Claim Status Prediction for OSIPTEL Using Neural Networks

Hugo D. Calderon-Vilca
Universidad Nacional Mayor de San Marcos
Lima, Perú
hcalderonv@unmsm.edu.pe

Kerly J. Quispe Quispe
Universidad Nacional Mayor de San Marcos
Lima, Perú
kerly.quispe@unmsm.edu.pe

Anthony B. Puitiza Lopez
Universidad Nacional Mayor de San Marcos287
Lima, Perú
anthony.puitiza@unmsm.edu.pe

Miguel A. Zuñiga Yamashita
Universidad Nacional Mayor de San Marcos
Lima, Perú
miguel.zuniga1@unmsm.edu.pe

Flor C. Cardenas-Mariño
Universidad Nacional Mayor de San Marcos
Lima, Perú
fcardenasm@unmsm.edu.pe

Reynaldo Sucari Leon
Universidad Nacional de Huanta
Ayacucho, Perú
rsucari@unah.edu.pe

***Abstract*—In recent years, the demand for complaints from users towards mobile operators has increased notably according to OSIPTEL indicators. Only for the year 2017, the operating companies registered 2 million 728 thousand 430 claims, of which 233 thousand 342 claims passed to second instance (Tribunal Administrativo de Solución de Reclamo - TRASU of OSIPTEL), putting OSIPTEL in big trouble. This research seeks to solve this problem by implementing a system to predict the meaning of the claim, making use of neural networks. It was decided to implement a Multilayer Perceptron, as a learning algorithm, Backpropagation of the error was chosen. As for the architecture of the Perceptron, several were tested, where the changing factor was the neurons in the hidden layer. The results show that the system has 86.969% accuracy.**

I. INTRODUCTION

In recent years, the demand for complaints from users towards their mobile operators has increased notably according to OSIPTEL indicators. Only in 2017, the operating companies registered 2’278,430 claims, of which 233,342 passed to second instance, that is, they reached the Administrative Tribunal for Settlement of Claims (known in Peru as TRASU) of OSIPTEL. Statistics show that 83.7% are mobile phone claims, 65% of these claims are handled by telephone and 63% of them were declared inadmissible or unfounded [10].

The excessive demand for complaints has put the OSIPTEL Claims System in serious trouble, so on February 28th of this year, a statement was made public on the official website of the entity, announcing the extraordinary measures that will be put in practice for a period of 18 months in order to avoid the collapse of the Claims System [8]. Among the measures taken one of the most critical is that the telephone mechanism for filing appeals and complaints has been suspended [10].

According to the analysis of OSIPTEL, the cause of the excessive increase in demand is due to the fact that users are making improper use of the claim filing channels. Since it is public knowledge that when a claim is filed against the operating company, it is obliged to suspend the collection of the amount of the receipt that is the subject of analysis, users would be making claims to avoid payment of their respective receipts [10].

However, not all are unfounded claims in order not to pay the receipts. There are many cases in which consumer rights are being violated. Institutions such as INDECOPI and the Consumer Protection Society have identified and taken corrective actions against mobile operators, such as Telefónica S.A.

In December 2017, INDECOPI ordered the suspension of the advertising of the unlimited plans, after OSIPTEL ordered the suspension of the sale of said plans. The reason behind these measures is that the main operators in the country did not provide accurate, correct, truthful and easily accessible information to users about the mobile service they offered [12].

Artificial intelligence algorithms allow solving different problems as well as can be analyzed in: heart disease detection [13], tuberculosis detection [17], information search and retrieval [14], augmented reality [15], video game recommendation [18] and other purposes.

The objective of this research is to present a predictor model that optimizes claims management within the regulatory entity OSIPTEL through the use of a multi-layer perceptron neural network with Backpropagation learning applied to a set of data obtained between 2016 and 2018. The solution presented will facilitate the rapid response by the regulatory entity OSIPTEL to those users who have a pending claim or file a new claim.

Artificial intelligence algorithms allow solving different problems as well as can be analyzed in: heart disease detection [13], tuberculosis detection [17], information search and retrieval [14], augmented reality [15], videogame recommendation [18], application of algorithms for routing questions and answers [16], among other contributions, they use algorithms such as neural networks to solve different problems.

This article is organized as follows: Section II represents the State of the Art making a study of the literature that is related to the management of claims and customer service. In Section III he describes the proposed solution, as well as the specification of the data that will be used in it. Section IV shows the construction of the system, system requirements, system actors, activity diagram or system flow, deployment diagram, data modeling and software architecture. In Section V, the results of the experiments carried out in the present investigation are presented. In Section VI, the comparison of the results obtained with studies that are part of our state of the art is presented; and finally, the conclusions.

II. State of the Art

Currently, hospitals have modern records that are called claims. These can be records of medical treatments by a provider (clinic / doctor), prescriptions, among others.

The study of [7] presents the problems that hospitals have with respect to the rejection and denial of health care claims, since they represent an important administrative burden and sources of losses of various health services for providers and consumers. The research seeks to automate the identification of claims likely to be denied by different factors.

The authors propose a method of classification based on Machine Learning, to completely automate the identification of such rejection-prone notifications, with high precision, and investigate the reasons for the denial of claims. In this way, they seek to develop a claim adjustment reason code (CARC) with high gain of information for Machine Learning.

These classifications are trained machine learning algorithms such as binary classification trees, neural networks, support vector machines (SVM) and Naive Bayes classification, to predict whether a given claim will likely involve rework (ie, have a high risk of rejection). or denial) through the binary classification.

In the construction of Machine Learning algorithms, we used a data set of 10,000 accepted claims and 3,000 denied claims from one of the best insurance companies in the United States.

The authors make a comparison between all the algorithms to identify which of them best meets the objective. Getting these results:

When the data patterns include multiple complex influences, better results are obtained with 77% using an SVM algorithm for the classification of the claims compared to 70% using the decision tree algorithm.

Neural networks did not produce good performance, they obtained poor precision measurements with 45% accuracy when classifying claims.

Training in SVM always finds a global minimum, while the performance of neural networks can sometimes suffer from the existence of multiple local minimum solutions.

The conclusion of the authors is to present a new classification framework for the identification of those medical claims that will probably be rejected during the submission process, in addition a significant improvement of the practice status of using Machine Learning is achieved to automate and improve the risks of denial of claims.

Another study conducted in [2], the authors identify the problem that has been presented in the health services. Health insurance companies have a medical claim system that serves to minimize the loss of resources, the medical claim system blocks the procedures that were erroneously ordered, being necessary to work 24 hours a day and this generates a high cost. The research proposes to improve the quality of the data through the Knowledge Discovery Database process, and then compare the classification techniques based on decision trees to learn about the behavior of the medical claims system, experts in the field will evaluate if the medical requests must or not be authorized.

The majority of databases have problems such as redundancy of data, duplicate values, among others, hence a process is necessary to improve the quality of the data that is why the authors propose the process of knowledge discovery in databases. data (KDD). Is worth noticing that this process is done before learning.

According to the study, a comparison between the Precision, Recall, Acura, F-Measure, AUC and Kappa results is shown for the algorithms tested using the attributes obtained after manual selection (DB-MS), automatic selection (DB-AS) and the attributes represented by (DB-Final).

Two classifiers based on decision trees were compared: C4.5, CART, RandomForest (RF) and RandomTree (RT). The results show that the RF and RT have a similar performance and higher than the other classifiers with 93% and 92% respectively for Precision, 98%, and 98% respectively for Recall, 95% and 95% for Acura, 95 % and 95% respectively for F-Measure, 99% and 96 respectively for AUC and 90% and 89% for Kappa.

As the conclusion of this study, the classifiers that have a better performance to learn from the behavior of the medical claims system, experts, and these evaluate them if the medical requests must be authorized or not are the RF and RT algorithms.

Both investigations [2] and [7] use classification trees (also called decision trees) to reach the conclusion of their work. While the first investigation concludes, although the vectorial support machines are better developed with a multiple pattern of complex data; decision trees are a better option in a simpler context, although it does not delve further into the subject. Later, in the investigation [7] a more in-depth study of the subject is given. A comparison of different classifiers based on decision trees is made, finding that RandomForest and RandomTree stand out above their competition, besides presenting many similarities.

The investigation [3] reveals the importance of customer satisfaction in the companies that provide communication services in Pakistan. The objective of the article was to conduct a survey of 200 clients based on their questions in their literature review on customer satisfaction. These were about: network coverage, call quality, dropped call rate, SMS sending time, internet service and duration of voice call setup.

When performing the analysis of the results, it was possible to verify that 3 of the 6 attributes surveyed are those that generate the most dissatisfaction among the users: Network coverage, Internet service and quality of calls.

In addition to the study, in spite of not having a neural network model for the analysis of the data with greater precision, these results will help the communication service providers to improve the satisfaction of their clients in general.

Another case study is conducted by the investigation [1] for the Westfield insurer, which intends to carry out a replacement of its inherited claim system. This resulted in an end to final simulation of the claims process, including adjudication, liquidation, litigation, subrogation, salvage and fraud claims.

In conclusion, unlike spreadsheet management, solutions that are often based on averages or gram static workflow diagrams; the simulation captures end-to-end interdependencies and variability between processes and sources.

In addition to the study, the participation of the executives is recommended, which was crucial and was obtained with the analysis based on facts provided by the simulation, as well as with the realistic description of the process in the animation.

In addition to the study, the participation of the executives is recommended, which was crucial and was obtained with the analysis based on facts provided by the simulation, as well as with the realistic description of the process in the animation.

The research article [4] proposes a knowledge model for customer relationship management (CRM), first identifying the problem between the user and the company, and the solution offered by the CRM strategy, this being the analysis of the context . The authors propose the collection of information through questionnaires about their level of satisfaction, loyalty to the company, administration and other complementary aspects.

These questionnaires allow to classify the clients in different categories according to their importance. Therefore, three important processes emerge: detect the problem, the root of the dissatisfaction, and the classification of clients. The proposed model is implemented in the Visirule software, classifying the users according to the questionnaires and, in turn, dividing the claims into different criteria.

The purpose of the research [5] is to document the approach, lessons learned and the impact on the business of a simulation in the Westfield Insurance call center.

The results show that hybrid resources minimize call abandonment. Removing these resources increased the abandonment rate. A staff model was given where only hybrid resources were present and resulted in the highest level of service improvement and the greatest reduction in abandonment. The authors reach the conclusion that, within a call center with hybrid resources, it has much more use than resources concentrated in a single task.

Both the research [1] and [5] take place at Westfield Insurance, studying, respectively, a simulation model and the impact of hybrid resources within the call center. Thus, it can be seen that the Westfield insurer takes great importance to the management of claims through its CRM tool detailing its influence in both investigations.

III. Neural Network Architecture Design

A. *Obtaining the data*

The first step was obtaining the data source to train the neural network.

The data that will be used in the present investigation originates from the OSIPTEL database. The data were downloaded legally, because OSIPTEL is a public entity and is governed by the Law on Transparency and Access to Public Information (Law No. 27806). The information obtained dates from the years 2016 - 2018 being its last update in the middle of this year. The number of records obtained is 388 734.

The data was processed and organized in an Excel file.

TABLE I. Input Variables to RNA

Variable	Description	Values	
Type of resource	Type of resource that is presented to Trasu.	0= Complaint 1 = Appeal	
Date entry 2nd. Instance	Date on which the resource is presented.	For year 1 = 2016 2 = 2017 3 = 2018 Fort he months 1= January 2= February 3= March 4= April 5= May	6= June 7= July 8= August 9= Septembre 10= October 11= November 12= December
Date entry 2nd. Instance Date Resolution of Trasu	Date on which the Trasu Resolution is issued	For year 1 = 2016 2 = 2017 3 = 2018 Fort he months 1= January 2= February 3= March 4= April 5= May	6= June 7= July 8= August 9= Septembre 10= October 11= November 12= December
Company	Name of the company from which the claim type comes	1 = Telefónica del Perú S.A.A. 2 = América Movil del Perú S.A.C 3 = Entel Perú S.A 4 = DirecTV Perú S.R.L	
District	Name of the district of Lima from which the claim comes.	They take values from 1 to 25. Districts of Metropolitan Lima	
Matter	Complaint matter	1= Activation 2= Assignment 3=Low 4=Unjustified drop in service 5=Quality 6= Change of ownership 7= Unsolicited ownership change 8= Payment 9= Unsolicited contracting 10=Cut 11= Any related 12= Unlock equipment 13= Contract ignorance 14= Contract discount	
Sub-matter	Sub-matter of the claim	1= Abuse of Authority and prevaricator 2= Additional service failure 3 = Service failure 4= Low not attended 5 = Not applicable	
Service	The type of service for which the claim is filed	1= Lease of circuits 2= Fixed access Internet 3= Prepaid mobile access Internet 4= Postpaid mobile access Internet	

B. *Architecture proposal*

The present research proposes the implementation of a neural network in order to solve the problem of claims for OSIPTEL. We opted for the Multilayer Perceptron because it has shown good performance in applications of this type.

The neural network that we propose has 3 layers, an input layer, a hidden layer and an output layer. In Fig. 1 we present the architecture of the neural network.

In the input layer, there are 12 neurons, in the hidden layer, 6 neurons and in the output layer, 4 neurons. The number of neurons in the hidden layer was calculated according to the following formula:

$$H = \sqrt[2]{mn}$$

Where H is the number of neurons in the hidden layer; m, number of neurons in the input layer; n is the number of neurons in the output layer.

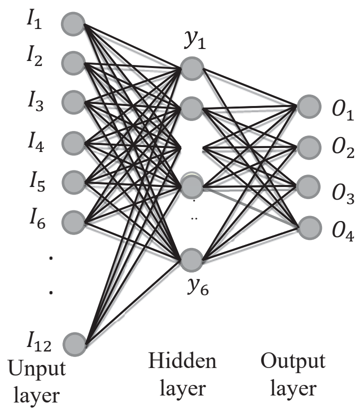

Fig. 1. Architecture of red neuronal proposal

We We will use the error backpropagation algorithm as a learning rule. This algorithm consists of taking the data entered by the user and converting them into the set of inputs, activating the neural network with random initial weights in the input layer.

In the hidden layer, the propagation rule that is used to integrate the information from the inputs and provide the value of the postsynaptic potential of the neurons is the weighted sum of the inputs with the synaptic weights, and on that weighted sum is applies a sigmoid-type transfer function, which is bounded in response.

Logarithmic sigmoidal transfer function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Error backpropagation: of GAN characteristics:

$$E_{(w_{ij},\theta_j,w'_{kj},\theta'_k)} = \frac{1}{2}\sum_p\sum_k\left[d_k^p - f(\sum_j w'_{kj}y_j^p - \theta'_k)\right]^2$$

Where dk are the desired outputs; f, the activation function; w_k the synaptic weights.

Depending on the backpropagation expression of the error, the following expressions are generated for the weight update.

Error signals:

$$\delta_{w'_{kj}} = -\epsilon\frac{\partial E}{\partial w'_{kj}}$$

$$\delta_{w'_{ji}} = -\epsilon\frac{\partial E}{\partial w'_{ji}}$$

$$\delta_{w'_{kj}} = \epsilon\sum_p \Delta_k'^p y_j^p \quad con \quad \Delta_k'^p = \left[d_k^p - f\left(v_k'^p\right)\right]\frac{\partial f\left(v_k'^p\right)}{\partial v_k^p}$$

$$\delta_{w_{ij}} = \epsilon\sum_p \Delta_j^p x_i^p \quad con \quad \Delta_j^p = \left(\sum_k \Delta_k'^p w'_{kj}\right)\frac{\partial f\left(v_k'^p\right)}{\partial v_k^p}$$

Where w_k are the synaptic weights that are between the hidden layer and the outputs; ϵ, the learning rate; ; w_j , the weights that are between the entrances and the hidden layer; E, the output of the feedforward phase (forward propagation); w_j, the outputs of the hidden layer; and, finally, x_i, the entry patterns.

Predicting the meaning of the claim

When the user answers the questionnaire, the system interprets each response as input values for the neural network. In order to test if the training of the network is correct, a web system will be implemented. Fig. 2 shows the general architecture of the system.

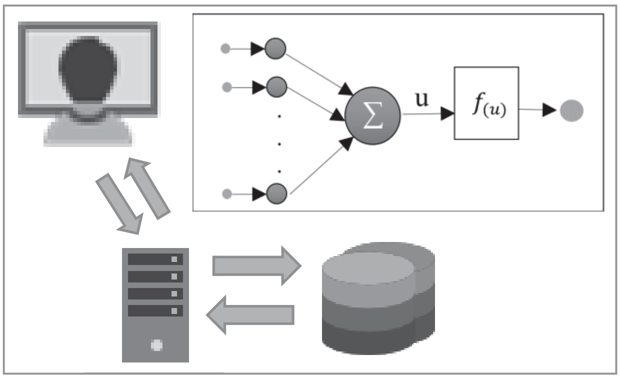

Fig. 2. General system architecture

C. General system architecture

The system takes the results generated by the neural network to be able to interpret them; with the data obtained it goes to the database where a table contains the interpretations. Table II shows the interpretations of the possible results of the neural networks, this takes the interpretation given according to the outputs that have been previously established in the neural network.

TABLE II. RNA OUTPUT

Variable	Description	Values
Sense	It refers to the meaning of the exit of the type of claim..	1 = Canceled 2=Archived 3=Terminate 4 = Founded

IV. Application Design

To manage the complaints that OSIPTEL receives, a web application will be developed that will allow the user to predict if his claim will proceed or not, in this way, the user will no longer generate a claim that he knows will not proceed.

A. System functional requirements

The first requirement is to record the responses of users (entries), the next requirement is to predict the meaning of the claim with the entries obtained from the user. The system is constructed and designed based on the functional requirements.

B. System actors

Any actor that interacts with it is defined as an actor of the system, Table III shows the actors of the system.

TABLE III. SYSTEM ACTORS

Actor	Description
User	The person makes the query about the meaning that your claim can take.
Administrator	It is the person in charge of supporting the platform.

C. System flow

The flow of the system starts when the client enters the web page, immediately the server loads the information of the page so that it is shown, this data is obtained from the database to which it is connected, this process is done in question of seconds. When the web page shows the information, the user enters the required data. The web page is responsible for sending them so that they are transformed into the set of input patterns that will be used in the neural network, the necessary operations and transformations are carried out to that, the network correctly predicts the meaning of the claim introduced. By the user, these results are sent and displayed on the web page.

The flow of the system is represented in Fig. 3, the UML notation was used in the elaboration of the activity diagram.

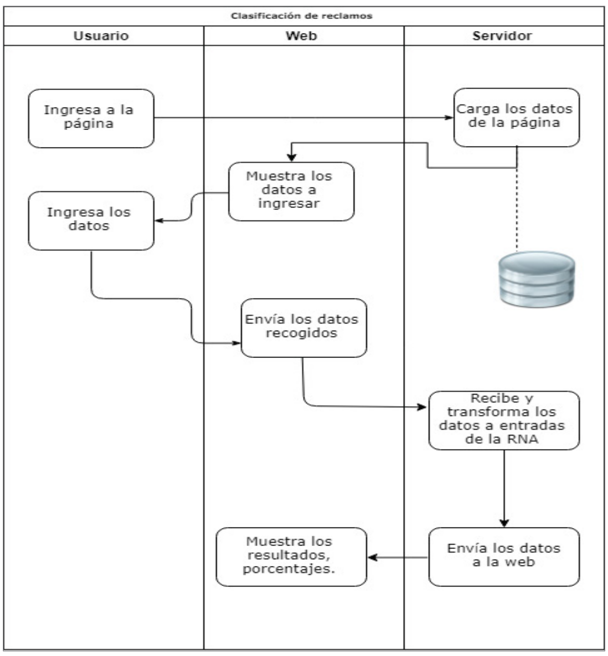

Fig 3. System flow

D. Application architecture

The web application with a 3-layer architecture: presentation, business, and persistence. In the first layer, the user interface is presented using technologies such as HTML, CSS and JavaScript for design and marking, and using Spring MVC with Thymeleaf annotations for data management. In the business layer, you have all the logic necessary for the proper functioning of the application. In addition, finally, in the persistence layer, Spring Data will be used integrated to the Hibernate framework.

As shown in Fig. 4, Maven is used as a management tool and Spring Boot for the quick configuration of the project.

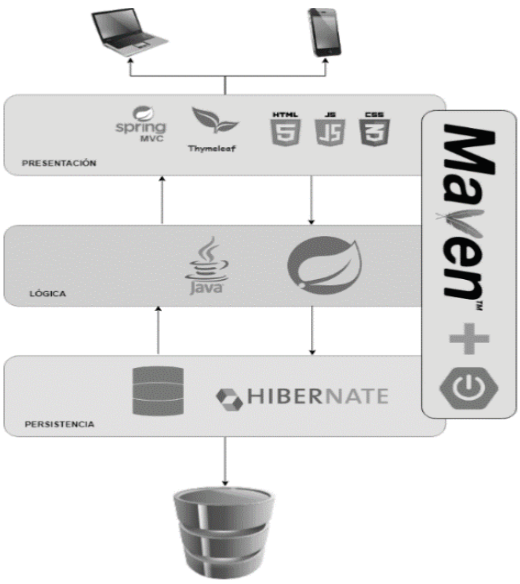

Fig 4. Application architecture

E. Deployment diagram

The system's deployment diagram provides a topological view of it, in addition, it shows the relationship of the software and hardware elements, that is, what software elements are implemented by what hardware elements. We use the UML notation to build the diagram shown in Fig. 5.

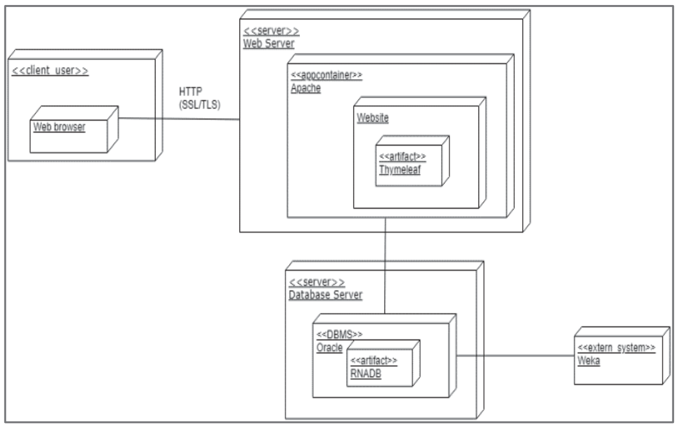

Fig. 5. Deployment diagram

F. Data model

Feature Fig. 6 represents the data model developed with the Oracle tool 12 c. The system consists of three tables in which the necessary data will be stored for its correct operation.

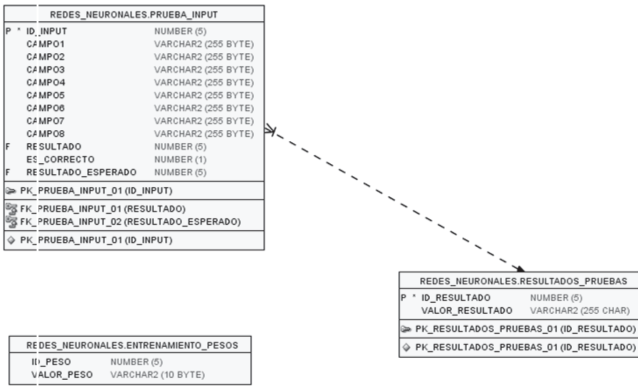

Fig 6. Data Model

G. System Interface

In Fig. 7 to 9, they are screenshots of the Claims management system. From the beginning login so you can enter the information by continuing, a training of the Neural network in case there is more data to analyze. In Fig. 8, a test of the Neural network is displayed to see how this prediction the data.

Fig 7. Login to the web Claims Management System

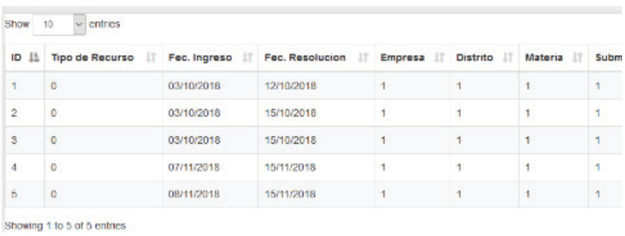

Fig 8. Main menu

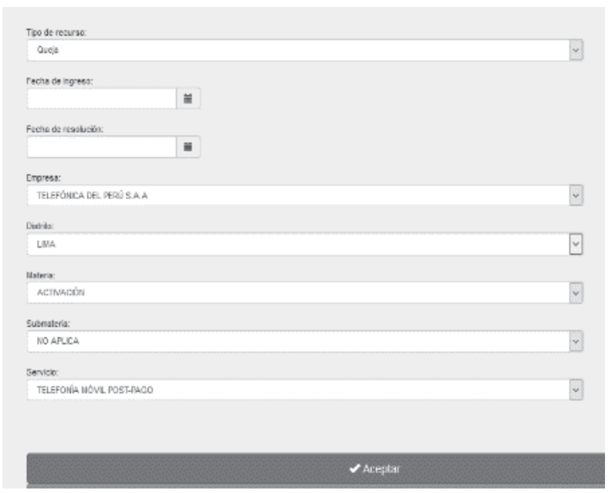

Fig 9. Test of the Neural network

V. Results and discussion

In this section, the results of the experiments are shown comparing three architectures for the neural network.

For the exploration of data and experimentation, the WEKA 3.8.3 tool was used, which is a software platform for automatic learning and data mining.

Multilayer Perceptron - 10 neurons in the hidden layer

For the first model, we considered 12 neurons in the input layer, 9 neurons in the output class, and 12 neurons in the hidden layer. The learning rate is 0.2 and the momentum is 0.1, reaching an average absolute error of 0.117 with 60% of instances correctly classified and 40% of instances classified incorrectly from 198112 instances.

TABLE IV. Training results

N°	Pattern	Prediction margin	Pred.	Class
1	1,24,4,1,4,10,1,14,30,19,45,9	96.7891%	4	4
2	1,11,7,2,1,2,3,14,40,15,14,10	81.6854%	4	4
3	1,14,5,2,14,11,2,1,14,9,45,10	94.1528%	4	4
4	1,7,12,1,28,3,2,1,25,15,9,10	45.8445%	3	4
5	0,27,1,2,13,7,2,1,4,42,105,10	85.4149%	4	4
6	0,9,4,2,11,1,3,1,5,24,41,10	43.0328%	3	4
7	1,21,12,1,15,12,2,1,14,47,45,10	46.6565%	3	4
8	1,31,10,1,23,2,2,1,32,15,19,10	53.2636%	3	4
9	1,18,1,1,28,1,1,16,8,15,9,10	49.9279%	3	3
10	1,11,8,1,10,11,1,1,42,15,10,10	57.2712%	3	3
…	…	…	…	…
…	…	…	…	…
75175	0,17,3,2,26,2,3,1,15,35,65,10	11.4377%	4	1

After this first model, although the number of correctly classified instances is high, more than 50% is still a very low number to be able to predict with greater precision as observed in the table of training results in Table IV. For this reason decided to shorten the registration number by eliminating classes that do not have a considerable number in order to train the network with greater precision. The departures were analyzed again, opting to remove those with the smallest amounts of records and those that had the greatest impact for solving the problem. From 198112 record was reduced to 75175 record and having only five types of outputs.

Multilayer Perceptron - 8 neurons in the hidden layer

For the second model, we considered 12 neurons in the input layer, 5 neurons in the output layer, 8 neurons in the hidden layer. The learning rate is 0.2 and the momentum is 0.1, reaching an average absolute error of 0.172 with 68% of instances correctly classified and 32% of instances classified incorrectly out of 75175 instances.

TABLE V. TRAINING RESULTS

N°	Pattern	Prediction margin	Predicted class	Class
1	0,1,4,1,14,4,1,1,40,24,41,10	97.9189%	4	4
2	0,17,4,1,18,8,1,14,18,24,41,10	86.5116%	4	4
3	1,24,3,2,24,12,2,1,37,15,9,10	93.5196%	4	4
4	1,7,12,1,28,3,2,1,25,15,9,10	45.8445%	3	4
5	0,4,10,1,10,4,2,1,42,32,45,10	96.7449%	4	4
6	0,14,8,1,11,10,1,1,35,24,41,10	97.0328%	4	4
7	1,21,12,1,15,12,2,1,14,47,45,10	76.6565%	4	4
8	1,17,10,1,24,1,2,1,14,15,14,10	42.2636%	3	4
9	1,18,1,1,28,1,1,16,8,15,9,10	49.9279%	3	3
10	1,3,11,2,16,1,3,1,17,15,9,10	64.2712%	3	3
…	…	…	…	…
…	…	…	…	…
75175	0,17,3,2,26,2,3,1,15,35,65,10	11.4377%	4	1

Multilayer Perceptron - 8 neurons in the hidden layer

TABLE VI. TRAINING RESULTS

N°	Pattern	Prediction margin	Predicted class	Class
1	1,2,10,2,23,1,3,1,40,15,9,10	96.9891%	4	4
2	1,29,5,2,30,1,3,1,37,15,9,10	79.9854%	4	4
3	1,4,10,1,10,1,2,1,12,15,9,10	93.5128%	4	4
4	1,8,8,1,25,10,1,12,24,15,9,10	48.8445%	3	4
5	1,25,1,1,11,2,1,12,37,15,14,10	85.4149%	4	4
6	1,24,10,1,10,2,2,2,24,15,19,10	42.0328%	3	4
7	1,16,11,1,7,3,2,1,11,15,9,10	79.6565%	4	4
8	1,23,9,1,10,1,2,1,15,15,9,10	53.2636%	3	4
9	1,8,11,1,14,2,2,16,43,15,9,10	49.9279%	3	3
10	1,30,10,2,6,2,3,1,32,15,9,10	57.2712%	3	3
…	…	…	…	…
…	…	…	…	…
75175	0,17,3,2,26,2,3,1,15,35,65,10	11.4377%	4	1

The network clearly improved 8% accuracy in contrast to the previous model and this was thanks to the reduction of outputs in the neural network. The network can focus on the most important types of output but for our purpose the network still continues being not enough, that is why we decided to add a better model as shown in the table of results of the training in Table V.

Multilayer Perceptron - 6 neurons in the hidden layer

For the third model, we considered 12 neurons in the input layer, 4 neurons in the output layer, and 6 neurons in the hidden layer. The learning rate is 0.2 and the momentum is 0.1, reaching an average absolute error of 0.0889 with 86.989% of instances correctly classified and 13.011% of instances classified incorrectly out of 75175 instances.

After this model, we continued testing more models for the network, but instead of improving the network, it was the opposite, the error grew more and the prediction was not accurate.

For this reason, he stayed with this time of architecture, for being better and more precise than the previous architecture models of neural networks mentioned above.

Results of the third training - Multilayer Perceptron with 6 neurons in the hidden layer

After choosing the architecture that gave the best results, more workouts were performed. Table VI shows the results of the third training. As can be seen, the network predicts with a precision greater than 70%, but still has faults in predicting classes 1 and 2.

TABLE VII. TRAINING RESULTS

In the research [7], the authors propose a solution based on Machine Learning algorithms. They make use of algorithms of binary classification trees, support vectors (SVM), neural networks and classification of Naive Bayes, for the construction of the algorithms they used 13 thousand records between accepted and rejected files. The results show that the performance of the SVM algorithm is better than the others, since it reaches 77% accuracy. We propose the use of the Multilayer Perceptron to predict the meaning of the claim, we use 75 175 records and we achieved 86.989% accuracy.

In [4], the authors aim to improve the relationship with the client through a CRM knowledge model. They use the Vesirule software to categorize and classify clients according to their importance, they use questionnaires to gather the information that will be the set of tickets for Vesirule; In our proposal, the neural network uses the data recorded through the website as a set of entries, in addition to being trained with 75 175 records previously.

In [2], the study states that the problem arises in terms of the quality of the data stored in the databases, the problem lies in the redundancy of these, that is why the authors propose to use the Discovery process of knowledge in databases, while, for our proposal, we have refined incomplete records as it hindered the learning of the network.

Our system uses the questionnaire as a tool to collect information from users through the website; in [3], they also make use of this tool to collect information on customer satisfaction. Despite not using a neural network to analyze the data with greater precision, they were able to identify the attributes that generate dissatisfaction and therefore are the subject of future claims.

Compared to [1], we predict the possible exit that the user's claim will have so that this will avoid generating a claim that he knows will not proceed, in this way the time can be optimized, while the study [1] uses a simulation model to capture interdependencies and the variability of processes and sources.

CONCLUSIONS AND FUTURE WORK

In the present investigation the use of the Multilayer Perceptron was proposed as a solution tool since this type of network is very useful in the prediction of results; However, if you do not have the right data, it will not yield the desired results.

In our case, we had to re-analyze the data and better categorize the inputs and outputs to improve the accuracy percentage. The result of the investigation is a system that predicts the meaning of the user's claim, which will avoid the excessive demand of claims in OSIPTEL.

For future research, we recommend re-training with more records and that the difference between classes be minimal.

REFERENCES

[1] Grabau, M., Conly, Q., & Marshal, M. (1981). Business Process Simulation for Claims Transformation. Journal of Chemical Information and Modeling, 53(9), 1689-1699.

[2] De Araujo, F. H. D., Santana, A. M., & De Alcantara Dos Santos Neto, P. (2015). Evaluation of classifiers based on decision tree for learning medical claim process. IEEE Latin America Transactions, 13(1), 299-306. https://doi.org/10.1109/TLA.2015.7040662

[3] Khan, N., Akram, M. U., Shah, A., & Khan, S. A. (2018). Important attributes of customer satisfaction in telecom industry: A survey based study. 4th IEEE International Conference on Engineering Technologies and Applied Sciences, ICETAS 2017, 2018-January, 1-7. https://doi.org/10.1109/ICETAS.2017.8277858

[4] Villanueva, H. E. A., Faed, A., & Truffa, M. A. M. (2017). Knowledge Model to Manage Customer Satisfaction Based on Claims. Proceedings - 14th IEEE International Conference on E-Business Engineering, ICEBE 2017 - Including 13th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 207, 195-200. https://doi.org/10.1109/ICEBE.2017.38

[5] Conley, Q., & Grabau, M. (2013). Simulating a modified hybrid approach to resource assignment in a shared billing and claims call center. Proceedings of the 2013 Winter Simulation Conference - Simulation: Making Decisions in a Complex World, WSC 2013, 1778-1783. https://doi.org/10.1109/WSC.2013.6721558

[6] Li, X., Tong, L., Hui, Z., & Hou, X. (2015). CSI-model-based research on factors affecting rate satisfaction of government and corporate customers-taking special-line products as an example. China Communications, 12(8), 183-189. https://doi.org/10.1109/CC.2015.7224700

[7] Saripalli, P., Tirumala, V., & Chimmad, A. (2017). Assessment of healthcare claims rejection risk using machine learning. 2017 IEEE 19th International Conference on E-Health Networking, Applications and Services, Healthcom 2017, 2017-December, 1-6. https://doi.org/10.1109/HealthCom.2017.8210758

[8] Diario el Peruano. (2018). Norma que aprueba Medidas Extraordinarias relacionadas al procedimiento de atención de reclamos de usuarios. Consultado 20 setiembre 2018. Available in: https://busquedas.elperuano.pe/normaslegales/norma-que-aprueba-medidas-extraordinarias-relacionadas-al-pr-resolucion-n-51-2018-cdOSIPTEL-1620019-1/

[9] OSIPTEL (2018). Reglamento para la Atención de Reclamos de Usuarios de Servicios Públicos de Telecomunicaciones. Consultado el 20 de setiembre 2018. Available in: https://www.movistar.com.pe/documents/10182/802004/3%29+RESOLUCION+DE+CONSEJO+DIRECTIVO+N%C2%BA++047-2015-CD+-OSIPTEL%28Reglamento+p....pdf/d9fb875a-36db-4cf4-8d23-ad54a701080f

[10] OSITEL (2018). OSIPTEL adopta medidas para evitar el colapso Del Sistema de reclamos de usuarios. Consultado el 19 de setiembre 2018. Available in: https://www.OSIPTEL.gob.pe/noticia/np-norma-extraordinaria-reclamos

[11] OSITEL (2018). Indicadores de Reclamos de Usuarios - Primera Instancia (Diversas Empresas). Consultado el 19 de setiembre 2018. Available in: http://www.OSIPTEL.gob.pe/documentos/11-indicadores-de-reclamos-de-usuarios-primera-instancia

[12] INDECOPI (2018). INDECOPI prohíbe que se publiciten "planes ilimitados" de operadores móviles. Consultado el 29 de stiembre 2018. Available in: https://www.indecopi.gob.pe/en/noticias,

[13] H. D. Calderon-Vilca, K. E. C. Callupe, R. J. I. Aliaga, J. B. Cuba and F. C. Mariño-Cárdenas, "Early Cardiac Disease Detection Using Neural Networks," 2019 7th International Engineering, Sciences and Technology Conference (IESTEC), Panama, 2019, pp. 562-567.

[14] M. A. Silva-Fuentes, H. D. Calderon-Vilca, E. F. Calderon-Vilca and F. C. Cárdenas-Mariño, "Semantic Search System using Word Embeddings for query expansion," 2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America), Gramado, Brazil, 2019, pp. 1-6

[15] H. D. Calderon-Vilca, E. G. Ocrospoma-Callupe, F. C. Cárdenas-Mariño and E. F. Calderon-Vilca, "Architecture and mobile application with augmented reality, visualizing videos and 3d objects in museums," 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Valparaiso, Chile, 2019, pp. 1-5.

[16] C. E. Serquen-Llallire, H. D. Calderon-Vilca and F. C. Cardenas-Mariño, "Comparison of two Algorithms for Routing Questions and Answers, Applied to Group of Students Software Engineering," 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Gudalajara, Mexico, 2018, pp. 1-6.

[17] H. D. Calderon-Vilca, Luis M. Ortega Melgarejo and F. C. Cardenas-Mariño, "Tuberculosis Detection Architecture with Image Processing using the SIFT and K-Means Algorithm", Journal Computación y Sistemas, Vol. 24, N° 3, 2020, pp. 989-997.

[18] H. Calderon-Vilca, N. M. Chavez and J. M. R. Guimarey, "Recommendation of Videogames with Fuzzy Logic," 2020 27th Conference of Open Innovations Association (FRUCT), Trento, Italy, 2020, pp. 27-37, doi: 10.23919/FRUCT49677.2020.9211082.