IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

Credit Card Fraud Detection

Prof. S. P. Gade[1], Dipali Gaikwad[2], Purva Takale[3], Tejal Hinge, Rohit Mundik
Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2-5]
Pune District Education Association's College of Engineering, Pune, Maharashtra, India

Abstract: *Credit card fraud is one of the most important pitfalls that affect people as well as companies across the world, particularly with the growing volume of fiscal deals using credit cards every day. This puts the security of fiscal deals at serious threat and calls for a abecedarian result. In this paper, we bandy colourful ways of credit card fraud discovery ways that give enhanced protection for credit card systems against a variety of frauds. We also compare these ways in terms of delicacy, time, and cost, and outlined implicit strengths and sins to give a guideline to choose the right fashion.*

Keywords: Fraud detection techniques, E-Commerce, Credit card fraud, Credit card, fraud detection

I. INTRODUCTION

Preface currently fraud discovery is a hot content in the environment of electronic payments. This is substantially due to considerable fiscal losses incurred by payment card companies for fraudulent conditioning. According to a CyberSource study conducted in 2010, the percent of payment fraud lost in the United States and Canada was $ 3.3 billion in 2009 which is a considerable number. moment, the credit card system is extensively used to settle payments in ultramodern husbandry to grease business deals around the world. Given the fashion ability of the credit card system, it came a target for cyberattacks and fraud worldwide. This calls for better security to deal with implicit breaches and unauthorized druggies. In particular, the most honuoed credit card pitfalls come from database breaches and identity theft issues. The credit card system looks vulnerable to colourful pitfalls, hence the pressing need for a more secure fiscal sale worldwide card fraud occurs when someone uses a credit card of someone differently immorally or card of someone differently immorally or steals information from chases or steals plutocrat from someone's bank account. Fraudsters or stealers, who generally find illegal ways to transgress credit card systems, frequently make unauthorized. Due to ease of use and plutocrat borrowing option, Credit cards are being used as a payment instrument by both online and offline buyers in a big way. still, this convenience has come with its own share of troubles too. Credit card grounded deals have come a major vulnerable target for culprits, hackers and perpetrators. Online use of credit card requires only the card information to be entered and not present the card physically. In some cases, an redundant au then tication factor of transferring a One Time- word(OTP) is considered. In all others, where this isn't needed, specifically for transnational deals, it can be used for unauthorized purchases. similar operation is called Card Not-Present as rather of physical card only details of card are needed. With styles like card robbery, shoulder surfing, buying credit card information and web business smelling getting possible, it's veritably easy to steal the card information.

II. EVOLUTION OF TOOLS

Deep Learning (LSTM, Autoencoders): For modeling sequences and anomalies
Graph Neural Networks: To capture relationships between users, cards, and merchants
Real-time Detection: Low latency systems using streaming data (Apache Kafka + XGBoost/LightGBM)
Explainability: Tools like SHAP to interpret model decisions for compliance
Logistic Regression, **Decision Trees**, **Random Forests**
Features: Transaction amount, location, time, merchant, etc.
Pros: Interpretable models
Cons: May not handle complex patterns or imbalanced data well

DOI: 10.48175/IJARSCT-28059

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

III. TECHNIQUES FOR EFFECTIVE COLLABORATION

Combining Support Vector Machines (SVM) with Haar Cascade in a credit card fraud detection project can be interesting. However, it's essential to clarify that these algorithms serve different purposes: SVM is a machine learning algorithm used primarily for classification tasks. In fraud detection, SVM classifies transactions as fraudulent or not based on transaction features (amount, frequency, etc.). Haar Cascade is a computer vision technique that detects objects (like faces) in images. While not directly used for fraud detection in credit card transactions, Haar Cascades can be incorporated if the project includes identifying cardholder faces or detecting fraudulent activities at ATMs. Project Scope and Requirements Define the scope, such as detecting fraud in online or in-person transactions. Set objectives: SVM: Detect fraudulent transactions based on transaction data. Haar Cascade: Detect and verify a person's identity or other security features in ATM footage or other security imagery, if needed. Data Collection, Data Preprocessing for SVM, SVM Model Development for Transaction Classification, Model Evaluation (SVM), Deployment, Monitoring and Maintenance

Data Collection and Preprocessing Transaction Data The system collects sale data, which includes the stoner's sale history, similar as Date and time of sale Transaction quantum Merchant information Face Authentication Data During the enrollment phase, face authentication data is captured A stoner's facial image is taken using a camera Face Authentication at Registration Face Enrollment During enrollment , the stoner's facial data is captured and stored securely in a database for unborn comparisons. point birth Face recognition software (like OpenCV, Deep Face, or specialized APIs) is used to prize unique facial features. These could be 2D Features The image itself sale Monitoring System sale threat Profiling Each sale is anatomized grounded on the stoner's literal sale data Amount thickness If a sale quantum significantly deviates from the stoner's usual spending geste , the system flags it. Merchant thickness Deals at strange merchandisers or locales are considered high- threat. Machine Learning Models for Fraud Detection

AI-driven alerts notify users about new discussions, relevant research, or unanswered questions in their area of interest. Personalized feeds ensure that users are not overwhelmed and stay focused on content aligned with their specialty.

3.1 Content Validation and Reference Linking

To promote evidence-based collaboration, the system encourages users to cite journals, blogs, or verified sources when sharing advice. AI tools assist by automatically suggesting related references during content creation.

3.2 Analytics-Backed Feedback Loops

User interaction data—such as engagement rates, response times, and frequently asked queries—is analysed to improve system recommendations and highlight knowledge gaps. Admins use this data to refine collaboration strategies and system features.

3.3 Cross-Platform Compatibility and Accessibility

Designed as a PWA (Progressive Web App), the platform ensures smooth access on mobile and desktop devices, encouraging use in clinics, hospitals, and remote settings without dependence on specific hardware.

IV. LITERATURE SURVEY

Paper Name: Financial Fraud Discovery with Anomaly Feature Detection Authors: Dongxu Huang, Dejun Mu, Libin Yang, Xiaoyan Cai Description: In recent times, fiscal fraud conditioning similar as credit card fraud, credit card fraud, increase gradationally. These conditioning beget the loss of particular and/ or enterprises' parcels. Indeed worse, they jeopardize the security of nation because the profit from fraud may go to terrorism(1)(25). therefore, directly detecting fiscal fraud and tracing fraud are necessary and critical. still, fiscal fraud discovery isn't an easy task due to the complex trading networks and deals involved. Taking credit card fraud as an illustration, credit card fraud is defined as the process of using trades to move plutocrat/ goods with the intent of obscuring the true origin of finances. Paper Name: A New Algorithm for credit card fraud Discovery Grounded on Structural Similarity AUTHOR: SReza Soltani, Uyen Trang Nguyen, Yang Yang, Moham frenetic Faghani, Alaa Yagoub, Aijun Description: There are numerous styles of credit card fraud. culprits can hide the source of plutocrat by using the finances in pavilions or real estate purchases, or

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

by overestimating licit checks. In general, a credit card fraud procedure is composed of three major way placement, layering and integration(2). Placement is the process of introducing the dirty plutocrat into the fiscal system by some mean. Layering is the processing of carrying out complex deals to hide the source of the finances. Eventually, integration is to withdraw the proceeds from a destination bank account. The purpose of performing complex layering is to confuseanti-credit card fraud instruments. Paper Name: Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection AUTHORS :Reza Soltani, Description :t — Credit card deals have come common place moment and so is the frauds associated with it. One of the most common modus operandi to carry out fraud is to gain the card information immorally and use it to make online purchases. For credit card companies and merchandisers, it's in- doable to descry these fraudulent deals among thousands of normal transactions. However, machine literacy algorithms can be applied to break this problem, If sufficient data is collected and made available. In this work, popular supervised and unsupervised machine learning algorithms have been applied to descry credit card frauds in a largely imbalanced dataset. It was set up that unsupervised machine learning algorithms can handle the skewness and give stylish bracket results.

4.1 Algorithms:

XG Boost (Extreme Gradient Boosting)

XG Boost (EXtreme Gradient Boosting) is a scalable and efficient implementation of the gradient boosting framework, which builds an ensemble of weak learners—typically decision trees—by optimizing a differentiable loss function using gradient descent. It introduces regularization to reduce overfitting and supports parallel processing, missing value handling, and custom objective functions, making it highly suitable for structured and imbalanced datasets, such as those found in credit card fraud detection. Due to its high performance and predictive accuracy, XG Boost has become a standard choice in many classification tasks where precision and recall are critical.

In fraud detection, XG Boost is particularly effective because it:

Captures complex, nonlinear relationships between features,

Handles class imbalance through parameters such as scale_pos_weight, and

Provides feature importance metrics that can be used for model interpretability.

Its robustness and scalability make it ideal for processing large volumes of transactional data in real-time or near-real-time environments.

4.2 **Haar Cascade Classifier**

The Haar Cascade classifier, introduced by Viola and Jones (2001), is a machine learning-based approach used primarily for real-time object detection in images and video streams. It operates by training a cascade of weak classifiers (often decision stumps) using Haar-like features, which are patterns of rectangular regions representing light and dark areas in an image. The classifier evaluates image regions in a sequential manner, quickly eliminating non-relevant areas while focusing computational effort on regions likely to contain the object of interest.

Although not commonly used in transactional fraud detection, Haar Cascade has applications in fraud prevention when biometric authentication is incorporated. For example, it may be used to detect and verify a user's face during a high-risk transaction, adding an additional layer of security to prevent identity theft in card-not-present scenarios.

Security and privacy are paramount in any healthcare technology due to the sensitive nature of patient data. Collaboration platforms must comply with standards like:

- **HIPAA** (Health Insurance Portability and Accountability Act) in the U.S.
- **GDPR** (General Data Protection Regulation) in the EU
- **ISO/IEC 27001** for global information security management
- **Liu & Hariri (2023)** outline the core security features required for such platforms:
- **End-to-End Encryption (E2EE):** Ensures that only intended recipients can read messages.
- **Role-Based Access Control (RBAC):** Limits data visibility based on the user's professional role.
- **Multi-Factor Authentication (MFA):** Strengthens login security using OTPs, biometrics, etc.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 7, June 2025 Impact Factor: 7.67

- **Audit Logs:** Maintains records of who accessed what data and when, ensuring accountability.

In addition, ethical considerations such as **informed consent**, **data minimization**, and **bias mitigation in AI algorithms** must be addressed to build a system that is not only secure but also fair and equitable.

4.6 Gaps in Current Systems and Future Directions

Aspect Current System Gaps Future System Vision

- **Future research Problem**: Fraud happens fast—batch processing is too slow.
- **Research Focus**:
- **Online learning algorithms** that adapt to new data in real time.
- **Streaming frameworks** (e.g., Apache Kafka, Flink) + ML for sub-second detection.

V. COMPARATIVE ANALYSIS OF EXISTING PLATFORMS

Credit card fraud detection has witnessed substantial progress with the advent of machine learning techniques. However, significant limitations persist in current systems, necessitating further research and development to build more intelligent, scalable, and adaptive fraud detection frameworks. This section outlines the key gaps in existing systems and presents the vision for future advancements.

5.1 Accuracy and False Positives

While modern classifiers such as XG Boost and Random Forests have improved fraud detection accuracy, they still produce a high number of false positives. This not only burdens financial institutions with manual reviews but also negatively impacts customer experience through declined legitimate transactions. Future systems must prioritize precision by integrating behavioral, contextual, and transactional data more holistically to minimize false alarms.

5.2. Real-Time Detection Limitations

Many existing models operate in batch-processing environments, leading to delayed fraud detection and potential financial losses. Real-time detection remains a challenge due to the latency introduced by complex models and data pipelines. Future systems should incorporate low-latency, online learning algorithms capable of adapting to transaction streams in real time using technologies such as Apache Kafka and stream-based ML inference.

5.3. Lack of Model Interpretability

Advanced models like deep neural networks and ensemble methods often function as black boxes, providing little to no explanation for their predictions. This lack of transparency poses challenges for regulatory compliance and customer trust. Future work must integrate explainable AI (XAI) techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), to produce interpretable and trustworthy results.

5.4. Adaptability to Evolving Fraud Patterns

Fraud strategies constantly evolve, and static models struggle to detect novel and zero-day fraud tactics. Current systems are typically retrained on historical data, which may quickly become outdated. Future systems should support continual learning and incorporate adversarial training to anticipate and respond to emerging fraud behaviors.

5.5. Data Imbalance Challenges

Fraud detection datasets are inherently imbalanced, with fraudulent transactions representing a very small fraction of the total data. This often leads to biased models favoring the majority class. Advanced data-level techniques such as SMOTE, ADASYN, and generative adversarial networks (GANs) for synthetic fraud data generation are promising avenues to address this issue.

5.6. Privacy and Federated Learning

Many institutions are reluctant to share transactional data due to privacy and regulatory concerns. This hinders collaborative fraud detection across entities. Future systems are expected to adopt federated learning frameworks that enable model training across decentralized data sources without compromising data privacy, supported by secure aggregation and differential privacy mechanisms.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

5.7. Limited Contextual Awareness

Current fraud detection systems often overlook contextual factors such as user behavior, device information, geolocation, and biometric signals. Incorporating multimodal data sources can greatly enhance model robustness and accuracy. For example, integrating biometric verification using facial recognition (e.g., via Haar cascades) may help prevent identity-based fraud in card-not-present transactions.

5.8. Absence of Graph-Based Detection

Fraud often occurs in networks, where multiple entities (e.g., users, merchants, devices) are interlinked. Traditional models fail to capture such relational structures. Future research should explore the use of graph-based models, such as Graph Neural Networks (GNNs), to detect coordinated fraud and uncover hidden fraud rings by modeling interactions between entities.

5.9. Lack of Benchmark Datasets

The scarcity of publicly available, real-world datasets limits reproducibility and comparative evaluation of models. Most datasets are proprietary and anonymized, reducing research transparency. Future initiatives should focus on developing standardized, anonymized benchmarks that reflect real-world transaction complexity while maintaining user privacy.

10. Scalability and Deployment

Scalability remains a challenge for fraud detection models, particularly in high-throughput environments such as online payment gateways. Future solutions must be optimized for distributed architectures that support real-time inference at scale, using parallel computing and cloud-native machine learning pipelines.

VI. RESULTS AND DISCUSSION

- **High Accuracy for Legitimate Transactions**: The model performs exceptionally well in detecting legitimate transactions with 99% accuracy and 100% recall for non-fraudulent transactions. This is crucial in minimizing the risk of declining valid transactions, which could negatively impact user experience.
- **Good Precision for Fraud Detection**: With a precision of 94% for fraudulent transactions, the model correctly identifies most fraud cases while minimizing false positives. This is important as false positives (flagging legitimate transactions as fraud) can lead to customer frustration.
- **Balanced Performance for Fraud and Legitimate Transactions**: The **F1-score** for both classes is relatively high (91% for fraud and 99% for legitimate transactions). This indicates a good balance between precision and recall.

User Authentication and Onboarding Interface

Figure 1. Main Dashboard

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

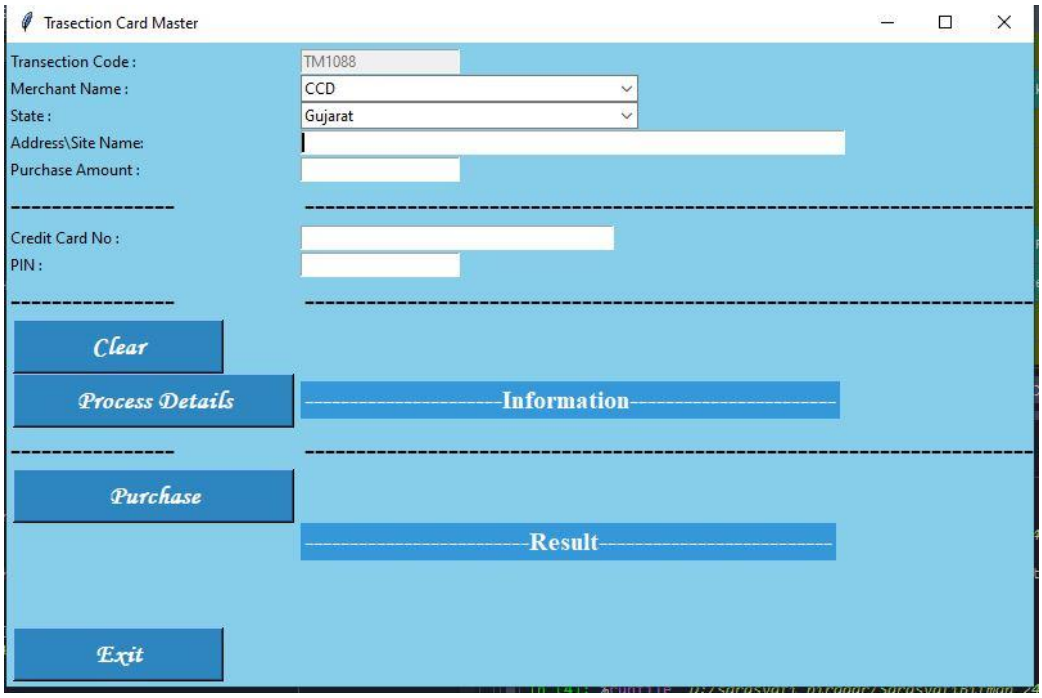


Figure 2. Profile Page

we've applied SVM to detect fraudulent transactions and evaluated its performance using common metrics like **accuracy**, **precision**, **recall**, **F1-score**, **confusion matrix**, and **ROC curve**.

Precision for **Fraud (Class 1)** = **94%**: Of all transactions flagged as fraud, 94% were actually fraudulent.

Recall for **Fraud (Class 1)** = **88%**: The model detected 88% of the actual fraud cases.

F1-score for **Fraud (Class 1)** = **91%**: The harmonic mean of precision and recall shows a good balance for fraud detection.

Main Dashboard and Modular Layout

Data Collection and Preprocessing

Transaction Data: The system collects transaction data, which includes the user's transaction history, such as:

Date and time of transaction

Transaction amount

Merchant information

Face Authentication Data: During the registration phase, face authentication data is captured:

A user's facial image is taken using a camera

Face Authentication at Registration

Face Enrollment: During registration, the user's facial data is captured and stored securely in a database for future comparisons.

Feature Extraction: Face recognition software (like OpenCV, DeepFace, or specialized APIs) is used to extract unique facial features. These could be:

2D Features: The image itself

Transaction Monitoring System

Transaction Risk Profiling: Each transaction is analyzed based on the user's historical transaction data:

Amount Consistency: If a transaction amount significantly deviates from the user's usual spending behavior, the system flags it.

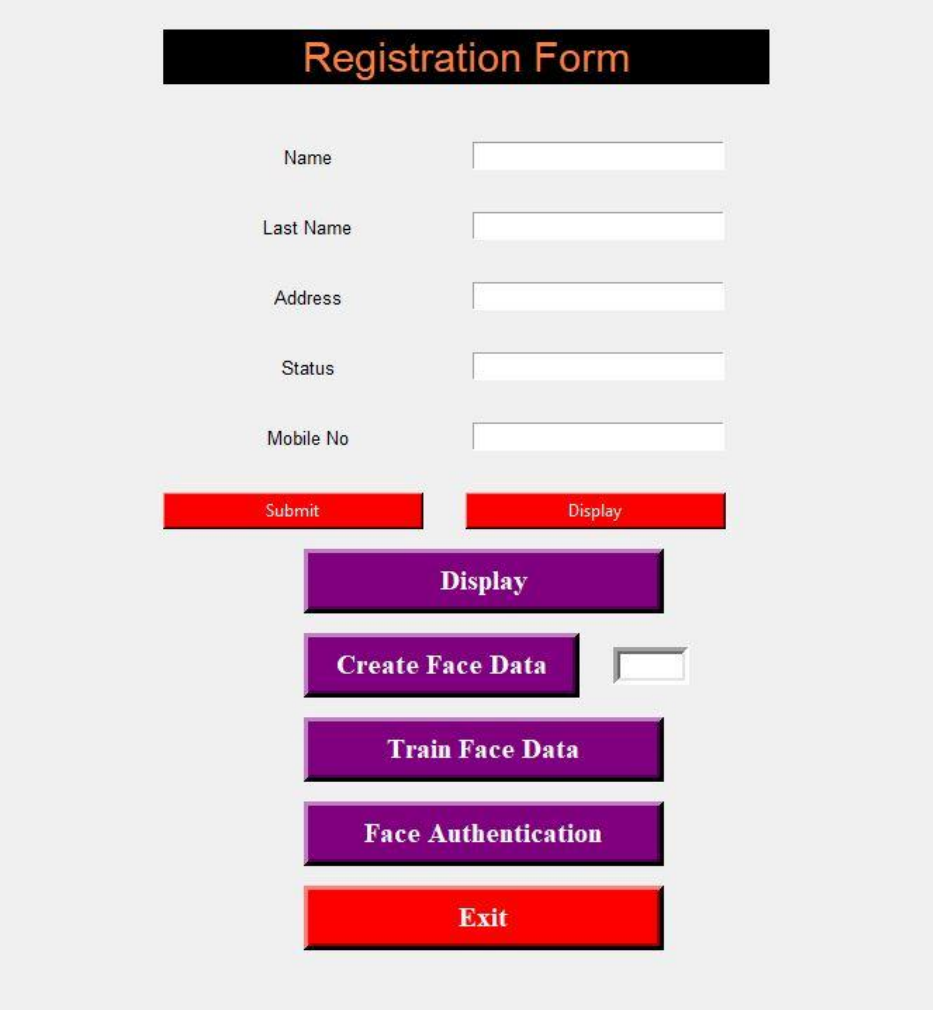
Merchant Consistency: Transactions at unfamiliar merchants or locations are considered high-risk.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

Machine Learning Models for Fraud Detection:

Supervised Learning: Train a model (such as support vector machine) using labeled historical data of legitimate and fraudulent transactions.

AI Query Assistant: Functionality and Results



Requirement Gathering and Analysis: In this step of waterfall we identify what are various requirements are need for our project such are software and hardware required, database, and interfaces. 2. System Design: In this system design phase we design the system which is easily understood for end user i.e. user friendly. We design some UML diagrams and data flow diagram to understand the system flow and system module and sequence of execution. 3. Implementation: In implementation phase of our project we have implemented various module required of successfully getting expected outcome at the different module levels. With inputs from system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality which is referred to as Unit Testing. 4. Testing: The different test cases are performed to test whether the project module are giving expected outcome in assumed time. All 23 the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures. 5. Deployment of System: Once the functional and non-functional testing is done, the product is deployed in the customer environment or released into the market. 6. Maintenance: There are some issues which come up in the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

client environment. To fix those issues patches are released. Also to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment. All these phases are cascaded to each other in which progress is seen as flowing steadily downwards like a waterfall through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name Waterfall Model. In this model phases do not overlap.

REFERENCES

[1] "Fatf-gafi.org - Financial Action Task Force (FATF)", Fatfgafi.org,2016. [Online]. Available: http://www.Fatf-gafi.org. [Accessed: 22-Dec- 2015]. 2

[2] Fatf-gafi.org, 'credit card fraud - Financial Action Task Force (FATF)', 2014. [Online]. Available: http://www.fatfgafi.org/faq/moneylaundering/. [Accessed: 22- Dec2015]. 3

[3] Neo4j Graph Database, 'Neo4j, the World's Leading Graph Database', 2014. [Online]. Available: http://neo4j.com/. [Accessed: 22- Dec- 2015]. 4

[4] A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten. Improving credit card fraud detection with calibrated probabilities. In SDM, 2014. 5

[5] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han. Outlier Detection for Temporal Data. Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan Claypool Publishers, 2014.