ABSTRACT

Detection and prevention of fraudulent transactions in e-commerce platforms have always been the focus of transaction security systems. However, due to the concealment of e-commerce, it is not easy to capture attackers solely based on the historic order information. Many researches try to develop technologies to prevent the frauds, which have not considered the dynamic behaviors of users from multiple perspectives. This leads to an inefficient detection of fraudulent behaviors. To this end, this paper proposes a novel fraud detection method that integrates machine-learning and process mining models to monitor real-time user behaviors. First, we establish a process model concerning the B2C e-commerce platform, by incorporating the detection of user behaviors. Second, a method for analyzing abnormalities that can extract important features from event logs is presented. Then, we feed the extracted features to a Support Vector Machine (SVM) based classification model that can detect fraud behaviors. We demonstrate the effectiveness of our method in capturing dynamic fraudulent behaviors in e-commerce systems through the experiments.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVATIONS

E-Commerce	Electronic Commerce
B2C	Business to Customer
SVC	Support Vector Classifier
API	Application Programming Interface
DPN	Deep Predictive Network
SVM	Support Vector Machine
AUC	Area Under the Curve

LIST OF SCREENSHOTS

CHAPETER-1

INTRODUCTION

With the increasing popularity of e-commerce platforms, more and more commercial transactions are now relying on web-based systems than the traditional cash-based approach. Although the entity economy is greatly impacted by the COVID-19 epidemic in recent years, e-commerce remains largely unaffected by the pandemic, aiding steady market growth. The sales volume of B2C (Business to Customer) e-commerce is expected to reach 6.5 trillion dollars by 2023.

Though the growth of e-commerce and the expansion of modern technologies offer better opportunities for online businesses, new security threats have emerged over the past few years. The significant increase in the number of online fraud cases costs billions of dollars worldwide every year. The dynamic and distributed nature of the Internet has made anti-fraud systems inevitable to ensure the security of online transactions.

Current fraud detection systems typically rely on detecting abnormal user behaviors, but they are often reactive, fragmented, and isolated in scope. These systems may fail to capture multi-user fraud scenarios where multiple entities (e.g., buyers and sellers) collaborate or manipulate system flows to exploit vulnerabilities. Moreover, many systems lack contextual awareness and cannot adapt to the evolving tactics used by sophisticated fraudsters.

This paper introduces a process-based method where user behaviors are recorded and analyzed in real-time. By modeling and analyzing the business process of e-commerce systems, this method dynamically detects changes in user behaviors, transaction processes, and noncompliance situations.

Main contributions include:

1. A conformance checking method using process mining to detect anomalies.
2. A user behavior detection system using Petri nets.
3. An SVM model embedded with process mining features to classify frauds automatically.

The objective of this project is to develop a robust fraud detection method that reduces false positives and enhances fraud identification in real-time using a multi-perspective strategy.

The approach aims to improve transparency, accountability, and decision-making within the transaction lifecycle. By applying a holistic view that captures all roles involved in a transaction, the system enhances trust and security in digital commerce ecosystems.

CHAPTER-2

SYSTEM STUDY

The system is designed to study behavioral patterns and transactional anomalies from multiple sources. This includes the collection of data from users' past transactions, login activities, browsing behavior, and communication patterns. The study involves both static rule-based systems and dynamic machine learning approaches for anomaly detection.

Additionally, the system incorporates log analysis, user profiling, and third-party reputation checks to cross-validate transaction legitimacy. Time-series trend analysis helps uncover fraud patterns spread over days or weeks, enhancing fraud anticipation rather than reaction.

Furthermore, the system integrates process mining techniques to map actual user behavior against predefined business workflows, helping detect inconsistencies and unauthorized activities within transactional processes. Context-aware analytics are employed to understand the role of each entity in the transaction chain. This allows the system to differentiate between accidental anomalies and intentional fraud attempts, reducing false alerts.

The system also emphasizes scalability and real-time responsiveness, ensuring effective performance in high-volume e-commerce environments. Security features such as data encryption, access control, and session tracking are incorporated into the system to support secure data handling and accountability.

2.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ♦ ECONOMICAL FEASIBILITY
- ♦ TECHNICAL FEASIBILITY

- SOCIAL FEASIBILITY

2.2. ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.3. TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.4. SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

CHAPTER-3

LITERATURE SURVEYS

3.1. EXISTING SYSTEM

The existing fraud detection systems primarily rely on rule-based mechanisms and anomaly detection algorithms that analyze historical transactional data. These systems often operate on predefined thresholds and static patterns to identify irregularities. While effective to some extent, they lack the adaptability and real-time responsiveness necessary to detect sophisticated or collaborative frauds. Additionally, these systems do not provide comprehensive behavioural analysis or process-level monitoring, making them insufficient in complex e-commerce ecosystems involving multiple participants.

Traditional fraud detection frameworks commonly utilize statistical metrics, blacklists, and known fraud signatures to flag abnormal transactions. However, they are unable to handle evolving fraud tactics such as synthetic identities, fake reviews, and collusion between users. Most legacy systems also rely on batch processing, resulting in delayed response times, which can cause damage before any preventive action is taken.

Moreover, conventional systems are typically siloed, focusing on a single entity (like the buyer) rather than considering the roles and interactions among all participants. They lack cross-user correlation analysis, which is essential for detecting fraud involving multiple actors or coordinated schemes.

Limitations include:

- Inability to handle multi-entity fraud scenarios.
- High false positive rates due to lack of contextual analysis.
- Limited use of real-time data processing.
- No incorporation of workflow or business process tracking.

3.1.1 DISADVENTAGES

1) Fraud mode one - an order is tempered by a malicious actor: The malicious actor may deceive the victim merchant by sending a fake formal payment order order F

A to the cashier server. The malicious actor obtained the order items that do not match the payment value by tampering with the order information, such as the total amount.

2) Fraud mode two - subcontract the order: The victim pays the malicious actor’s order instead of his order. To achieve their goals, the malicious actors impersonate the duties of sellers and buyers. The order information changes before and after the payment.

3.2 PROPOSED SYSTEM

The proposed system overcomes the limitations of traditional methods by employing a multi-perspective approach. It integrates process mining techniques with machine learning models to analyse user behaviour in real-time and across different layers of the transaction flow. By capturing contextual and temporal data, the system can detect both isolated and coordinated fraudulent activities more accurately.

Key features of the proposed system:

- Real-time fraud detection using dynamic process analysis.
- Conformance checking to identify deviations from expected business workflows.
- Machine learning classifiers (e.g., SVM) integrated with process mining for high accuracy.
- Behavioural profiling and anomaly detection at both user and system levels.
- Scalable architecture for high-volume e-commerce applications.
- Enhanced cross-user interaction tracking to identify coordinated fraud schemes.
- Adaptive learning mechanisms to recognize and respond to emerging fraud tactics.
- Integration with external verification APIs for identity and transaction validation.
- Visual dashboards for administrators to monitor, analyse, and respond to alerts quickly.

This innovative architecture ensures proactive fraud detection with enhanced accuracy and reduced false positives. By considering multiple perspectives—including buyer, seller, and system behaviour—the model offers comprehensive visibility into fraudulent transactions.

3.2.1 ADVANTAGES

- ➢ To arrive at a clearer result, the plug-in Multi-Perspective Process Explorer and Conformance Checking are used to match and analyze the event log and the DPN. The result is shown in this system, where each action is represented with different colors. For instance, green represents the move both on model and log, purple means move on the model only, and grey represents invisible actions, that is, skipped actions.

- By clicking on a given action, we can obtain the matching information between the model and the event log in the data flow of each action. The data marked in red indicates a mismatch. We extract these suspicious anomalies and use them as the basis for subsequent training using machine learning models.

Class Diagram :

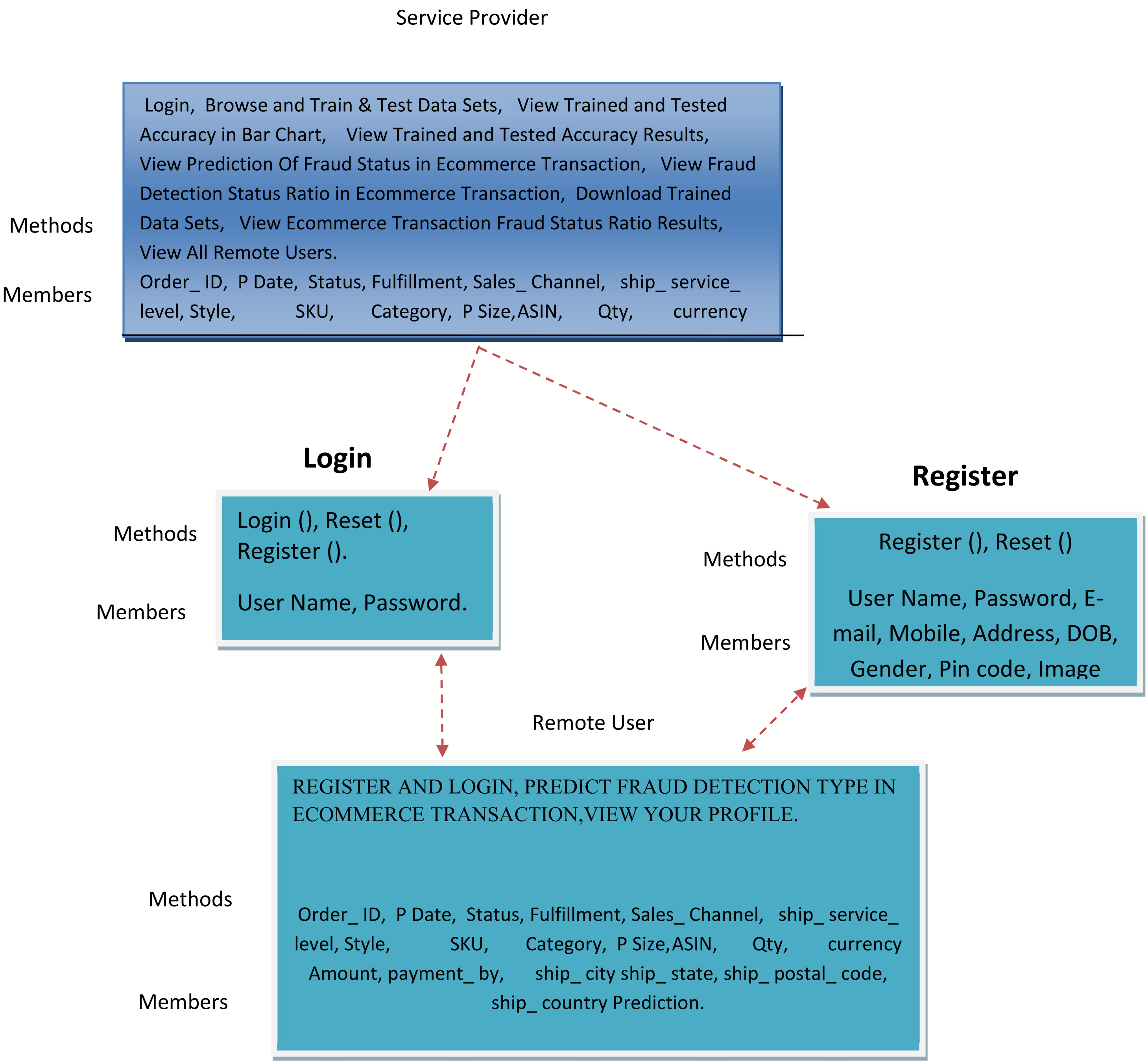

Service Provider
Methods
Members
Login, Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Fraud Status in Ecommerce Transaction, View Fraud Detection Status Ratio in Ecommerce Transaction, Download Trained Data Sets, View Ecommerce Transaction Fraud Status Ratio Results, View All Remote Users.
Order_ ID, P Date, Status, Fulfillment, Sales_ Channel, ship_ service_ level, Style, SKU, Category, P Size, ASIN, Qty, currency
Login
Methods
Members
Login (), Reset (), Register ().
User Name, Password.
Register
Methods
Members
Register (), Reset ()
User Name, Password, E-mail, Mobile, Address, DOB, Gender, Pin code, Image
Remote User
Methods
Members
REGISTER AND LOGIN, PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION,VIEW YOUR PROFILE.
Order_ ID, P Date, Status, Fulfillment, Sales_ Channel, ship_ service_ level, Style, SKU, Category, P Size, ASIN, Qty, currency Amount, payment_ by, ship_ city ship_ state, ship_ postal_ code, ship_ country Prediction.

FIG 1- Service Provider

CHAPTER-4

ARCHITECHTURE DIAGRAM

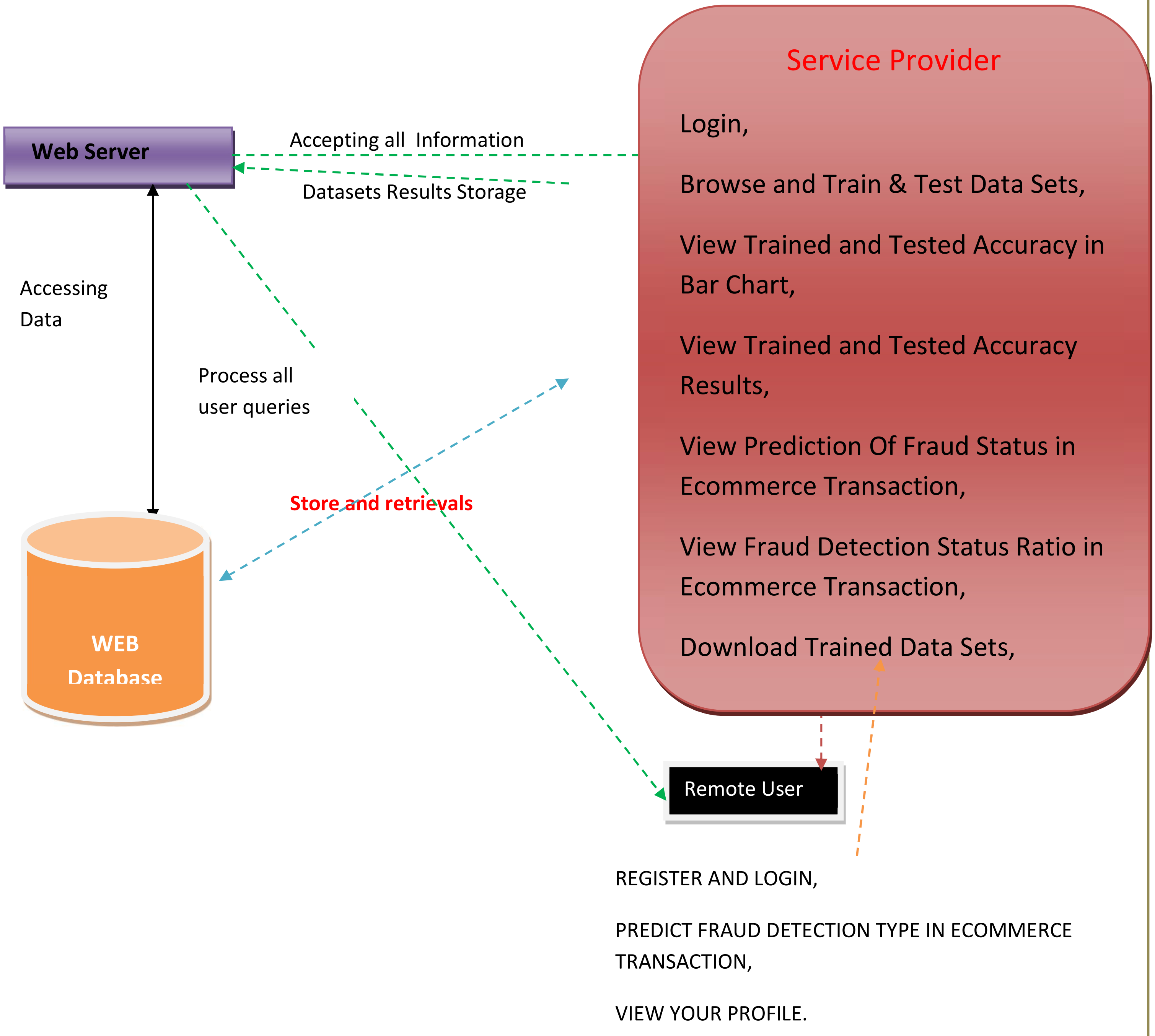

Service Provider
Login,
Browse and Train & Test Data Sets,
View Trained and Tested Accuracy in Bar Chart,
View Trained and Tested Accuracy Results,
View Prediction Of Fraud Status in Ecommerce Transaction,
View Fraud Detection Status Ratio in Ecommerce Transaction,
Download Trained Data Sets,
Web Server
Accepting all Information
Datasets Results Storage
Accessing Data
Process all user queries
Store and retrievals
WEB Database
Remote User
REGISTER AND LOGIN,
PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION,
VIEW YOUR PROFILE.

FIG 2- ARCHITECTURE STRUCTURE

CHAPTER-5

ANALYSIS

5.1 SYSTEM REQUIRMENTS

- **H/W System Configuration: -**

Processor	-	Pentium –IV
RAM	-	4 GB (min)
Hard Disk	-	20 GB
Key Board	-	Standard Windows Keyboard
Mouse	-	Two or Three Button Mouse
Monitor	-	SVGA

5.2 SOFTWARE REQUIRMENTS

Operating system	:	Windows 7 Ultimate.
Coding Language	:	Python.
Front-End	:	Python.
Back-End	**:**	Django-ORM
Designing	:	Html, css, javascript.
Data Base	**:**	MySQL (WAMP Server)

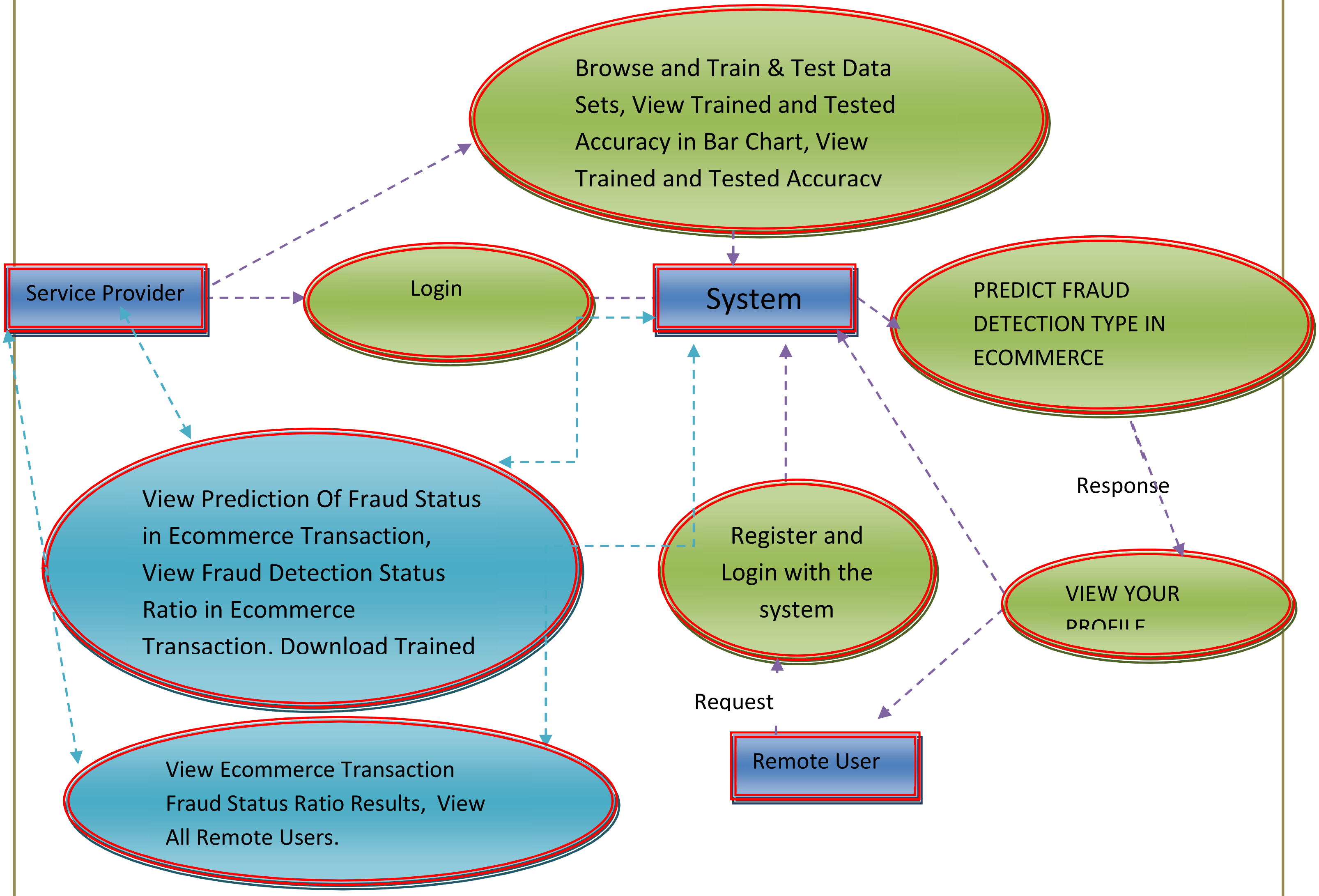

Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracv
Service Provider
Login
System
PREDICT FRAUD DETECTION TYPE IN ECOMMERCE
Response
View Prediction Of Fraud Status in Ecommerce Transaction, View Fraud Detection Status Ratio in Ecommerce Transaction. Download Trained
Register and Login with the system
VIEW YOUR PROFILE
Request
Remote User
View Ecommerce Transaction Fraud Status Ratio Results, View All Remote Users.

FIG 3- STEPS FOR DETECTION PROCESS

CHAPTER-6

ALGORITHMS

6.1 DICISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

6.1.1.GRADIENT BOOSTING

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms forest. A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

6.2 K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm

- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given

- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

6.3.LOGISTIC REGRESSION CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

6.4.NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

6.5.RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

6.6.SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptron's*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination , whereas the perceptron and GA classifier models are different each time training is initialized.

CHAPTER-7

SYSTEM DESIGN

7.1 INTRODUCTION FOR SYSTEM DESIGN

In the dynamic realm of e-commerce, ensuring transaction integrity is a critical challenge, especially in multi-participant ecosystems involving buyers, sellers, logistics providers, and platforms. With the growing complexity and volume of digital transactions, fraudulent behaviors have become more sophisticated, often exploiting the lack of holistic oversight across all participants. Conventional fraud detection approaches are typically siloed, focusing on individual actors and missing out on coordinated or contextual fraud patterns.

To overcome these limitations, we introduce DIGIN (Detection through Integrated Graph-based Intelligence Network), a multi-perspective fraud detection framework that captures, analyzes, and correlates data from all participants in e-commerce transactions. DIGIN employs a graph-based approach to model inter-entity relationships and utilizes advanced machine learning techniques to uncover hidden fraud patterns across diverse data sources.

7.2 MODULES

7.2.1SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Fraud Status in Ecommerce Transaction, View Fraud Detection Status Ratio in Ecommerce Transaction, Download Trained Data Sets, View Ecommerce Transaction Fraud Status Ratio Results, View All Remote Users

7.2.2VIEW AND AUTHORIZE USERS

In this module, the admin can view the list of users who all registered. In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

7.3.3REMOTE USER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION, VIEW YOUR PROFILE.

7.3 SEQUENCE DIAGRAM

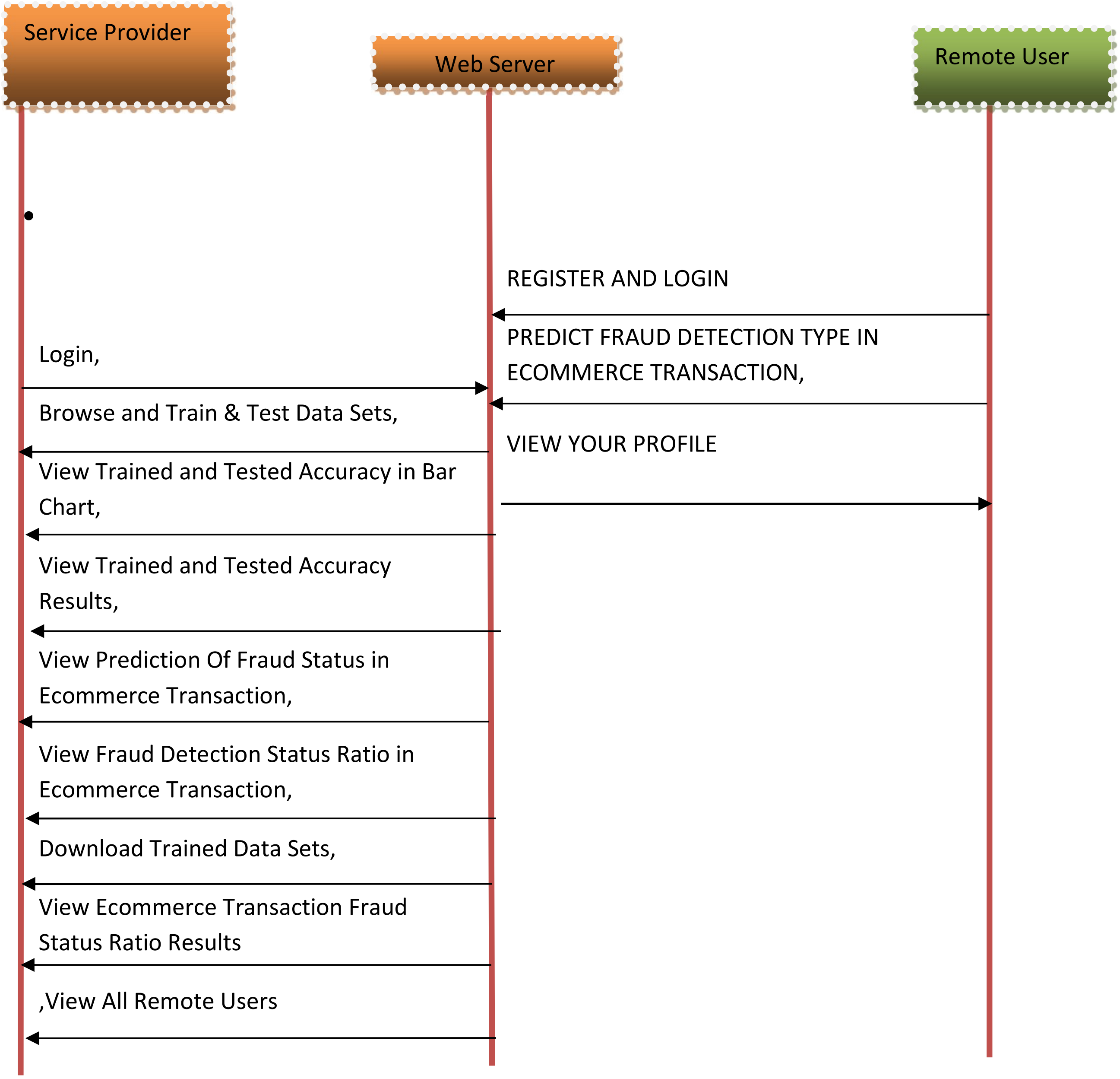

Service Provider
Web Server
Remote User
REGISTER AND LOGIN
PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION,
VIEW YOUR PROFILE
Login,
Browse and Train & Test Data Sets,
View Trained and Tested Accuracy in Bar Chart,
View Trained and Tested Accuracy Results,
View Prediction Of Fraud Status in Ecommerce Transaction,
View Fraud Detection Status Ratio in Ecommerce Transaction,
Download Trained Data Sets,
View Ecommerce Transaction Fraud Status Ratio Results
,View All Remote Users

FIG 4- SEQUENCE DIAGRAM

CHAPTER-8

SYSTEM TESTING

8.1 TESTING METHODOLOGIES

The following are the Testing Methodologies:

- **Unit Testing.**
- **Integration Testing.**
- **User Acceptance Testing.**
- **Output Testing.**
- **Validation Testing.**

8.1.1 UNIT TESTING

Unit testing focuses verification effort on the smallest unit of Software design that is the module. Unit testing exercises specific paths in a module's control structure to

ensure complete coverage and maximum error detection. This test focuses on each module individually, ensuring that it functions properly as a unit. Hence, the naming is Unit Testing.

During this testing, each module is tested individually and the module interfaces are verified for the consistency with design specification. All important processing path are tested for the expected results. All error handling paths are also tested.

8.1.2 INTEGRATION TESTING

Integration testing addresses the issues associated with the dual problems of verification and program construction. After the software has been integrated a set of high order tests are conducted. The main objective in this testing process is to take unit tested modules and builds a program structure that has been dictated by design.

The following are the types of Integration Testing:

1.Top Down Integration

This method is an incremental approach to the construction of program structure. Modules are integrated by moving downward through the control hierarchy, beginning with the main program module. The module subordinates to the main program module are incorporated into the structure in either a depth first or breadth first manner.

In this method, the software is tested from main module and individual stubs are replaced when the test proceeds downwards.

2. Bottom-up Integration

This method begins the construction and testing with the modules at the lowest level in the program structure. Since the modules are integrated from the bottom up, processing required for modules subordinate to a given level is always available and the need for stubs is eliminated. The bottom up integration strategy may be implemented with the following steps:

- The low-level modules are combined into clusters into clusters that perform a specific Software sub-function.
- A driver (i.e.) the control program for testing is written to coordinate test case input and output.
- The cluster is tested.
- Drivers are removed and clusters are combined moving upward in the program structure

The bottom up approaches tests each module individually and then each module is module is integrated with a main module and tested for functionality.

8.1.3 USER ACCEPTANCE TESTING

User Acceptance of a system is the key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system users at the time of developing and making changes wherever required. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.

8.1.4 OUPUT TESTING

After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format. Asking the users about the format required by them tests the outputs generated or displayed by the system under consideration. Hence the output format is considered in 2 ways – one is on screen and another in printed format.

8.1.5 VALIDATION CHECKING

Validation checks are performed on the following fields.

Text Field:

The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables. Incorrect entry always flashes and error message.

Numeric Field:

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform. Each module is subjected to test run along with sample data. The individually tested modules are integrated into a single system. Testing involves executing the real data information is used in the program the existence of any program defect is inferred from the output. The testing should be planned so that all the requirements are individually tested.

A successful test is one that gives out the defects for the inappropriate data and produces and output revealing the errors in the system.

Preparation of Test Data

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data. While testing the system by using test data errors are again uncovered and corrected by using above testing steps and corrections are also noted for future use.

Using Live Test Data:

Live test data are those that are actually extracted from organization files. After a system is partially constructed, programmers or analysts often ask users to key in a set of data from their normal activities. Then, the systems person uses this data as a way to partially test the system. In other instances, programmers or analysts extract a set of live data from the files and have them entered themselves.

It is difficult to obtain live data in sufficient amounts to conduct extensive testing. And, although it is realistic data that will show how the system will perform for the typical processing requirement, assuming that the live data entered are in fact typical, such data generally will not test all combinations or formats that can enter the system. This bias toward typical values then does not provide a true systems test and in fact ignores the cases most likely to cause system failure.

Using Artificial Test Data:

Artificial test data are created solely for test purposes, since they can be generated to test all combinations of formats and values. In other words, the artificial data, which can quickly be prepared by a data generating utility program in the information systems department, make possible the testing of all login and control paths through the program.

The most effective test programs use artificial test data generated by persons other than those who wrote the programs. Often, an independent team of testers formulates a testing plan, using the systems specifications.

The package “Virtual Private Network” has satisfied all the requirements specified as per software requirement specification and was accepted.

8.2 USER TRAINING

Whenever a new system is developed, user training is required to educate them about the working of the system so that it can be put to efficient use by those for whom the system has been primarily designed. For this purpose the normal working of the project was demonstrated to the prospective users. Its working is easily understandable and since the expected users are people who have good knowledge of computers, the use of this system is very easy.

8.3 MAINTAINENCE

This covers a wide range of activities including correcting code and design errors. To reduce the need for maintenance in the long run, we have more accurately defined the user’s requirements during the process of system development. Depending on the requirements, this system has been developed to satisfy the needs to the largest possible extent. With development in technology, it may be possible to add many more features based on the requirements in future. The coding and designing is simple and easy to understand which will make maintenance easier.

8.4TESTING STRATEGY :

A strategy for system testing integrates system test cases and design techniques into a well planned series of steps that results in the successful construction of software. The testing strategy must co-operate test planning, test case design, test execution, and the resultant data collection and evaluation .A strategy for software testing must accommodate low-level tests that are necessary to verify that a small source code segment has been correctly implemented as well as high level tests that validate major system functions against user requirements.

Software testing is a critical element of software quality assurance and represents the ultimate review of specification design and coding. Testing represents an interesting anomaly for the software. Thus, a series of testing are performed for the proposed system before the system is ready for user acceptance testing.

8.5 SYSTEM TESTING:

Software once validated must be combined with other system elements (e.g. Hardware, people, database). System testing verifies that all the elements are proper and that overall system function performance is

achieved. It also tests to find discrepancies between the system and its original objective, current specifications and system documentation.

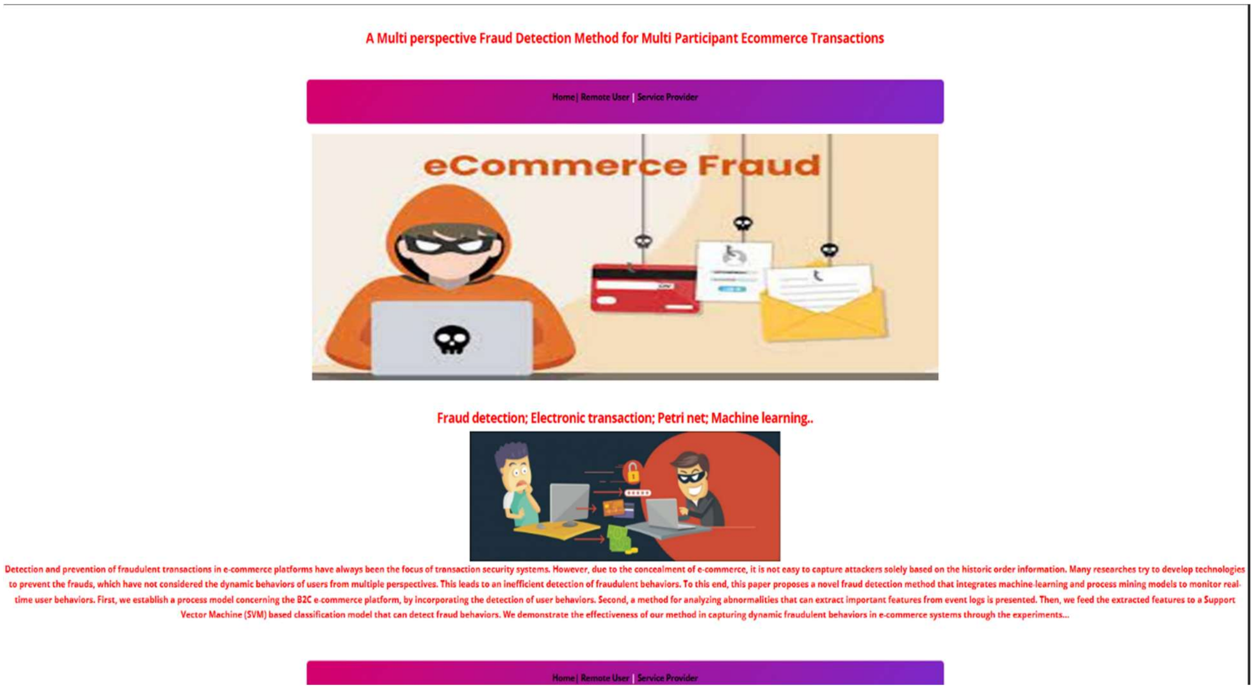
8.6 UNIT TESTING:

In unit testing different are modules are tested against the specifications produced during the design for the modules. Unit testing is essential for verification of the code produced during the coding phase, and hence the goals to test the internal logic of the modules. Using the detailed design description as a guide, important Conrail paths are tested to uncover errors within the boundary of the modules. This testing is carried out during the programming stage itself. In this type of testing step, each module was found to be working satisfactorily as regards to the expected output from the module.

In Due Course, latest technology advancements will be taken into consideration. As part of technical build-up many components of the networking system will be generic in nature so that future projects can either use or interact with this. The future holds a lot to offer to the development and refinement of this project.

CHAPTER 9

SCREENSHOTS



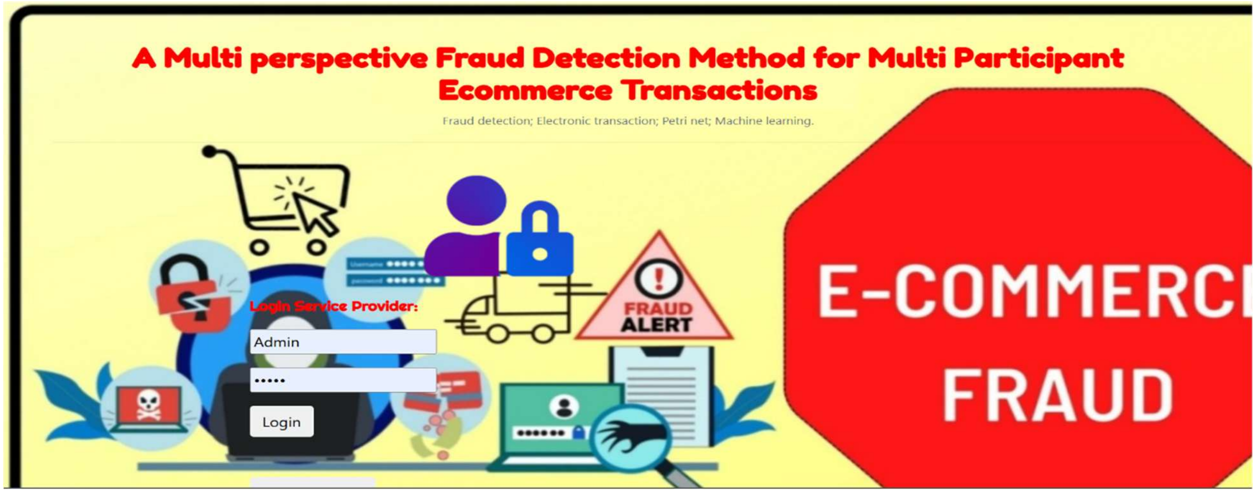
A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions

Fraud detection; Electronic transaction; Petri net; Machine learning..

Detection and prevention of fraudulent transactions in e-commerce platforms have always been the focus of transaction security systems. However, due to the concealment of e-commerce, it is not easy to capture attackers solely based on the historic order information. Many researches try to develop technologies to prevent the frauds, which have not considered the dynamic behaviors of users from multiple perspectives. This leads to an inefficient detection of fraudulent behaviors. To this end, this paper proposes a novel fraud detection method that integrates machine-learning and process mining models to monitor real-time user behaviors. First, we establish a process model concerning the B2C e-commerce platform, by incorporating the detection of user behaviors. Second, a method for analyzing abnormalities that can extract important features from event logs is presented. Then, we feed the extracted features to a Support Vector Machine (SVM) based classification model that can detect fraud behaviors. We demonstrate the effectiveness of our method in capturing dynamic fraudulent behaviors in e-commerce systems through the experiments...

Home| Remote User | Service Provider
eCommerce Fraud
Home| Remote User | Service Provider

SCREENSHOT 1 – HOME PAGE



A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions

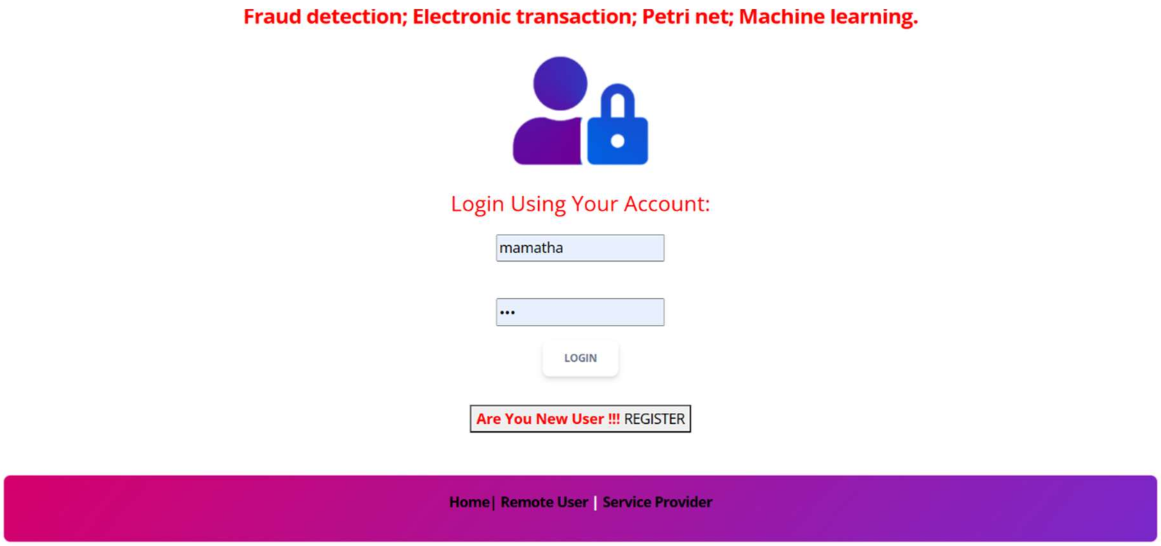
Fraud detection; Electronic transaction; Petri net; Machine learning.

Login Service Provider:

Admin

•••••

SCREENSHOT 2 – ADMIN PAGE



Fraud detection; Electronic transaction; Petri net; Machine learning.

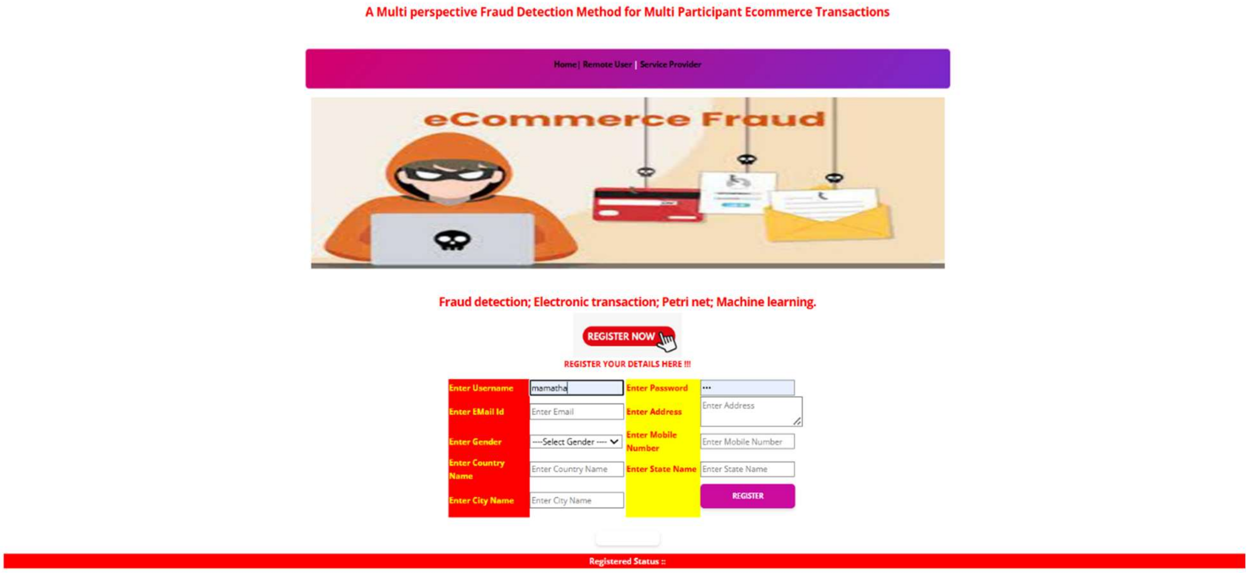
Login Using Your Account:

mamatha

•••

LOGIN
Are You New User !!! REGISTER
Home| Remote User | Service Provider

SCREENSHOT 3 – LOGIN PAGE



A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions

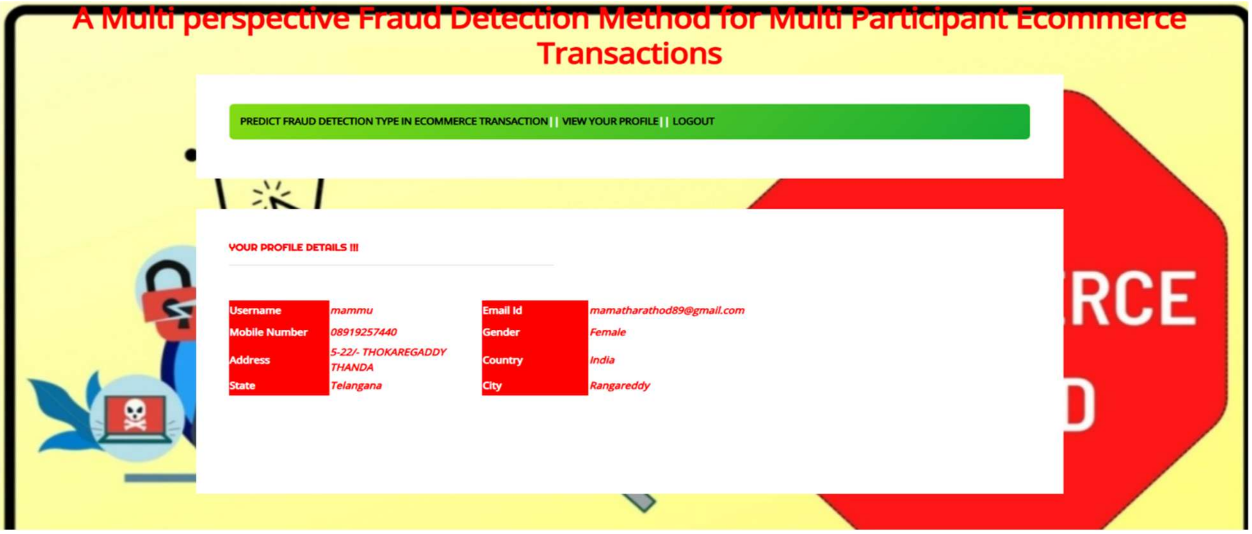
Fraud detection; Electronic transaction; Petri net; Machine learning.

REGISTER YOUR DETAILS HERE !!!

Enter Username	mamatha	Enter Password	•••
Enter EMail Id	Enter Email	Enter Address	Enter Address
Enter Gender	----Select Gender ----	Enter Mobile Number	Enter Mobile Number
Enter Country Name	Enter Country Name	Enter State Name	Enter State Name
Enter City Name	Enter City Name		REGISTER

Registered Status ::

SCREENSHOT 4 – REGISTERING PAGE



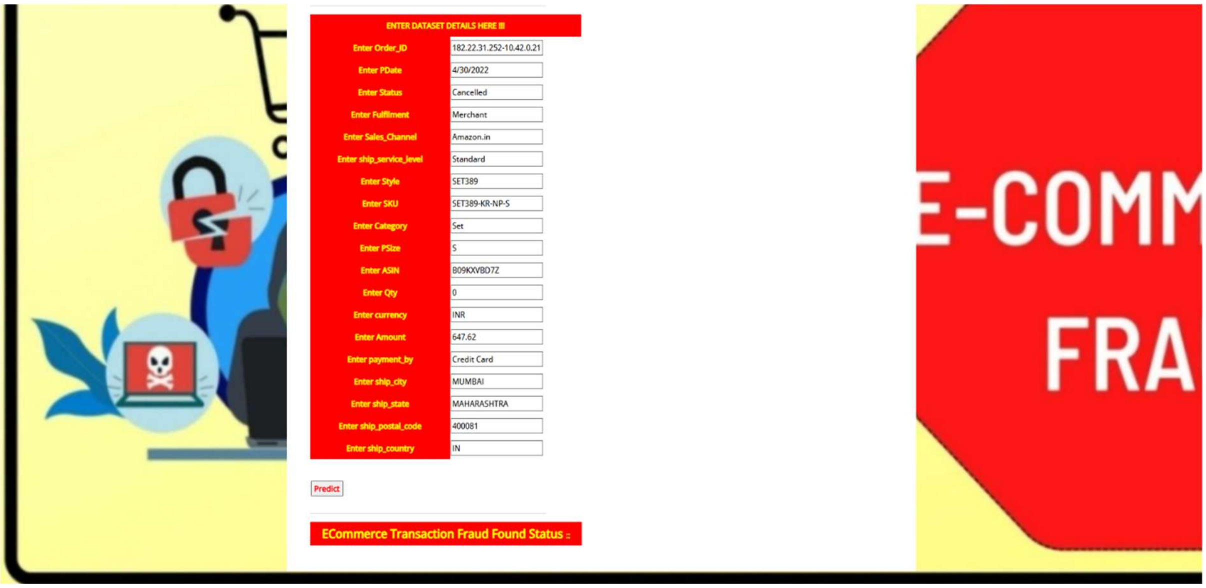
A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions

PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION || VIEW YOUR PROFILE || LOGOUT
RCE
D

YOUR PROFILE DETAILS !!!

Username	mammu	Email Id	mamatharathod89@gmail.com
Mobile Number	08919257440	Gender	Female
Address	5-22/- THOKAREGADDY THANDA	Country	India
State	Telangana	City	Rangareddy

SCREENSHOT 5 – PROFILE VIEW PAGE

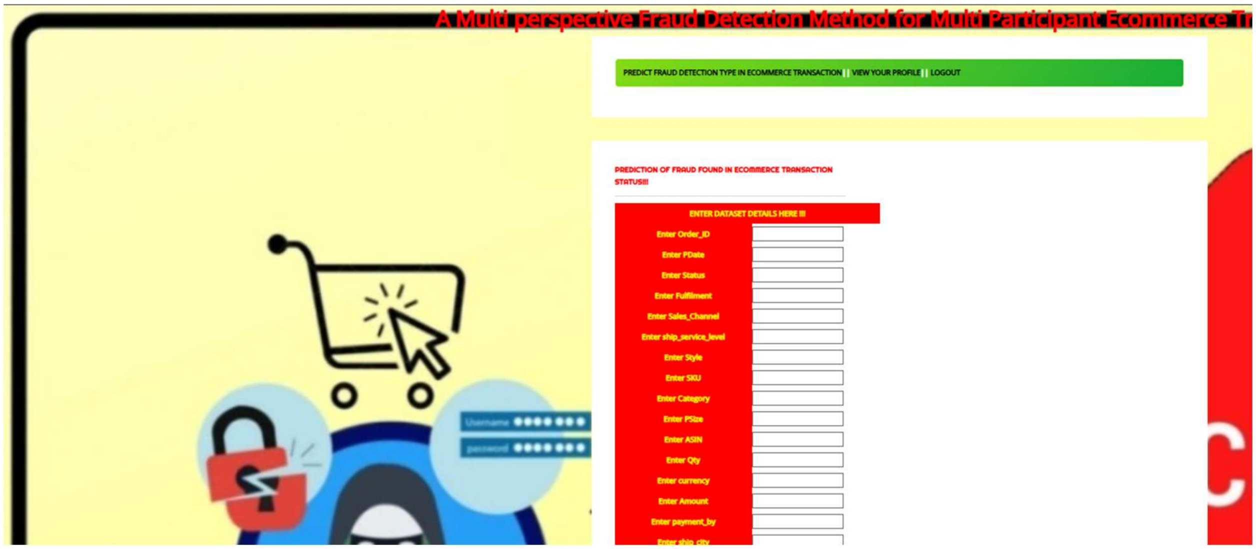


ENTER DATASET DETAILS HERE !!!

Enter Order_ID	182.22.31.252-10.42.0.21
Enter PDate	4/30/2022
Enter Status	Cancelled
Enter Fulfilment	Merchant
Enter Sales_Channel	Amazon.in
Enter ship_service_level	Standard
Enter Style	SET389
Enter SKU	SET389-KR-NP-S
Enter Category	Set
Enter PSize	S
Enter ASIN	B09KXVBD7Z
Enter Qty	0
Enter currency	INR
Enter Amount	647.62
Enter payment_by	Credit Card
Enter ship_city	MUMBAI
Enter ship_state	MAHARASHTRA
Enter ship_postal_code	400081
Enter ship_country	IN

ECommerce Transaction Fraud Found Status ::

SCREENSHOT 6 – FRAUD PREDICTION PAGE

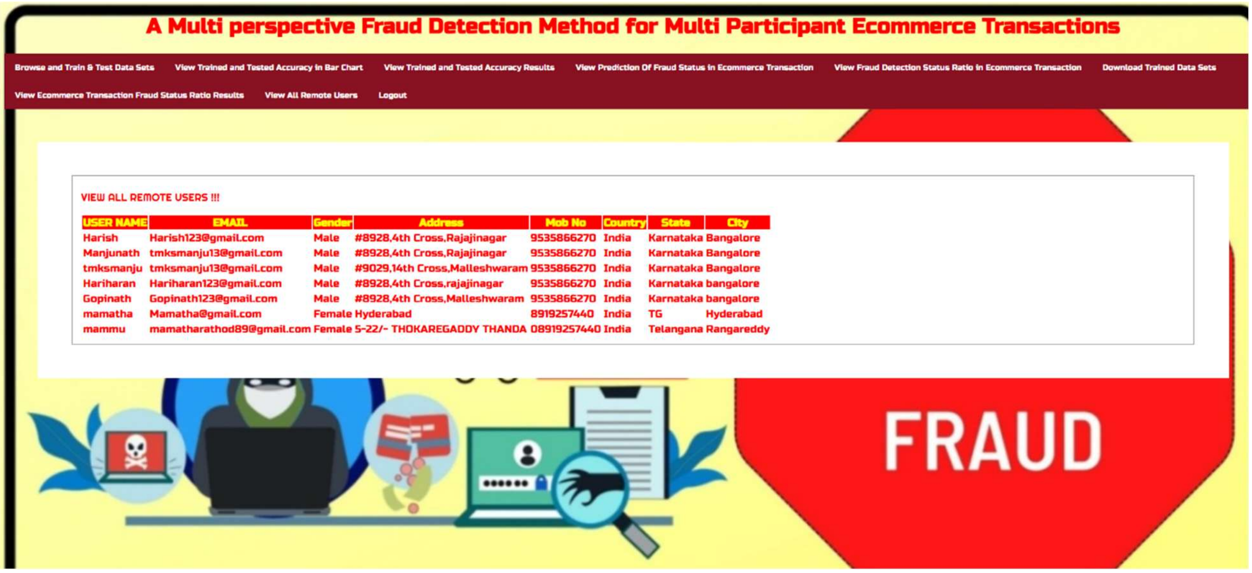


A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Tr
PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION || VIEW YOUR PROFILE || LOGOUT
Username
password

PREDICTION OF FRAUD FOUND IN ECOMMERCE TRANSACTION STATUS!!!

ENTER DATASET DETAILS HERE !!!	
Enter Order_ID	
Enter PDate	
Enter Status	
Enter Fulfilment	
Enter Sales_Channel	
Enter ship_service_level	
Enter Style	
Enter SKU	
Enter Category	
Enter PSize	
Enter ASIN	
Enter Qty	
Enter currency	
Enter Amount	
Enter payment_by	
Enter ship_city	

SCREENSHOT 7 – FRAUD PREDICTED STATUS

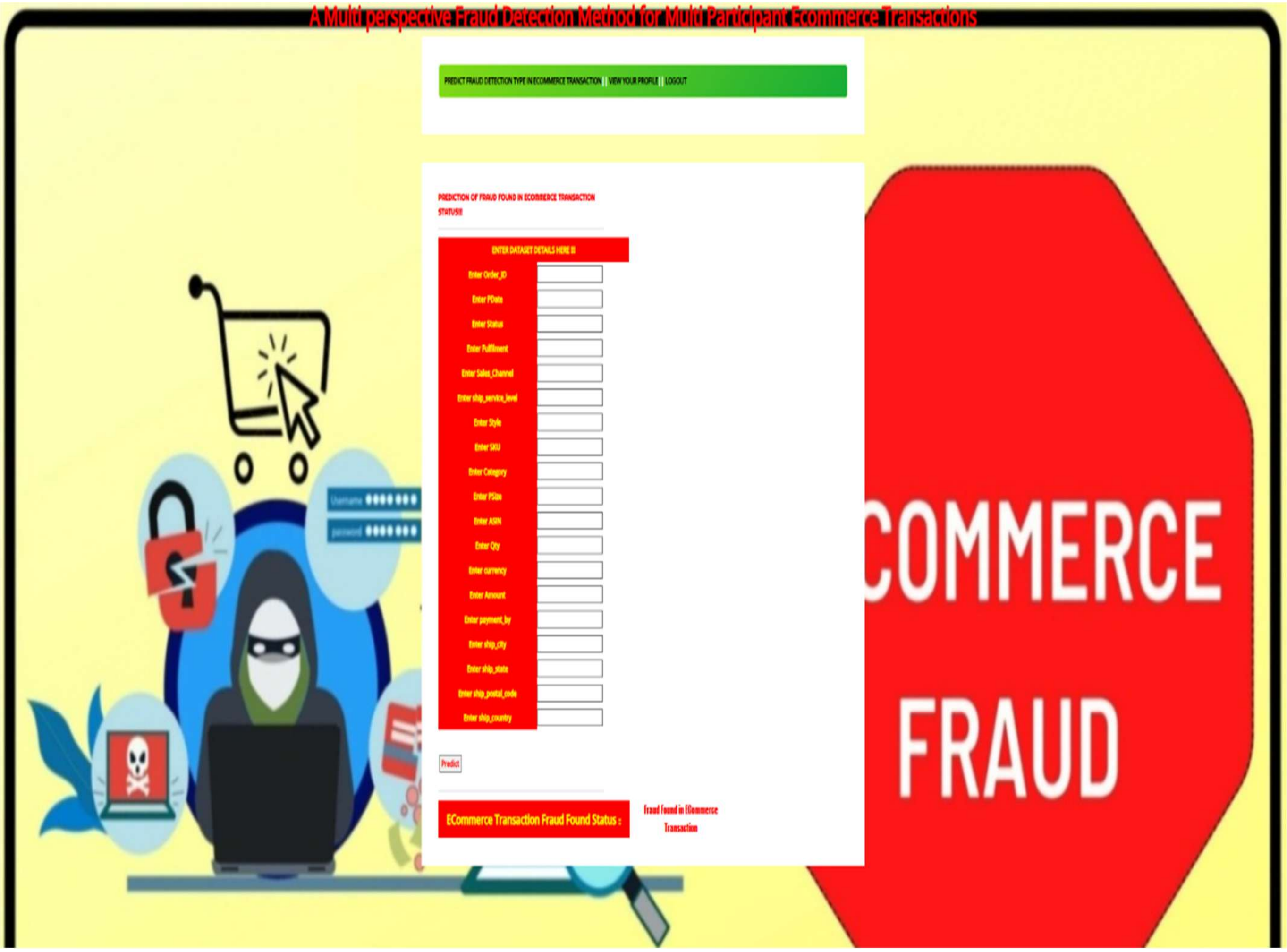


A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Fraud Status in Ecommerce Transaction
View Fraud Detection Status Ratio in Ecommerce Transaction
Download Trained Data Sets
View Ecommerce Transaction Fraud Status Ratio Results
View All Remote Users
Logout
FRAUD

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Gender	Address	Mob No	Country	State	City
Harish	Harish123@gmail.com	Male	#8928,4th Cross,Rajajinagar	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	Male	#8928,4th Cross,Rajajinagar	9535866270	India	Karnataka	Bangalore
tmksmanju	tmksmanju13@gmail.com	Male	#9029,14th Cross,Malleshwaram	9535866270	India	Karnataka	Bangalore
Hariharan	Hariharan123@gmail.com	Male	#8928,4th Cross,rajajinagar	9535866270	India	Karnataka	bangalore
Gopinath	Gopinath123@gmail.com	Male	#8928,4th Cross,Malleshwaram	9535866270	India	Karnataka	bangalore
mamatha	Mamatha@gmail.com	Female	Hyderabad	8919257440	India	TG	Hyderabad
mammu	mamatharathod89@gmail.com	Female	5-22/- THOKAREGADDY THANDA	08919257440	India	Telangana	Rangareddy

SCREENSHOT 8 – REMOTE USER DATA PAGE



A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions

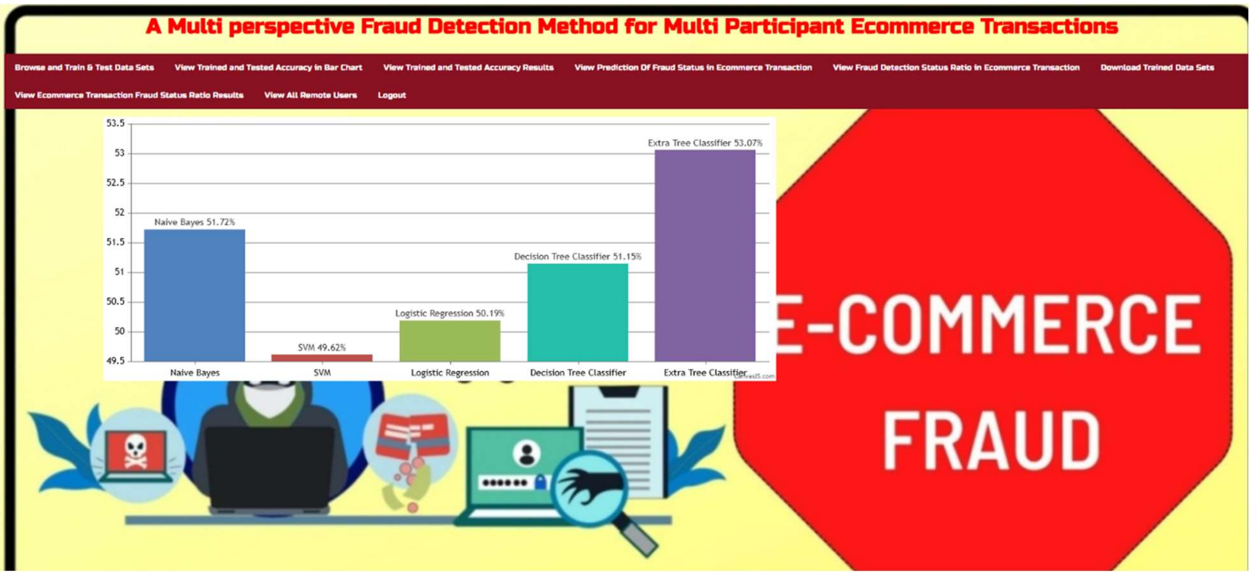
PREDICT FRAUD DETECTION TYPE IN ECOMMERCE TRANSACTION || VIEW YOUR PROFILE || LOGOUT

PREDICTION OF FRAUD FOUND IN ECOMMERCE TRANSACTION STATUS!!!

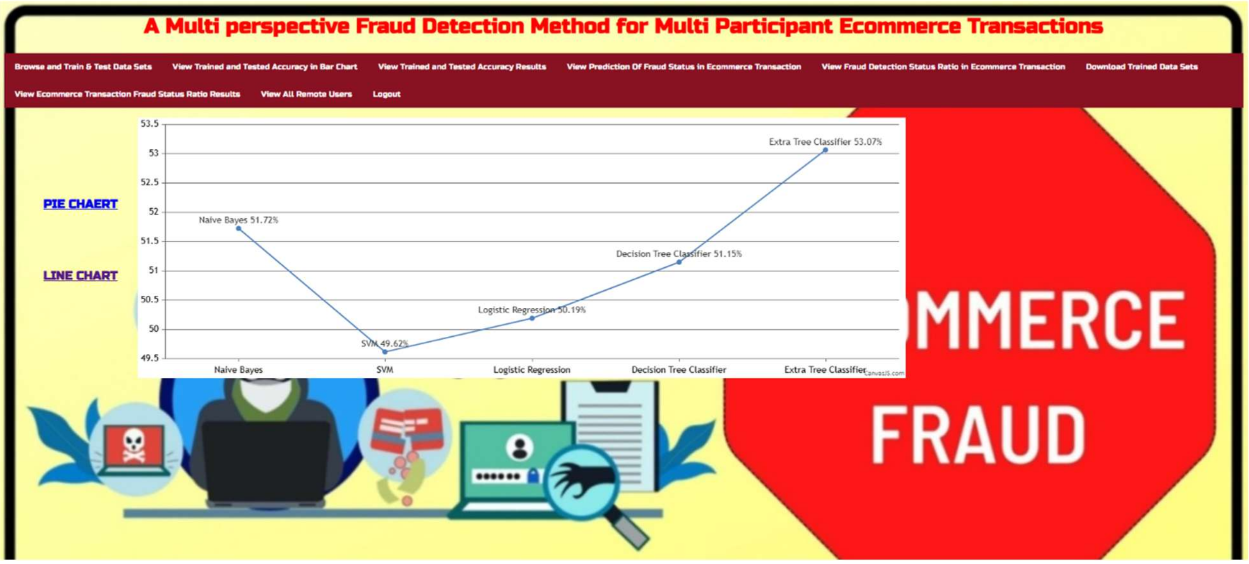
ENTER DATASET DETAILS HERE !!	
Enter Order_ID	
Enter PDate	
Enter Status	
Enter Fulfilment	
Enter Sales_Channel	
Enter ship_service_level	
Enter Style	
Enter SKU	
Enter Category	
Enter PSize	
Enter ASIN	
Enter Qty	
Enter currency	
Enter Amount	
Enter payment_by	
Enter ship_city	
Enter ship_state	
Enter ship_postal_code	
Enter ship_country	

ECommerce Transaction Fraud Found Status : Fraud Found in ECommerce Transaction

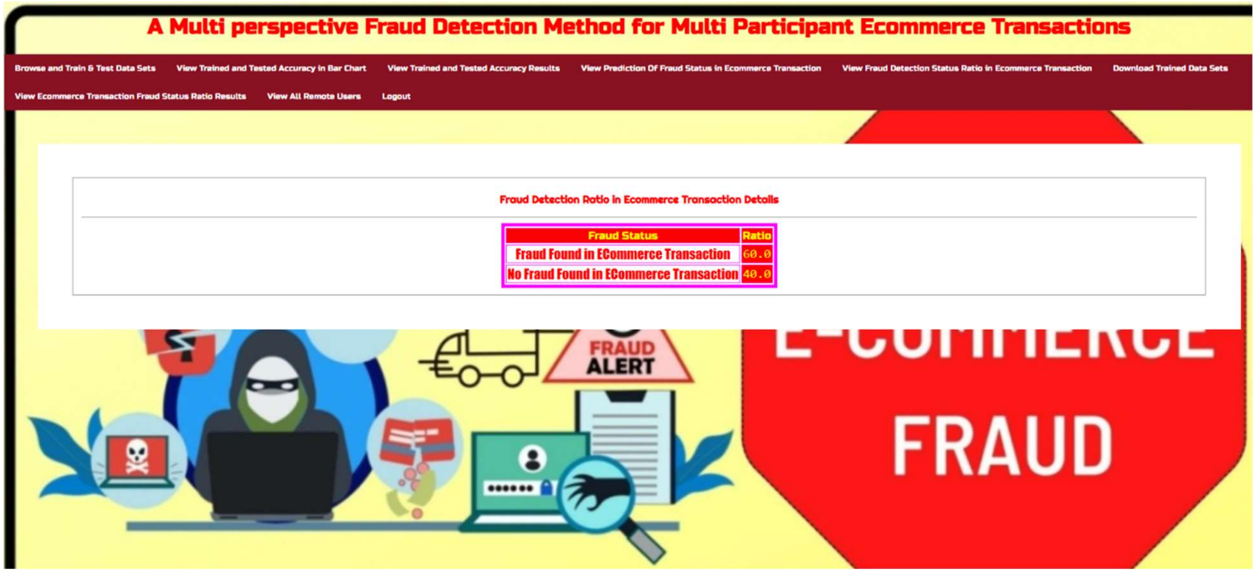
SCREENSHOT 9 – FRAUD DETECTED SAMPLE


A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Fraud Status in Ecommerce Transaction
View Fraud Detection Status Ratio in Ecommerce Transaction
Download Trained Data Sets
View Ecommerce Transaction Fraud Status Ratio Results
View All Remote Users
Logout
53.5
53
52.5
52
51.5
51
50.5
50
49.5
Naive Bayes 51.72%
SVM 49.62%
Logistic Regression 50.19%
Decision Tree Classifier 51.15%
Extra Tree Classifier 53.07%
Naive Bayes
SVM
Logistic Regression
Decision Tree Classifier
Extra Tree Classifier
CanvasJS.com
E-COMMERCE
FRAUD

SCREENSHOT 10 – BARCHAT VIEW PAGE


A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Fraud Status in Ecommerce Transaction
View Fraud Detection Status Ratio in Ecommerce Transaction
Download Trained Data Sets
View Ecommerce Transaction Fraud Status Ratio Results
View All Remote Users
Logout
PIE CHAERT
LINE CHART
53.5
53
52.5
52
51.5
51
50.5
50
49.5
Naive Bayes 51.72%
SVM 49.62%
Logistic Regression 50.19%
Decision Tree Classifier 51.15%
Extra Tree Classifier 53.07%
Naive Bayes
SVM
Logistic Regression
Decision Tree Classifier
Extra Tree Classifier
CanvasJS.com
MMERCE
FRAUD

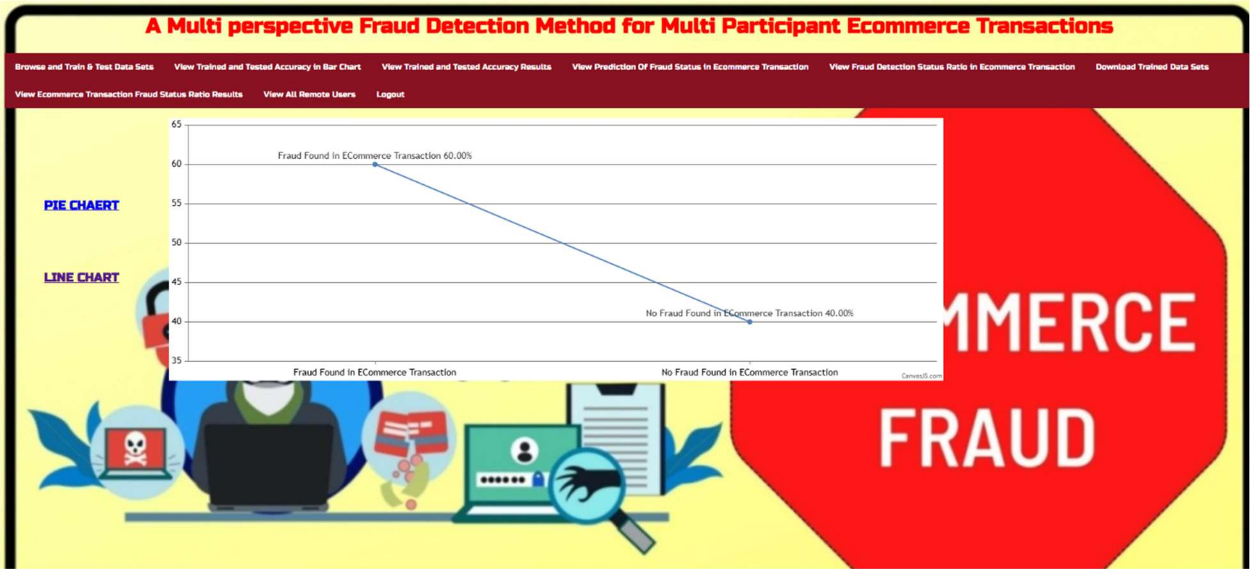
SCREENSHOT 11 – ANALYSED FRAUD PAGE


A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Fraud Status in Ecommerce Transaction
View Fraud Detection Status Ratio in Ecommerce Transaction
Download Trained Data Sets
View Ecommerce Transaction Fraud Status Ratio Results
View All Remote Users
Logout
FRAUD ALERT
E-COMMERCE
FRAUD

Fraud Detection Ratio in Ecommerce Transaction Details

Fraud Status	Ratio
Fraud Found in ECommerce Transaction	60.0
No Fraud Found in ECommerce Transaction	40.0

SCREENSHOT 12 – RATIO OF PREDICTED FRAUD PAGE


A Multi perspective Fraud Detection Method for Multi Participant Ecommerce Transactions
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Fraud Status in Ecommerce Transaction
View Fraud Detection Status Ratio in Ecommerce Transaction
Download Trained Data Sets
View Ecommerce Transaction Fraud Status Ratio Results
View All Remote Users
Logout
PIE CHAERT
LINE CHART
65
60
55
50
45
40
35
Fraud Found in ECommerce Transaction 60.00%
No Fraud Found in ECommerce Transaction 40.00%
Fraud Found in ECommerce Transaction
No Fraud Found in ECommerce Transaction
CanvasJS.com
MMERCE
FRAUD

SCREENSHOT 13 – RANGE OF PREDICTED FRAUD PAGE

CHAPTER-10

RESULTS

The experimental results demonstrate that the proposed multi-perspective fraud detection model, which integrates both control flow and data flow features, achieves superior performance compared to single-perspective models. Specifically, the model combining control and data flow achieved a precision of 0.946, recall of 0.852, F1-score of 0.895, and an AUC of 0.935. In contrast, using only data flow features resulted in a precision of 0.912, recall of 0.837, F1-score of 0.871, and AUC of 0.892, while using only control flow features yielded a precision of 0.889, recall of 0.812, F1-score of 0.849, and AUC of 0.842. These results indicate that integrating multiple perspectives leads to more accurate and comprehensive fraud detection in e-commerce transactions.

CHAPTER-11

CONCLUSION

This paper proposed a hybrid method to capture fraud transactions by integrating the formal process modeling and the dynamic user behaviors. We analyzed the e-commerce transaction process under five major perspectives: control flow perspective, resource perspective, time perspective, data perspective, and user behavior patterns. This paper utilized high-level Petri nets as the basis of process modeling to model the abnormal user behaviors and created an SVM model to perform fraudulent transaction detection. Our extensive experiments showed that the proposed method can effectively capture fraudulent transactions and behaviors. The overall index of our proposed multi-perspective detection method outperformed the single-perspective detection method. As our future work, related deep learning [38-42] and model checking methods [43-45] would be incorporated in the proposed framework for higher accuracy. Additionally, it's also a future work to incorporate more time features to the behavior patterns so as to make the risk identification more accurate. Furthermore, we will conduct research on constructing a standard fraud mode library, and apply the proposed methodology to other malicious behavior areas by coordinating the models.

CHAPTER-12

REFERENCES

[1] R. A. Kuscu, Y. Cicekcisoy, and U. Bozoklu, *Electronic Payment Systems in Electronic Commerce*. Turkey: IGI Global, 2020, pp. 114–139.

[2] M. Abdelrhim, and A. Elsayed, "The Effect of COVID-19 Spread on the e-commerce market: The case of the 5 largest e-commerce companies in the world." *Available at SSRN 3621166*, 2020, doi: 10.2139/ssrn.3621166.

[3] P. Rao et al., "The e-commerce supply chain and environmental sustainability: An empirical investigation on the online retail sector." *Cogent. Bus. Manag.*, vol. 8, no. 1, pp. 1938377, 2021.

[4] S. D. Dhobe, K. K. Tighare, and S. S. Dake, "A review on prevention of fraud in electronic payment gateway using secret code," *Int. J. Res. Eng. Sci. Manag.*, vol. 3, no. 1, pp. 602-606, Jun. 2020.

[5] A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *J. Netw. Comput. Appl.*, vol. 68, pp. 90-113, Apr. 2016.

[6] E. A. Minastireanu, and G. Mesnita, "An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection," *Info. Econ.*, vol. 23, no. 1, 2019.

[7] X. Niu, L. Wang, and X. Yang, "A comparison study of credit card fraud detection: Supervised versus unsupervised," *arXiv preprint arXiv*: vol. 1904, no. 10604, 2019, doi: 10.48550/arXiv.1904.10604. [8] L. Zheng et al., "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 3, pp. 796-806, 2018.

[9] Z. Li, G. Liu, and C. Jiang, "Deep Representation Learning With Full Center Loss for Credit Card Fraud Detection," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 2, pp. 569-579, 2020.