ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 2, 551-560 (2020)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Approach to Combine an Ontology-Based on Payment System with Neural Network for Transaction Fraud Detection

Ahmed EL Orche*, Mohamed Bahaj

[1]*Faculty of Sciences and Technologies, Hassan 1st University, Settat, Morocco*

ARTICLE INFO

Article history:
Received: 17 November, 2019
Accepted: 18 March, 2020
Online: 08 April, 2020

Keywords:
Ontology
Machine learning
Neural network
Payment system
Fraud

ABSTRACT

Fraud, as regards means of payment, means the behavior of any legal or natural one that makes an abnormal or irregular use of a way of payment, elements of it or information contained therein, to improperly obtain an honest, service or enrichment, and or causing financial damage to the one that has distributed the means of payment to a user or a 3rd party; Contests in bad faith a legitimate payment order of which she is that the initiator. during this paper we are getting to propose an approach to managing the risks, it consists to mix a machine learning with an ontology-based on a payment system to succeed in this objective. Machine learning may be a field of study that improves their performance in solving tasks without being explicitly programmed by each. An ontology is that the structured set of terms and ideas that represent the meaning of an information field, whether by the metadata of a namespace, or the elements of a domain of knowledge.

1. Introduction

Artificial Intelligence (AI) has come to the fore, with many companies using them to develop their solutions and/or services. If AI is a global concept, Machine Learning (ML) is a technology that allows machines to access data so that they can learn, predict, and categorize information. ML is a branch of artificial intelligence, which is mainly based on the automatic construction of statistical models based on the widest possible body of learning.

Deep Learning is a sub-branch of this discipline, which uses as model neural networks, very complex with many layers. This approach, which has been made popular by the availability of low-cost computing power.

With AI and ML, companies can enrich and leverage this information. Finally, via pre-established models, they will be able to test the results obtained and reiterate them throughout the life cycle.

The adoption of ML has been accelerated by increased data processing power, the development of Big Data and advances in statistical modeling.

It relies on complex statistical methods and high computing power. At the heart of this concept, however, is a very simple idea. By identifying relationships, the most influential cause-and-effect of the past, a machine can learn to make accurate predictions for the future. The ML is based on powerful computers that are guided by human intelligence to sift through billions of data and identify cause-and-effect relationships. Then all this information is introduced in a variety of algorithms to come up with predictions. With time, computers improve in identifying these cause-and-effect relationships, they exploit the knowledge they have acquired and use it to refine the algorithms. It is "learning" that takes place and with a much faster processing speed than that of the human brain.

The fraud tracked and resolved by ML?

Fraud detection is a big challenge. However, Fraudulent transactions are few and represent a very small part of the activity within an organization. Nevertheless, a small percentage of the business can quickly turn into significant financial losses without the right tools and systems in place to deal with. Cybercriminals are smart. The traditional fraud schemes are no longer effective, they have made them evolve. The good news is that with Machine Learning advances, systems can learn, adapt and discover new ways to prevent fraud.

On the other hand, many strengths make ML such a powerful and effective tool in the fight against fraud:

* Ahmed EL Orche, Email: ahmed.elorche@gmail.com

- Facilitate real-time decision-making: Rule-based systems, to determine the types of orders to accept or reject, require a lot of time and manual interaction. The ML can help evaluate a large number of transactions in real-time.
- Improving Accuracy: Cybercriminals have become more sophisticated and more skilled at hiding fraud. ML can often be more effective than humans in detecting subtle or unintuitive patterns to identify fraudulent transactions. It can also help avoid "false positives", good orders that are mistakenly identified as fraudulent.
- React quickly to change: Fraudsters constantly changing tactics. ML continually analyzes and processes new data, and then updates its models autonomously to reflect the latest trends.
- Reduce Costs: Technological innovations have reduced the costs associated with ML solutions and the computer systems that can make them work. As ML improves accuracy, it also reduces costly false positives and minimizes the time and cost of manual revisions.

In general, fraud management solutions rely on two types of ML models to combat payment fraud. On one side are static models that learn to identify fraud at a given moment by sifting through millions of past transactions. Static models are effective in identifying historical patterns of fraud and tend to work well after they are implemented. The problem is that there is no way to update or to adjust these models as new patterns of fraudulent activity emerge. On the other side are ML models based on self-learning that continually integrate data from new transactions to adapt and recognize evolving fraud patterns.

Self-learning models are very effective in identifying the latest fraud techniques. However, the "black box" nature of these models makes it almost impossible for a human to follow, control or adjust what the machine learns, which means that the model can suddenly cause enormous problems if he makes bad choices and starts blocking reliable customers.

All ML solutions are not based solely on static models or self-learning models. There is a middle ground that can compensate for ML weaknesses by combining an automated system with a rules-based approach. The rules serve as a guideline for companies to better control fraud decisions in real-time.

An ontology is an explicit specification of a conceptualization and a conceptualization is an abstract and simplified worldview that one wants to represent for a given goal (Gruber, 1993). The explicit representation of information occupies a place important in software development.

In the case of classical software development, we are interested in data structures centered on algorithms that enable ease of understanding and maintenance as well as efficiency in terms of memory and execution time. What interests the developer of the ontology is rather a representation centered on a conceptualization answering the aims of the ontology. This conceptualization is based on the real world because it consists of a simplified vision of it.

The representation of knowledge by ontology has several advantages that make this technology interesting. The attributes that are retained here are its simplicity, its flexibility, the possibility of applying reasoning and the possibility of questioning it at various levels of abstraction.

Describing data in an ontology is an important phase in the approach proposed in this paper, [1] presents a method to migrate a relational database to an ontology by taking into account the semantics of the data, this method is based on two levels ontology model (TBOX) and individuals (ABOX), it is the same method as we adopted. While the [1] presents this method, the aim of the [2] is the synchronization between the RDB and the ontology.

The history of AI shows that knowledge is important for intelligent systems. In many cases, better knowledge could also be more important in solving a task than better algorithms. to possess truly intelligent systems, knowledge must be captured, processed, reused and communicated. The ontologies support of these tasks.

The term "ontology" can be defined as an explicit specification of conceptualization. The ontologies capture the structure of the domain, which is to say the conceptualization. This includes the domain model with possible restrictions. The conceptualization describes the knowledge of the domain, not the particular situation of the domain. In other words, the conceptualization does not change or changes very rarely. The ontology is then the specification of this conceptualization - the conceptualization is specified using a particular modeling language and particular terms. A formal specification is necessary to be able to process ontologies and operate on ontologies automatically.

An ontology describes an area, while a knowledge base (based on an ontology) describes a particular situation. Each knowledge-based system or agent has its knowledge base. Only what can be expressed using an ontology can be stored and used in the knowledge base. When an agent wants to communicate with another agent, he uses the constructions of certain ontologies. To understand communication, ontologies must be shared between agents.

This paper is only an extension of a work already started in papers [2]-[5], and the added value is how we combine an ontology-based on a payment system with a machine learning to detect and prevent frauds on a payment system.

In [3] and [4], an approach has been proposed and described for detecting and preventing suspicious transactions on a payment system using an ontology. the approach shares and adopts rules to prevent cases of fraud on an ontology-based on a payment system. In [2] and [5], we described the way to migrate a relational database to an ontology, this approach allows to explain well the first steps of our work, because most of the payment systems in the world use this type of databases, algorithms, and rules are well described in papers to migrate schema and data. The other papers in the references are all discussing how to fight or prevent fraud in any system via machine learning and/or how to design and present an ontology. Each paper enriches the current paper with the following. For large multi-stakeholder systems such as the electronic payment systems that we are trying to study, [6] proposes to what degree the ontology must be heavy to meet the

needs of the different players in this system. the paper presents a set of important requirements in this paper and will be pre-requisites in the design in the current ontology. in addition to that, we will also take into consideration the modeling of the graphic means used in [7], guide and help the experts in the field to determine the "best" adaptation to existing rules to capture all fraudulent transactions and, respectively, omit all legitimate transactions as presented in [8], how semantic technologies could make investigations of cybercrime more efficient in [9]. this paper also builds on the strength of BIM on semantic web technology to establish an ontology of risk knowledge [10]. in another way, our approach takes into consideration the study which presents an approach lies in the use of the technique-driven by the ontology which not only minimizes the data the cost of modeling but also makes the expert system extensible and reusable for different applications [11]. and finally, it takes into consideration the possibility of arriving at a formal integrative ontology and sufficiently general generic primitives to describe the semantics of the concepts of specialized knowledge domains is far from being acquired [12].

2. Payment systems

The Internet allows us to shop, pay bills, make transfers and buy everything without having to move. With the growth of purchases made on the web, the concern about the security of electronic payments is still present. Online trading platforms ensure secure payment methods to gain the trust of customers. To check if online payments are secure, you first need to understand things like how purchases and electronic payments work.

2.1. Payment systems architecture

A payment system is an infrastructure of the financial market dedicated to the transfer of funds by clearing and/or settlement based on a or several means of payment. It is made up of elements below:

- A formal multilateral agreement between an operator which can be a central bank, or a structure interbank and financial institutions called "participants";
- Operating rules and procedures standardized.
- A technical infrastructure agreed between operator and participants.
- A risk management system both at the level of the operator than of the participants.
- One or more means of payment. Way to payment is an instrument for transferring funds, whatever the medium or process used. As an illustration, cash, check, transfer, bill of exchange, promissory note as well that direct debit is means of payment.

Electronic payment systems involve the six participants as Figure 1 shows.

- Cardholder designates the holder of a bank card, of which he is the bearer. It is the bank that issues it, on the one hand to its name and with its logo, and on the other hand to the name of the holder.
- The payment card is a small plastic rectangle, measuring 85.60 × 53.98 mm, offered by financial institutions. It is equipped with a magnetic card and a chip, which allows electronic reading from a distance.

  Synonymous with a credit card, the payment card is both in the name of the issuing bank and in the name of its holder. It allows the latter to make withdrawals from ATMs, to pay for purchases from a large number of merchants, as well as on virtual terminals on the Internet. The amount is then debited from his bank account.
- An acceptor is a commercial or service establishment that accepts, for its account or that of its network, the payment of goods or services via an electronic money instrument.
- An acquirer is an entity or entities that hold deposit accounts for the acceptor (merchant) card and to whom the card acceptor transmits transaction data. The acquirer is responsible for collecting information on transactions and settlement with the acceptor.
- The payment system network is an institution that transmits information and funds through a payment system network. It may operate as an agent or a principal.

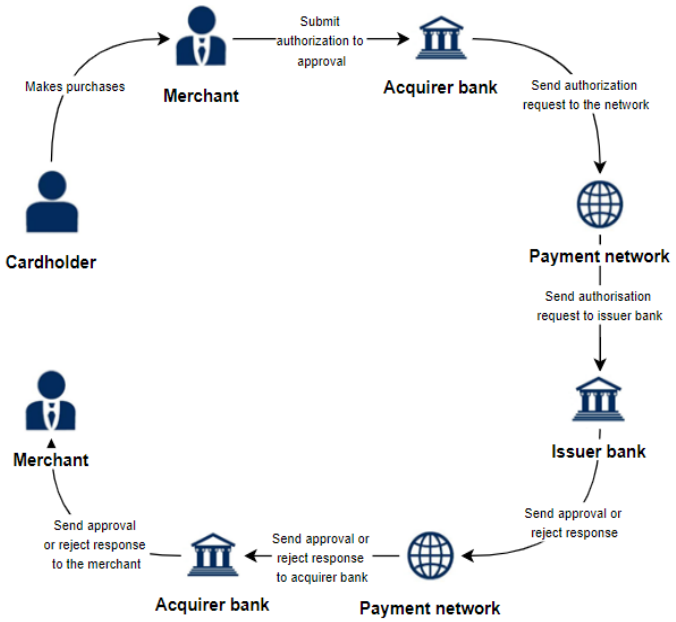

Figure 1: Payments system process

2.2. Electronic payment: principle and operation

Electronic payment replaces traditional payment in internet transactions. Today, with businesses opening up to e-commerce and Internet users increasingly connected, electronic payment is used by a majority of consumers in countries around the world. The advantage of online transactions is accessibility and speed. There is no need to move or queuing in department stores.

2.3. Electronic payment methods

To make an online payment, it should have a powered account. Several online payment methods have been developed in recent

years. The smart card can pay for purchases in the store using the TPE, but also to make payments online. With digitalization, the wallet or electronic wallet facilitates transactions. No need to enter each time the figures of a card so that a payment is validated.

Today, with the opening on electronic payments, several directives are available that help ensures that transactions are secure:

- SSL: All sites on which you are trading, must use SSL to ensure the security of sensitive information and data on their site. SSL is represented by the small padlock in the URL bar and the web address that starts with HTTPS.
- PCI compliance: it protects against fraud. The data is encrypted and is not visible. PCI DSS (Data Security Standards in the Payment Card Industry) is a standard that tells merchants who use the web the requirements they must meet. These are designed to provide data protection, detect vulnerabilities, control access, monitoring, and information security policy. PCI compliance also considers the creation of tokens in a future release.
- Tokenization: This is a way to encrypt payment data. When a person enters the details of his credit card, they will be stored in the form of a token. In other words, the token will replace this information.
- 3D Secure: this is to consider a password that will be sent to the owner of the card for example before validating the payment. In this case, every time a transaction is made with this card, the password is required.

In the same logic of these means of payment, we are confronted with a dynamic and changing market, with many new means of payment that proliferate, and we see the importance and growth of the use of payment methods more and more secure.

3. Payment system fraud

3.1. How an online transaction works

To understand the risks of fraud, one must first understand who is involved in an online transaction. There are you, your client, the payment gateway provider and the payment corporate or company (Visa, MasterCard, American Express, etc.).

When a customer pays for their order it is not PayPal for example that transfers the money, but the customer's credit company that funds. The advance will be made in the measures where the card number is valid and the holder of it has enough credit for the transaction.

3.2. Risks and consequences

In the financial world, an online transaction is considered as no card transaction. This is the most difficult type of transaction to protect, mainly because it is not possible to verify the identity of the cardholder. As a merchant, it is not necessary to be alarmed, but it is necessary to be aware that the current legal provisions favor the customer. Unfortunately, it has no guarantee that it will receive the money or any real remedy in the event of fraud. Even if the credit company authorizes the transaction if this card is declared stolen, the credit company will reimburse the legitimate cardholder and it will lose the money.

Unfortunately, payment gateway providers cannot offer recourse when the credit company reimburses its customers for fraud. However, it is possible to configure gateways so that warn the customer when a transaction seems fraudulent. People who shop online are not recognized for their patience. it should make a compromise between a platform with a quick and easy payment process, which could have vulnerabilities and a safer but more complex platform that could discourage some customers.

3.3. Fraud at the instant payment time, between tracking and profiling, the parade is called "artificial intelligence"

The issue of fraud takes on a new dimension with instant payment.

Artificial intelligence, which has been used to secure and prevent payment fraud, is even more justified for instant payment. One of its great strengths is to have flexible algorithms, which can lead to both millisecond controls on a mass of data, to adapt to changing behaviors, and profile the fraudsters. A human is unable to integrate these data in the time of his reasoning. Today, only an algorithm is adapted to block and advise the human in his decision making.

Although artificial intelligence demonstrates its ability to diagnose, anticipate and stop some of the fraud, it will remain complementary to the traditional parades implemented. The first step is to secure the hardware or acquisition channel, to block the entrance. As a priority, secure the smartphone, which is currently experiencing massive cyber-attacks, by downloading ad hoc software, listening parade and remote phone handshake. The second step, the enrollment to the service then the registration of the immediate beneficiaries, by strong authentication using application, token or biometric device based on artificial intelligence. In the third step, after entering the payment data, the payer confirms his payment by strong authentication. Fourth step, the bank or the payment service provider gives its agreement after the traditional controls of the type primo-use, ceilings, lists of beneficiaries, destination country, by integrating finer indicators generated by tools of scoring.

The success of building the Fraud Risk Management Building on the instant payment will go through a mixed.

4. Combining ontology and neural networks in the payment system.

An ontology-based on an electronic payment system is defined as a set of knowledge describing this domain, Once the ontology has been developed, it must also be managed. For example, building an ontology-based on another shared top ontology and using a modular design usually means use and easier maintenance. In this chapter, we focus on a part of the ontology related to the payment transactions to study and manage fraud cases.

This is the aim of which we will combine machine learning with this part of the ontology.

Figure 2 describes the approach using an ontology-based on the payment system and machine learning. creating and maintaining

an ontology-based on this system is not easy. and there are several ways to proceed. to adopt an approach to:

- Migrate all databases and data into ontologies according to a well-defined structure in an automatic way.
- Create ontology silently but this approach requires investigation of more resources for migration, synchronization, and maintenance.
- Follow a semi-automatic approach, by creating an ontology-based in a first step on the databases then add an extension to add more semantics and for describing non-existing cases in the database, functional and decisional cases Etc. the data migration must be automatic for the data of the database and by user intervention for external data.

Using machine learning to detect or prevent cases of fraud in this approach is done by an ontology data analysis by machine learning, extraction is generally done by the language Sparql with a data preparation in the form of machine learning.

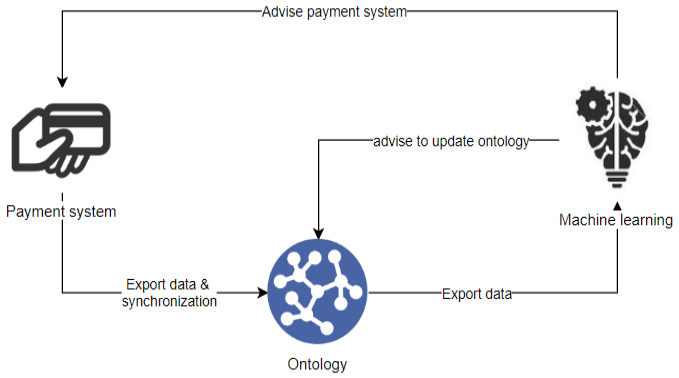

Figure 2: Combining ontology with Machine learning

In [2] and [5], we proposed an approach to migrate a relational database towards an ontology in 2 levels. the first level focuses on extracting the RDB schema, then creating an ontology model that is made up of a set of classes with data type properties and link to each other by object properties. This model constitutes the TBOX part of the ontology. The second level aims at extracting the data from the RDB and using it for assertions of the different elements of the model obtained at the first level. All of these assertions constitute the ABOX part of the ontology.

4.1. Neural network for fraud detection

The detection of payment fraud has two very particular characteristics. The first is the very limited time in which the decision of acceptance or rejection must be made. The second is the huge amount of credit card transactions that need to be processed at some point. For example, millions of card transactions take place every day.

The Operating Principle of Neural Network Based Fraud Detection relies entirely on the operating principle of the human brain. The technology of neural networks has made a computer capable of thinking. While the human brain learns from past experiences and uses its knowledge or experience to decide in everyday life, the same technique is applied with payment fraud detection technology.

When a consumer uses their credit card, there is a fixed pattern of credit card usage, created by the way the consumer uses their credit card. The use of the data neuron network for last years makes it possible to better understand the model of the use of a credit card by a consumer. The neural network contains information on the different categories of cardholders, such as cardholder occupation, income, occupation can fall into one category, while in another category, related information, this information includes the number of major purchases, the frequencies of major purchases, the place where these types of purchases take place, etc.

Within a specified time. Despite the pattern of card use, neural networks are also trained within the varied card frauds previously encountered by a bank. supported the card usage model, the neural network uses a prediction algorithm on this model data to classify the fact that a selected transaction is fraudulent or genuine. When the card is used by an unauthorized user, the neural network-based fraud detection system verifies the pattern employed by the fraudster and matches the rationale of the primary cardholder on whom the neural network has formed, if the pattern matches the neural network. declare the transaction ok.

The neural network layer during a card when a transaction arrives for authorization, it's characterized by a stream of authorization data fields that contain information identifying the cardholder and thus the characteristics of the transaction. Other data fields are often saved during a feed of the authorization system.

In most cases, banks don't archive logs of their authorization files. Only transactions sent by the merchant for payment are archived by the cardboard processing system of the bank. Thus, a gaggle of transaction data has been composed of an extract of the data stored within the Bank's settlement file.

Fraud detection model doesn't suggest that the transaction should exactly match the model, but the neural network to determine how far there is a difference if the transaction is on the brink of the model, the transaction is ok otherwise if there's a huge difference then the prospect of being an illegal transaction increases and thus the neural network declares the transaction by default. The neural network is supposed to provide a real value output between 0 and 1. There are cases where the transaction made by a legal user is different and it is also possible that the illegal person uses a card that corresponds to the model used for the formation of the neural network. Transaction OK probably fraudulent Age, income, occupation, cardholder Number of great purchases on the cardboard Frequency of major purchases Location where the transaction was made Detection of card fraud via a neural network.

4.2. Neural network

A neural network is the association of elementary objects, in a complex graph. The main networks are distinguished by the organization of the graph, that is to say, their architecture, their level of complexity, by the type of neurons and finally the objective: supervised learning or not, optimization, dynamic systems, etc.

In summary, a biological neuron is a cell that is characterized by:

- Synapses, connection points with other neurons, fibers nervous or muscular.
- Dendrites or neuron inputs.
- Axons, or exits the neuron to other neurons or muscle fibers.
- The kernel activates the outputs according to input stimulations.

By analogy, the formal neuron is a model that is characterized by a state internal s ∈ S, input signals x_1, ... , x_p and an activation function:

$$s = h(x_0, \dots, x_p) = g(\alpha_0 + \sum_{j=1}^{n} \alpha_j x_j) = g(\alpha_0 + \alpha' x) \quad (1)$$

The activation function performs a transformation of an affine combination input signals, α0, constant term, being called through the neuron.

This affine combination is determined by a vector of weight $[\alpha_0, \dots, \alpha_p]$ associated with each neuron and whose values are estimated in the learning phase. They constitute the memory or distributed knowledge of the network.

The different types of neurons are distinguished by the nature of their activation function. The main types are:

- Linear g is the identity function,
- Seuil: $g(x) = 1_{[0,+\infty[}(x)$
- Sigmoïde: $g(x) = 1/(1 + e^x)$
- Rectified linear unit: $g(x) = max(0, x)$
- Softmax: for each k ∈ {1 ... K}, $g(x)_j = \frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}}$
- Radiale: $g(x) = \sqrt{1/2\pi} e^{-x^2/2}$
- Stochastic: $g(x) = 1$ with probability $\frac{1}{1 + e^{-x/H}}$, 0 if not (H intervenes as the temperature in a simulated annealing algorithm).

Linear, sigmoidal, Rectified linear unit, softmax models are well adapted to learning algorithms involving (see below) a backpropagation gradient because their activation function is differentiable; these are the most used. The threshold model is probably more in line with the biological reality but poses learning problems. Finally, the stochastic model is used for global optimization problems of disturbed functions or again for analogies with particle systems (Bolzman machine).

4.3. Multilayer Perceptron

The multilayer perceptron (PMC) is a network composed of successive layers. A layer is a set of neurons that have no connection between them. An input layer reads incoming signals, a neuron by input x_j, an output layer responds to the system. According to the authors, the input layer that does not introduce any changes is not counted. One or more hidden layers participate in the transfer.

In a perceptron, a neuron of a hidden layer is input connected to each of the neurons of the previous layer and output to each neuron of the next layer.

For the sake of consistency, the same notations have been kept through the different chapters. Thus, the inputs of a network are still denoted $X_1, \dots, X_p$ as the explanatory variables of a model, while the weights of the inputs are parameters α, β to be estimated during the learning procedure and the output, is the variable Y to explain or target the model.

A multilayer perceptron thus realizes a transformation of the input variables: $Y = f(x_1, \dots, x_p; \alpha)$ where α is the vector containing each of the parameters *αjk`* of the jth input of the kth neuron of the `layer; the input layer *(`= 0)* is not parameterized, it only distributes the inputs on all the neurons of the next layer

A so-called universal approximation theorem shows that this elementary structure with a single hidden layer is sufficient to take into account the classic problems of modeling or statistical learning. Indeed, any regular function can be approached uniformly with arbitrary precision and in a finite domain of the space of its variables, by a network of neurons comprising a layer of finite number hidden neurons all having the same function of activation and a linear output neuron. Attention, this result, which seems contradictory to the structures of deep learning, is theoretical, it masks difficulties of learning and stability for complex problems in a very big dimension.

In the usual way and regression (quantitative Y), the last layer consists of a single neuron equipped with the identity activation function whereas the other neurons (hidden layer) are equipped with the sigmoid function.

In binary classification, the output neuron is also equipped with the sigmoid function, whereas in the case of discrimination with m classes, the output neuron integrates a softmax activation function with values in R m and sum unit. These m values are comparable to the probabilities of belonging to a class. Thus, in regression with a perceptron at a hidden layer of q neurons and an output neuron, this function is written:

$$y = f(x; \alpha, \beta) = \beta_0 + \beta'_z \quad (2)$$

where $z_k = g(\alpha_{k0} + \alpha'_k x); k = 1, \dots, q$

Suppose that we have a learning base of size n of observations ($x1_i$, ..., xp_i, y_i) explanatory variables $X_1, \dots, X_p$ and the variable to predict Y. Consider the simplest case of regression with a network consisting of a linear output neuron and a q-layer neuron whose parameters are optimized by least squares. This is generalized to any differentiable loss function and therefore to m class discrimination. The learning is the estimation of the parameters $\alpha_j = 0, p$; k = 1, q and βk = 0, q by minimization of the quadratic loss function or that of a classification entropy function:

$$Q(\alpha, \beta) = \sum_{i=1}^{n} Q_i = \sum_{i=1}^{n} [y_i - f(x; \alpha, \beta)]^2 \quad (3)$$

In elementary networks, a simple option to avoid over-learning is to introduce a penalization or regulation term, as in ridge

regression, into the criterion to be optimized. This then becomes $Q(\theta) + \gamma||\theta||^2$. The higher the value of the γ (decay) parameter, the lower the weight of the neuron inputs can take chaotic values, thus helping to reduce the risk of over-learning.

The user must therefore determine:

1. The input variables and the output variable; to submit to them as for all statistical methods, possible transformations, normalizations.
2. The architecture of the network: the number of hidden layers that correspond to an ability to deal with problems of non-linearity, the number of neurons per hidden layer. These two choices directly affect the number of parameters (weight) to be estimated and therefore the complexity of the model. They participate in the search for a good compromise bias/variance that is to say the balance between quality of learning and quality of forecasting.
3. Three other parameters are also involved in this compromise: the maximum number of iterations, the maximum error tolerated and a possible term of a regulation ridge (decay).
4. The learning rate and a possible strategy of the evolution of it.
5. The size of the sets or batches of observations considered at each iteration.

In practice, all these parameters cannot be adjusted simultaneously by the user. This one is confronted with choices mainly concerning the control of the over-learning: to limit the number of neurons or the duration of learning or to increase the coefficient of penalization of the standard of the parameters. This requires determining an error estimation mode: sample validation or test, cross-validation or bootstrap.

4.4. Combining the ontology with a neural network

Figure 3 describes an example of the ontology proposed in this paper, and that will be used as a data source for machine learning.

The ontology as proposed in [2] is defined as follow:

$$\boldsymbol{O} = \{C, DO, DT, I\} = \{TBOX, ABOX\} \quad (4)$$

Where:

$\boldsymbol{TBOX}: \{C, dataType, \{objectProperty\}/$

$C = (className, classeParent),$

$dataType = (dataTPDomain, dataTPName, dataTPRange),$

$ObjectProperty = (objectPDomain, objectPName, objectPRange)\}$

And:

$\boldsymbol{ABOX}: \{I\ /$

$I = (IName, IType, dataTypeList, objectPropertyList)/$

$dataTypeList = \{(dTPName, dTPtype, dTPvalue)\},$

$objectPropertyList = \{(object\ PropertyName, ITarget)\}\}$

Weka is open source software issued under the GNU General Public License. It is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization. all experiences done for this paper are implemented through it.

Attribute-Relation File Format (ARFF) are files developed by a machine learning project from the Computer Science Department of Waikato University for use with the Weka machine learning tool. it is an ASCII text file describing a list of instances sharing a set of attributes.

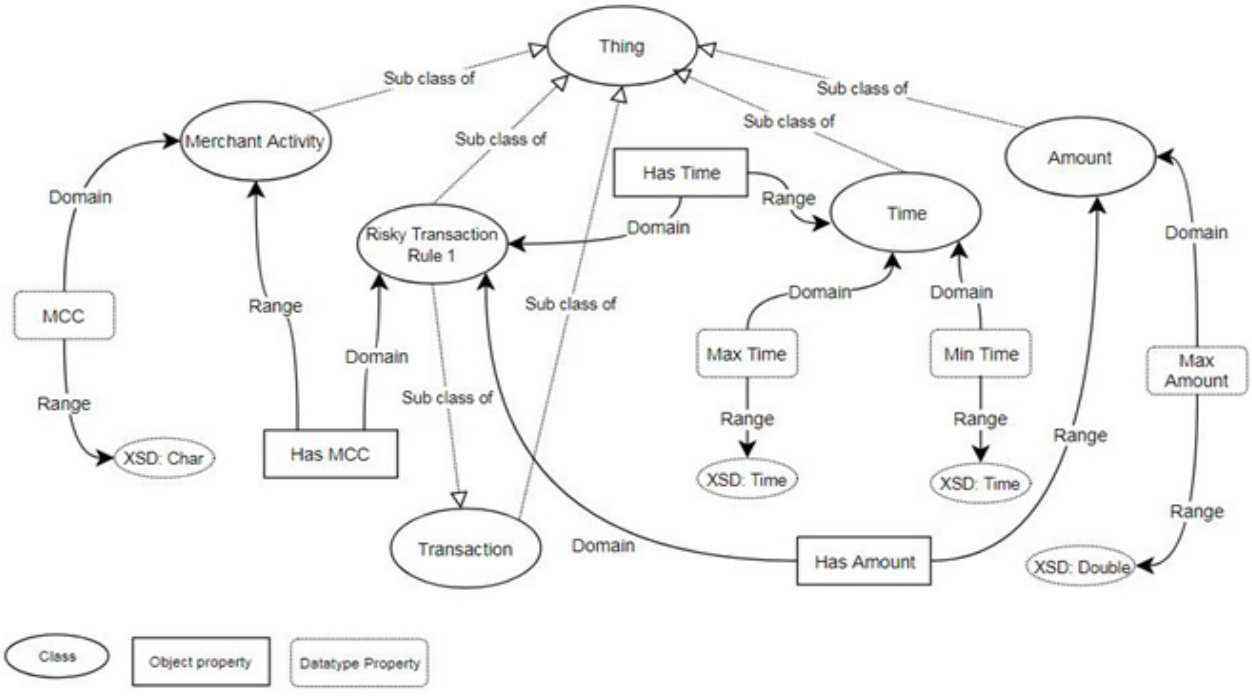

Figure 3: Part of ontology-based on a payment system

ARFF files contain two sections. The first is the Header information, the second is the Data information.

ARFF Header Section

The Header of the ARFF file contains the relation name and a list of the attributes with their types. The relation name is defined as the first line.

The format is:

@relation <relation-name>

where <relation-name> is a string and must be quoted if the name includes spaces.

The declarations of the attributes are presented in the form of an ordered sequence of attribute instructions. The attribute in the dataset has its declaration which uniquely defines the name of that attribute and its data type. The order of the attributes indicates the position of the column in the data section of the file.

The format is:

@attribute <attribute-name> <datatype>

where the <attribute-name> must start with an alphabetic character. If spaces are to be included in the name, then the entire name must be quoted.

The <datatype> can be any of the four types (version 3.2.1) supported by Weka:

- String.
- Numeric.
- <Nominal-specification>.
- Date [<date-format>].

where <nominal-specification> and <date-format> are defined below. The keywords numeric, string and date are case insensitive.

The ontology-based on a payment system is very large, but in our case, we are only interested in the part of the transactions to try to take and detect the cases of fraud, that's why we will migrate via Sparql to an ARFF file the data we need and that are related to the transaction. The extraction of data is done via the requests Sparql in the form of triples (subject, predicate, object).

Then the format of the attribute will in the case of Nominal-specification as:

@attribute <attribute-name> <'Predicate;Subject;Object'>

Example:

@relation AEO_PAYMENT_SYS

@attribute transaction_ref real

@attribute card_number real

@attribute bank_network {'BNKNWK;100001;10','BNKNWK;100001;11','BN, …

@attribute bank_bin {'BNKBIN;100001;200001','BNKBIN;100001;200002', …

@attribute bank_country {'BNKCRY;100001;301','BNKCRY;100001;302',', …

…

@attribute class {'H','M','L'}

ARFF data Section

The ARFF Data section of the file contains the data declaration line and the actual instance lines.

The data Declaration

The data declaration is a single line denoting the start of the data segment in the file. The format is: @data

The instance

Each instance is represented on a single line, with carriage returns denoting the end of the instance.

Attribute values for each instance are delimited by commas. They must appear in the order that they were declared in the header section.

@data

'230000001556231730830 12','1400007623528765','BNKNWK;100005;11','BNKBIN;100005;200009',…,L

'230000001556231383746 41','1400007626995405','BNKNWK;100005;10','BNKBIN;100005;200004',…,M

'230000001556231613823 86','1400007606073039','BNKNWK;100003;11','BNKBIN;100003;200005',…,L

'230000001556231957942 87','1400007637878184','BNKNWK;100003;10','BNKBIN;100003;200004',…,H

Figure 4 below shows an example or ARFF file that generated from ontology and will be executed into the weka tool:

```
@relation AEO_PAYMENT_SYS
@attribute transaction_ref real
@attribute card_number real
@attribute bank_network {'BNKNWK;100001;10','BNKNWK;100001;11','BNKNWK;100002;10','BNKNWK;100002;11','BNKNWK;100003;10','BNKNWK;100003;11','BNKNWK;100004;10','BNKNWK;100004;
@attribute bank_bin {'BNKBIN;100001;200001','BNKBIN;100001;200002','BNKBIN;100001;200003','BNKBIN;100001;200004','BNKBIN;100001;200005','BNKBIN;100001;200006','BNKBIN;100001
@attribute bank_country {'BNKCRY;100001;301','BNKCRY;100001;302','BNKCRY;100001;303','BNKCRY;100001;304','BNKCRY;100001;305','BNKCRY;100001;306','BNKCRY;100001;307','BNKCRY;
...
...
...
@attribute bank_currency {'BNKCCY;100001;401','BNKCCY;100001;402','BNKCCY;100001;403','BNKCCY;100001;404','BNKCCY;100001;405','BNKCCY;100001;406','BNKCCY;100001;407','BNKCC
@attribute marchant_bank {'MERBNK;500001;100001','MERBNK;500001;100002','MERBNK;500001;100003','MERBNK;500001;100004','MERBNK;500001;100005','MERBNK;500002;100001','MERBNK;5
@attribute merchant_country {'MERCRY;500001;301','MERCRY;500001;302','MERCRY;500001;303','MERCRY;500001;304','MERCRY;500001;305','MERCRY;500001;306','MERCRY;500001;307','MER
@attribute city_country {'CTYCRY;6001;301','CTYCRY;6001;302','CTYCRY;6001;303','CTYCRY;6001;304','CTYCRY;6001;305','CTYCRY;6001;306','CTYCRY;6001;307','CTYCRY;6001;308','CTY
@attribute class {'H','M','L'}
@data
'230000001556231730830 12','1400007623528765','BNKNWK;100005;11','BNKBIN;100005;200009','BNKCRY;100005;306','BNKCCY;100005;403','MERBNK;500004;100005','MERCRY;500004;306','CT
'230000001556231383746 41','1400007626995405','BNKNWK;100005;10','BNKBIN;100005;200004','BNKCRY;100005;308','BNKCCY;100005;409','MERBNK;500010;100005','MERCRY;500010;308','CT
'230000001556231613823 86','1400007606073039','BNKNWK;100003;11','BNKBIN;100003;200005','BNKCRY;100003;306','BNKCCY;100003;402','MERBNK;500005;100003','MERCRY;500005;306','CT
'230000001556231957942 87','1400007637878184','BNKNWK;100003;10','BNKBIN;100003;200004','BNKCRY;100003;309','BNKCCY;100003;404','MERBNK;500015;100003','MERCRY;500015;309','CT
'230000001556231628039 68','1400007639636759','BNKNWK;100002;11','BNKBIN;100002;200005','BNKCRY;100002;305','BNKCCY;100002;402','MERBNK;500015;100002','MERCRY;500015;305','CT
'230000001556231825036 13','1400007679584734','BNKNWK;100001;10','BNKBIN;100001;200010','BNKCRY;100001;310','BNKCCY;100001;405','MERBNK;500002;100001','MERCRY;500002;310','CT
'230000001556231481612 45','1400007601683083','BNKNWK;100004;10','BNKBIN;100004;200007','BNKCRY;100004;305','BNKCCY;100004;404','MERBNK;500013;100004','MERCRY;500013;305','CT
'230000001556231577002 27','1400007637217291','BNKNWK;100002;10','BNKBIN;100002;200008','BNKCRY;100002;305','BNKCCY;100002;407','MERBNK;500005;100002','MERCRY;500005;305','CT
'230000001556231422636 14','1400007660310361','BNKNWK;100003;10','BNKBIN;100003;200007','BNKCRY;100003;306','BNKCCY;100003;403','MERBNK;500004;100003','MERCRY;500004;306','CT
```

Figure 4: AREF file

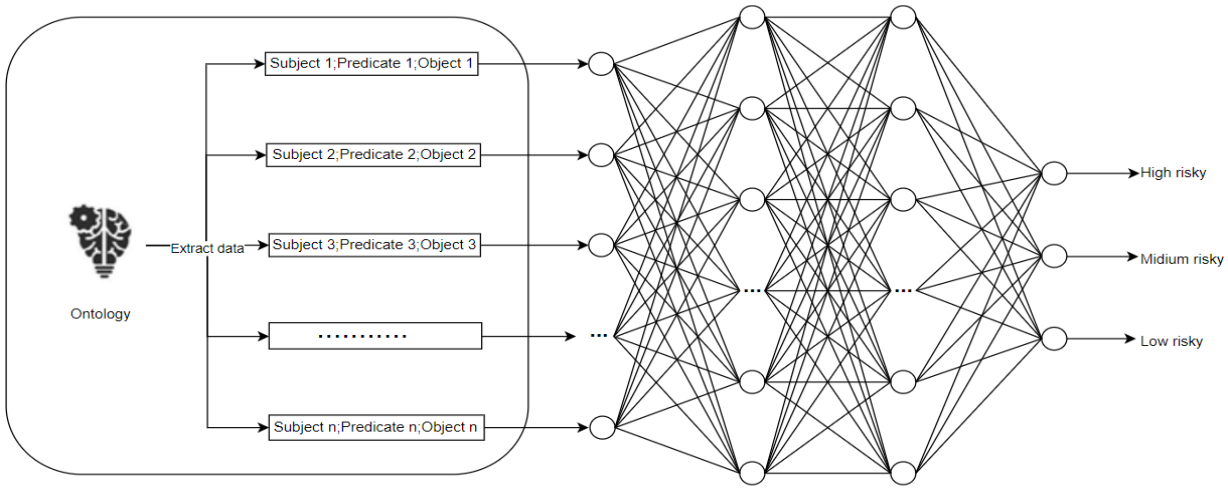

Figure 5: Combining ontology and neural network process

Figure 5 shows the extraction of data from ontology, creation an ARFF file, then combining this data with a neural network for classification according to the class levels defined by users.

4.5. Comparison and statistics:

The statistics below show the results of a comparison study to live the performance of the neural networks and ontology to detect cases of fraud. within the initiative, we applied the neural network directly on data downloaded from the payment system, within the second step, we created an ontology-based on these data, then we applied the same neural network on this ontology. Table 1 shows the statistics and percentages obtained. as a conclusion, it's seen that the mixture of neural networks and ontology allows having an improvement of precision on the detection of fraud.

Table 1: Statistics for Detection fraud

Class	Neural network on normal data		Neural network on Ontology data	
H(High)	27	0,27%	35	0,35%
M(Medium)	88	0,88%	100	1%
L(Low)	9885	98,85	9865	98,65%

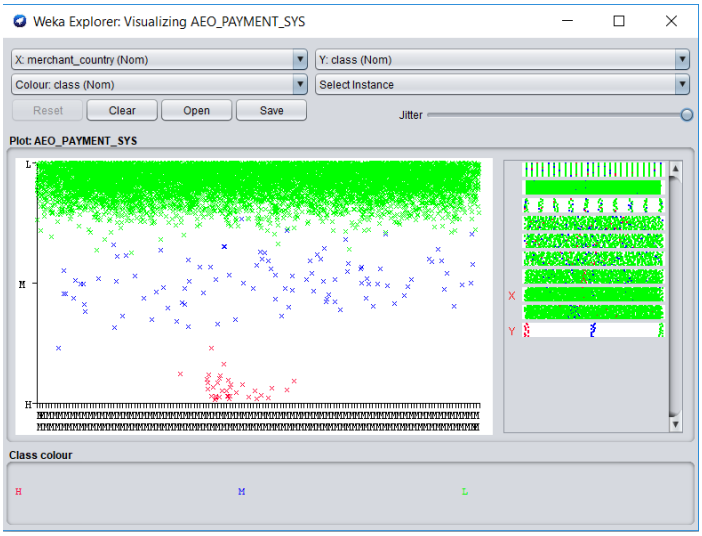

Figure 6: Results visualizing

The results are shown in figures 6 and 7 visualize a classification with a multilayer perceptron where we distingue if the transaction is high, medium or low risk done by neural network on ontology data.

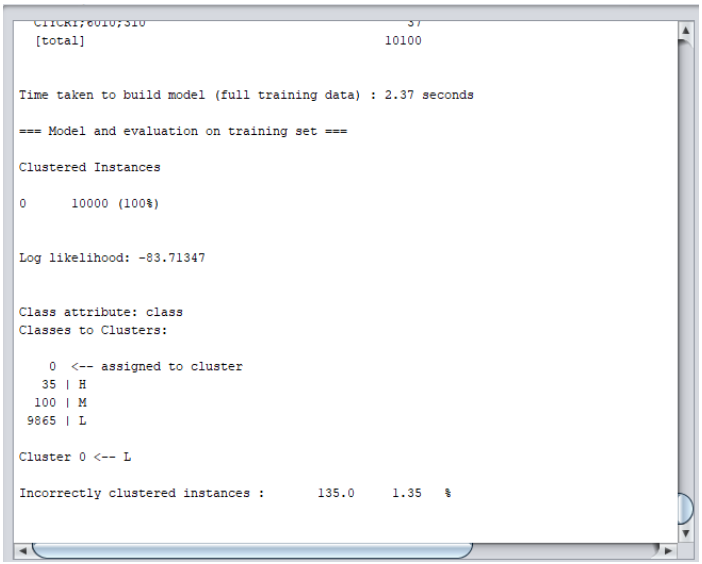

Figure 7: Results visualizing

5. Conclusions

The growth of payment card fraud and the evolution of artificial intelligence is the purpose of this paper. Our overall goal is to identify anomalies in a payment system to detect the largest number of fraudulent activities with a reasonable number of false positives.

There are several approaches and techniques used to detect the frauds in the payment system, which transactions are normal or fraudulent. The main advantage of the proposed approach is to combine semantics concepts with machine learning to achieve the goal of the paper.

In this paper, we have proposed the use of an ontology with machine learning for the detection of payment system frauds. The performance of this approach is evaluated on an ontology with the neural network method. The experimental phase proves that the

use of structured data in an ontology is very effective in detecting anomalies. In future work, we will focus on including SVM, Decision Tree, and KNN in terms of precision, AUC on building and implement a new architecture capable of detecting fraudulent transactions, then make a comparison and deduce the best method to combine with the ontology.

References

[1] Jamal Bakkas, Mohamed Bahaj, Abderrahim Marzouk ; "Direct Migration Method of RDB to Ontology while Keeping Semantics"; International Journal of Computer Applications (0975 – 8887) Volume 65– No.3, March 2013.

[2] Ahmed EL ORCHE, Mohamed BAHAJ; "Using framework to synchronize ontology with relational database"; Journal of Theoretical and Applied Information Technology 31st May 2016. Vol.87. No.3; ISSN: 1992-8645; E-ISSN: 1817-3195

[3] Ahmed EL ORCHE and Mohamed BAHAJ; "Approach to use ontology based on electronic payment system and machine learning to prevent Fraud"; The 2nd International Conference on Networking, Information Systems & Security, March 27-29, 2019, Rabat, Morocco.

[4] Ahmed EL ORCHE, Mohamed BAHAJ; "Ontology-based on electronic payment fraud prevention"; Proceeding of 5th International Congress on Information Science and Technology; ISBN: 978-1-5386-4385-3; IEEE Catalog Number: CFP1867R-ART; October 21-27, 2018; P.143

[5] Ahmed EL ORCHE, Mohamed BAHAJ; "A Method for Updating RDB of Ontology while keeping the Synchronization between the OWL and RDB"; ARPN Journal of Systems a and Software; ISSN 2222-9833; VOL. 4, NO. 4, July 2014.

[6] Kingston, J., Schafer, B., & Vandenburghe, W.; "Ontology Modelling in the Legal Domain - Realism Without Revisionism"; Proceedings of the KI2003 Workshop on Reference Ontologies and Application Ontologies, Hamburg, Germany, September 16, 2003.

[7] Ali Ahmadian Ramaki, Reza Asgari, and Reza Ebrahimi Atani; "Credit card fraud detection based on ontology graph", International Journal of Security, Privacy and Trust Management (IJSPTM), Vol. 1, No 5, October 2012.

[8] Tova Milo, Slava Novgorodov, Wang-Chiew Tan; "Interactive Rule Refinement for Fraud Detection"; Proceedings of the VLDB Endowment, v.9 n.13, p.1465-1468, September 2016.

[9] Rodrigo Carvalho, Michael Goldsmith, Sadie Creese; "Applying Semantic Technologies to Fight Online Banking Fraud"; European Intelligence and Security Informatics Conference; 2015.

[10] L.Y. Ding, B.T. Zhong, S.Wu, H.B. Luo; "Construction risk knowledge management in BIM using ontology and semantic web technology"; Safety Science 87 (2016) 202–213.

[11] Quratulain Rajput, Nida Sadaf Khan, Asma Larik, Sajjad Haider; "Ontology-Based Expert-System for Suspicious Transactions Detection"; Computer and Information Science; Vol. 7, No. 1; 2014 ISSN 1913-8989 E-ISSN 1913-8997.

[12] Gunnar Declerck, Audrey Baneyx, Xavier Aimé, Jean Charlet; "A quoi servent les ontologies fondationnelles ?"; 23èmes Journées francophones d'Ingénierie des Connaissances (IC 2012), Jun 2012, Paris, France. pp. 67-82, 2012.