A design methodology for web-based services to detect fraudulent transactions in credit card

Debachudamani Prusti
National Institute of Technology Rourkela Odisha India
debaprusti@gmail.com

Abhishek Kumar
National Institute of Technology Rourkela Odisha India
219cs3466@nitrkl.ac.in

Ingole Shubham Purusottam
National Institute of Technology Rourkela Odisha India
19cs3469@nitrkl.ac.in

Santanu Kumar Rath
National Institute of Technology Rourkela Odisha India
skrath@nitrkl.ac.in

ABSTRACT

Financial fraud associated with the transactions of credit card leads to unauthorized access of performing credit card transactions in different platforms by intercepting important card credentials. In order to curb this problem, an effective fraud detection system is of primary importance for any financial institution. In the proposed methodology, a web-based fraud detection system has been designed considering two different protocols for the web-based services such as simple object access protocol (SOAP) and representational state transfer (REST). Further, for detecting the fraudulent transactions, these services are associated with five different machine learning techniques such as support vector machine (SVM), multilayer perceptron (MLP), random forest regression, autoencoder and isolation forest. The performance analysis of each machine learning algorithm associated with SOAP and REST services are critically assessed. The web services have been designed based on concepts of service oriented architecture (SOA) by considering a middleware family of software products i.e., Oracle SOA suite which is very often used by the software architects.

KEYWORDS

Machine learning, Ensemble methods, Simple object access protocol (SOAP), RESTful web service

ACM Reference Format:
Debachudamani Prusti, Abhishek Kumar, Ingole Shubham Purusottam, and Santanu Kumar Rath. 2021. A design methodology for web-based services to detect fraudulent transactions in credit card. In *14th Innovations in Software Engineering Conference (formerly known as India Software Engineering Conference) (ISEC 2021), February 25–27, 2021, Bhubaneswar, Odisha, India.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3452383.3452410

1 INTRODUCTION

Fraudulent activities in credit card transactions have been an ever-growing critical issue for the financial institutions as well as for the customers [5][16]. It is very much crucial for the financial institutions to improve the quality of fraud detection system to protect the customers from any type of financial insecurity as the fraudsters adapt various strategies constantly. By applying knowledge discovery, different techniques such as decision tree, knowledge-based reasoning, neural network, ensemble techniques etc. are used by a good number of researchers to evolve various fraud detection systems[8][25]. To learn the fraud pattern, the above techniques mostly require a larger number of normal as well as fraudulent transactions. For a particular financial institution, the ratio of normal to fraudulent transactions is very less [8]. If a financial institution fails to obtain the updated fraud patterns in due time, it suffers from fraudulent attacks. Due to the legacy formats of storing the historical transactional data, the fraudulent transactions create some sort of blockage in exchanging the information across heterogeneous platforms in the financial institutions [9][12].

In present-day scenario, the fraud detection system is desired to be web service based due to many advantages like service extensibility, web service interoperability (WS-I), which establishes a smooth channel to exchange data across heterogeneous applications [9]. Apart from this, it also comes up with various advantages such as loose coupling, composability, autonomy, statelessness, reusability and discoverability [11]. Various researchers have often considered different methodologies to design web services such as Extensible Markup Language (XML), Simple Object Access Protocol (SOAP), Web Services Description Language (WSDL), Representational State Transfer (REST), Universal Description, Discovery and Integration (UDDI) for smooth channelization of data exchange across heterogeneous applications [19].

.In this proposed study, a comparative analysis among the performances of machine learning algorithms incorporated with SOAP and REST based web services has been presented and their performances are critically assessed. Both the protocols such as SOAP and REST have their own standard features of strengths and weaknesses to define their suitability for specific applications to find the fraud pattern [2]. SOAP uses an enveloping technique to transfer the transactional data (fraudulent or legitimate) through heterogeneous platform; whereas, RESTful web service is lightweight in the absence of envelope and is used for faster data transfer with a greater variety of data formats. The existing approaches were using a single machine learning model for detecting fraudulent transactions. In this approach, we have used web service concepts with SOA based architecture to detect the fraudulent transactions by using ensemble machine learning techniques.

ISEC 2021, February 25–27, 2021, Bhubaneswar, Odisha, India
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-9046-0/21/02…$15.00
https://doi.org/10.1145/3452383.3452410

For the classification of the fraudulent and legitimate transactional data in the dataset, different machine learning algorithms have been applied by various researchers as well as practitioners [22][3]. Machine learning techniques are applied by using both supervised and unsupervised techniques to detect the fraudulent transactions. Various supervised techniques such as Naïve Bayes, k-nearest neighbor (k-NN), SVM, MLP, neural network, random forest, decision tree, linear regression and ensemble methods are applied to detect the fraudulent transactions in credit card [3][17][1]. Similarly, to monitor the real time behavior of the transactions various unsupervised techniques such as k-mean clustering, Gaussian mixture model (GMM), autoencoder, isolation forest and local outlier factor (LOF) have been considered [1]. In this proposed study, supervised techniques such as SVM, MLP and random forest regression and unsupervised techniques such as autoencoder and isolation forest are implemented with the inclusion of SOAP and REST web services to compare and assess the performance.

2 LITERATURE REVIEW

Continuous research investigation on fraud detection in credit card has been carried out by various researchers and fraud analysis practitioners by applying several algorithms and different methodologies. Since, fraudsters change their strategies in a frequent manner it is a very critical challenge for identifying the fraud in real-time as well as enacting fast and securely to retain the amount before hijacked from customer's account.

Chiu et al., (2004) have discussed about a collaborative scheme for the fraud detection based on web services, in which the financial institutions share fraud patterns in heterogeneous as well as distributed platform [9]. This study has applied the SOAP with other web services for the smooth channelization of transactional data transfer in heterogeneous platform. Also, this paper has discussed about the fraud pattern extraction from fraudulent transactions by applying Fraud Patterns Mining (FPM) algorithm, based on an apriori algorithm.

TSelykh et al, (2015) have a discussion about the REST API implementation to achieve a highly scalable, fault-tolerant anti-fraud service [23]. The objective of their paper is to construct an anti-fraud service system to minimize the chance of fraudulent payments over the online transactions.

Stolfo et al., (1997) have given focus on investigating the credit card fraud detection system by applying various techniques like Classification and Regression Tree (CART), Iterative Dichotomizer 3 (ID3), Ripper and Bayes method with heterogeneous applications [22]. They concluded Bayes method has a higher true positive rate as compared to other meta learners. But still it does not reflect any real-world impact on detecting fraudulent transactions when legitimate transactions are higher.

Maes et al., (2002) have considered Bayesian Belief Network and artificial neural network (ANN) algorithms to find the total number of true positives and false positives by classifying the fraud instances. It has been studied by them that the performance evaluation of Bayesian network is higher than ANN with lesser processing time of the Bayesian network [17]. Rather analyzing the use of traditional methods, their investigation has focused on cost sensitive credit card fraud detection based on Bayes Minimum Risk technique.

Demla, Nancy, and A. Aggarwal (2016) have proposed the application of SVM in credit card fraud detection by grouping similar instances and employed the incremental learning technique to reduce the misclassification rate and ultimately to minimize the false alarm [13].

John et al., (2019) have applied LOF and isolation forest to analyze the dataset and classify the transactions as fraudulent or legitimate [14]. This study made a brief comparison for these algorithms to investigate which algorithm provides better results and can be widely trusted by credit card merchants and users to detect fraudulent transactions.

Objective of the proposed study This study intends to evaluate the performance of the model empirically by classifying the transactions as fraudulent or legitimate with consideration of the transactional dataset. With minimization of the false positive rate, the model will be more effective to detect the class labels and ultimately identify the detection rate. The training and testing on data samples have been carried out effectively by the selected classification algorithms. For the better channelization of the data web services such as SOAP and REST are incorporated with the machine learning algorithms which provides a better predictive accuracy as well as other performance metrics. It is also further intended to develop a web-based model that acts as a set of tasks for the BPEL process manager with proper orchestration to provide a single model for the end user to detect whether the transaction is legitimate or fraudulent one.

3 WEB SERVICE-BASED APPLICATIONS

It is intended to apply both the SOAP and RESTful Web Services to implement various machine learning techniques to detect fraudulent transactions [2]. SOAP and RSET frameworks are implemented to provide the web services with a greater extent of distinguished features. Web-based services integrate the web applications such as Extensible markup language (XML), SOAP and REST over an Internet Protocol and provide a platform independent solution.

Web services are the latest embodiment of a long line of technologies for making API requests over a network. Enterprise JavaBeans (EJB) and Java's RMI are limited to Java. The Distributed Component Object Model (DCOM) is limited to Microsoft platforms. The Common Object Request Broker Architecture (CORBA) is excessively complex and does not provide backward or forward compatibility. But using SOAP and REST all the issues of Interoperability can be overcome by these Web services.

3.1 Simple object access protocol (SOAP)

SOAP helps to exchange structured data by using enveloping technique over the network [11]. Its standards are defined by W3C (World Wide Web consortium) [24]. SOAP works with two protocols such as HTTP and SMTP for exchanging data using remote procedure call (RPC). SOAP is constructed using XML specifications for encoding the messages. Metadata exchange is used to find fraudulent transactions with XML format based on SOAP web service on the consumer's request. SOAP envelope encapsulates the metadata with different performance metrics, evaluated by implementing

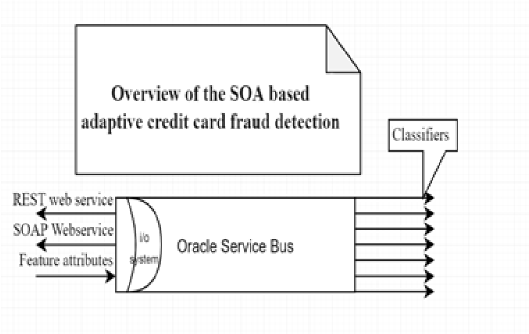

Figure 1: Overview of SOA based adaptive credit card fraud detection system

the classification algorithms. It facilitates platform independent message exchange with smooth channelization of heterogeneous applications [2].

3.2 Representational state transfer (REST)

REST represents the web services known as RESTful web services and helps to build loosely coupled web-based services as it is lightweight, faster and scalable [4]. To design RESTful services, the architectural constraints include uniform interface, client-server separation, statelessness, cacheable resources and layered system.

REST relies upon the HTTP standard to perform its work and is a very useful web service which doesn't need any strict API definition. It further updates the API definitions without ignoring the legacy system. It essentially requires HTTP (also it considers the use of XML, JSON, HTML etc.) to provide format-agnostic solutions. REST helps to decompose larger complex problems into smaller units to get the solution and later composes the individual agnostic solutions to provide the final metadata solution.

3.3 Designing services of Oracle SOA suite

This study aims at developing a web-based application with the help of 'Oracle SOA suite' is a middleware family of software products that enables to build, deploy and manage service integrations using the concepts of service-oriented architecture (SOA) [15][26]. It fulfills the enterprise organization's desire to integrate applications with mobile technologies, cloud and internet of things; all within a single a single platform. The latest release is Oracle SOA suite 12c (12.2.1.4.0) has been considered to develop the proposed model for credit card fraud detection system. An overview of SOA based adaptive credit card fraud detection system has been shown in Figure 1.

As shown in Figure 1, the input-output system takes care of the feature attributes and makes prediction based on the attributes if a transaction is a legitimate or fraudulent one. The input-output system can be exposed to end users in various formats.

- As SOAP web service: It provides feature attributes in the form of XML and takes the output in form of XML.
- As RESTful web service: It dispenses with feature attributes in any form such as XML, JSON, HTML etc. and takes response as XML or JSON or HTML
- As comma separated value (CSV) file: It provides feature attributes in .csv format for legacy system and takes the response also in the form of .csv format.

OSB (oracle service bus) comprises of a good number of supervised as well as unsupervised models and are trained on a single dataset. These models are deployed in different cloud as web services, since OSB uses the platform.

The Oracle SOA Suite components such as Oracle metadata repository, Service infrastructure, Oracle adapters, Oracle mediator, Oracle business rules, event delivery network, business events, Oracle BPEL process manager, Spring context, Oracle B2B, Human workflow, Oracle business activity monitoring, Oracle user messaging service, Oracle WSM policy manager, Oracle JDeveloper and Oracle enterprise manager. Out of these components, three major components such as Oracle Business process execution language (BPEL) Process Manager, Oracle Mediator and Oracle Adapters are considered as they determine the business structure and control the business behavior.

The focus of the paper is to propose a new methodology based on Service oriented architecture to detect the fraudulent transactions, where different exposed machine learning models are as rest service orchestrated using BPEL to provide a singular result.

3.3.1 Oracle BPEL process manager. Process (Placing a successful order) consists of different tasks like checking the inventory for availability of the ordered product, managing the delivery of product, managing the bill payment and so on. These all services are needed to be executed in a particular order (Orchestration, which is also called business process logic) to deliver a successful business process. To do this, the Oracle SOA suite provides the constructs in forms of activities using the Oracle BPEL Process Manager Component.

3.3.2 Oracle mediator. This is one of the major components in Oracle SOA suite having the mediation capability. It is useful for selective routing, validation capabilities, transformation along with the idea of synchronous, asynchronous and event publishing message exchange patterns.

3.3.3 Oracle adapters. It uses Java component architecture (JCA) to connect external system from Oracle SOA suite.

Threats to validity: For processing large XML documents, it should be assigned as a part otherwise; it might be deleted from the memory as well as from the database.

4 MACHINE LEARNING ALGORITHMS

For detecting the fraudulent transactions from the dataset, the classification algorithms are implemented with the association of web-based services. In this proposed study, three number of supervised techniques and two number of unsupervised techniques are applied with the web services after performing the majority voting method among the classification models [14][21].

4.1 Support vector machine (SVM)

SVM is a supervised machine learning technique that is widely used for classification of data samples as positive or negative instances [6]. To classify the objects or samples SVM algorithm develops a hyperplane as a decision boundary and separates the samples as positive or negative class by maximizing the margin of separation. It separates the samples either with linearly separable case or non-linearly separable case. In linearly separable case, one or more hyperplanes separate the positive or negative classes by training all the data samples. Whereas, in non-linearly separable case, it is difficult to locate a linear hyperplane that can separates the samples [7]. To resolve this, the maximization technique for margin of separation is being relaxed by permitting the data access points on the reverse side of the margin to classify the positive and negative samples.

4.2 Multilayer perceptron (MLP)

MLP has the capability to approximate the non-linear functions to a higher extent for achieving the predictive accuracy [17]. It is also a supervised learning system used to train the classification model. It comprises of three of layers such as input layer which takes the values as input variables, hidden layer that captures the non-linear relationships among all the feature attribute variables and the output layer provides the predictive values of the result by taking the input from the hidden layers. The neurons of a lower layer are connected to all nodes of the upper layer through a set of connecting weights. After computing the weighted sum of the lower layer, it passes to the continuous nonlinear function. While training the model, MLP input pattern is applied and the output of all the nodes in each layer is computed.

4.3 Random forest regression

The behavior of Random forest (RF) algorithm works as ensemble learning process, where classification and regression method involves in grouping the data into classes [16]. In RF method, the predictive result is achieved by applying a group of decision trees. While training the model, the decision trees are built to predict the class label and this is achieved by using voting technique for each individual trees. The predictive value of the output depends on the class with highest votes. Independently building the trees make random forest computationally efficient and faster. Classification and regression are performed by constructing a series of decision trees during the training process and in output the classification is the mode of the classes and regression is the mean for prediction of individual trees. In random forest, the predictions are attended by one argument and response with another argument. If the response acts as a factor, then random forest operates with classification and if the response is not a factor, random forest behaves as a regression model.

4.4 Autoencoder

Autoencoder belongs to the artificial neural network family where it learns the data coding in an unsupervised approach [20]. The objective of autoencoder is to represent a set of data for dimensionality reduction by training the network and ignores noisy signals. Along with the reduction side, a reconstructing side is learnt in parallel, where the autoencoder attempts to develop the reduced encoding representation close to its original input. The hidden layer describes the code to represent the input with two main part such as encoder and decoder. The encoder maps the input into the code and the decoder maps the code to construct the original input. Autoencoder is capable to learn the representation of higher dimensional data to lower dimensional data.

4.5 Isolation forest

Isolation forest is originally a tree-based model that helps to detect the outliers [14]. This algorithm considers various anomalies as different data points. It operates with some susceptible mechanism to isolate the anomalies. It is quite different from other existing methods to detect the anomalies on the basis of isolation rather than distance and density measures. This algorithm needs a small memory and low linear time complexity to build a good performing model with lesser number of trees and uses minimum sub-samples of fixed size, regardless the dataset size. ***Ensemble of Classification Algorithms*** Ensemble technique combines a group of weak learners iteratively to construct a strong learner which can classify the data samples more efficiently [11]. As compared to random guessing, weak learners are better since they are added one by one iteratively to form a strong learner. The performance of the ensemble model increases gradually, where the classification accuracy is nearer to the correct value. It has been observed from research investigation; instead of a single classifier an ensemble classifier achieves best accuracy for a unique set of problems. Also, it improves the scalability and efficiency to classify the fraudulent and normal class with proper model training.

To provide an effective performance the classifiers are combined to form a meta-classifier by considering the majority voting technique [18]. The classifier votes to a specific class label and the class label more than half of the votes is considered to be final class label. The classification model with majority of the votes generates a better predictive model and ultimately develops a hybridized classification model having strong generalization ability.

5 DATASET USED FOR EXPERIMENT

The credit card fraud detection dataset considered in this study based on the transactions carried out by European credit card holder (https://www.kaggle.com/mlg-ulb/creditcardfraud). The dataset contains a total of 284807 transactional data samples. Out of which total normal transactions are 284315 and fraudulent transactions are 492. The dataset is highly unbalanced with fraudulent class sums to 0.1727%. The dataset is a principal component analysis (PCA) transformation of the original data due to confidentiality issues. PCA is used to explain the variance-covariance structure of a set of variables through linear combinations. The dataset contains a total number of 31 features with last column as class. The features are amount (V0), time (V1) and V2, V3…V29 correspond to the attributes of the transactional dataset. The features i.e., amount (V0) and time (V1) are not obtained from the PCA transformed method whereas, V2, V3…V29 are PCA transformed. The training and testing set split is done by reshuffling the dataset for multiple times and the best result is presented. The dataset is split in 75% and 25% ratio for training and testing respectively, since in this ratio the accuracy and other performance parameters are quite improved for the classification models using web-based services. In the binary classification problem, the class feature takes two values 0 and 1, which represents the normal and fraudulent class respectively. It is observed that the performance of the considered ML algorithms is quite satisfactory on the particular dataset.

6 PROPOSED METHODOLOGY

Different web services based on SOAP as well as REST are designed in such a manner that each web service caters to a particular machine learning algorithm which is collaboratively implemented considering the transactional data from the dataset. The dataset is applied for both training as well as for testing in such a manner that the dataset is split in 75% and 25% ratio correspondingly. The performances of the models based on SOAP and REST web services are subsequently analyzed. Their performance depends on the types of API used, whether SOAP API or REST API.

In this proposed model as shown in Figure 2, five machine learning algorithms have been considered out of which three algorithms are based on supervised machine learning algorithms and two algorithms are based on unsupervised machine learning algorithms. Three supervised algorithms are SVM, MLP and random forest regression model. Also, the two unsupervised techniques used are autoencoder and isolation forest model.

As shown in Figure 2, five classifiers have been applied are SVM as RestService, MLP as RestService, autoencoder as RestService, random forest regression as RestService and isolation forest as RestService. Considering the model as RestService, it is configured as to take Features attributes using the Input placeholder of the invoke activity and generates output in the form i.e. received by the output place holder of the invoke activity.

The functionalities of various web services catering to different machine learning algorithms are tested by preprocessing the data initially. They are trained with a classifier as well as a regressor by considering the performance parameters independently. After the development of these machine learning models, they are deployed as a REST webservice on WebLogic application server provided by the Oracle service suite. These RSET based models act as a set of tasks for the BPEL process manager and they are orchestrated to provide a single model for the end user to detect whether the transaction is legitimate or fraudulent one. As shown in Figure 3, there are three swim lanes such as left lane, right lane and middle lane.

Left swim lane consists of SOAP or REST based web service i.e. exposed to credit card fraud detection service which takes the reduced attributes and extract the right features from the transactional data [10]. It is to be used by the end user or the other applications. The left lane of the system consists of a REST based service named a ClientRestService in the proposed model. It takes features attributes (transaction) from the end user, stores them as the variables in the request section of the REST method definition as V0, V1....... V29 and generates response (whether transaction is legitimate or fraudulent) in form of XML. These attributes of features which are stored in the variables named as V (0...29) are permitted to pass onto the BPEL Process Manager, that exists in the Middle Swim lane.

Right Swim lane consists of external services used by the fraud detection system in order to determine the class associated with the given transaction. In this framework, five different classifiers such as random forest regression, isolation forest, autoencoder, SVM and MLP are deployed on a RESTful protocol, coupled with WebLogic application server. Since the oracle SOA suite is a tool for service-based integration, it comprises of various WebLogic servers needed for different business activities. The output of the classification models is combined with the fraud detection system through the business process that lies in the middle lane. As mentioned above, in the right swim lane, we have external services available to help in achieving the goal of business process, where the objective is to detect the fraud in the transaction. The attributes of features from the right swim lane are passed onto the BPEL manager available in the middle lane.

Middle Swim lane: The attributes of the features obtained from classifiers available in the right swim lane are received by the activity manager for orchestration of different tasks by the BPEL manager. Different tasks carried out by the activity manager are process activity, service activity, flow activity for parallel processing, which are being sequenced by five model-processors (such as Model1processor, Model2Processor, Model3Processor, Model4Processor and Model5Processor). Each sequence has a responsibility to invoke a model (namely Model1Invoker) using Invoke activity. A REST based Model is available in right swim lane that receives the output produced by the middle swim lane in the output placeholder (Model1InvokeOut).

The flow of output sequence of different classifiers using the concept of majority voting are passed from the right swim lane to the middle swim lane as shown in Figure 4. The BPEL process manager executes all the activities after receiving all the features. The DetectionBusinessProcess BPEL component receives input from the exposed REST service using receive activity of the BPEL component. Different tasks carried out by the activity manager are process activity, service activity, flow activity for parallel processing etc.

7 RESULT ANALYSIS AND DISCUSSION

Various performance parameters like accuracy, recall, precision, F-measure, Matthews correlation coefficient (MCC), fall-out are evaluated using confusion matrix and their importance for the model development is explained. MCC is used as a measure for the quality of the binary classes and is considered as a balanced measure for different class size. Fall-out is defined as the false positive rate evaluated for the probability of falsely rejecting the misclassified data samples.

Confusion matrix The performances of the classification models considered in this proposed study are critically assessed with the help of the elements associated with confusion matrix, which indicate as to whether the classified data samples are positive or negative instances. A confusion matrix framework of two class classification is presented as a 2 X 2 table designed by counting the quantity of the four results of a binary classifier, being denoted as true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

- True positive (TP): The predicted fraudulent transactions are the actual fraudulent transactions.
- False positive (FP): The normal transactions are predicted as fraud and warned with a false alarm. With a smaller number of false positive counts, the false alarm minimizes.

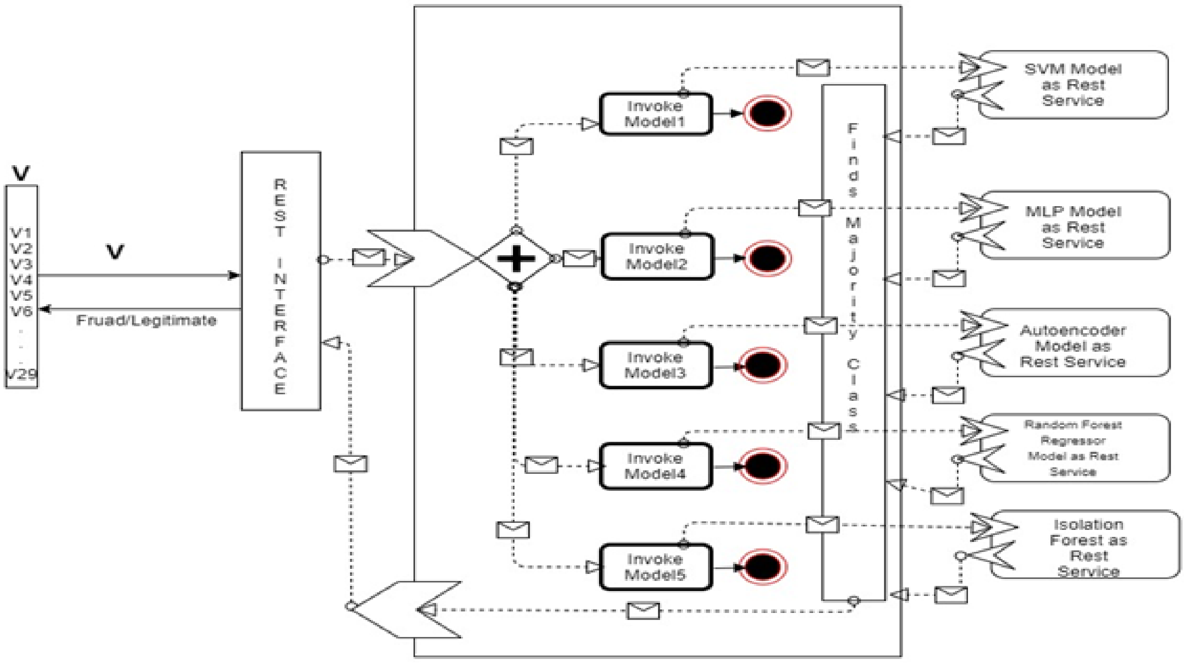
V
V1
V2
V3
V4
V5
V6
.
.
.
V29
V
Fruad/Legitimate
REST INTERFACE
Invoke Model1
Invoke Model2
Invoke Model3
Invoke Model4
Invoke Model5
Finds Majority Class
SVM Model as Rest Service
MLP Model as Rest Service
Autoencoder Model as Rest Service
Random Forest Regressor Model as Rest Service
Isolation Forest as Rest Service

Figure 2: Proposed architectural model with REST web service interface

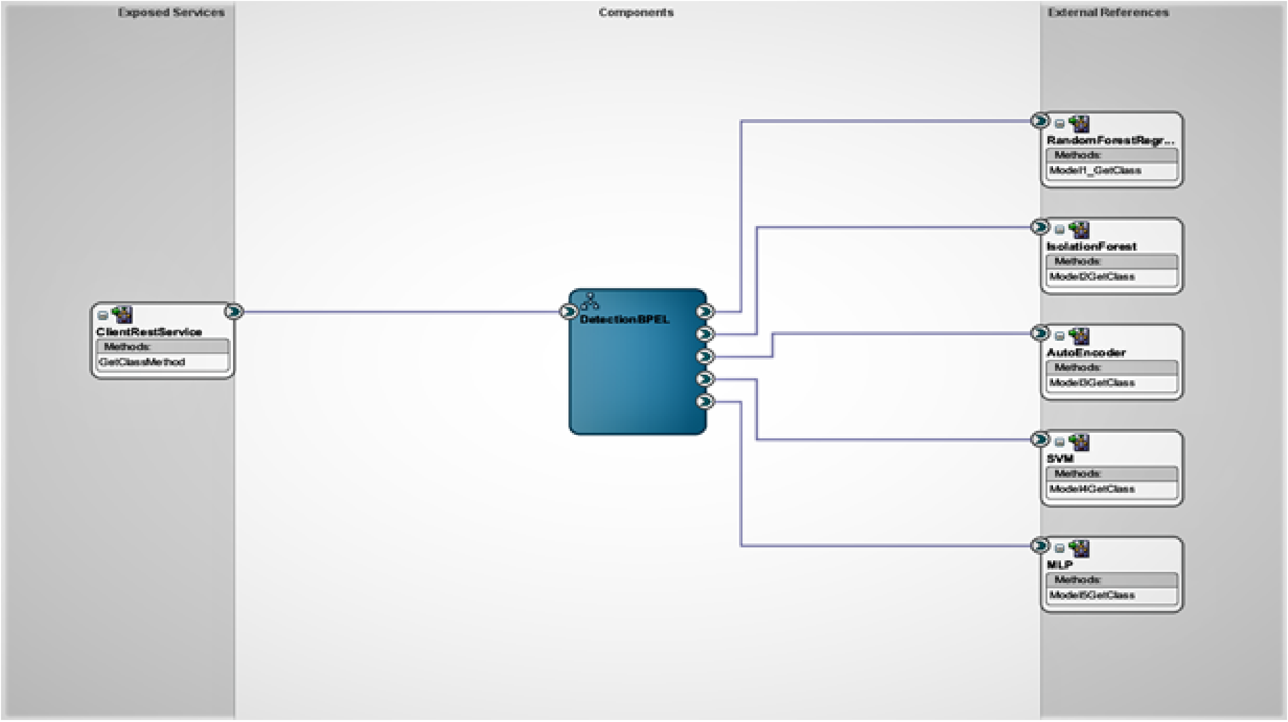

Figure 3: Components used in different swim lanes for orchestration of services to achieve business goal

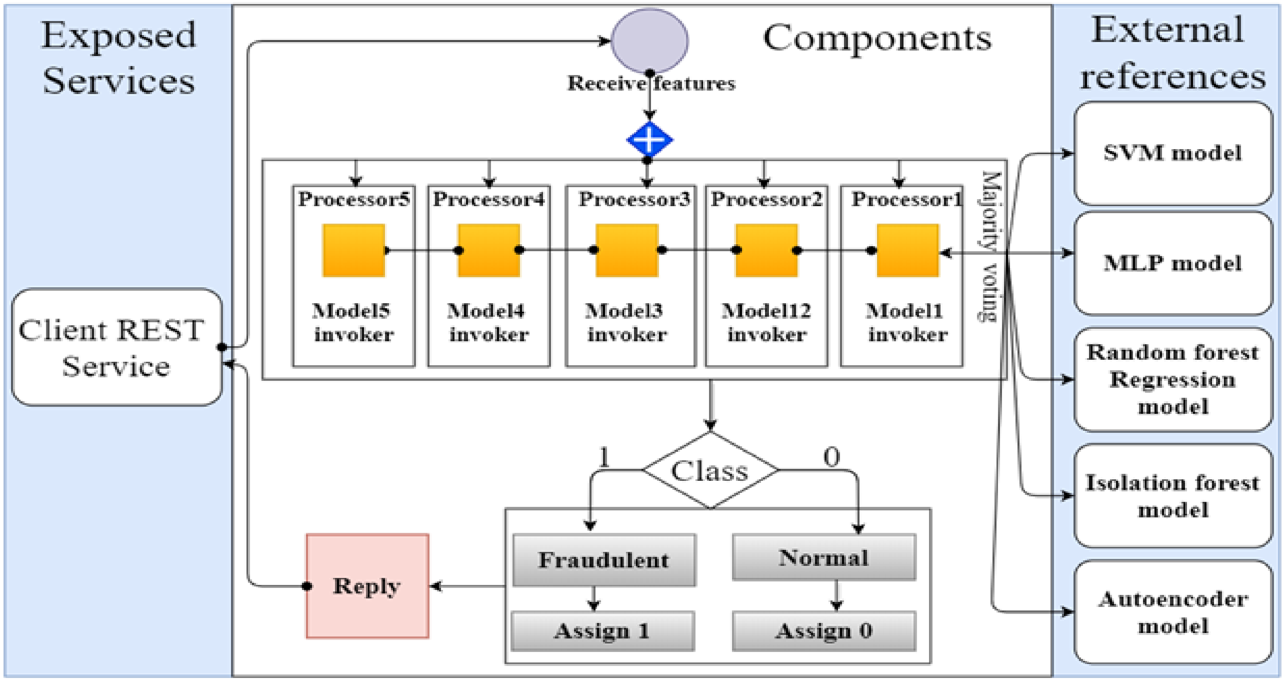

Figure 4: BPEL Component of the middle swim lane

- True negative (TN): Number of normal transactions are predicted as a normal one.
- False negative (FN): It identifies fraudulent transactions as a normal transaction.

The performance results evaluated empirically are analyzed by using both SOAP and REST web services. Five classification algorithms are trained with the dataset with 75% data instances and testing is performed with 25% of transactional data. The algorithms are implemented through majority voting technique with SOAP as well as REST web services separately. REST works in a faster lite infrastructure mode with a greater variety of data formats by utilizing lesser bandwidth that helps to improve the result of the models, whereas SOAP uses only XML. By considering the SOAP web service with five classification techniques, the performances of each algorithm have been empirically assessed as shown in Table 1. With SOAP web service, the transactional data transfer is made with the enveloping technique in a secure channel. The performance measures for the machine learning algorithms are considered for further analysis and comparison. The accuracy percentage of random forest regression model with the SOAP web service is observed to be 99.01%. The performance evaluation of different models using

Table 1: Performance evaluation of the machine learning techniques (in %) by using SOAP web service

Methodology used	Accuracy	Recall	Precision	F-measure	MCC	Fall-out
SVM	97.64	97.43	98.02	97.72	88.65	31.07
MLP	98.45	97.27	98.21	97.74	84.33	14.58
Random forest regression	99.01	98.65	98.73	98.69	91.33	08.22
Autoencoder	94.16	98.39	97.89	98.14	81.44	27.64
Isolation forest	91.97	93.77	98.64	96.14	56.91	29.83

Table 2: Performance evaluation of the machine learning techniques (in %) by using REST web service

Methodology used	Accuracy	Recall	Precision	F-measure	MCC	Fall-out
SVM	98.94	99.21	98.91	99.06	85.96	24.98
MLP	99.14	99.72	98.36	99.04	92.87	09.11
Random forest regression	99.97	99.86	98.91	99.38	99.76	00.01
Autoencoder	97.10	98.43	98.47	98.45	76.57	21.65
Isolation forest	95.66	95.97	99.59	97.75	42.39	22.61

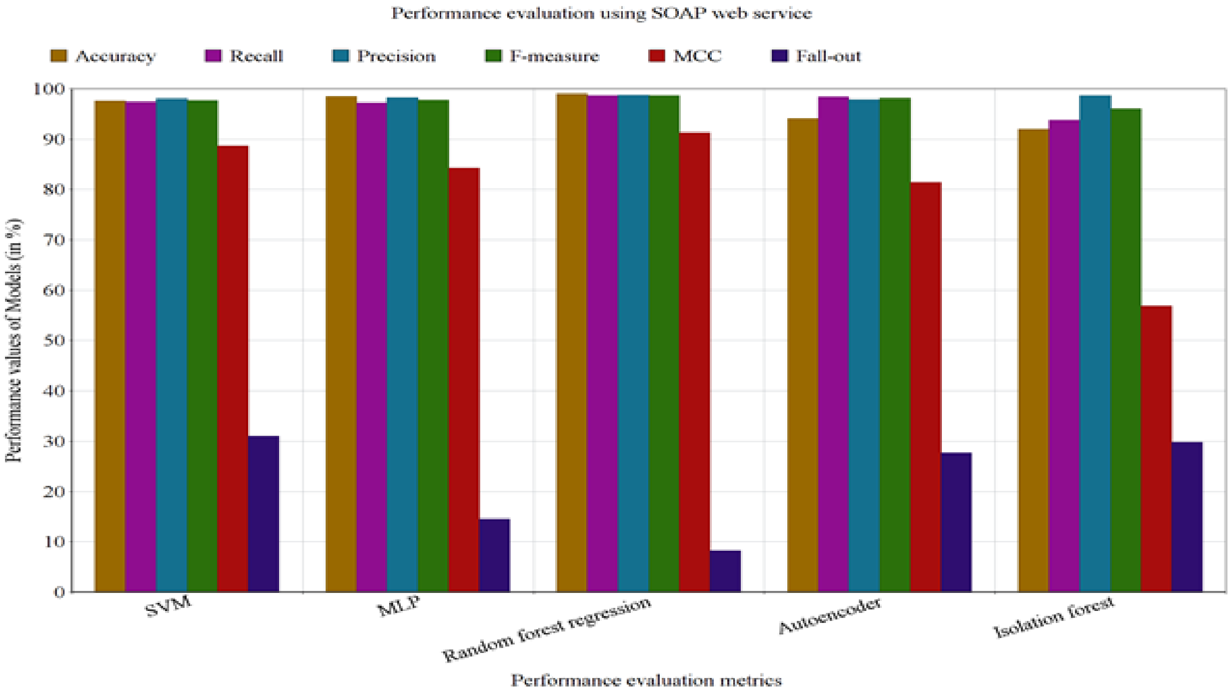

Figure 5: Comparison of various machine learning algorithms by considering the SOAP web service

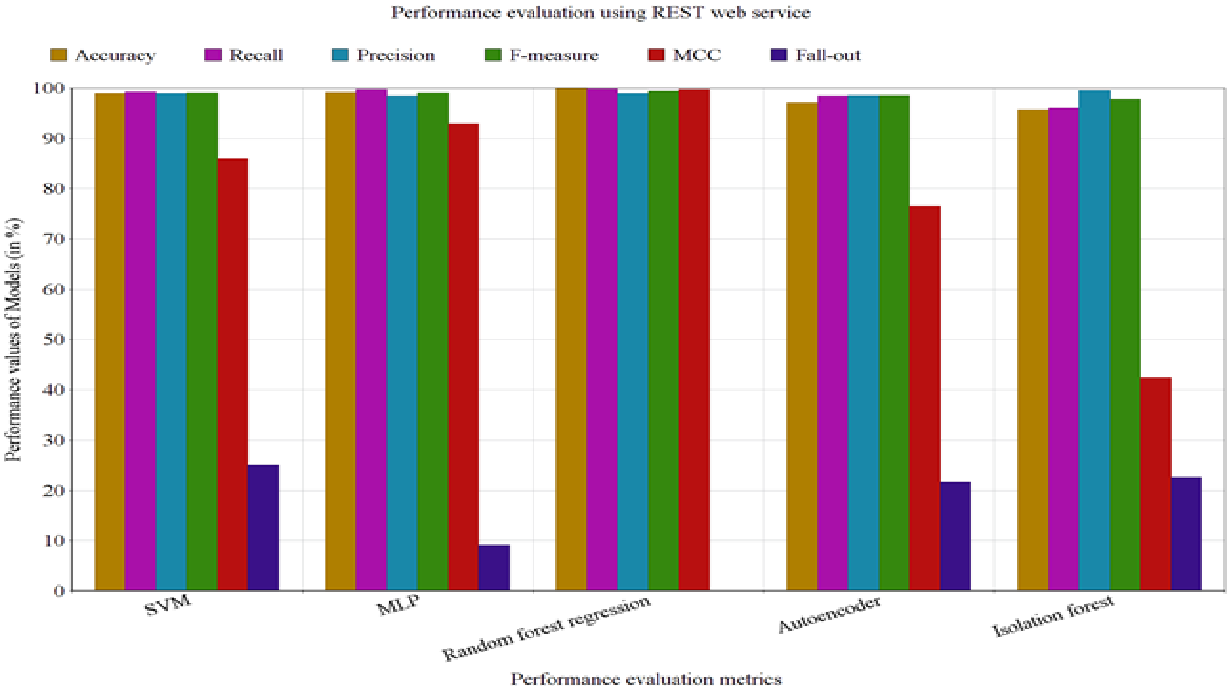

Figure 6: Comparison of various machine learning algorithms by considering the REST web service

SOAP based web service has shown in Figure 5. With the consideration of REST web service collaborated with five machine learning techniques, the performances of each algorithm have been empirically evaluated and represented in Table 2. With the inclusion of REST web service, the transactional data transfer is faster with a smooth channelization. It has been observed that there is a significant improvement in the performances. The accuracy percentage of random forest regression model with REST web service is observed to be 99.97%, which is highest among all the classification models. The performance evaluation of different classification models using REST based web service has been shown in Figure 6.

8 CONCLUSION AND FUTURE WORK

For the web-based fraud detection system, two different web services such as SOAP and REST have been considered in this study, where the services are associated with different classifiers based on various machine learning algorithms for channelization of the transactional data transfer in heterogeneous and distributed environment. The performance metrics of the machine learning algorithms are critically assessed in accordance with various parameters based on the European transactional dataset.

The performance of the classification algorithms with REST based web services is observed to be improved significantly as compared to the performance of classification algorithms with SOAP based web services. REST has a good number of advantages over SOAP and popularly accepted because of greater variety of data formats whereas, SOAP only allows XML data format. Although SOAP is more secured than REST, still REST is preferred because of fastness and wide flexibility. In future work, other categories of web services can be considered with different supervised and unsupervised machine learning algorithms for critical assessment.

REFERENCES

[1] Vikrant Agaskar, Megha Babariya, Shruthi Chandran, and Namrata Giri. 2017. Unsupervised learning for credit card fraud detection. *International Research Journal of Engineering and Technology (IRJET)* 4, 3 (2017), 2343–2346.

[2] Feda AlShahwan and Klaus Moessner. 2010. Providing soap web services and restful web services from mobile hosts. In *2010 Fifth International Conference on Internet and Web Applications and Services*. IEEE, 174–179.

[3] John O Awoyemi, Adebayo O Adetunmbi, and Samuel A Oluwadare. 2017. Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 International Conference on Computing Networking and Informatics (ICCNI)*. IEEE, 1–9.

[4] S Kishore Babu, S Vasavi, and K Nagarjuna. 2017. Framework for Predictive Analytics as a Service using ensemble model. In *2017 IEEE 7th International Advance Computing Conference (IACC)*. IEEE, 121–128.

[5] Alejandro Correa Bahnsen, Djamila Aouada, Aleksandar Stojanovic, and Björn Ottersten. 2016. Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications* 51 (2016), 134–142.

[6] Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, and J Christopher Westland. 2011. Data mining for credit card fraud: A comparative study. *Decision support systems* 50, 3 (2011), 602–613.

[7] Rüdiger Brause, T Langsdorf, and Michael Hepp. 1999. Neural data mining for credit card fraud detection. In *Proceedings 11th International Conference on Tools with Artificial Intelligence*. IEEE, 103–106.

[8] Philip K Chan, Wei Fan, Andreas L Prodromidis, and Salvatore J Stolfo. 1999. Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications* 14, 6 (1999), 67–74.

[9] Chuang-Cheng Chiu and Chieh-Yuan Tsai. 2004. A web services-based collaborative scheme for credit card fraud detection. In *IEEE International Conference on e-Technology, e-Commerce and e-Service, 2004. EEE'04. 2004*. IEEE, 177–181.

[10] Dahee Choi and Kyungho Lee. 2017. Machine learning based approach to financial fraud detection process in mobile payment system. *IT CoNvergence PRActice (INPRA)* 5, 4 (2017), 12–24.

[11] Francisco Curbera, Matthew Duftler, Rania Khalaf, William Nagy, Nirmal Mukhi, and Sanjiva Weerawarana. 2002. Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI. *IEEE Internet computing* 6, 2 (2002), 86–93.

[12] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi, and Gianluca Bontempi. 2017. Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE transactions on neural networks and learning systems* 29, 8 (2017), 3784–3797.

[13] N Demla and A Aggarwal. 2016. Credit card fraud detection using SVM and Reduction of false alarms. *International Journal of Innovations in Engineering and Technology* 7, 2 (2016), 176–182.

[14] Hyder John and Sameena Naaz. 2019. Credit card fraud detection using local outlier factor and isolation forest. *Int. J. Comput. Sci. Eng.* 7, 4 (2019), 1060–1064.

[15] Matjaz B Juric. 2010. *WS-BPEL 2.0 for SOA Composite applications with oracle SOA Suite 11g*. Packt Publishing Ltd.

[16] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.

[17] Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, and Bernard Manderick. 2002. Credit card fraud detection using Bayesian and neural networks. In *Proceedings of the 1st international naiso congress on neuro fuzzy technologies*. 261–270.

[18] K Philip and SJS Chan. 1998. Toward scalable learning with non-uniform class and cost distributions: A case study in credit card fraud detection. In *Proceeding of the Fourth International Conference on Knowledge Discovery and Data Mining*. 164–168.

[19] Debachudamani Prusti and Santanu Kumar Rath. 2019. Web service based credit card fraud detection by applying machine learning techniques. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, 492–497.

[20] Apapan Pumsirirat and Liu Yan. 2018. Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. *International Journal of advanced computer science and applications* 9, 1 (2018), 18–25.

[21] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim, and Asoke K Nandi. 2018. Credit card fraud detection using AdaBoost and majority voting. *IEEE access* 6 (2018), 14277–14284.

[22] Salvatore Stolfo, David W Fan, Wenke Lee, Andreas Prodromidis, and P Chan. 1997. Credit card fraud detection using meta-learning: Issues and initial results. In *AAAI-97 Workshop on Fraud Detection and Risk Management*. 83–90.

[23] Alexey Tselykh and Dmitry Petukhov. 2015. Web service for detecting credit card fraud in near real-time. In *Proceedings of the 8th International Conference on Security of Information and Networks*. 114–117.

[24] Aaron E Walsh. 2002. *Uddi, Soap, and WSDL: the web services specification reference book*. Prentice Hall Professional Technical Reference.

[25] Richard Wheeler and Stuart Aitken. 2000. Multiple algorithms for fraud detection. In *Applications and Innovations in Intelligent Systems VII*. Springer, 219–231.

[26] Matt Wright and Antony Reynolds. 2009. *Oracle SOA suite developer's guide*. Packt Publishing Ltd.