plt.plot(df.V19)
plt.subplot(5, 6, 6) ; plt.plot(df.V6) ; plt.subplot(5, 6, 20) ; plt.plot(df.V20)
plt.subplot(5, 6, 7) ; plt.plot(df.V7) ; plt.subplot(5, 6, 21) ; plt.plot(df.V21)
plt.subplot(5, 6, 8) ; plt.plot(df.V8) ; plt.subplot(5, 6, 22) ; plt.plot(df.V22)
plt.subplot(5, 6, 9) ; plt.plot(df.V9) ; plt.subplot(5, 6, 23) ; plt.plot(df.V23)
plt.subplot(5, 6, 10) ; plt.plot(df.V10) ; plt.subplot(5, 6, 24) ; plt.plot(df.V24)
plt.subplot(5, 6, 11) ; plt.plot(df.V11) ; plt.subplot(5, 6, 25) ; plt.plot(df.V25)
plt.subplot(5, 6, 12) ; plt.plot(df.V12) ; plt.subplot(5, 6, 26) ; plt.plot(df.V26)
plt.subplot(5, 6, 13) ; plt.plot(df.V13) ; plt.subplot(5, 6, 27) ; plt.plot(df.V27)
plt.subplot(5, 6, 14) ; plt.plot(df.V14) ; plt.subplot(5, 6, 28) ; plt.plot(df.V28)
plt.subplot(5, 6, 29) ; plt.plot(df.Amount)
plt.show()
```



```
In [13]:
class_names = ['not_fraud', 'fraud']
matrix = confusion_matrix(y_test, pred)
# Create pandas dataframe
dataframe = pd.DataFrame(matrix, index=class_names, columns=class_names)
# Create heatmap
sns.heatmap(dataframe, annot=True, cbar=None, cmap="Blues", fmt = 'g')
plt.title("Confusion Matrix"), plt.tight_layout()
plt.ylabel("True Class"), plt.xlabel("Predicted Class")
plt.show()
```



1. True Positive Rate, which can be defined as the number of fraudulent transactions that are even classified by the system as fraudulent.
2. True Negative Rate, which can be defined as the number of legitimate transactions that are even classified as legitimate by the system.
3. False Positive Rate, which can be defined as a number of the legal transactions which are wrongly classified as fraud.
4. False Negative Rate is defined as the transactions that are fraud but are wrongly classified as legal.

For a financial institution dealing with identifying fraud, Sensitivity and F1 - Score might be more important metrics. F1- Score represents a more balanced result as it is the harmonic mean between Precision and Recall. Sensitivity is more important in the sense that we are more interested in identifying fraud than than identifying legitimate customers.

In [14]:

```
from sklearn.metrics import f1_score, recall_score
f1_score = round(f1_score(y_test, pred), 2)
recall_score = round(recall_score(y_test, pred), 2)
print("Sensitivity/Recall for Logistic Regression Model 1 : {recall_score}".format(recall_score
= recall_score))
print("F1 Score for Logistic Regression Model 1 : {f1_score}".format(f1_score = f1_score))
```

```
Sensitivity/Recall for Logistic Regression Model 1 : 0.58
F1 Score for Logistic Regression Model 1 : 0.68
```

## **10. CONCLUSION AND FUTURE ENHANCEMENT**

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results. While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions. Since the entire dataset consists of only two days' transaction records, its only a fraction of data that can be made available if this project were to be used on a commercial scale. Being based on machine learning algorithms, the program will only increase its efficiency over time as more data is put into it.

While we couldn't reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here. The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project. More room for improvement can be found in the dataset. As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives. However, this requires official support from the banks themselves.

## **11. REFERENCES**

- [1] “Credit Card Fraud Detection Based on Transaction Behavior -by John Richard D. Kho, Larry A. Vea” published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017.
- [2] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 “ A Comprehensive Survey of Data Mining-based Fraud Detection Research” published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800.
- [3] “Survey Paper on Credit Card Fraud Detection by Suman” , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014.
- [4] “Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang” published by 2009 International Joint Conference on Artificial Intelligence.
- [5] “Credit Card Fraud Detection through Parenclitic Network Analysis By Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral” published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages.
- [6] “Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy” published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018