Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 3, June 2023, pp. 1596~1608
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i3.pp1596-1608

Design of a optimization algorithm for binary classification

Miguel Angel Cano Lengua[1,2], Erik Alex Papa Quiroz[1,3,4], Marco Antonio Alvarado Cifuentes[2], Carlos Antonio Alvarado Cifuentes[2]
[1]Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú
[2]Facultad de Ingeniería de Sistemas e Informática, Universidad Tecnológica del Perú, Lima, Perú
[3]Instituto de Matemática y Estadística, Universidad Federal de Goiás, Goiania, Brazil
[4]Departamento de Ciencias, Universidad Privada del Norte, Lima, Perú

Article Info

Article history:

Received Dec 27, 2022
Revised Jan 17, 2023
Accepted Jan 19, 2023

Keywords:

Binary classification
Proximal distances
Proximal multiplier
Scrum methodology
Support vector machine

ABSTRACT

In the present work, the design of a system to classify data is carried out, using the Scrum methodology. The validation was carried out by expert judgment, having favorable results in terms of different criteria such as; integrity, ease of use, innovation, and scalability. Regarding the development of the functional elements of the system, it was obtained; he developed the architecture of the system, the database, and the prototypes, among other points considered. From the implementation of the system, the equation of a classifying plane in three-dimensional space will be obtained, as well as the number of internal iterations that the algorithm develops, the estimated execution time, and the graph of the plane. This system is based on a recently introduced symmetric cone proximal multiplier algorithm to solve separable optimization problems, this algorithm made an application for classification-related support vector machines.

Corresponding Author:

Miguel Angel Cano Lengua
Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos
Lima, Perú
Email: mcanol@unmsm.edu.pe

1. INTRODUCTION

At present, due to the demands that exist in different areas such as marketing for customer segmentation, in medicine for the classification of cancer cell images, and in the personalizing preferences. For this, efficient classification algorithms are required. The different classification models start from a data set which has a set of characteristics, and we know the class to which it belongs by calling this training or learning set, a set of rules is created which allows us to validate with a set of different data. Classification techniques arise thanks to the advancement of science. In the field of artificial intelligence, one of the important disciplines is learning machines, which are divided into two areas: supervised learning machines and unsupervised learning machines, as can be seen in the Figure 1.

Supervised learning machines are subdivided into regression and classification techniques. With respect to classicication tecjniques there are some techniques that have important contributions in different areas, within the techniques we have, for example:

- Logistic regression.
- Decision tres.
- Neural networks and deep learning.
- Support vector machines.

Among others, at present, data classification algorithms are an important contribution regarding the treatment of information. They are used when the required result has a discrete label. That is, they are useful

when the response to the company's requirements is housed within a finite set of possible results. We have as an example, the verification if an email is unwanted or unwanted, in this case there are only two options and it is known as binary classification, we can see in works related to the detection of lung cancer to classify the nodules in benign and malignant [1], in addition to image processing for automatic discrimination of bananas [2]. On the other hand, you can have information in which its classification is given by multiple categories. This is very useful in different fields, for example for the detection of cervical cancer [3], other applications can be seen in the detection of brain tumors [4], to segment the client [5], [6], categorize images [7], classify to categorize audios [8] and analyze text to optimize customer sentiment [9].

Data classification techniques have great relevance in different areas, in particular there are abundant information on the technique of support vector machines related to classification [10], because there are not many software related to data classification that use the technique. of separable optimization, we present the design of a symmetric cone proximal multiplier algorithm for separable optimization, application to support vector machine (SVM) for binary classification, which is theoretically based on finite-dimensional vector spaces, as well as an inner product, a Euclidean Jordan Algebra, proximal distances, among other concepts,

The algorithm from which the system design will be developed solves an optimization problem with a separable structure, for this we consider $\mathbb{V}_1$ and $\mathbb{V}_2$ be two linear vectorial spaces of finite dimensions on $\mathbb{R}$, $\mathbb{R}$ is considered as a real field. On these spaces we define two inner products: $\langle\cdot,\cdot\rangle_{\mathbb{V}_1}$ for $\mathbb{V}_1$ and $\langle\cdot,\cdot\rangle_{\mathbb{V}_2}$ for $\mathbb{V}_2$ with the Jordan product $\circ_1$ and $\circ_2$ respectively. Based in these tools we can define the Euclidean Jordan Algebra $\mathbb{V}_1 = (\mathbb{V}_1, \circ_1, \langle\cdot,\cdot\rangle_{\mathbb{V}_1})$ and $\mathbb{V}_2 = (\mathbb{V}_2, \circ_2, \langle\cdot,\cdot\rangle_{\mathbb{V}_2})$, respectively, see subsection 2.1 for a strict definition of this concept.

Our interest is to present the design of the system based on an optimization algorithm to solve the optimization problem of the convex symmetric cone, considering variables and linear constraints:

$$\min\{f(x) + g(z) : \mathbb{A}x + \mathbb{B}z = b, x \in \mathcal{K}_1, z \in \mathcal{K}_2\}, \tag{1}$$

problem (1) can be solved using the theory of the multiplier symmetrical cone algorithm, for this we will give some concepts below.

Let $H_1: \mathbb{V}_1 \times \mathbb{V}_1 \to \mathbb{R} \cup \{+\infty\}$ be a function satisfying $H_1 \in \mathcal{D}\big(int(\mathcal{K}_1)\big)$.

Consider another function $H_2: \mathbb{V}_2 \times \mathbb{V}_2 \to \mathbb{R} \cup \{+\infty\}$ with $H_2 \in \mathcal{D}\big(int(\mathcal{K}_2)\big)$ and $\theta_1, \theta_2 > 0$ positive parameters. We define (1) and (2).

$$H_{\theta_1}(x_1, x_2) \coloneqq H_1(x_1, x_2) + \frac{\theta_1}{2}\|x_1 - x_2\|^2 \tag{2}$$

$$H_{\theta_2}(z_1, z_2) \coloneqq H_2(z_1, z_2) + \frac{\theta_2}{2}\|z_1 - z_2\|^2 \tag{3}$$

It is easy to show that for each $i = 1,2, H_{\theta_i}$ is also a proximal distance with respect to $int(\mathcal{K}_i)$, that is $H_{\theta_i} \in \mathcal{D}\big(int(\mathcal{K})\big)$. The proposed algorithm, called SC-PMA which mean symmetric cone proximal multiplier algorithm, for solving the problem (1) is defined by:

Let $H_i \in \mathcal{D}\big(int(\mathcal{K}_i)\big), \theta_i > 0, i = 1,2\,, tol > 0$ and $\{\varepsilon_k\}, \{\zeta_k\}, \{\lambda_k\}$ be sequences of positive scalars.

Step 0: start with some initial point $\omega^0 = (x^0, z^0, y^0) \in int(\mathcal{K}_1) \times int(\mathcal{K}_2) \times \mathbb{R}^m$. Set $k = 0$.

Step 1: compute

$$p^{k+1} = y^k + \lambda_k(\mathbb{A}x^k + \mathbb{B}z^k - b), \tag{4}$$

Step 2: find $(x^{k+1}, z^{k+1}) \in int(\mathcal{K}_1) \times int(\mathcal{K}_2)$ and $(g^{k+1}, g^{k+2}) \in \mathbb{V}_1 \times \mathbb{V}_2$, such that

$$g_1^{k+1} \in \partial_{\varepsilon_k} f(x^{k+1})$$

$$g_1^{k+1} + \mathbb{A}^* p^{k+1} + \frac{1}{\lambda_k}\nabla_x H_{\theta_1}(x^{k+1}, x^k) = 0 \tag{5}$$

$$g_2^{k+1} \in \partial_{\zeta_k} g(z^{k+1})$$

$$g_2^{k+1} + \mathbb{B}^* p^{k+1} + \frac{1}{\lambda_k}\nabla_z H_{\theta_2}(z^{k+1}, z^k) = 0 \tag{6}$$

Step 3: compute

$$y^{k+1} = y^k + \lambda_k(\mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b) \quad (3.11)$$

Step 4: set $w^{k+1} = (x^{k+1}, z^{k+1}, y^{k+1})$. If $\|w^{k+1} - w^k\| \leq tol$, stop; otherwise replace k by $k+1$ and go to Step 1.

From the SC-PMA algorithm, a concrete application for classification-related support vector machine can be obtained. This algorithm can work for more general fields, due to its global convergence.

Given a set of instances with their respective variables, (x_i, y_i), where $x_i \in \Re^n, i = 1, \dots, m$ y $y_i \in \{-1, +1\}$, the SVM determine an optimal hyperplane.

$$H(w, b) = \{x \in \Re^n : w^T x + b = 0\}$$

The central idea of data classification is to locate an optimal hyperplane, from which it is possible to separate the properly labeled data, for this the support vector machines use a technique which is to solve the following quadratic programming problem:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i \; s.t. \; y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, m, \quad (7)$$

where $\xi = (\xi_1, \dots, \xi_m) \in \Re^m$ y $C > 0$ is a penalty meter. Let us denote for $z = (w, b) \in \Re^{n+1}$ y $Diag(y)$ a mixed diagonal, whose main diagonal is the vector $y = (y_1, \dots, y_m)$, of 1 a vector$\Re^m$, and of:

$$\hat{X} = \begin{pmatrix} x_1^T & 1 \\ \vdots & \vdots \\ x_m^T & 1 \end{pmatrix} \in \Re^{m \times n+1}$$

then the first inequality constraint can be written as $Diag(y)\hat{X}z + \xi - \mathbf{1} \geq 0$. Introducing the variables $u = Diag(y)\hat{X}z + \xi - \mathbf{1}$ y $v = (\xi, u)$, problem (7) can be rewritten as (8).

$$\min_{z,v}\{f(z) + g(v) : \mathbb{A}\mathbb{z} + \mathbb{B}\mathbb{v} = \mathbb{b}, \mathbb{v} \geq \mathbb{0}\}, \quad (8)$$

Where $f(z) = \frac{1}{2}\|w\|^2, g(v) = C\mathbf{1}^T\xi, \mathbb{A} = Diag(y), \hat{X} \in \Re^{m \times n+1}, \mathbb{B} = (I - I) \in \Re^{m \times 2m}$ y $b = \mathbf{1}$. Therefore, this model has the form of problem (1).

For more details see Quiroz *et al*. [11]. In Figure 2 we can see the flowchart of the SC-PMA algorithm, in which we have as input data an arbitrary initial point, a parameter ρ of acceleration, another parameter λ which is the step size of each iteration, then we look for the form w^{k+1} and follow the conditionals, this process is done iteratively until finding the optimal point.

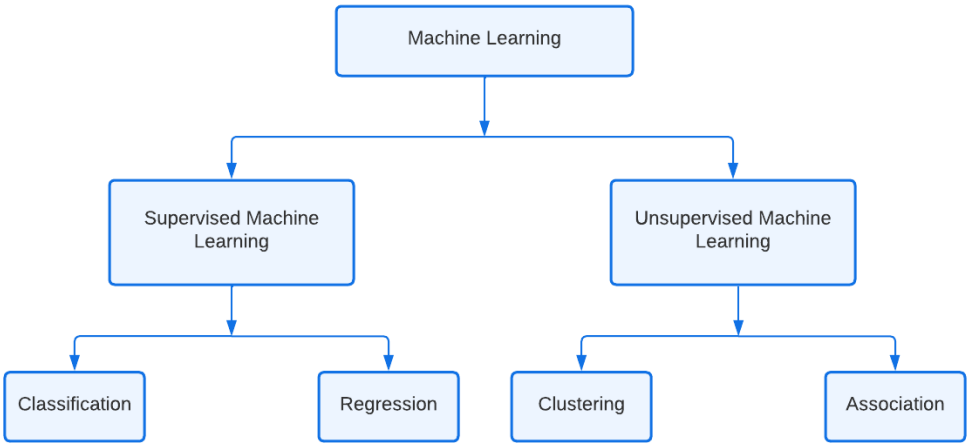


Figure 1. Studies of machine learning

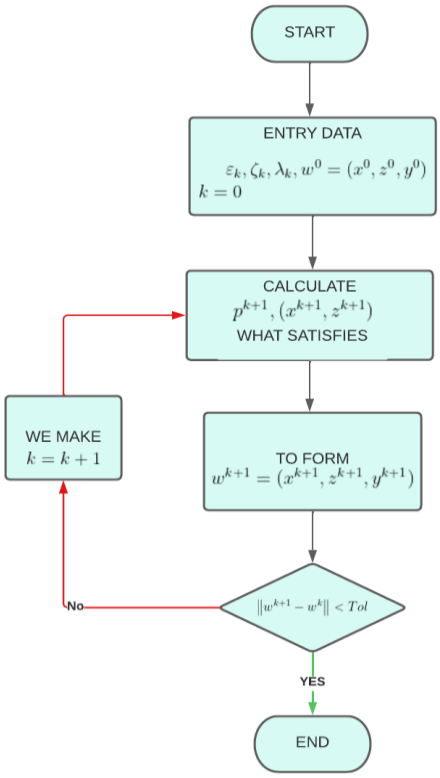


Figure 2. Flowchart of the SC-PMA algorithm

2. LITERATURE REVIEW

In the field of classification, there are a variety of jobs which provide solutions to different areas such as medicine, economics, engineering, and data science. According to the research, we have that different data classification techniques are used, the same ones that show a considerable percentage that the techniques or methods are appropriate, so we present some contributions of works related to the classification theme.

Regarding the type of algorithm that will be used in the system design, it is motivated by the study by Sarmiento *et al*. [12], who works with an inexact proximal multiplier method using proximal distances to solve convex minimization problems with a separable structure. There are important papers using a supervised autocoder approach for the classification of clinical metabolomic data, which are efficient in supporting diagnostic or prognostic assessment based on a comprehensive metabolomic analysis [13]. In addition, there is another technique based on a neural network and that is applied to resist the identity theft attack [14]. We also have the method that classifies quality kaffir lime oil into two groups: high and low quality, based on its significant chemical compounds [15], on the other hand there is work related to human activity recognition (HAR) based on in mobile phones, which consists of inferring the type of user activity from the analysis of inertial mobile sensor data [16].

On the other hand, we have the next important contribution related to caring for the environment is the one referred to recycling. Shrink sleeves interfere with the mechanical recycling of plastic bottles, plus the state of printing and shrinkage of the sleeve affect sorting, bottles with printed sleeves are more difficult to sort correctly than those with non-printed sleeves, and sorting accurately decreases with increasing shrinkage of other materials, some important results such as avoiding printing or coloring sleeves black and sleeves with large area printing and high shrinkage help the sorting process, see [17]. There is also research related to the selection of hybrid characteristics to improve the performance of cancer classification, for this he uses different bootstrap which is a representative sample of resampling with replacement before the selection step, in his research he concludes that the sets high-dimensional data and the recursive function elimination (RFE) algorithm face many problems in cancer classification performance. To do this, he proposes positioning the first step of bootstrap (PFBS), random forest for selection (RFS), recursive function elimination (RFE) that will solve problems with different positions [18], as well as other research related -to health, especially with skin. It is the diseases generated by melanoma and carcinoma are usually quite difficult to detect, one of the types of skin cancer is caused by melanoma, in this investigation a method is established to identify if a certain sample is affected by melanoma, for this purpose, collects data from labeled preprocessed images, the results of the samples have an accuracy of 90% of classification [19]. Also, there are researches related to image

classification, this article presents a general framework to estimate the parameters of these distributions using the maximum likelihood approach. Thus, the experimental results are obtained, which show that the proposed technique increases the effectiveness of SVM cores for different artificial vision tasks, such as recognition of natural scenes, classification of satellite images, and recognition of human actions in videos [20].

3. METHOD

For the development of the work, it has been decided to implement the agile Scrum methodology, because it adapts quite well to our framework and that the composition of our team focuses on the multifunctionality of all our members, as well as that it is made up of a small group to which different functions are assigned and the members are Todistas, according to what Lasa has seen regarding the versatility of its members see [21], for this reason the agile Scrum methodology has been used for the development of this.

In addition, there are some important results with this type of method regarding the implementation of systems, Ramirez *et al.* [22], whose work was related to developing a virtual store, another work with this methodology, the research related to the monitoring of people affected by COVID-19, through an application [23]. Finally, another important work related to electronic commerce to promote the purchase and sale of products online, is given through the implementation of a mobile application in a Peruvian company [24]. As shown in Figure 3, we can see a life cycle of the projects carried out with this agile method.

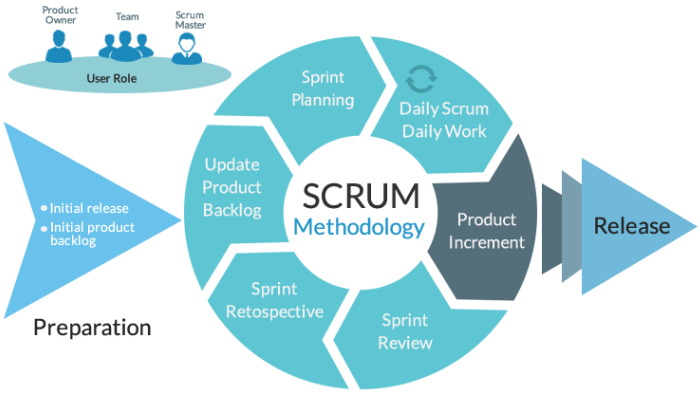


Figure 3. Scrum life cycle

Unlike a more traditional work development, which is based on order and hierarchy, the work team in a Scrum project must be capable of self-organizing and the members capable of assuming any activity necessary to complete and develop the project. For this, the 5 phases defined in the Scrum will be detailed, according to Salpathy, Melorose *et al.* [25], which are: start; planning and estimation; implementation, review and retrospective; finally, the Launch phase. The following Figure 4 we show a sample of the designation of the work team.

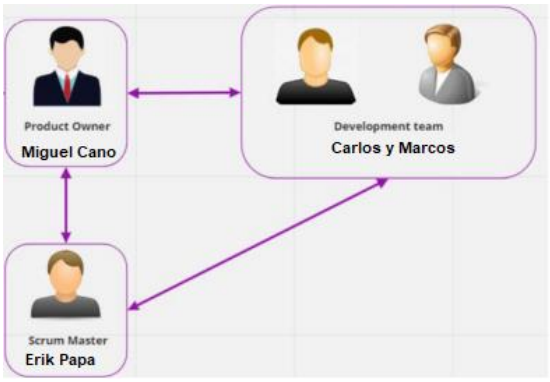


Figure 4. Scrum team diagram

3.1. Scrum

3.1.1. Initiation of the scrum proyect

Here the people in charge of the project will be defined, as well as the basic needs for the sprints in general. We will focus on detailing the main objectives of our project, as well as the specific objectives. We will analyze how many sprints the project will have, which will begin as soon as the previous one ends, as well as what will be delivered in each spint.

3.1.2. Planification

In this present section, certain goals of our project will be established, as well as understanding the relevant needs that allow the development of the software design, these needs are translated into requirements, which will become defined tasks for the Scrum team and will be ordered in a sprint backlog. In addition, we will have the sprint planning. It is in this phase of sprint planning that the sprint backlog is created with all its content. Tasks are divided into sprints to make them easier to tackle and better manageable. This makes it easier to obtain previous results to locate errors or deviations from the requirements, allowing you to work with less uncertainty and have clearer objectives.

3.1.3. Implementation

During this phase, the execution of the tasks generated in the planning will be carried out, in addition, the activities to be carried out to generate the products in the project will be detailed. This phase has two important processes, the maintenance, and updating of pending issues through the daily Scrum that we will describe. The daily Scrum is a meeting that takes place daily between the work team, as well as the Scrum master who is in charge of maintaining the pace of work and promoting communication among team members. In this meeting, all team members have the ability to voluntarily comment on what activities they have carried out, what activities they will carry out next, and what impediments they have encountered to continuing with their work. These are the points that you must treat and none more. There is no need to go into the details of what you have done; it is not necessary to discuss the solutions that have been or could be adopted; don't ramble on where the impediment originates, just how to fix it and who can do it. Focus on the purpose of the meeting and keep it short and productive. If there are pending issues, the people directly affected will deal with them calmly, thus avoiding distracting the entire team. If it affected the entire team, it would still be covered outside of the Journal to keep it short and focused on its purpose.

3.1.4. Review and retrospective

In the review and retrospective stage, it is in which the entire work team comes together to hold a meeting and review everything that has been worked on so far, in order to have a better organization that allows us to work more efficiently in the future. Through the review, the result of the work carried out is analyzed and the retrospective allows us to improve the way in which the project has been developed. The right time to perform these 2 processes is always at the end and beginning of an iteration to achieve constant improvement in the workflow.

3.1.5. Deploy

In the development of the Scrum project, the final product generated will be delivered to the client, to later document it, in addition, the corresponding tests will be carried out to verify that the system is working correctly as planned. Finally, the possibility of putting it on the web will be finalized so that this form has greater access to the use of the system and the ratios of use of the product are increased.

3.2. Case study

3.2.1. User history

This process is part of a formula to capture requirements and its purpose is to explain in a general way a function that the software will fulfill, in addition to the acceptance criteria of each of them. In each story, the functionalities established by those interested in the project and product owners will be briefly defined, which will be defined by their requirements and will be easily understood by all those involved. The process for defining these stories typically involves the product owner, Scrum master, and Scrum team, in order to facilitate the breakdown and transformation of prioritized products, ensuring quality in the delivered product even before reaching the process planning stage to create the product in question.

This tool allows us to focus on the innate value of each part of the deliverable and how these generate an effect on other components of the project, instead of going directly to the development phase of product details [26]. The usefulness of this tool is reflected in situations where the product requirements are changing, which happens in most projects. This helps so that the Scrum team helps to resolve the doubts that the stakeholders may have throughout this stage. With this we can conclude that the use of this tool provides us with multiple advantages when defining the product requirements, since it encourages constant collaboration

between members of the development team or Scrum team, as well as allows the client to have a more wide and real of the processes used to develop the project and to be able to define more precisely the requirements that are not clear or are volatile. In this way we will simplify the project to small partial deliverables that are easy to estimate, these deliverables will have an approximate duration of a few weeks and will be easy to maintain, according to [27]. In Tables 1, 2 and 3 we will see the user stories according to the requirements.

Table 1. User history I

Users history	
Number: 1	User: Product owner
Name: Login by credentials	
Priority: Medium	Risk: Medium
Estimated Points: 2	Iteration: 1
Responsible: Miguel Cano Lengua	
Description: users who wish to use the application must be able to log into it through a credential validation interface, entering those previously designated by the administrator.	
Validation: the user will be able to access the system by logging in by entering the credentials assigned by the administrator, which will be validated through a query and authentication service with the database.	

Table 2. User history II

Users history	
Number: 2	User: Product owner
Nombre: Linear Data Generation	
Priority: High	Risk: Low
Estimated Points: 5	Iteration: 2
Responsible: Miguel Cano Lengua	
Description: i want users to be able to generate linearly separable data through an interface that accepts user-defined variables, this data will only be visible through a three-dimensional plane	
Validation: in this screen the user will be able to enter data of the minimum and maximum quadrant value, number of observations and make the decision to generate or not the graph in R3.	

Table 3. User history III

Users history	
Number: 3	User: Product owner
Nombre: Graph generation in R3 and data processing	
Priority: High	Risk: High
Estimated Points: 5	Iteration: 2
Responsible: Erik Papa Quiroz	
Description: in an interface, the user must be able to process the data and obtain a hyperplane that divides the space in R3 and thus classify the data, thereby facilitating decision making.	
Validation: the user will be able to see the data previously processed and delimited in a hyperplane, which will be temporarily stored locally.	

3.2.2. Entity relation–diagram (DER)

This diagram created in the 70's by Dr. Peter Chen aims in our research to generate a logical model that helps us in the process of capturing information within the topic that we develop, generating an adequate framework for the development of our database. This provides us with both a data model and independent accesses, so that we can opt for a subsequent implementation technique of possible software in the future, which has the ability to be scalable [28]. In Figure 5 we show the database diagram of the system.

3.2.3. Product backlog

As mentioned by Sedano, J. *et al*, see [29], the product backlog or also known as the product stack, comes to be all the ideas defined in the vision of the stakeholders of the project materialized, so it is essentially made up of all the planned functionalities that the final product must have in order for it to be considered finished, as well as to achieve the development objective. As can be seen in the following Table 4.

3.2.4. Software architecture

This pattern oriented to the development of applications, will be part of our project, since it will allow us to separate the logic of our project from the design, which will be the visible part for the client. Providing ease of maintenance and scalability at the programming level to our software, In this way, the user will be able

to carry out their queries without the need to directly interact with the internal data of the database server, according to [28]. In Figure 6 we can see the architecture of the system.

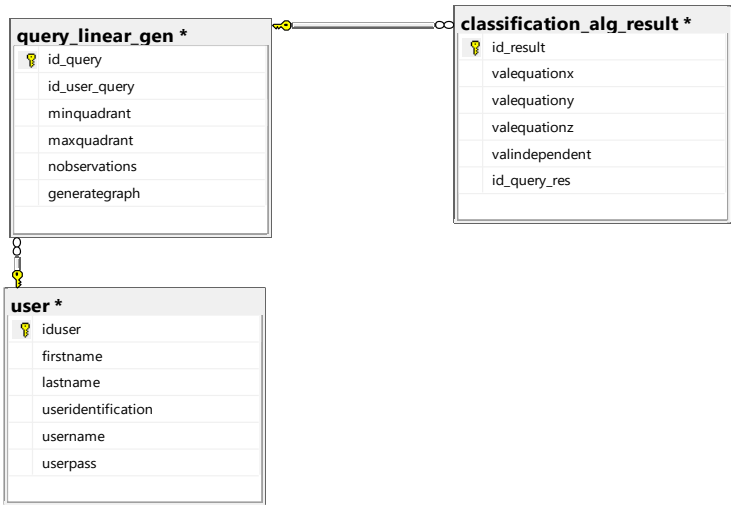


Figure 5. Database diagram

Table 4. Product backlog

User History	Priority	Estimation
H3: In an interface, the user must be able to process the data and obtain a hyperplane that divides the space in R3 and thus classify the data, thereby facilitating decision making.	High	13
H2: I want users to be able to generate linearly separable data through an interface that accepts user-defined variables, this data will only be visible through a three-dimensional plane.	High	5
H1: Users who wish to use the application must be able to log into it through a credential validation interface, entering those previously designated by the administrator.	Medium	2

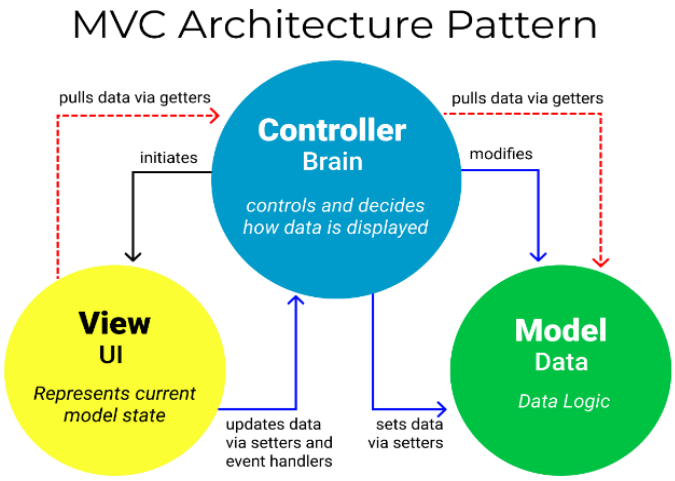


Figure 6. System atchitecture

4. RESULTS AND DISCUSSION

4.1. Prototype development

Finally, for the development of our deliverable, the prototype of our desktop application is generated, which will be ordered and related to the aforementioned requirements. Additionally, the images that were created during the design phase will be published. The images were developed in the figma software (Good results have been obtained from this software due to previous work), which meet the requirements established in the user stories, starting the first deliverables, that is, the first Sprint.

4.2. First sprint

The first sprint will result in the answer to user story 1, which will be explained in detail later with its respective application design prototype. H1: users who wish to use the application must be able to log into it through a credential validation interface, entering those previously designated by the administrator. In Figure 7 you can see the login design where the administrator user will be able to enter with the assigned credentials, when these are confirmed by a database query service. After confirmation, the user will be able to access all the functionalities of the system.

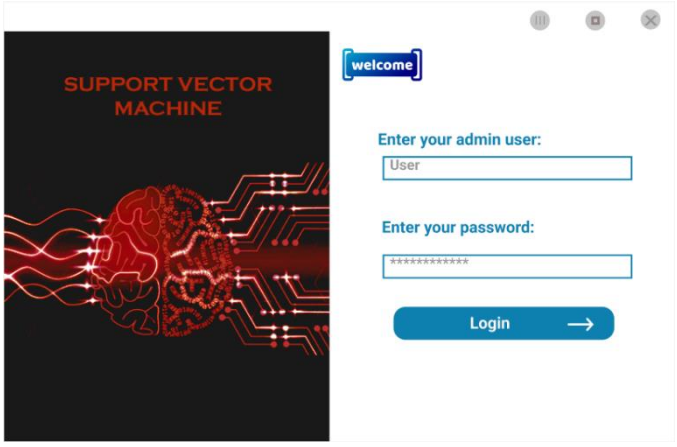
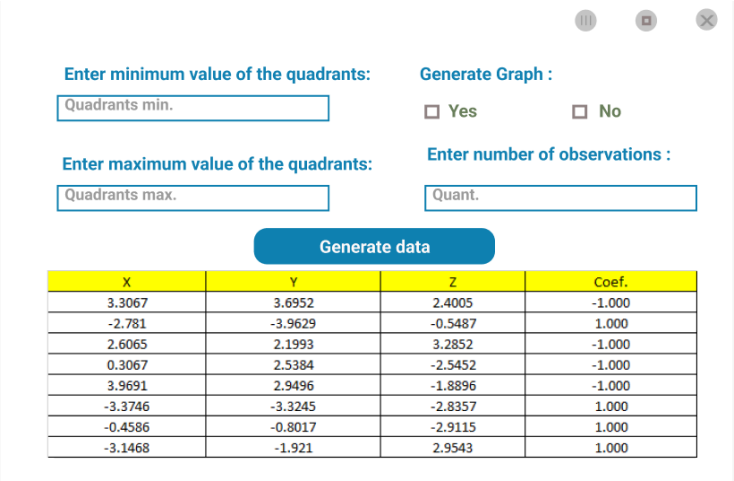


Figure 7. User login

4.3. Second sprint

The second sprint is made up of user stories 2 and 3. Next, you can see the design of the system functionalities defined by the corresponding user stories. In other words, some parameters can be entered to start the execution of the algorithm, which will provide us with a graph in three-dimensional space. H2: i want users to be able to generate linearly separable data through an interface that accepts user-defined variables, this data will only be visible through a three-dimensional plane. As can be seen in Figure 8, the user through the presented interface can enter the data of the minimum and maximum value of the quadrants, as well as select whether to generate a graph in the final interface and finally enter the number of observations and then proceed to press the generate button. data that starts the algorithm execution process.

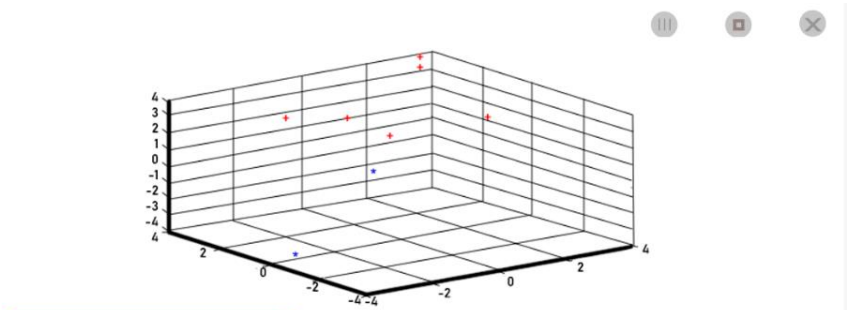


X	Y	Z	Coef.
3.3067	3.6952	2.4005	-1.000
-2.781	-3.9629	-0.5487	1.000
2.6065	2.1993	3.2852	-1.000
0.3067	2.5384	-2.5452	-1.000
3.9691	2.9496	-1.8896	-1.000
-3.3746	-3.3245	-2.8357	1.000
-0.4586	-0.8017	-2.9115	1.000
-3.1468	-1.921	2.9543	1.000

Figure 8. System parameter requirements

H3: in an interface, the user must be able to process the data and obtain a hyperplane that divides the space in R3 and thus classify the data, thereby facilitating decision making. Finally, in Figure 9 you can see

the graph that would be generated if the "Generate graph" option was selected in the previous process, as well as the data generated by the algorithm. In Figure 10 we can see the separator Euclidean plane, which divides the cloud of points which have a label and which were entered into the system, later the system provides us with the plane, the same one that is developed in three-dimensional space, we can also observe the equation generated by the system.



Process

iter	N(z_{k})	N(v_{k})	||z_{k}-z_{k-1}||	||v_{k}-v_{k-1}||	|w_{k}-w_{k-1}|_max	Objective
1	13	60	21.63938	4.0295	103.35049	200.07207
2	31	81	329.53588	39.57013	810.41018	47883.0261
3	15	140	1882.58656	209.13038	2744.32983	1235693.14
4	17	141	4883.01888	547.14145	4883.01888	5470100.2
5	20	141	5728.23446	761.86147	5728.23446	2920965.5
6	12	41	2083.58367	278.22211	2083.58367	56059.8634
7	13	21	585.88425	81.71181	585.88425	31595.1951
8	9	23	117.80751	19.06915	236.57306	68069.7291
9	16	17	57.56287	5.44092	176.79915	48510.0862
10	14	18	111.19339	13.68227	111.19339	20104.5262

Figure 9. Result of the data and graph of Lo data in R^3

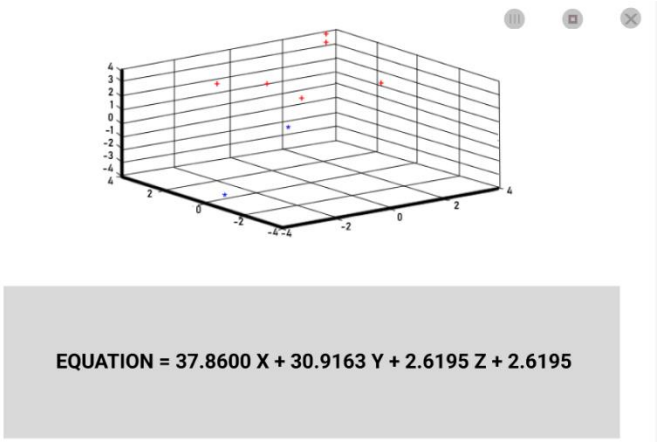


Figure 10. Graph of the separating plane and its general equation

4.4. Expert judgment

This method will be relevant in our investigation since it will allow us to validate the reliability of our investigation, there are works related to applications of web systems [29], according to what was seen by the investigation [30], for this investigation 5 experts from the area of software and database development were summoned so that they could make a value judgment about our investigation work. The prototypes made in the free software Figma were presented to them, so that they can be judged under criteria of: Integrity, ease of use, innovation and scalability. In such a way that they provide their qualification to each of the sections, with qualifications on a scale of 1 to 20, with 1 being the lowest score and 20 the best, see Figure 11. With this we obtained an average score of 18.2 for all the sections, thus giving as a result that the implementation of this desktop software met its objective. The following Table 5 shows the results of the expert judgment and in Table 6 we see the weighted results.

Figure 11. Evaluation results of the expert jury

Table 5. Results of the evaluation of expert judgment

Criterion	Jurors				
	1	2	3	4	5
Integrity	18	18	17	17	15
Easy of use	16	16	17	17	18
Innovation	19	20	19	18	17
Escalability	16	18	17	18	18

Table 6. Weighted results

Task	Average	Compliance
Integrity	17	High
Easy of use	17.4	High
Innovation	18.8	High
Escalability	15.8	Medium
Average	17.25	High

Figure 11 shows the graph of the judges' evaluation results, which were developed in accordance with the demanding criteria considered by the people who were in charge of developing the design evaluation. This graph shows that the innovation criterion is the one with the highest score followed by ease of use, integrity and scalability respectively. Table 7 shows us the validation of the reliability of the measurement instrument that was used, which was Cronbach's Alpha, whose result is 0.948, which is within the evaluation range that has this reliability coefficient of excellent. This validation is important because it tells us that the instrument created, developed, and validated is correct. As additional information we can also mention that this coefficient will help us determine the reliability of the scale used to measure our variables under study, which were also applied to the criteria evaluated by expert judgment, obtaining the mean and standard deviation shown in Table 8. This information is important for the validation of the design by expert judgment.

Table 7. Results of Cronbach's Alpha

Reliability statistics		
Cronbach's Alpha	Cronbach's Alpha based on standardized elements	N of Elements
,948	,901	8

Table 8. Measurement of criteria by Cronbach's Alpha

Item statistics			
Parameter	Mean	Standard deviation	N
Integrity1	3,40	1,342	5
Integrity2	4,40	,548	5
EasyofUse1	3,60	1,673	5
EasyofUse 2	3,80	1,643	5
Innovation1	3,40	1,342	5
Innovation2	4,00	1,732	5
Scalability1	3,80	1,643	5
Scalability2	3,60	1,517	5

5. CONCLUSION

In conclusion, the present work consists in the elaboration of a desktop application, which seeks to generate a vector support machine. With the use of the Netbeans development IDE and the figma design tool, both powerful tools for prototyping and creation. of software models, with which the designs and deliverables are elaborated, covering the needs of a public of experts in the subject. In addition, the software architecture with which our project was conceived, allowed the creation of a functional and detailed product, facilitating the possibility of future development modules. In the future we plan to complement this research with the development of a graphic module, which allows the user to visualize the cloud of results obtained, in striking 3D graphics, as well as the possibility of embedding planes in it, to visualize the division of the data.

REFERENCES

[1] I. Naseer, T. Masood, S. Akram, A. Jaffar, M. Rashid, and M. A. Iqbal, "Lung cancer detection using modified AlexNet architecture and support vector machine," *Computers, Materials and Continua*, vol. 74, no. 1, pp. 2039–2054, 2023, doi: 10.32604/cmc.2023.032927.

[2] M. Senthilarasi, S. M. M. Roomi, and M. R. H. Prasanna, "Shape based approach for detecting Musa Species in fruit industry," in *6th International Conference on Advanced Computing, ICoAC 2014*, Dec. 2015, pp. 157–160, doi: 10.1109/ICoAC.2014.7229765.

[3] S. Priya, N. K. Karthikeyan, and D. Palanikkumar, "Pre screening of cervical cancer through gradient boosting ensemble learning method," *Intelligent Automation and Soft Computing*, vol. 35, no. 3, pp. 2673–2685, 2023, doi: 10.32604/iasc.2023.028599.

[4] M. Kolla, R. K. Mishra, S. Z. U. Huq, Y. Vijayalata, M. V. Gopalachari, and K.-A. Siddiquee, "CNN-based brain tumor detection model using local binary pattern and multilayered SVM classifier," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–9, Jun. 2022, doi: 10.1155/2022/9015778.

[5] J. V. S. L. Jyostna, B. Jameema, N. Anusha, and P. Ramva, "A classification model for mall customers data," in *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, Sep. 2021, pp. 883–888, doi: 10.1109/ICIRCA51532.2021.9544515.

[6] S. J. J. Rathnayaka, C. J. Ranathunga, R. Navarathna, A. Kaneswaran, and Y. Balathasan, "Predicting movie ratings from audience behaviors on movie trailers," in *2021 10th International Conference on Information and Automation for Sustainability, ICIAfS 2021*, Aug. 2021, pp. 489–493, doi: 10.1109/ICIAfS52090.2021.9606066.

[7] S. D. C. S. E. Araujo, V. S. Malemath, and K. M. Sundaram, "Advanced processing and classification of plant disease," in *Proceedings of Seventh International Congress on Information and Communication Technology*, vol. 447, 2023, pp. 127–136, doi: 10.1007/978-981-19-1607-6_11.

[8] S. R. Zaman, D. Sadekeen, M. A. Alfaz, and R. Shahriyar, "One source to detect them all: gender, age, and emotion detection from voice," in *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jul. 2021, pp. 338–343, doi: 10.1109/COMPSAC51774.2021.00055.

[9] J. Luo, S. Qiu, X. Pan, K. Yang, and Y. Tian, "Exploration of spa leisure consumption sentiment towards different holidays and different cities through online reviews: implications for customer segmentation," *Sustainability (Switzerland)*, vol. 14, no. 2, Jan. 2022, doi: 10.3390/su14020664.

[10] M. A. C. Lengua and E. A. P. Quiroz, "A systematic literature review on support vector machines applied to classification," in *2020 IEEE Engineering International Research Conference (EIRCON)*, Oct. 2020, pp. 1–4, doi: 10.1109/EIRCON51178.2020.9254028.

[11] E. A. P. Quiroz, J. L. Luis, and M. C. Lengua, "A proximal multiplier method for convex separable symmetric cone optimization," in *ACM International Conference Proceeding Series*, May 2020, pp. 92–97, doi: 10.1145/3404716.3404734.

[12] O. Sarmiento, E. A. P. Quiroz, and P. R. Oliveira, "A proximal multiplier method for separable convex minimization," *Optimization*, vol. 65, no. 2, pp. 501–537, Feb. 2016, doi: 10.1080/02331934.2015.1062009.

[13] D. Chardin, C. Gille, T. Pourcher, O. Humbert, and M. Barlaud, "Learning a confidence score and the latent space of a new supervised autoencoder for diagnosis and prognosis in clinical metabolomic studies," *BMC Bioinformatics*, vol. 23, no. 1, Sep. 2022, doi: 10.1186/s12859-022-04900-x.

[14] J. Xue and H. Zhou, "Physiological-physical feature fusion for automatic voice spoofing detection," *Frontiers of Computer Science*, vol. 17, no. 2, p. 172318, Apr. 2023, doi: 10.1007/s11704-022-2121-6.

[15] N. S. J. Jailani, Z. Muhammad, M. H. F. Rahiman, and M. N. Taib, "Intelligent grading of kaffir lime oil quality using non-linear support vector machine," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6716-6723, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6716-6723.

[16] M. B. Abidine, M. Oussalah, B. Fergani, and H. Lounis, "Activity recognition on smartphones using an AKNN based support vectors," *Sensor Review*, vol. 42, no. 4, pp. 384–401, Jun. 2022, doi: 10.1108/SR-05-2021-0157.

[17] X. Chen, N. Kroell, M. Althaus, T. Pretz, R. Pomberger, and K. Greiff, "Enabling mechanical recycling of plastic bottles with shrink sleeves through near-infrared spectroscopy and machine learning algorithms," *Resources, Conservation and Recycling*, vol. 188, Jan. 2023, doi: 10.1016/j.resconrec.2022.106719.

[18] N. M. Abdelwahed, G. S. El-Tawel, and M. A. Makhlouf, "Effective hybrid feature selection using different bootstrap enhances cancers classification performance," *BioData Mining*, vol. 15, no. 1, 2022, doi: 10.1186/s13040-022-00304-y.

[19] N. V. Kumar, P. V. Kumar, K. Pramodh, and Y. Karuna, "Classification of skin diseases using image processing and SVM," in *Proceedings - International Conference on Vision Towards Emerging Trends in Communication and Networking, ViTECoN 2019*, Mar. 2019, pp. 1–5, doi: 10.1109/ViTECoN.2019.8899449.

[20] M. H. Rahman and N. Bouguila, "Efficient feature mapping in classifying proportional data," *IEEE Access*, vol. 9, pp. 3712–3724, 2020, doi: 10.1109/ACCESS.2020.3047536.

[21] C. Lasa, A. Álvarez, and R. Heras, *Manual Imprescindible Métodos Ágiles Scrum, Kanban, Lean*, 1st ed. Anaya Multimedia, 2018.

[22] A. S. B. Ramirez, B. A. S. Diestra, and M. A. C. Lengua, "Implementation of a virtual store to exponentiate the flow of product sales in a private company in the city of Lima," in *2021 IEEE Engineering International Research Conference (EIRCON)*, Oct. 2021, pp. 1–4, doi: 10.1109/EIRCON52903.2021.9613146.

[23] J. C. P. Gonzales, M. A. C. Lengua, and L. A.-Arenas, "Multiplatform mobile: Post-COVID-19 patient management in Lima, Peru," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 415–427, Oct. 2022, doi: 10.14445/22315381/IJETT-V70I10P241.

[24] J. R. L. Motta, E. T. P. Ipanaque, M. A. C. Lengua, and L. A.-Arenas, "Implementation of a mobile application: sales optimization in a peruvian company," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 58–69, Oct. 2022, doi: 10.14445/22315381/IJETT-V70I11P206.
[25] J. Melorose, R. Perroy, and S. Careas, "Statewide agricultural land use baseline," in *Climate Change 2013 - The Physical Science Basis*, 2015.
[26] L. S. J. Abad, *User Stories: A Pragmatic View*. Independent Publication, 2018.
[27] A. Menzinsky, G. López, J. Palacio, M. Á. Sobrino, R. Álvarez, and V. Rivas, "Historia de Usuario," in *de Scrum Manager, Madrid, Creative Commons*, 2020, p. 62. Available: https://scrummanager.net/files/scrum_manager_historias_usuario.pdf . ID: 2009135322450.
[28] R. Barker. "El modelo entidad - relación". *Copublicacion Edciones Días de Santos*¸1997.
[29] M. A. Jabar, M. F. Azmi, and F. Sidi, "Integration of mobile and web application: An implementation of diabetic management system," *Journal of Theoretical and Applied Information Technology*, vol. 55, no. 2, pp. 168–173, 2013.
[30] J. E.-Pérez and C. U. El Bosque, "Content validity and expert judgment: an approach to its use," *Avances en Medición*, vol. 6, no. 27–36, 2008,

BIOGRAPHIES OF AUTHORS

Miguel Angel Cano Lengua is a professor at the Technological University (UTP) of Peru and the National University of San Marcos (UNMSM), has a degree in Mathematics, a Master's in Systems Engineering from the National University of Callao (UNAC) and is a candidate for a doctorate in Engineering of Systems and Computer Science from the Universidad Nacional Mayor de San Marcos. He works on continuous optimization, artificial intelligence algorithms, conical programming, numerical methods, methodology, and software design. He can be contacted at email: mcanol@unmsm.edu.pe.

Erik Alex Papa Quiroz has a degree in mathematics from the National University of Callao (UNAC). He studied masters and doctorate at the Federal University of Rio de Janeiro (UFRJ), Brazil, obtaining the degree of Doctor in Systems and Computer Engineering, he is a Professor at the San Marcos National University and a researcher at the Universidad Privada del Norte. He is a researcher of mathematical optimization, study new methods to solve problems in continuous optimization, variational inequalities and equilibrium problems in vector spaces and Riemannian manifolds. His research interests include also conic programming, semidefinite optimization, quasiconvex optimization and Criptography. He can be contacted at email: epapaq@unmsm.edu.pe.

Marco Antonio Alvarado Cifuentes is a graduated of Systems Engineering from the Technological University of Peru (UTP), graduated of computing and informatics from the Institute for Research and Development of Administration and Technology (IDAT), a statistics student from the Universidad Mayor de San Marcos (UNMSM). Experience focused on the development of native mobile applications for the improvement and innovation of industrial processes and problems of society. With some software development projects on Github. He can be contacted at email: u20309837@utp.edu.pe.

Carlos Antonio Alvarado Cifuentes is a graduated of Systems Engineering from the Technological University of Peru (UTP), a graduate of computing and informatics from the Institute for Research and Development of Administration and Technology (IDAT). Experience focused on the development of Web services in Spring for business applications, searching for improvement and automation of processes. With publications in a journal indexed in Scopus. He can be contacted at email: u20308845@utp.edu.pe.