

—  
**Vol. 1**  
**Núm. 1**

2024  
—



**REVISTA** \_\_\_\_\_  
Datos, Políticas  
e Innovación Pública



**IIEG**  
Instituto de Información  
Estadística y Geográfica  
de Jalisco

## **2024. Instituto de Información Estadística y Geográfica de Jalisco**

Datos, Políticas e Innovación Pública (DP+IP), es una revista internacional de difusión científica de revisión entre pares que busca explorar las implicaciones, alcances, retos, prácticas y cuestionamientos alrededor de las innovaciones tecnológicas, metodológicas y epistemológicas relacionadas con la administración pública en el marco de la sociedad de la información y los procesos de datificación.

En los últimos años se ha visto un creciente interés en el uso de tecnologías digitales como la inteligencia artificial o la integración de la ciencia de datos en todos los órdenes de gobierno, siendo implementadas para abordar una amplia gama de desafíos, desde la planificación del transporte, hasta la lucha contra el cambio climático.

Esta revista busca presentar investigaciones, estudios de caso y análisis críticos sobre el uso de datos y tecnología al interior de la administración pública.

Esperamos que esta publicación sirva como una fuente valiosa de información y perspectivas para aquellos interesados en el papel de los datos y la tecnología al interior de la administración pública.

Revista Datos, Políticas e Innovación Pública, editada por el Instituto de Información Estadística y Geográfica de Jalisco, Calzada de los Pirules # 71, Colonia Ciudad Granja, 45010, Zapopan, Jal., Teléfono 33 3777 1770.

El contenido de los artículos, así como sus títulos y, en su caso, fotografías y gráficos utilizados son responsabilidad del autor, lo cual no refleja necesariamente el criterio editorial institucional. Asimismo, la Revista se reserva el derecho de modificar los títulos de los artículos, previo acuerdo con los autores. La mención de empresas o productos específicos en las páginas de la Revista no implica el respaldo del IIEG.

Se permite la reproducción total o parcial del material incluido en la revista, siempre y cuando se cite la fuente, incluyendo el nombre del autor, el nombre y número de la revista, y la dirección electrónica de la misma.

## Presentación

Para el Instituto de Información Estadística y Geográfica, es de vital importancia la promoción de la innovación como uno de los ejes transversales con el que abordamos nuestras labores y actividades sustantivas. Esta apuesta no solo radica en la generación de nuevos productos informativos o de consulta, sino también en la generación y promoción de conocimiento que abone a la consolidación de un ecosistema de saberes públicos, y que permita abordar los problemas públicos de mejor manera.

Poco explorado en el Instituto, son las posibilidades que abren el hecho de ser reconocidos como un Centro de Investigación Público, distinción que nos obliga a ser enfáticos en la forma en la que dialogamos y hacemos equipo con el sector académico. Esta es la razón por la que apostamos en crear una revista de revisión entre pares llamada Datos, Política e Innovación Pública, que busca explorar las implicaciones, alcances, retos, prácticas y cuestionamientos alrededor de las innovaciones tecnológicas, metodológicas y epistemológicas relacionadas con la administración pública en el marco de la sociedad de la información y los procesos de datificación.

En los últimos años se ha visto un creciente interés en el uso de tecnologías digitales como la inteligencia artificial o la integración de la ciencia de datos en todos los órdenes de gobierno, siendo implementadas para abordar una amplia gama de desafíos, desde la planificación del transporte, hasta la lucha contra el cambio climático. Esta revista busca presentar investigaciones, estudios de caso y análisis críticos sobre el uso de datos y tecnología al interior de la administración pública. Los temas cubiertos incluyen:

- Uso y protección de datos
- Ética de la inteligencia artificial
- Gobernanza de los datos
- Uso de metodologías mixtas y datos densos (big data)
- Mecanismos de participación ciudadana
- Gobiernos abiertos
- Ciencia ciudadana
- Tecnologías ancestrales para abordar problemas públicos
- Entre otros

Agradecemos a todos nuestros revisores y a todas las personas que enviaron su artículo, un gesto importante para nosotros, ya que sin él, no podríamos aspirar a seguir los procesos para lograr que nuestra revista cumpla con los formalismos para la obtención del ISSN y su posterior indexación.

## Índice

1. Towards the use of technology in an open, complex and popular democracy	<b>6</b>
2. Más allá de la protección de datos personales: ¿Y la reparación del daño?	<b>13</b>
3. Diagnóstico educativo de Jalisco: una mirada desde el registro histórico de indicadores	<b>20</b>
4. Control de acceso biométrico a espacios públicos con base en tecnologías emergentes	<b>30</b>
5. Revisión de literatura: análisis de sentimientos en X para medir la aceptación de políticas públicas	<b>39</b>
6. Aplicación Web para identificar personas a partir de la similitud semántica en descripciones físicas textuales utilizando técnicas de Procesamiento del Lenguaje Natural	<b>53</b>



Datos, Políticas e Innovación Pública, vol. 1, núm. 1, marzo; 2024.  
Editada por el Instituto de Información Estadística y Geográfica de Jalisco, Calzada de los  
Pirules # 71, Colonia Ciudad Granja, 45010, Zapopan, Jal., Teléfono 33 3777 1770.

## Towards the use of technology in an open, complex and popular democracy

Juan Pablo Martínez Díaz

Gestión Pública y Políticas Globales, Instituto Tecnológico y de Estudios Superiores de Occidente, Tlaquepaque, Jalisco, México.  
juanpablomartd@gmail.com

**Summary.** The essay delves into the transformative impact of rapid technological and economic changes, wealth concentration, climate emergency, and the escalating complexity of global issues on contemporary societies. Specifically, it explores the sway of technocracy and neoliberalism on public administration, honing in on the New Public Management paradigm prevalent in the final quarter of the 20th century. The consequence has been a diminishing governmental capacity and an inadequacy in addressing societal challenges, notably the issue of inequality. The narrative reflects on the intricacies and constraints of representative democracy, advocating for an interdisciplinary and technological approach to confront public problems effectively. Furthermore, the essay underscores the potential conviviality of technology as a pivotal means of addressing these challenges and emphasizes its role in fostering democratic participation, especially at the local government level. Illustrative examples, such as the Decidim system in Barcelona and other international cases, serve to highlight the positive impact of incorporating technology. In essence, the essay puts forth an argument for the imperative of inclusive, open, and popular governance models that leverage technology for the collective good and the enhancement of societal realities.

**Keywords.** *Popular participation, technology, democracy, complexity, local governments.*

**Resumen.** El ensayo aborda el impacto transformador de los rápidos cambios tecnológicos y económicos, la concentración de riqueza, la emergencia climática y la creciente complejidad de los problemas globales en las sociedades contemporáneas. Específicamente, explora la influencia de la tecnocracia y el neoliberalismo en la administración pública, centrándose en el paradigma de la Nueva Gestión Pública predominante en el último cuarto del siglo XX. La consecuencia ha sido una disminución de la capacidad gubernamental y una incapacidad para abordar desafíos sociales, especialmente la desigualdad. La narrativa reflexiona sobre las complejidades y limitaciones de la democracia representativa, abogando por un enfoque interdisciplinario y tecnológico para enfrentar eficazmente los problemas públicos. Además, el ensayo destaca el potencial de convivencia de la tecnología como un medio fundamental para abordar estos desafíos y enfatiza su papel en fomentar la participación democrática, especialmente a nivel gubernamental local. Ejemplos ilustrativos, como el sistema Decidim en Barcelona y otros casos internacionales, sirven para resaltar el impacto positivo de incorporar la tecnología. En esencia, el ensayo presenta un argumento sobre la necesidad de modelos de gobierno inclusivos, abiertos y populares que aprovechen la tecnología en beneficio colectivo y para mejorar las realidades sociales.

**Palabras clave.** *Participación popular, tecnología, democracia, complejidad, gobiernos locales.*

### 1. Introduction

“The technique at the service of a better homeland”

Jesús Robles

A reality as changing as it is complex requires rethinking our democracies to solve the diverse crisis we live in. Within a process of increasing technification, incorporating digital tools present an alternative which makes possible a better resolution of social issues. Nevertheless, the hegemonic paradigms in public administration and the concepts mostly used in the public opinion seems not to adequately incorporate these tools for the generation of welfare and popular empowerment. The suggested alternative is reducing the cost of participation in the institutional political system through means which simplifies, traduce and bring the politics closer to every person, contributing to distinct spheres for the understanding of complexity, making possible the solution of social conflicts.

Adapting to the reality of Jalisco, Mexico, according to data from the ENDUTIH, in 2020 there were 5 '820,823 internet users, who represented 75.6% of the state's population aged over 6 years old, of which 9 out 10 utilize it everyday. In turn, the population who uses the internet the most, with 19.5% of the users, was between 25 and 34 years old. This data serves as a foundation for advocating the use of technology in the creation of political pedagogy tools. There exists a propitious ground for the use of technologies as infrastructures that complements the popular participation, given that the technological gap isn't wide, just as its use is a fertile sector for the creation of tools of political pedagogy. Not forgetting that these tools do not replace territorial work, mainly for the participation of sectors with no or limited access to these technologies, as well as the dynamics of face-to-face activities to approach the understanding of complexity and facilitating the resolution of social conflicts.

Along this essay there's a reflexion over complexity and the boundaries of representative democracies on the contemporary societies, in the frame of climate emergency, the hyperconcentration of wealth and the interrelation of diverse global issues, emphasized by the neoliberalism and the New Public Management over the last decades. The argument goes for the necessity of an interdisciplinary and technological approach on the solving of social problems through mechanisms of participatory and popular democracy within local governments. In essence, this holistic perspective seeks to align democratic principles with the fluid and multifaceted nature of our present reality.

## **2. The arrival of technocracy and neoliberalization**

The world isn't what once's was: the societies grow faster, the consumption of products and commodities are limitless, the desires and interests of the persons have been multiplying, the hyperconcentration of wealth has been increasing, some multinational corporations own more power than many of the government of the G20, a climate emergency just around the corner, the financialization of housing, the growing influence of high tech and big data and the complexity of cross-borders problems that difficult the multilateral cooperation. Facing a volatile and vertiginous world, democracies seem lethargic regarding the necessities of contemporary societies.

Aguilar Villanueva (2010) identifies the origin of the present paradigm of government management at the edge of the Welfare State model and the appearance of multiple social actors who demand an expansion of participation on public decisions, just as the corporate sector, during a process of expanding global innovation, privatization and transnationalization, the so called process of neoliberalization and the arrival of the New Public Management at the principles of many of the governments around the world. In this global period many opaque and inefficient public programs were canceled, there were a large number of privatizations, models of design and evaluation cost-efficiency were introduced in the public policy methods and many autonomous public organizations were found. In the Western societies it seemed there were no alternatives after the fall of the Berlin wall. In Mexico, following the Miguel de la Madrid government, sectors ideologically related took control of the Partido Revolucionario Institucional (PRI), therefore, the government and model of the country: it was a moment of management and growth on purpose of placing the Republic as an emerging, modern and global power.

However, this process, such as in Mexico as the rest of the communities it was implemented, turned into the loss of the government's directive capacity due to the decrease on their faculties and finance, letting them as complementary actors faced with the international firms, becoming on the public impossibility of solving social problems just as inequality, social rights, medicare, education, mobility and infrastructure.

Despite the overall weakening of the government capabilities, efforts have been made to strengthen the organizations and methods related to the electoral sphere of democracy. Despite several advancements, these measures have not entirely resolved the technical issues surrounding elections. Furthermore, the presidential elections in Mexico in both 1988 and 2006 were overshadowed by methodological issues that cast doubt on the transparency and legitimacy of the processes. In 1988, the contest between Carlos Salinas de Gortari and Cuauhtémoc Cárdenas was affected by questioning of the vote count, sparking controversies and raising doubts about the authenticity of the results with the "caída del sistema" (system crash). The lack of clarity in the process underscored the need for improvements in electoral supervision. In 2006, the competition between Felipe Calderón and Andrés Manuel López Obrador faced methodological challenges, ranging from establishing equitable conditions to monitoring irregularities. Lack of fairness and a perception of opacity led to tensions on the validity of the results. These issues emphasized the urgency of addressing deficiencies in the design and execution of technologies to strengthen public trust and preserve the integrity of the democratic system.

Years later, crisis like the one on 2008, added to multitudinous protests unleashed around the world, just as Occupy Wall Street in the United States or los Indignados in Spain and particularly in Mexico with the #YoSoy132 movement, predicted the necessity of more open models of democracy, not only focused on technologies within the electoral process, but also adding to all spheres of public decision, incorporating the desires of the few but guaranteeing them not overcome the complex will of the many. Those movements sought to democracies where the responses to the complexity don't result in technocracies which make legislation and policies away from the peoples problems, where the institutions don't mean distance and opacity.

## **3. A complex reality: the boundaries of representatives democracies**

Contemporary societies with a vast diversity of ideas and interest require thinking about an approach which permits facing a diversified and fluctuating reality with the implementation of participation models for expression, deliberation and decision making about that host of ideas and interes, models which propose different forms of construction of individual and collective knowledge.

At the first conference of his course “The truth and the judicial forms” (1984) Michel Foucault presented his theory of knowledge, which defines from the “domains of knowledge”. He postulates that the subject is constructed based on knowledge, just as on control and surveillance. Hence, for him, knowledge isn’t only imposed, it is created. As being generated, the subject of knowledge emerged from the emergency of subjectivity, meaning it came out “of games of action and reaction, of question and answer, of domination and also of struggle” that respond to a specific historical context (Foucault, 1984). In this way, there are distinct subjects who grasp in different ways, not a subject predetermined by immobile superstructures. Foucault argues that to understand the knowledge of the subjects is fundamental studying the history of its creation with a political approach in the search of deciphering power relations between persons and groups of a determined community.

Furthermore, for the french sociologist, politics and economics are not an obstacle for knowledge rather are “through which the subjects of knowledge are formed and consequently, the relations of truth” (Foucault, 1984). For his part, Antonio Gramsci express that determine forms of knowledge come through power to possess the “hegemony” of the “common sense” (Errejón, 2011) for sustaining a project of society, a “historic bloc” (Gramsci, 2009). For these authors, also on science and technology are truth models, “paradigms” in Thomas Kuhn terminology, product as at the same time constituents of political structures. Therefore, knowledge, on a complex reality, is constructed and disputed.

Then, it seems that the understanding of public problems gets more complicated for their resolution. Nevertheless, if we follow the definition suggested by the philosopher and sociologist Edgar Morin (1990) the adjective “complex” doesn’t mean “complicated”, rather means an interdependent system in which distinct logics subsist and connect (Barberousse, 2008). In this sense he postulates that “simplified” thought, in which a single discipline or a specific way of approaching the reality owns an absolute priority over the others, is what truly makes world analysis more difficult, because of its insufficiency for approaching all the interconnections behind the problems. For example, the neoliberal assumption of the supremacy of the market and the overutilization of economical variables on policy making and evaluation. Thus, it’s necessary to act on the basis that knowledge is open, unfinished and in continued reconstruction, which leads to the conclusion that the decision making, product of that complex knowledge, also has to be interconnected, open and in perpetual reconstruction, a process in which technology and science play a central role. Complexity paradigm establishes real interrelations and intercommunications among disciplines, specialists, methodologies, realities and specific languages.

The austrian physicist and philosopher, Paul Feyerabend (2004), expressed “a free society is a society in which all traditions have equal rights and equal access to the centers of power [...] is not only fair, but the most convenient”. In consequence, comprehending reality as a complex system, implicates the necessity of multidisciplinary and interdisciplinarity for the acquisition of knowledge, therefore its relevance for the decision making process, being a central value on societies democratization. The dispute and the construction of the public need to be released on articulated spaces, that permits their complementation and management of their dissents.

#### **4. On the use of technology in an open, popular and participatory democracy**

A technology is a tool, a dispositive designed to achieve a purpose, a means to fulfill a goal. Anthropologist Roger Bartra (2014) postulates that the human being is a “social-technified” being, that is a species, on the basis of social contract, capable of creating technology and recreating itself with the use of that technology. The use of tools for the solution of our necessities is something natural: from instruments for fire generation, the register and articulation of a specific language, to the creation of the wheel, the banking models, the pencil, the hydroelectric plants and the quantum computers. All designed for solving specific problems and in which process of implementation have changed the way people relate to each other and to the environment. In this sense, thinking about technology beyond its contemporary reality, focused on its digital possibilities, allows us to understand it critically, as well as to understand it as a substantial factor of the social being of our communities.

The austrian theologian, biologist and philosopher Ivan Illich pointed out that there is a process that represents a danger in the utilization of tools, that of “counterproductivity”, “the fact that a tool, when it exceeds a certain intensity, inevitably leads more people away from the purpose for which it was created than allows them to take advantage of its benefits” (Beck, 2016), that is, it goes from being a means to becoming the goal of the activity. As a case in point of everyday life is the use of video games, which supposedly are used for entertaining and enjoyment of leisure time, to share quality time with loved ones or even professionally as a way of working to obtain the economic resources that allow a better life. The process of counterproductivity occurs, for example, when the use of video games interpose with the time of fundamental activities, also when video games disconnect relationships and, in the labour case, when it produces an overexploitation of working days.

Comprehending technology as a substantial part of human beings and their societies allow us to realize the repercussions of counterproductivity on major scales. The neoliberalization process impulsed the counterproductivity phenomena on technology, where the idea of community and the importance of it in the evolution of societies got debilitated, passing to a paradigm, a common sense, a subjectivity where technologies are product of exceptional minds in moments of brilliance to satisfied individual demands from whom can afford that specific product in a free market economy. If



technologies have a role in government during neoliberalism, it's under the parameters of market models: cost-efficiency decisions negotiations between technocracies and economic structures.

Additionally, in the face of this scenario, documentaries like "The Great Hack" (2019), produced and directed by Jehane Noujaim and Karim Amer, depict the case of Cambridge Analytica as an other facet of the use of technology in the political-electoral process in modern societies. This company, a provider of data-driven political communication strategies, played a crucial role in Donald Trump's presidential campaign and the Leave.EU campaign during the Brexit referendum. These examples show the practice of data exploitation of unauthorized collection of vast amounts of personal data from Facebook and other social media users. Cambridge Analytica exploited this data to create detailed voter profiles, enabling them to craft highly targeted and personalized political messages via microtargeting and psychological manipulation; exhibiting the importance of Technology Giants Corporations role in the data ecosystem, shaping public opinion, and influencing political landscapes. These are presented as paradigmatic examples of the discretionary use of Big Data and the role of technological transnational corporations in shaping public opinion as a product and not using it to empower popular participation to solve public problems.

To get out of the crisis of counterproductivity, Illich proposed another concept related to the use of technologies: the "conviviality" making the tool being at the service of the community, coming back as means. In civilizations like ours it's necessary a new goal: that the means to return to being just that, a means. High technological progress, contemporary crises and the complexity of society demand it or we will continue to be surpassed.

Humberto Beck (2020), raised the question: "could the Marx of the future be, perhaps, a programmer?". Beyond the disputes around the figure of the prussian philosopher, the important question is to ask ourselves if the next great criticisms and mobilizations against the status quo will be carried out by people who handle Big Data, Artificial Intelligence and the programming languages. The Marx of the future needs to know programming or to have the facility to relate with programming communities, but also, the future dictator could be a programmer. Technology, it's only a tool in front of a reality as complex as technified. The importance of the implementation of these technologies is urgent, just as is the politicization and democratization of high-qualified persons and processes for their contribution to the common good.

Starting from educational institutions: let's imagine that in engineering they also give courses on critical coding and datafication, digital mobilization, algorithms and gender; or that those who study social sciences have courses where the Internet as a political space, programming as a fundamental tools in the public policies cycles or digital rights as human rights are addressed. Let's imagine, even better, that we can strengthen this knowledge with those who do not have the option of studying in a formal institution, creating community Internet networks, generating communal workshops on taking care of our information on social networks or public programming courses. Pluralize knowledge and politicize tools for the improvement of daily life, as well as the deepening of our democracies.

Moreover, we must establish democratic institutions and organizations with control mechanisms to manage the vast volumes of data and the technological progress surrounding this phenomenon. Digital training for individuals in public office and the inclusion of technical experts in the field should be a necessary action to accompany this process. The significance of datafication and digitization should be cut across all levels of government. Noteworthy initiatives, such as the Digital Agency of Mexico City, could serve as models for creating similar bodies nationwide, along with open data portals that are gaining traction, particularly due to the COVID-19 pandemic. Furthermore, providing economic, human, and technical resources to electoral institutions is essential for expanding digital monitoring, moving beyond mere discussion to identifying bots, trolling, fake news, astroturfing, and enhancing scrutiny systems, as well as electoral counting.

Efforts should aim to extend these legislations and organizations globally, given that major technology corporations transcend national borders. Consequently, resolving this issue entails a paradigm shift in human rights and cooperation. Promoting educational and civic processes beyond governmental institutions is crucial to fostering a societal culture where the importance of the individual's relationship with society, technology and information is understood.

## **5. Mechanism for popular participation, a brief route from local realities**

The predilect space to prone innovation and experimentation of tools for the resolution of social problems are the municipalities. The local government is the structure of public power more accessible to the people, because it's related to their immediate reality and facilitates the process of popular participation. In a democracy located in a complex reality, participatory mechanisms allow the different knowledge and interests to be shared, also they allow conflict management and the appropriation of solutions exercised, in part, by governments. This does not imply that the simple use of any participation mechanism generates a public benefit.

It is necessary to think about its specific purposes, its temporalities, its relations with the different stages of the processes of public policy and legislation cycles, carrying out a pedagogy about its importance and its use, encouraging broad and binding participatory budgets or advisory bodies. There is no democracy without constant deliberation, limiting

participation to one vote in a referendum or a recall is the same as limiting democracy to the election of popular representatives. In this sense, technology has an opportunity to facilitate all of these processes beyond the increase in voting mechanisms -such as electronic ballot boxes or e-voting- in a greater range of government decisions.

There are some cases that allow people to imagine new horizons, either for their successes or for their failures, around the world. Barcelona en Comú, the ruling political party in the Catalan capital from 2015 to 2023, made its Municipal Action Plan, the strategic document that set the guidelines for local government, collaborative. This participatory process “has lasted more than two months, has combined almost 400 face-to-face debates held in different parts of the city with online participation through the Decidim.barcelona website, where residents have contributed their ideas” (Llort, 2016). With the implementation of Decidim within the management ecosystem of the Barcelona City Council, a vision of the use of technology in participation mechanisms has been introduced, not as a substitute for territorial activities and work groups, but as facilitators and complements, since it has inputs for collaborative mapping, preparation of work groups, discussion forums, collaborative writing, voting mechanisms, activity registration and delivery of educational materials. Functions that require accompaniment, instruction and learning process, without replacing the dynamics of participation in the communities.

Decidim defines itself as “a public, free and open digital infrastructure for participatory democracy” (Decidim, s.f.). The term “digital infrastructure” refers to a set of tools, data sets, documents, codes and interfaces accessible by digital means (Decidim, s.f.). Decidim is a medium that enables participatory democracy, it does not replace it, nor does it make it better, much less solve the public problems of its community, but it provides a multipurpose, adaptable, transparent and accessible tool to empower people. There is a commitment to conviviality, not to the romanticization or alienation of the tool from its social context. After its success, cities such as Rosario, Helsinki, Zurich, New York, Milan, Kakogawa and Mexico City have joined the use of this digital infrastructure.

There are some other examples of local government initiatives. The Open University of Recoleta (OUR) (Universidad Abierta de Recoleta, UAR) which focuses on giving knowledge a practical use in everyday life to the inhabitants of the communities. The institution does not give formal grades, nonetheless it offers free courses, workshops and seminars in charge of academics of the best universities in Chile and social leaderships specialized in technical areas, prying the possibility of democratizing knowledge in diverse areas from quantum physics, passing through philosophy and economy, to practical local gastronomy (Ramírez, 2018). The OUR gives courses at municipal educational centers in vespertine and matutine hours all days of the week, also in a modality at a digital platform. After their year it had more than 11 thousand students all over the country.

La Comunicadora is a public bidding and procurement system of Barcelona's Ajuntament that prioritizes cooperatives and social economy companies, reprioritizing contractual relationships of the government, moving from relationships with Microsoft and Cisco to software and open source companies (Arboleda, 2021). An additional case is Gewobag, the public real estate company of Berlin State, remunicipalized six thousand houses and invested 920 millions euros to establish caps on rental prices based on public referendums and collaborative spaces to fight housing speculation. Furthermore, Gewobag generates more than 10,000 megawatt-hours of solar energy and passes the savings on to its tenants, linking social justice with climate justice (Arboleda, 2021).

Other paradigmatic cases, although with less success, are those implemented in the last decade in some political parties around the world. The best known is the case of the 5 Star Movement, an Italian political party, which, beyond being able to agree or not with its ideology, has incorporated an innovative online tool within its party: “Rousseau”. The platform carried out continuous consultations and discussion forums for more than 115,000 members until 2021, although approximately half did so actively, about the decisions of their political organization (Bertacche & Follain, 2019). Parties that followed the path of including consultation and decision tools are also the Partido de la Red in Argentina and the Piratpartiet (Swedish Pirate Party) who have formed international cooperation networks, which although they have not been the most successful in the electoral contest, are parties that have managed to position issues related to digital rights, participatory budgets, e-government, open data and direct democracy.

Social movements such as the Digital 3D Replication Movement led by Adrian Bowyer emerged with the purpose of designing low price 3D printers for communal design and the replication of tools, furniture and daily life components. Precedents such as Bioleft, Farm Hack y Slow Tools, produce agricultural supplies and spare tractor parts showing the potentialities in rural areas of open code and digitalization, finding common ground areas between engineers, designers, architects, farmers and programmers based on the paradigm that agricultural interest are a collective effort (Arboleda, 2021).

Urban gardens, community skate parks, self-managed Wi-Fi networks and alternative energy microstations demonstrate the new sociotechnical relationships that are imprinted on the territory throughout the world, facilitated by municipalist movements, generating new concepts to define itself as platform municipalism (Arboleda, 2021). The Fearless Cities network has sought to articulate this movement as an alternative to the various international crises we are experiencing. Faced with distant multinational organizations with little capacity for impact, flexible and fast networks of cities, although limited, have a transformative will.

This experience served to transfer, as well as learn, knowledge from Barcelona and Berlin to the Kurdish communities in the Middle East, with ecological and community experiments in France and Italy, with autonomous movements in Hong Kong, and with labor struggles in New York. Across the planet, the municipalist movement has progressed. In Latin America, cases like Recoleta have shared knowledge with Belo Horizonte, Valparaíso, Santiago, and Rosario in Argentina. Seeking to ensure that regional knowledge and particularities do not hinder the democratizing potential of technology use with popular participation, Fearless Cities has seen the design of its local initiatives under the paradigm of scaling devices aiming for replicability and adaptation accompanied by the groups that are part of the organization.

These examples show the global need for close and innovative governments, as well as political forces and public institutions with the capacity to adjust to the information obtained, as well as to changing circumstances, with the capacity to generate consensus and strategic projects among various actors to generate welfare and public goods. In the words of the economist Mariana Mazzucato (2021), professor at University College London, Entrepreneurial States are required, that is, guiding agents who articulate and direct an environment of social and economic innovation to produce social and environmental well-being. In her book, *The Entrepreneurial State* (2021) she takes up cases such as the role of the State in the research and development processes of the iPhone, NASA, clean energy and the internet, seeking with her text to encourage the government role in the economy and the possibility of promoting the innovations of society in the process. In other words, the State generates the ecosystem of creation and redistribution for tools which produce public-social innovation and common benefit of the society.

In a context like that of Jalisco, Mexico and the Latin American region is important to acknowledge the digital gap, which despite having been reducing, still has been an important factor in the process of digitalization and the use of technology in political activities, not only focusing on the access to infrastructure, but also, even more important, the appropriation of tools and competences by the people, as well as the adaptation to local necessities of the communities in which they are implemented. Lay Arellano (2023) argues that even the measures implemented by federal and state governments in Mexico to expand the coverage and quality of infrastructure, there are still important deficiencies in the continuity of its quality monitoring, not permitting the correct appropriation and use of communities.

Broad, open and popular governments that take advantage of technology for the sake of collective dignity require, as the first steps, basic technology, the critical mass of communication networks, open data portals and transparent vinculation with technology already present in the life of the person. Then applications are needed, where the information is transferred to concrete actions for common use and decision making. The latest is the public use of such applications as complementary tools for the improvement, design and implementation of social projects (McKinsey Global Institute, 2018).

It is necessary to open dialogue and spaces, so that public policies and social projects have a more complete vision, that goes beyond “technical” or consultative and restrictive tools for popular participation, but rather takes into account other branches of knowledge, differentiated realities and trust in people’s capacity to understand, as well as to solve their common problems. Technology requires to be democratized and democracy needs to be technified under its own logic: the plural dialogue between praxis and reason for the various ways of living. This is a broad, deep, innovative and dignified democracy.

## **6. Conclusions**

Through this essay it has been argued that representative democracies have multiple limitations to favor the resolution of the various contemporary crises and their complexity, especially after the neoliberalization process in Western societies, as well by the paradigm of the New Public Management in governments. The need for an interdisciplinary, participatory, and technological approach in solving public problems has been argued, primarily in local governments due to its accessibility and potential for the implementation of such tools, closeness to people, and relevance in the global problem-solving chain.

Technology as a means capable of channeling and potentiating the ideas, feelings and actions of a community as a whole, accompanied by technological processes and territorial participation mechanisms, open up new horizons of possibility in the governmental and collective imagination. Reclaiming the means, through conviviality, a revolution is accomplished: the revolution of the common sense.

## 7. References

- Arboleda, M. (2021) *Gobernar la utopía: sobre la planificación y el poder popular*. Caja Negra
- Aguilar Villanueva, L. (2010). *Gobernanza: el nuevo proceso de gobernar*. Fundación Friedrich Naumann para la Libertad.
- Amer, K. & Noujaim, J. (Film producers). (2019). *The Great Hack* [documentary]. United States of America: Netflix.
- Barberousse, P. (2008). Fundamentos teóricos del pensamiento complejo de Edgar Morin. *Revista Electrónica Educare*, 12(2), 95–113.
- Bartra, R. (2014). *Antropología del cerebro: Conciencia, cultura y libre albedrío* (2.a ed.). Fondo de Cultura Económica.
- Beck, H. (2016). *Otra modernidad es posible: El pensamiento de Iván Illich*. Mal Paso Editorial.
- Beck, H. (2020, 2 de enero). Dos décadas del nuevo siglo: breves apuntes para un balance. *Otros Diálogos*. [Enlace] <https://otrosdialogos.colmex.mx/dos-decadas-del-nuevo-siglo-breves-apuntes-para-un-balance>
- Bertacche, M., & Follain, J. (2019, 2 de septiembre). Italy's New Coalition Faces Hurdle in Five Star Direct Democracy. Recuperado el 7 de noviembre, 2019, de [Enlace] <https://www.bloomberg.com/tosv2.html?vid=&uuid=61880410-0165-11ea-b56f-d588bdd9119a&url=L25ld3MvYXJ0aWNsZXRMvMjAxOS0wOS0wMy9kaXJlY3QtZGVtb2NyYWN5LWNvdWxkLWJlLWxhc3QtaHVyZGxILWZvci1pdGFseS1zLW5ldy1nb3Zlc>
- Decidim. (s.f.). *Decidim: A brief overview*. Decidim Docs. [Enlace] <https://docs.decidim.org/en/develop/whitepaper/decidim-a-brief-overview.html>
- Errejón Galván, Í. (2011). ¿Qué es el análisis político? Una propuesta desde la teoría del discurso y la hegemonía. *Revista Estudiantil Latinoamericana de Ciencias Sociales*. 10.18504/ri0101-2011-2012.
- Feyerabend, P. (2004). *La Ciencia en una sociedad libre*. Siglo XXI.
- Foucault, M. (1984). *La verdad y las formas jurídicas*. Gedisa.
- Gramsci, A. (2009). *La política y el Estado moderno*. Diario Público.
- Kuhn, T. S. (2019). *La estructura de las revoluciones científicas* (4.a ed.). Fondo de Cultura Económica.
- Morín, E. (1990). *Introducción al pensamiento complejo*. Recuperado de [Enlace] [http://cursoenlineasincostoedgarmorin.org/images/descargables/Morin\\_Introduccion\\_al\\_pensamiento\\_complejo.pdf](http://cursoenlineasincostoedgarmorin.org/images/descargables/Morin_Introduccion_al_pensamiento_complejo.pdf)
- Mazzucato, M. (2021). *The Entrepreneurial State*. Penguin.
- McKinsey Global Institute. (2018). *Smart Cities: Digital solutions for a more livable future*.
- Lay Arellano, I. T. (2023). *La red de internet gratuito del gobierno del estado de Jalisco*. ETIUS. Recuperado 10 de enero de 2024, de <https://informedemedios.iteso.mx/2023/03/29/la-red-de-internet-gratuito-del-gobierno-del-estado-de-jalisco/>
- Llort, M. (2016, 27 de abril). *La nueva política ensaya su modelo de democracia en Barcelona*. Recuperado el 7 de noviembre, 2019, de [Enlace] <https://ctxt.es/es/20160427/Politica/5657/Barcelona-Colau-democracia-participativa-Subirats.htm>
- Ramírez, N. (2018, 26 noviembre). *Universidad abierta: Cómo funcionará el nuevo proyecto popular de Recoleta*. Emol. <https://www.emol.com/noticias/Nacional/2018/11/26/928708/Universidad-Abierta-Como-funcionara-el-nuevo-proyecto-popular-de-Recoleta.html>

# Más allá de la protección de datos personales: ¿Y la reparación del daño?

Rafael Ríos Nuño

Consejo de la Judicatura del Poder Judicial del Estado de Jalisco, Guadalajara, México,  
rafael.rios.nuno@gmail.com

**Resumen.** En la legislación mexicana, ha sido comúnmente aceptado que las víctimas del delito o de probables violaciones a sus derechos humanos, tienen el derecho a que se le repare el daño. Constitucionalmente, ha sido reconocido el derecho a la protección de datos personales como un derecho humano fundamental. Sin embargo, aún no se cuentan con mecanismos claros para garantizar dicha reparación.

Las causas a la posible problemática, pueden ser varias, tal vez se deba a la novedosa creación de la ley especializada en materia de víctimas y a las comisiones que derivan de ésta; o bien, al reciente reconocimiento del derecho a la protección de datos personales que data, apenas, de 2009. Sea cual sea la razón, en el presente artículo de investigación, se busca hacer un análisis breve del derecho a la protección de datos personales, sus orígenes, así como del derecho de daños, para intentar justificar la reparación.

**Palabras clave.** *Derechos humanos, protección de datos personales, derecho de daños, víctima, reparación del daño.*

## 1. El derecho humano a la privacidad y a la protección de datos personales

La Suprema Corte de Justicia de la Nación (en adelante SCJN), en su labor de interpretar la ley, ha elaborado una serie de criterios respecto al derecho a la vida privada.

La SCJN ha dicho que lo privado, es lo que no constituye vida pública; el ámbito reservado frente a la acción y el conocimiento de los demás; lo que se desea compartir únicamente con aquellos que uno elige; o aquello que las personas no desempeñan con el carácter de servidores públicos.

En el derecho interno mexicano, la privacidad se encuentra reconocida implícitamente en el artículo 6° de la Constitución Política de los Estados Unidos Mexicanos (CPEUM), como un límite al derecho a la información, así como en los artículos 7, 14 y 16 (De la Parra, 2014); además, de conformidad con el artículo 1 y 133 constitucionales, en relación con la Convención de Viena sobre el Derecho de los Tratados; el citado derecho, se encuentra estipulado en la Declaración Universal de los Derechos Humanos (en adelante DUDH) en su artículo 12, en el Pacto Internacional de Derechos Civiles y Políticos (en adelante PIDCP) en su artículo 17, en la Convención Americana sobre Derechos Humanos (en adelante CADH) en su artículo 11 y en la Convención sobre los Derechos del Niño de 1989 en su numeral 16.

Los citados instrumentos legales, han sido igualmente interpretados por la Corte Interamericana de Derechos Humanos (en adelante Corte IDH) y la SCJN, quienes han destacado la vinculación de la privacidad con un amplio abanico de otros derechos, como la inviolabilidad de la correspondencia y de las comunicaciones en general, la inviolabilidad del domicilio, las garantías respecto de los registros personales y corporales, las relaciones con la recopilación y registro de información personal en bancos de datos y otros dispositivos; el derecho a una vivienda adecuada, a la salud, y a la igualdad; los derechos reproductivos, o la protección en caso de desalojos forzados.

Por lo que, tanto la SCJN y la Corte IDH, sostienen que las personas tienen derecho a gozar de un ámbito de protección de su existencia que quede reservado de la invasión y la mirada de las demás, que les concierna solo a ellas y les provea de condiciones adecuadas para el despliegue de su individualidad –para el desarrollo de su autonomía y libertad–. La misma idea puede describirse apelando al derecho de las personas a mantener fuera del conocimiento de los demás ciertas manifestaciones o dimensiones de su existencia que los demás no las invadan sin su consentimiento. Por lo tanto, la vida privada implica poder conducir parte de la vida de uno, protegido de la mirada y las injerencias de los demás, y guardar conexiones de variado tipo con pretensiones más concretas que los textos constitucionales actuales reconocen a veces como derechos conexos: el derecho de poder tomar libremente ciertas decisiones atinentes al propio plan de vida, el derecho a ver protegidas ciertas manifestaciones de integridad física y moral, el derecho al honor o reputación, el derecho a no ser presentado bajo una falsa apariencia, el derecho a impedir la divulgación de ciertos hechos o la publicación no autorizada de cierto tipo de fotografías, la protección contra el espionaje, la protección contra el uso abusivo de las comunicaciones privadas, o la protección contra la divulgación de informaciones comunicadas o recibidas confidencialmente por una persona particular.

Ahora, por lo que respecta a la protección de datos personales, tema central de este artículo de investigación, es sabido que éste evolucionó del derecho a la privacidad, y en algunas ocasiones se llegó a considerar una extensión de este último. Sin embargo, las opiniones de las personas especialistas<sup>1</sup>, coinciden que, en la actualidad, el derecho humano a la protección de datos personales es un derecho humano autónomo, y, que, debido, a la interdependencia de los derechos humanos, la protección de datos personales puede seguir manteniendo nexos con la privacidad, la intimidad, el honor o la propia imagen, por mencionar algunos ejemplos.

<sup>1</sup> Eduardo de la Parra, Isabel Davara, Diego García Ricci, Joel Gómez Treviño.

Uno de los países pioneros en el tema de privacidad y protección de datos personales, es Estados Unidos, quien tomó como primer elemento la libertad a la esfera íntima, luego por extensión la jurisprudencia fue reconociendo otros derechos como en los casos de Mayer vs. Nebraska y las leyes que prohibían la enseñanza antes del noveno grado de otra lengua que no fuera el inglés; más tarde a través de otros casos relevantes en materia de sexualidad, como la prohibición de anticonceptivos, esterilizaciones a criminales, así como el de la creación de un registro computarizado y centralizado para mantener bases de datos de pacientes y sus enfermedades que permitían identificar al paciente (Gregorio, 2004, p. 304).

Alemania, por su parte, fue una de las primeras en reconocer la autodeterminación informativa, traducido en el derecho a controlar la información de la persona y la capacidad de determinar si esa información puede ser recogida y de cómo debe hacerse (Martínez, 2007, p. 48).

En virtud de lo anterior, Cuervo (2014), en su obra «Autodeterminación informativa», asevera que, de la resolución del Tribunal Constitucional de Alemania, este reconoció el derecho a la autodeterminación informativa como un derecho derivado de la personalidad.

La mencionada resolución, tiene su antecedente con el grupo de «los verdes» y la Ley de Censo, donde, en un primer momento, se obtuvo una suspensión provisional en tanto se resolvía el fondo del asunto: en el cual se reconoce este nuevo derecho, pero al mismo tiempo también le impone límites en el marco del interés general y bajo un supuesto establecido en la Constitución (Martínez, 2007, pp. 48 a 51).

Ahora bien, España, hizo lo propio, en la sentencia del Tribunal Constitucional 53/1985<sup>2</sup> se reconoce expresamente que el derecho a la intimidad informática, dejó su estatus pasivo para convertirse en un estatus activo, el cual faculta al titular la posibilidad de ejercer un control sobre el caudal de la información, mismo que se traduce en el constitucionalismo actual del Estado social de Derecho (Cuervo, 2014).

El desarrollo histórico de la protección de datos personales en México comenzó en 2009, cuando se elevó a rango constitucional el citado derecho en el artículo 6º, A., fracción II y el 16, segundo párrafo; en el año 2010 se promulgó la Ley Federal de Protección de Datos Personales en Posesión de Particulares y en 2011 su Reglamento; finalmente en 2017 se aprobó la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados, e inmediatamente después surgieron sus homologas en las demás entidades federativas.

Ahora bien, la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados, en su artículo 3º, fracciones IX y X, señala:

*Artículo 3. Para los efectos de la presente Ley se entenderá por:*

*IX. Datos personales: Cualquier información concerniente a una persona física identificada o identificable. Se considera que una persona es identificable cuando su identidad pueda determinarse directa o indirectamente a través de cualquier información;*

*X. Datos personales sensibles: Aquellos que se refieran a la esfera más íntima de su titular, o cuya utilización indebida pueda dar origen a discriminación o conlleve un riesgo grave para éste. De manera enunciativa más no limitativa, se consideran sensibles los datos personales que puedan revelar aspectos como origen racial o étnico, estado de salud presente o futuro, información genética, creencias religiosas, filosóficas y morales, opiniones políticas y preferencia sexual;*

Por su parte, la SCJN sostiene que el derecho a la protección de datos personales y la autodeterminación informativa, es:

*«(...) el derecho a poseer la intimidad a efecto de disponer del control sobre la publicidad de la información tanto de la persona como de su familia; lo que se traduce en el derecho de la autodeterminación de la información que supone la posibilidad de elegir qué información de la esfera privada de la persona puede ser conocida o cuál debe permanecer en secreto, así como designar quién y bajo qué condiciones puede utilizar esa información. En este contexto, el derecho a la intimidad impone a los poderes públicos, como a los particulares, diversas obligaciones, a saber: no difundir información de carácter personal entre los que se encuentran los datos personales, confidenciales, el secreto bancario e industrial y en general en no entrometerse en la vida privada de las personas; asimismo, el Estado a través de sus órganos debe adoptar todas las medidas tendentes a hacer efectiva la protección de este derecho»*

*(Tesis: I.3o.C.695 C.).*

Martínez (2007, p. 51), a su vez, define el derecho a la protección de datos personales como aquel que “garantiza la facultad del individuo de decidir básicamente por sí mismo sobre la difusión y la utilización de sus datos personal”.

En esa sintonía Gregorio (2004, p. 303), de su obra titulada Protección de Datos Personales: Europa vs Estados Unidos, todo un dilema para América Latina, define los conceptos de público y privado, limitando al primero como la necesidad de dejar determinado documento accesible al público con la finalidad del control de la ciudadanía de los actos de gobierno; el último, lo describe como toda decisión personal en la que el Estado no puede intervenir.

<sup>2</sup> Para más información consulte la sentencia en la siguiente liga: <http://hj.tribunalconstitucional.es/ca/Resolucion/Show/433>

Escalante (2004, p. 186), abunda a la conceptualización, afirmando que la estructura constitucional depende de la existencia de dos ámbitos claramente distintos, uno privado y otro público, que no se explican ni justifican, salvo donde hay excepciones. El límite que se refiere a la privacidad, se justifica debido a la dignidad humana que implica el poder elegir en libertad; dicho de otra manera, se trata de elegir el propio plan de vida.

Garzón (1998, p. 227), define la privacidad como el ámbito donde imperan exclusivamente los deseos y preferencias individuales. Es condición necesaria del ejercicio de la libertad individual. Por otro lado, advierte que lo público está caracterizado por la libre accesibilidad de los comportamientos y decisiones de las personas en sociedad. Así pues, tenemos que, si lo íntimo está caracterizado por su total opacidad, lo que caracteriza a lo público es la transparencia (Garzón, 1998, p. 226).

En otro orden de ideas, Martínez (2007, p. 55) advierte que, para que el derecho a la protección de datos personales ceda frente al derecho de acceso a la información, es importante tomar en consideración que la información sea relevante o de interés público y, por tanto, contribuya al debate público o a la formación de una opinión pública libre. También será relevante cuando la persona objeto de la noticia posea un carácter público y la noticia ilustre sobre algún aspecto relevante de ese perfil.

## 2. El origen y breve introducción al derecho a la reparación integral del daño

Es un principio de derecho que toda violación a una obligación que haya producido un daño, se traduce en un deber de repararlo adecuadamente<sup>3</sup>, el citado principio, está consagrado en la CADH, en su artículo 63.1. Asimismo, en el derecho moderno, muchos Estados, al igual que, en los instrumentos creados por los organismos internacionales, prevén la reparación del daño, no solo por parte de los particulares, sino del propio Estado o de quienes fungen como titulares del poder.

El derecho a la reparación del daño comprende dos premisas fundamentales:

1. Toda aquella persona que cause un daño a otra, tiene la obligación de repararlo; y
2. Toda aquella persona que haya sido víctima, o ha sufrido un daño, tiene el derecho a que se le repare.

En relación con lo anterior, surgen los conocidos Principios de Joinet, presentados en el informe de 1997 del Relator Especial de la Comisión de Derechos Humanos de la Organización de las Naciones Unidas (en adelante ONU). Mismos que se enfocaban en la impunidad, Louis Joinet, estableció como derechos elementales de cualquier víctima de abuso de poder, los siguientes:

1. *El derecho a saber: es la prerrogativa inalienable de la víctima a conocer la verdad sobre las violaciones de derechos humanos ocurridas.*
2. *El derecho a la justicia: consiste en que se integre y resuelva, por parte de un tribunal o instancia competente, sobre los derechos que se han vulnerado, las personas agentes o servidoras públicas involucradas y las sanciones que correspondan a las mismas; y*
3. *El derecho a obtener reparación: contiene principios sobre los procedimientos de reparación y el ámbito de aplicación de este derecho, así como las garantías para la no repetición de las violaciones.*

En el año 2000, el Relator Especial sobre el Derecho de Restitución, Indemnización y Rehabilitación de las Víctimas de Violaciones de los Derechos Humanos y las Libertades Fundamentales, M. Cherif Bassiouni, presentó un informe final ante la Comisión de Derechos Humanos de la ONU, y adjuntó a su informe una propuesta de Principios y Directrices Básicos sobre el Derecho de las Víctimas de Violaciones de Derechos Humanos y del Derecho Internacional Humanitario a Interponer un Recurso y a Recibir Reparación (conocidos como Principios van Boven-Bassiouni). En dichos principios se reconocen como formas de reparación: la restitución, indemnización, rehabilitación, satisfacción y garantías de no repetición.

Resulta importante mencionar que, como medida de restitución para daños que resultan irreparables, la Corte IDH, ha establecido la indemnización con el pago de una reparación que aspire a ser integral para las personas ofendidas y víctimas de una violación. La autoridad violadora, como representante de las personas y garante de la seguridad de sus habitantes en el Estado, de manera proporcional al daño que ha provocado el acto perpetrado por una de sus personas agentes o servidoras públicas, debe retribuir a la persona ofendida, en numerario, el derecho violado. Asimismo, empleará los medios a su disposición para que el hecho no se repita. Dicha acción, conocida como la «garantía de no repetición», implica previamente el reconocimiento público de la violación de un derecho.

La reparación del daño es un mecanismo reconocido por el derecho internacional para enfrentar la impunidad en la violación de los derechos humanos; es la justa reparación. La adecuada reparación del daño, según los criterios fijados por la Corte IDH y otros organismos internacionales, debe incluir:

**Daño emergente:** afectación al patrimonio, derivada inmediata y directamente de los hechos. En la legislación

3 Jurisprudencia de la Corte Interamericana de Derechos Humanos, cuyo sustento jurídico se encuentra previsto en el artículo 63.1 de la CADH, ratificada por nuestro Estado Mexicano, el 24 de marzo de 1981, y por lo tanto, es ley suprema para México. Dicho principio es invocado en numerosas jurisprudencias. Como ejemplo se cita: Caso Yvon Neptune vs Haití, sentencia del 6 mayo de 2008.

mexicana suele equipararse el daño en sentido amplio.

Lucro cesante: implica la ganancia o el beneficio que se dejó o dejará de percibir como consecuencia del hecho que causó el daño. En la legislación mexicana equivale al perjuicio.

Daño físico: es la lesión que sufre la persona en su cuerpo, evaluada en este caso de acuerdo con la tarifa establecida por la legislación aplicable para el caso de la pérdida de una vida humana.

Daño inmaterial: es la lesión sufrida en los bienes no materiales que forman parte del patrimonio de las personas. Puede consistir en un daño jurídico, en un daño moral, en un daño al proyecto de vida o en un daño social, por los vínculos afectivos y sociales que sostenía con sus seres queridos.

Hasta el momento, se puede identificar los siguientes aspectos:

*1. Daño jurídico: es la lesión que sufren las personas en sus derechos. Este daño se ve especialmente reflejado por la violación de las garantías individuales y sociales previstas en la Constitución y en los instrumentos internacionales sobre la materia.*

*2. Daño moral: la lesión sufrida en el aspecto psíquico de la persona, más precisamente, en el emocional. Puede tener una dimensión individual o social. Tal afectación se ve reflejada en los sentimientos de impotencia y de susceptibilidad que tienen las personas habitantes frente a las autoridades encargadas de velar por sus derechos.*

*3. Daño al proyecto de vida: es el que afecta la realización de la persona que ha sido víctima de la violación, considerando su vocación, aptitudes, circunstancias, potencialidades y aspiraciones que le permitían fijarse razonablemente expectativas determinadas y cumplirlas. Implica, por tanto, la pérdida grave o menoscabo de oportunidades futuras de desarrollo personal.*

*4. Daño social: es el que se provoca en quienes pertenecen a la comunidad y entorno en que se perpetró el hecho, debido a que la muerte de una persona en la que tuvo intervención, ya sea por acción o por omisión, alguna autoridad o servidor público, trasciende a quien o quienes sufrieron la afectación, impacta en la sociedad y representa un precedente que implica la pérdida de confianza y de consolidación de los encargados de velar por el cumplimiento de una función pública fundamental como la seguridad y la aplicación de justicia.*

Para garantizar el cumplimiento de cada uno de estos elementos, la Corte IDH, ha establecido, entre otras, las siguientes medidas para restituir a las personas ofendidas en el ejercicio de sus derechos:

*1. Gastos y costas: constituyen los pagos que se originen por los daños que se hayan causado tanto a la víctima como a las personas ofendidas.*

*2. Medidas de satisfacción y garantía de no repetición: acciones que efectúa el Estado para modificar prácticas administrativas o mecanismos de protección inadecuados.*

*3. Medidas preventivas: medidas tomadas como respuesta a un incidente, para prevenir, minimizar mitigar pérdidas o daños a las personas gobernadas.*

*4. Determinación y reconocimiento de responsabilidad: el objetivo es que exista la aceptación del Estado de la falta que hayan cometido sus autoridades o personas servidoras públicas. Es una medida significativa de satisfacción para las víctimas por los daños morales sufridos.*

*Al respecto, en el derecho interno mexicano, la Ley General de Víctimas señala:*

*Artículo 1. [...]*

*La reparación integral comprende las medidas de restitución, rehabilitación, compensación, satisfacción y garantías de no repetición, en sus dimensiones individual, colectiva, material, moral y simbólica. Cada una de estas medidas será implementada a favor de la víctima teniendo en cuenta la gravedad y magnitud del hecho victimizante cometido o la gravedad y magnitud de la violación de sus derechos, así como las circunstancias y características del hecho victimizante.*

*Artículo 7. Los derechos de las víctimas que prevé la presente Ley son de carácter enunciativo y deberán ser interpretados de conformidad con lo dispuesto en la Constitución, los tratados y las leyes aplicables en materia de atención a víctimas, favoreciendo en todo tiempo la protección más amplia de sus derechos.*

*Las víctimas tendrán, entre otros, los siguientes derechos:*

*I. A una investigación pronta y eficaz que lleve, en su caso, a la identificación y enjuiciamiento de los responsables de violaciones al Derecho Internacional de los derechos humanos, y a su reparación integral;*

*II. A ser reparadas por el Estado de manera integral, adecuada, diferenciada, transformadora y efectiva por el daño o menoscabo que han sufrido en sus derechos como consecuencia de violaciones a derechos humanos y por los daños que esas violaciones les causaron;*

*[...]*

*VII. A la verdad, a la justicia y a la reparación integral a través de recursos y procedimientos accesibles, apropiados, suficientes, rápidos y eficaces;*

*VIII. A la protección del Estado, incluido el bienestar físico y psicológico y la seguridad del entorno con respeto a la dignidad y privacidad de la víctima, con independencia de que se encuentren dentro un procedimiento penal o de cualquier otra índole. Lo anterior incluye el derecho a la protección de su intimidad contra injerencias ilegítimas, así como derecho a contar con medidas de protección eficaces cuando su vida o integridad personal o libertad personal sean amenazadas o se hallen en riesgo en razón de su condición de víctima y/o del ejercicio de sus derechos;*

*[...]*



*Artículo 65. Todas las víctimas de violaciones a los derechos humanos serán compensadas, en los términos y montos que determine la resolución que emita en su caso:*

- a) Un órgano jurisdiccional nacional;*
- b) Un órgano jurisdiccional internacional o reconocido por los Tratados Internacionales ratificados por México;*
- c) Un organismo público de protección de los derechos humanos;*
- d) Un organismo internacional de protección de los derechos humanos reconocido por los Tratados Internacionales ratificados por México, cuando su resolución no sea susceptible de ser sometida a la consideración de un órgano jurisdiccional internacional previsto en el mismo tratado en el que se encuentre contemplado el organismo en cuestión.*

*Lo anterior sin perjuicio de las responsabilidades civiles, penales y administrativas que los mismos hechos pudieran implicar y conforme lo dispuesto por la presente Ley.*

De lo mencionado hasta el momento, se observan los orígenes del derecho a la reparación del daño, y su evolución al considerarse, ya como un derecho a una reparación integral del daño, al incluir elementos como: restitución, rehabilitación, compensación, satisfacción y garantías de no repetición, en sus dimensiones individual, colectiva, material, moral y simbólica.

En otro orden de ideas, vale la pena mencionar que las teorías del daño moral en el sistema jurídico mexicano, se centran en las teorías objetivas, porque desde un punto de vista jurídico, debe ser suficiente que se haya realizado un ataque a un bien no pecuniario (como la vulneración al derecho a la protección de datos personales) para que se configure el daño moral, sin ser necesario que en realidad se cause un dolor a una persona física, puesto que esa situación de dolor es algo subjetivo, extremadamente difícil o generalmente imposible de probar objetivamente en juicio (De la Parra, 2014, p. 43). Esa opinión es compartida por diversas personas especialistas como la jurista chilena Carmen Domínguez Hidalgo, quien afirma que “el solo menoscabo efectivamente acreditado de esos bienes genera el derecho a obtener una reparación por parte del perjudicado, al margen de sus consecuencias espirituales” (De la Parra, 2014).

Esto es lo que, en materia de prueba del daño moral, se le llama teoría objetiva, la cual estima como prueba antijurídica el tener que demostrar algo tan subjetivo e interno como el daño moral, de forma que, más que acreditarse el daño moral (sufrimiento) propiamente dicho, lo que se debe de probar es el ataque a un bien moral, es decir, el hecho ilícito dirigido contra un bien moral, con independencia de que ese ataque haya causado o no una aflicción a la víctima. A la luz de la teoría objetiva, exigir la prueba del daño moral en los términos de la teoría subjetiva, se traduciría en una denegación de justicia y en una imposibilidad de reparar el daño moral, dada la carga de probar lo que no se puede probar (De la Parra, 2014, pp. 43 y 44).

Por tal motivo, se cita un criterio interpretativo de la SCJN, para reforzar lo anterior:

#### **DAÑO MORAL EN EL DERECHO POSITIVO MEXICANO.**

*En el dictamen de la Cámara Revisora del decreto de veintinueve de diciembre de mil novecientos ochenta y dos, que reforma, entre otros, el artículo 1916 del Código Civil para el Distrito Federal, se establece, en lo que interesa: “... La iniciativa se fundamenta en la doctrina civilista contemporánea de los derechos de la personalidad, la cual tiende a garantizar a la persona el goce de sus facultades y el respeto al desenvolvimiento de su personalidad física y moral.”. Los bienes que tutela esa figura son, de manera sólo enunciativa: a) afectos; b) creencias; c) sentimientos; d) vida privada; e) configuración y aspectos físicos; f) decoro; g) honor; h) reputación; e, i) la consideración que de uno tienen los demás. Estos derechos no pueden ser tasables o valorables perfecta ni aproximadamente en dinero, por referirse a la persona en su individualidad o intimidad. Por esa razón, la legislación mexicana adopta la teoría de la comprobación objetiva del daño y no la subjetiva; es decir, basta la demostración de: 1) la relación jurídica que vincula al sujeto activo con el agente pasivo o agraviado, y 2) la existencia de un hecho u omisión ilícitos que lesione uno o varios de los bienes que tutela la figura, enunciados con anterioridad. Entonces, no se requiere la justificación de la existencia efectiva ni la extensión o gravedad del daño, lo cual conduciría a una prueba imposible, y esa demostración y tasación se dejan al prudente arbitrio del juzgador.*

#### **CUARTO TRIBUNAL COLEGIADO EN MATERIA CIVIL DEL PRIMER CIRCUITO.**

*Amparo directo 14424/2002. El Espectáculo Editorial, S.A. de C.V. y otras. 13 de enero de 2003. Unanimidad de votos. Ponente: Gilda Rincón Orta. Secretaria: Carmina S. Cortés Pineda.*

*Véase: Semanario Judicial de la Federación y su Gaceta, Novena Época, Tomo XVI, noviembre de 2002, página 1131, tesis I.3o.C.368 C, de rubro: “DAÑO MORAL. PRESUPUESTOS NECESARIOS PARA LA PROCEDENCIA DE LA ACCIÓN RELATIVA (LEGISLACIÓN DEL DISTRITO FEDERAL).”.*

*Nota: Por ejecutoria de fecha 25 de noviembre de 2009, la Primera Sala declaró inexistente la contradicción de tesis 328/2009 en que participó el presente criterio.*

### **3. Problemática**

La presente investigación se centra en la probable problemática a la cual se deben enfrentar las personas que

busquen que la autoridad les garantice la reparación integral del daño, en caso de considerar que algún Sujeto Obligado (dependencia, autoridad, órgano u organismo de gobierno), ha violentado su derecho a la protección de datos personales (daño moral).

En la actualidad, las leyes especializadas en materia de protección de datos personales, contienen sanciones que buscan, entre otras cuestiones, imponer multas u otras medidas de apremio para disuadir a posibles personas infractoras. Sin embargo, no se cuenta con un pronunciamiento claro respecto a la reparación integral del daño, aunque la misma ley abre la puerta para que se puedan iniciar, en su caso, otras acciones, como penales, administrativas o civiles, esta última materia, especialmente, relacionada con el tema objeto de estudio, es decir, la reparación del daño moral.

Derivado de lo anterior, han surgido algunos movimientos sociales de diversa índole, donde se han presentado algunas propuestas, para subsanar la probable omisión legislativa. En septiembre de 2022, se presentó una iniciativa que pretendía modificar la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, para que el Servicio de Administración Tributaria, se encargara de ejecutar las multas y destinar un porcentaje, de lo cobrado, a la persona titular de los datos personales. Sin embargo, pareciera ser que, se les ha olvidado proponer una modificación similar a su homóloga la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados, sea como fuere, al final, la citada iniciativa no prosperó.

Asimismo, se considera que todavía falta mucho camino por recorrer en materia del derecho a la reparación integral de daño, pues, la costumbre apunta, casi siempre, a una reparación económica, como ejemplo de la iniciativa del párrafo que antecede. No obstante, ésta, es solo la punta del iceberg, pues la legislación y la jurisprudencia han reconocido expresamente otras formas de reparación, como ya se vio en el punto anterior, tales como las medidas de restitución, rehabilitación, compensación, satisfacción y garantías de no repetición, en sus dimensiones individual, colectiva, material, moral y simbólica.

Derivado de lo anterior, en el presente documento de investigación busca evidenciar que el derecho a la protección de datos personales es un derecho humano autónomo, y, por lo tanto, en caso de acreditarse una vulneración o un ataque al daño moral, éste podría ser susceptible a una utópica reparación integral.

#### 4. Conclusiones y propuestas de solución

En atención al principio de fuerza normativa de la Constitución y con fundamento en el artículo 1º, párrafo tercero de la CPEUM, los artículos 51.2 y 63.1 de la CADH, en relación con la obligación general de respetar los derechos humanos establecida en el artículo 1.1 de dicha CADH, el Estado Mexicano está obligado a reparar de manera integral las violaciones de los derechos humanos, incluido el derecho a la protección de datos personales.

Con base en estos preceptos, se derivan obligaciones constitucionales y de derecho internacional en materia de reparación a las violaciones de derechos humanos que son de naturaleza erga omnes, lo que significa, es que no se trata de una obligación de reciprocidad frente a otros Estados, sino que se configura dentro del orden jurídico interno como un derecho de las víctimas a acceder a la reparación, cuando han sufrido violaciones a sus derechos humanos, como el derecho a la protección de datos personales.

Algunos de los medios específicos de reparación, son los siguientes: restitución, indemnización, rehabilitación, satisfacción y garantías de no repetición. En derecho internacional, por restitución se entiende devolver a la víctima a la situación anterior a la violación. La indemnización considera los perjuicios económicamente evaluables. La rehabilitación incluye la atención médica y psicológica, así como servicios jurídicos y sociales. La satisfacción cubre, de ser posible, las medidas eficaces para conseguir que no continúen las violaciones. Finalmente, las garantías de no repetición tienen como finalidad contribuir a la prevención de violaciones futuras por medio de la adopción de medidas legislativas, judiciales o administrativas.

En este sentido, la Corte IDH ha establecido que el concepto de «reparación integral» (restitutio in integrum) implica “el reestablecimiento de la situación anterior y la eliminación de los efectos que la violación produjo, así como una indemnización como compensación por los daños causados.”<sup>4</sup> De igual manera, la Corte IDH, ha considerado:

*...que la naturaleza y monto de la reparación ordenada dependen del daño ocasionado en los planos tanto material como inmaterial. Las reparaciones no pueden implicar ni enriquecimiento ni empobrecimiento para la víctima o sus familiares, y deben guardar relación directa con las violaciones declaradas. Una o más medidas pueden reparar un daño específico sin que éstas se consideren una doble reparación<sup>5</sup>.*

Que la CPEUM, en su artículo 6º, apartado A, fracción VIII considera:

*Artículo 6º.*

*A. Para el ejercicio del derecho de acceso a la información, la Federación y las entidades federativas, en el ámbito de sus respectivas competencias, se regirán por los siguientes principios y bases:*

*VIII. La Federación contará con un organismo autónomo, especializado, imparcial, colegiado, con personalidad jurídica y patrimonio propio, con plena autonomía técnica, de gestión, capacidad para decidir sobre el ejercicio de*

<sup>4</sup>Ver, Caso González y otras «Campo Algodonero». Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia del 16 noviembre de 2009, párr. 450.

<sup>5</sup> Idem.

*su presupuesto y determinar su organización interna, responsable de garantizar el cumplimiento del derecho de acceso a la información pública y a la protección de datos personales en posesión de los sujetos obligados en los términos que establezca la ley.*

En consecuencia, el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), así como sus homólogos de las demás entidades federativas, se consideran organismos públicos de protección de los derechos humanos (en particular del derecho de acceso a la información pública y el de protección de datos personales), por lo que, de conformidad con el citado artículo 65, inciso c) de la Ley General de Víctimas, el INAI y sus homólogos de las demás entidades federativas, estarían facultados para emitir resoluciones a fin de garantizar la reparación integral del daño a las presuntas víctimas, y abriendo la oportunidad, para que trabajen en colaboración con la Comisión Ejecutiva de Atención a Víctimas y a sus homologas estatales, respectivamente.

Como dato adicional, y como ya se mencionó, el 20 de septiembre de 2022, fue presentada, por el Senador Alejandro Armenta Mier, la iniciativa con proyecto de decreto que reforma, adiciona y modifica diversas disposiciones de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, en la que se buscaba, entre otras cuestiones, por cuanto hace a la imposición de sanciones para el caso del incumplimiento, ya que no se prevé una ruta específica para el cobro, ni se establece quién será el encargado de hacer efectivas las multas respectivas, se proponía a) establecer al Servicio de Administración Tributaria, como la entidad encargada de hacer efectivas las multas impuestas por el INAI y b) destinar un 50% adicional de la multa impuesta a favor de la persona titular afectada. Se concluye, que, aunque han surgido algunas propuestas como la señalada en el párrafo anterior, para dejar «expresamente» en el texto legislativo la reparación integral del daño en materia de protección de datos personales; las Comisiones de Atención a Víctimas y los órganos garantes del derecho de acceso a la información pública y del de protección de datos personales, podrían comenzar a desarrollar, en colaboración, un papel importante en el tema. Por lo que, se considera la posibilidad de contar con otros mecanismos para garantizar lo que «implícitamente», reconoce la propia CPEUM, los tratados internacionales, las leyes en la materia y la jurisprudencia.

Finalmente, se reconoce la utopía de la propuesta. Sin embargo, se considera que bien vale la pena ponerla sobre la mesa de discusión.

## **5. Referencias**

- Araujo, E. (2009). El derecho a la Información y a la protección de datos personales en México. México: Porrúa.
- Cantorale, K. (2012). Derecho de protección de datos personales de la salud. México: Novum.
- De La Parra, E. (2014). El derecho a la propia imagen. México: Tirant lo Blanch.
- Flores, M. (2009). Diccionario de Derechos Humanos. México: FLACSO.
- García, P. (2014). Derechos y libertades, internet y tics. Valencia: Tirant lo Blanch.
- Garzón, V. (1998). Privacidad y publicidad, Revista Doxa, número 21, Cuadernos de filosofía del Derecho, Centro de Estudios políticos y Constitucionales, Alicante, España, pp. 223-244.
- Garzón, V. (2008). Lo íntimo, lo privado y lo público. Cuadernos de transparencia, número 6, Quinta Edición. México: INAI, 2008.
- González, I. (2020). Protección de Datos Personales. México. Tirant Lo Blanch.
- Gregorio, C. (2004). Protección de Datos Personales: Europa vs. Estados Unidos, todo un dilema para América Latina.
- Transparentar al Estado: la experiencia mexicana de acceso a la información. México: Instituto de Investigaciones Jurídicas de la UNAM. JIMÉNEZ, J. (1999). Derechos Fundamentales: Concepto y Garantías. España: Trotta.
- Murga, J., Fernández, M. Y Espejo, M. (2021). Cuestiones actuales sobre protección de datos en España y México. Valencia. Tirant Lo Blanch.
- Murillo, P. Y Piñar, J. (2011). El derecho a la autodeterminación informativa. México: Fontomara.
- Troncoso, A. (2010). La protección de datos personales. En busca del equilibrio. España: Tirant lo Blanch.
- Villanueva, E. (2009). Diccionario de Derecho a la información. México: Porrúa.
- Vuelvas, M. (2015). Letras libres vs. La Jornada: la libertad de expresión ante los tribunales. México: Universidad de Colima.

# Diagnóstico educativo de Jalisco: una mirada desde el registro histórico de indicadores

Rosalba Gascón Pérez, Itzia Gollás Núñez, Martha Pereira Moncayo.  
Mexicanos Primero Jalisco, Guadalajara, Jalisco, México.  
rgascon@mpj.org.mx, mpereira@mpj.org.mx, igollas@mpj.org.mx

**Resumen:** La pandemia de COVID-19 impactó diversos sectores, incluido el educativo. Ante la falta de información educativa oficial pública, comprensible, actualizada y accesible, esta investigación busca conocer el estado del derecho a la educación en Jalisco, México, antes, durante y después de la pandemia.

El estudio emplea una metodología documental para analizar indicadores educativos históricos. A través de fuentes oficiales, se recopilaron y analizaron (a nivel descriptivo e interpretativo) 11 indicadores educativos, agrupados en cuatro categorías: Numeralia educativa, Aprendizaje, Acceso y trayectoria, y Socioemocional. El micrositio resultante [www.diagnosticoeducativojalisco.org](http://www.diagnosticoeducativojalisco.org) integra el análisis de los indicadores y otras herramientas, como la consulta interactiva de indicadores desagregados por región y municipio, para garantizar la conservación de datos independientemente de los cambios administrativos.

Los resultados muestran tendencias favorables en indicadores de primaria y secundaria, así como en analfabetismo, grado promedio de escolaridad y nivel de escolaridad. Por el contrario, preescolar y media superior presentan tendencias menos favorables, posiblemente afectados por hechos de la pandemia.

Se destaca la necesidad de contar con mecanismos ágiles y sólidos para la recolección, análisis y difusión de datos a lo largo del tiempo, adaptando indicadores a necesidades contextuales, como la inclusión de indicadores socioemocionales, y revisando la metodología utilizada para generar información de indicadores debido a la variación de datos encontrada. También se resalta la importancia de mantener principios de calidad, relevancia, oportunidad y transparencia en los procesos de generación y comunicación de información educativa relevante, para fomentar la participación ciudadana y la toma de decisiones oportuna basadas en evidencia.

**Palabras clave:** *derecho a la educación, diagnóstico educativo, indicadores educativos, histórico de datos, plataforma ciudadana.*

## 1. Introducción

La pandemia de enfermedad por coronavirus (COVID-19) tuvo un impacto significativo en varios sectores, entre ellos el educativo. Estudiantes de todo el mundo se vieron obligados a abandonar las aulas para continuar sus estudios desde sus hogares. No obstante, esta situación generó diversas problemáticas, entre las que se destacan el aumento en la inequidad educativa, el rezago y la pérdida de aprendizaje, los efectos socioemocionales adversos, el abandono escolar, inasistencia y comunicación intermitente con niños y niñas (Comisión Estatal para la Mejora Continua en Jalisco [CEMEJ], 2021; Instituto Mexicano para la Competitividad [IMCO], 2022).

Según la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) (Organización de las Naciones Unidas para la Educación, 2020), más de mil 200 millones de estudiantes en todo el mundo dejaron de asistir físicamente a las escuelas debido a la pandemia. En el estado de Jalisco, México, se implementaron diversos esfuerzos para permitir un regreso gradual de los estudiantes a las aulas. Estas estrategias incluyeron asesorías individuales, grupos de seguimiento académico y el retorno voluntario a clases presenciales, incluso antes de que se implementara esta medida a nivel federal. Sin embargo, aquellos estudiantes que no tuvieron acceso a estas alternativas tuvieron que experimentar de 200 a 225 días de educación remota (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2021). Esto también implicó que los docentes recibieran formación en metodologías educativas diferentes a las habitualmente utilizadas, especialmente en lo que respecta a la integración de tecnología en los procesos de enseñanza y aprendizaje. Asimismo, las familias se vieron obligadas a involucrarse más en la educación de sus hijos e hijas para brindarles apoyo en sus tareas escolares. En este sentido, es conocido que muchos tuvieron que realizar gastos adicionales para adquirir dispositivos electrónicos o acceso a Internet, permitiéndoles así la comunicación con sus maestros (Miranda López et al., 2020).

Dada esta situación, es crucial comprender el estado actual de la educación. Contar con datos e información sobre nuestra situación actual nos proporciona antecedentes y un punto de partida para evaluar, tomar decisiones, establecer metas e implementar acciones (Marrero García, 2013). Si no tenemos un entendimiento oportuno y preciso de nuestras necesidades, difícilmente podremos avanzar hacia la mejora de la educación, el aprendizaje y el bienestar de las niñas, niños y adolescentes (NNA).

Desde Mexicanos Primero Jalisco, una organización civil que busca articular esfuerzos para que la educación pública habilite a cada NNA a alcanzar su máximo potencial y llegar a ser la mejor versión de sí mismos, hemos trabajado en la recolección de datos relacionados con la educación básica estatal. Hemos obtenido esta información a través de fuentes oficiales, solicitudes de información y el seguimiento de noticias en medios de comunicación. Nuestro propósito

es comprender la realidad y las necesidades presentes en las escuelas, aspectos que influyen en la garantía del derecho a la educación. A partir de este trabajo, hemos identificado dos problemáticas principales:

1) Falta información oficial sobre indicadores educativos, presentada de manera pública, actualizada y accesible de manera permanente. La información disponible corresponde a ciclos escolares pasados, lo cual no contribuye a la toma de decisiones oportuna. O los sitios web de fuentes oficiales dejan de estar disponibles para consulta pública, por ejemplo, la página web de estadística educativa de Jalisco <https://indicadores.sej.gob.mx/publicaciones/Ini2021-2022/index.htm> la cual ha estado caída desde inicios del 2023.

2) La información existente sobre indicadores educativos resulta de difícil acceso para la sociedad. Comprender gran parte de esta información requiere de un nivel de especialización en temas educativos, además de que la información no siempre es clara y se encuentra dispersa en diversos sitios web o documentos de difícil localización.

Estas problemáticas dificultan que la sociedad cuente con herramientas para expresar opiniones, reflexionar, dialogar y participar en la toma de decisiones basadas en evidencia. Esto, a su vez, obstaculiza los esfuerzos pertinentes y oportunos para garantizar el derecho a la educación.

Desde la sociedad civil, hemos identificado la posibilidad de llevar a cabo actividades de investigación y divulgación de información educativa dirigidas a la ciudadanía. Esto tiene como objetivo fomentar la participación ciudadana y la rendición de cuentas con el fin de mejorar la educación. Desde nuestra posición como organización civil, consideramos fundamental poner a disposición de la sociedad información sobre indicadores educativos como una herramienta valiosa para la consulta, la investigación, la participación y la formación. Contar con información proporciona elementos necesarios para comprender la realidad y establece el punto de partida para tomar acción y medidas por una mejor educación para NNA (Arias Pérez & Aristizábal Botero, 2011).

Es por esta razón que hemos llevado a cabo la presente investigación. En la siguiente sección, presentamos los objetivos de la investigación, seguidos por el marco de referencia, en el cual hacemos referencia a diversas fuentes que conforman el marco teórico y conceptual del estudio. Posteriormente, describimos la metodología empleada para la recopilación y análisis de datos. Finalmente, presentamos los resultados de la investigación, así como una sección destinada a la discusión y conclusiones.

## 2. Objetivo

El propósito de la presente investigación consistió en conocer el estado del derecho a la educación en Jalisco, es decir, cómo está la educación en el estado y qué resultados ha tenido en diferentes indicadores educativos, en el periodo anterior, durante y posterior a la pandemia por COVID-19. Para ello, se llevó a cabo la recopilación de datos provenientes de fuentes públicas oficiales, abarcando diversos indicadores. Esto tuvo el objetivo de identificar los avances y desafíos del sistema educativo, además de proporcionar a la ciudadanía información actualizada y de fácil acceso.

Como complemento a lo anterior, con el fin de garantizar que la información esté al alcance de la sociedad de manera sencilla y accesible, como resultado de esta investigación se creó un micrositio [www.diaagnosticoeducativojalisco.org](http://www.diaagnosticoeducativojalisco.org). Dicho micrositio consiste en una página web que integra información de distintos indicadores educativos, recopilada de fuentes oficiales, el cual es accesible desde cualquier lugar y dispositivo con conexión a internet. Esta iniciativa aborda la problemática de la fragmentación de plataformas y documentos con información dispersa y dificultad para su localización.

Este micrositio se identifica como una herramienta especialmente útil para diversos actores según su rol:

1. Académicos, investigadores y estudiantes de diversas áreas: Constituye una herramienta valiosa para la investigación.
2. Medios de comunicación: Sirve como fuente de consulta para obtener datos relevantes en temas educativos, permitiendo su divulgación y puesta en la agenda mediática.
3. Escuelas y funcionarios públicos: Incluyendo regidores, alcaldes, diputados y personal de diversas dependencias gubernamentales, especialmente aquellos que aspiran a ser elegidos, encuentran en este micrositio una herramienta que fomenta la toma de decisiones basadas en evidencia y la planificación.
4. Organizaciones de la sociedad civil y empresas: Constituye un recurso consultivo útil para la planificación y desarrollo de proyectos.
5. Sociedad en general: Se convierte en un medio formativo y de consulta, brindando más elementos para opinar y actuar.

En síntesis, el propósito fundamental de esta investigación es destacar la importancia de contar con información técnicamente sólida y oportuna para posicionar la mejora de la educación en el centro de la agenda pública, involucrando distintos sectores y esferas sociales.

### 3. Marco de referencia

La palabra diagnóstico se define como el proceso mediante el cual conocemos el estado o situación en la que se encuentra algo o alguien. En este caso, un diagnóstico educativo permite contar con información adecuada y pertinente sobre los fines y el objeto educativo de estudio a través de un proceso sistemático de recolección de información y su valoración. Un diagnóstico posibilita identificar las características y necesidades de aquello que se diagnostica y definir un punto de partida para mejorar, planear, tomar decisiones y/o intervenir (Sánchez Espinoza, 2021; Sobrado Fernández, 2005).

Aunado a lo anterior, cuando abordamos el estado de la educación, resulta necesario hacer referencia a los indicadores educativos, ya que estos son “instrumentos que nos permiten medir y conocer la tendencia o desviación de las acciones educativas, con respecto a una meta o unidad de medida esperada o establecida; así como plantear previsiones sobre la evolución futura de los fenómenos educativos” (Secretaría de Educación Pública [SEP], 2019). En otras palabras, los indicadores proporcionan información sobre la realidad educativa, ofrecen elementos de comparación para evaluar resultados y avances, y sirven como guía para planificar y preparar pasos futuros alineados con la visión ideal de la educación.

En líneas generales, la evaluación nos proporciona información para emitir juicios informados y tomar decisiones basadas en evidencia. Adicionalmente, los indicadores educativos nos ayudan a detectar el progreso o retroceso de los aspectos evaluados en la educación. Esto resulta crucial para el diagnóstico de necesidades, el establecimiento de objetivos, la planificación y ejecución de políticas públicas, así como para medir resultados e impactos. Según la Secretaría de Educación Pública (SEP), “en la medida en que la información estadística y los indicadores se utilicen para la planeación y la toma de decisiones, se podrá avanzar hacia una mayor calidad” (Secretaría de Educación Pública [SEP], 2019).

De acuerdo con la Ley General de Transparencia y Acceso a la Información Pública, tanto el gobierno federal, como los gobiernos estatales y municipales tienen la obligación de hacer pública información de manera accesible para la ciudadanía, incluida la información educativa. La Ley General de Educación, en este contexto, establece que se debe brindar a la ciudadanía información, sobre planes de estudio, resultados de evaluaciones educativas, gastos en educación y otros datos relevantes para evaluar la calidad educativa (Ley General de Educación, 2019).

Asimismo, la Ley del Sistema Nacional de Información Estadística y Geográfica estipula la creación de un conjunto de indicadores clave para abordar distintos temas, entre estos la educación. Y, además, afirma que se debe establecer un Subsistema Nacional de Información por parte de la SEP, basado en principios como accesibilidad, transparencia, objetividad e independencia (Ley Del Sistema Nacional de Información Estadística y Geográfica, 2015).

La transparencia en la información educativa es un derecho de los ciudadanos que permite entender el funcionamiento del sistema educativo y exigir responsabilidad por parte de los gobiernos. Es una herramienta que fomenta la participación ciudadana para la mejora de la educación y es algo fundamental para asegurar que la educación sea un derecho de todos los mexicanos.

En adición a esto, diversas instancias públicas han compartido información educativa a lo largo de los años. En el ámbito de la SEP, la Dirección General de Planeación, Programación y Estadística Educativa (DGPPyEE) se encarga de coordinar la planificación, programación y evaluación del sistema educativo nacional. Además, se encarga de generar y divulgar información estadística e indicadores educativos, junto con la gestión de una plataforma de datos centralizada para fomentar la transparencia y la rendición de cuentas (Secretaría de Educación Pública [SEP], 2020). Sumado a lo anterior, la DGPPyEE publica anualmente un documento con las cifras educativas principales del ciclo escolar más reciente, con indicadores desglosados por entidad federativa (P. y E. E. [DGPPyEE] Dirección General de Planeación, 2022a).

La Comisión Nacional para la Mejora Continua de la Educación (MEJOREDU), por su parte, mide y comparte los comportamientos de los indicadores educativos a nivel nacional, buscando promover mejoras en la educación (Comisión Nacional para la Mejora Continua de la Educación [MEJOREDU], 2022).

Previo a la existencia de MEJOREDU, existía el Instituto Nacional para la Evaluación de la Educación (INEE), el cual generaba sólida información sobre indicadores para guiar las decisiones de las instituciones educativas. Esta entidad publicaba estudios anuales sobre el panorama educativo de México y el estado de la educación obligatoria en el país entre 2004 y 2019, detallando diversos indicadores educativos (Instituto Nacional para la Evaluación de la Educación [INEE], 2017a, 2019).

El INEE también diseñó y administró pruebas estandarizadas, como el Plan Nacional para las Evaluaciones de los Aprendizajes (PLANEA) y el Programa Internacional para la Evaluación de Estudiantes (PISA), con el objetivo de evaluar el nivel de aprendizaje de los estudiantes mexicanos (Instituto Nacional para la Evaluación de la Educación

[INEE], 2015; Instituto Nacional para la Evaluación de la Educación [INEE, 2018). Además, llevó a cabo evaluaciones para comprender las condiciones básicas del derecho a la educación, como la Evaluación de Condiciones Básicas para la Enseñanza y el Aprendizaje (ECEA) (Instituto Nacional para la Evaluación de la Educación [INEE], 2017b), el Estudio Internacional de Educación Cívica y Ciudadana, el Estudio Internacional sobre la Enseñanza y el Aprendizaje (TALIS), el Laboratorio Latinoamericano de Evaluación de la Calidad (LLECE) y Evaluaciones de Currículo.

En el contexto del estado de Jalisco, la Secretaría de Educación Jalisco y la Comisión Estatal para la Mejora Continua en Jalisco (CEMEJ) son fuentes cruciales de información en cuanto a indicadores educativos. La CEMEJ, establecida en 2020, emite diversos informes que contienen datos sobre indicadores educativos y otros temas a nivel estatal, en ocasiones con desgloses municipales y/o regionales. En 2021, la CEMEJ publicó el informe Retos educativos en Jalisco para el logro académico y la permanencia en pandemia [2](Comisión Estatal para la Mejora Continua en Jalisco [CEMEJ], 2021). Por otra parte, la Dirección General de Planeación de la Secretaría de Educación Jalisco (SEJ) presenta información estatal sobre algunos indicadores educativos en su portal (Secretaría de Educación Jalisco, 2022), aunque, relacionado al apartado de Introducción, al momento del cierre de esta investigación, el acceso a dicho portal se encuentra inhabilitado. Además, la Dirección de Evaluación Educativa comparte los resultados de evaluaciones, como el Diagnóstico Socioemocional 2020 (Secretaría de Educación Jalisco [SEJ], 2021), y la SEJ opera el sitio web “Escuela Transparente”, que ofrece un mapa interactivo con las escuelas de Jalisco organizadas por turno, nivel, sostenimiento y municipio (Secretaría de Educación Jalisco [SEJ], 2019).

En el ámbito estatal, se cuenta además con la estrategia del Gobierno del Estado, Monitoreo de Indicadores del Desarrollo (MIDE) Jalisco. Esta estrategia busca coordinar y dar seguimiento periódico a indicadores y metas vinculadas a su Plan Estatal de Gobernanza y Desarrollo Jalisco 2018-2024 Visión 2030 (Monitoreo de Indicadores del Desarrollo [MIDE] Jalisco, 2016). Entre alguno de los indicadores que monitorea y de los cuales reporta su tendencia histórica, se encuentran los indicadores de la temática sectorial educativa dentro del eje de “Desarrollo Social” (Monitoreo de Indicadores del Desarrollo [MIDE] Jalisco, 2022).

Por último, otro organismo público autónomo que recoge datos sobre algunos indicadores educativos es el Instituto Nacional de Estadística y Geografía (INEGI), el cual levanta información de manera censal e intercensal sobre asistencia escolar, alfabetismo y nivel de escolaridad (Instituto Nacional de Estadística y Geografía [INEGI], 2021).

Es importante destacar que para la presente investigación se han considerado las fuentes oficiales citadas en este apartado tanto para el marco conceptual y teórico, como para la selección de indicadores y la recopilación y análisis de información.

Por otro lado, es relevante mencionar que la sociedad civil desempeña un papel fundamental en el monitoreo de la información educativa y la exigencia de transparencia en los datos. Las organizaciones de la sociedad civil (OSC) tienen la capacidad de realizar investigaciones y análisis para identificar problemas en el sistema educativo y promover la participación ciudadana para la mejora de la sociedad.

Entre las actividades que las OSC pueden llevar a cabo se encuentran: solicitar información educativa a los gobiernos mediante la Ley General de Transparencia y Acceso a la Información Pública, analizar los datos educativos para identificar fortalezas y áreas de mejora, difundir información educativa entre la ciudadanía para impulsar la participación, denunciar casos de corrupción o malversación de fondos en la educación, entre otras. La sociedad civil puede desempeñar un papel de corresponsabilidad, defensa de derechos y contrapeso importante frente a los gobiernos, fomentando la transparencia y la rendición de cuentas a favor de la mejora de la educación.

Entre algunas OSC que han realizado esfuerzos para medir los resultados educativos se encuentra Mexicanos Primero, que en su informe “Índice de Cumplimiento de la Responsabilidad Educativa Estatal (ICRE)” recopiló información sobre diversos indicadores educativos. Este estudio tenía como objetivo comparar las entidades federativas mediante la medición de indicadores educativos (Mexicanos Primero, 2018).

## 4. Metodología

Para el presente estudio se utilizó una metodología cualitativa, es decir que los hallazgos de la investigación no se produjeron “por medio de procedimientos estadísticos u otros medios de cuantificación [...]. Algunos de los datos pueden cuantificarse, por ejemplo, con censos o información sobre los antecedentes de las personas u objetos estudiados, pero el grueso del análisis es interpretativo” (Strauss & Corbin, 2002).

Bajo esta metodología, para cumplir con el objetivo de esta investigación, se llevó a cabo un análisis de datos históricos educativos públicos oficiales mediante una investigación documental. Según Tancara (Tancara, 1993), esta metodología consiste en una “serie de métodos y técnicas de búsqueda, procesamiento y almacenamiento de la información contenida en los documentos, en primera instancia, y la presentación sistemática, coherente y suficientemente argumentada de nueva información en un documento científico, en segunda instancia” (p.94). Por lo tanto, se recopilaron y analizaron

principalmente datos cuantitativos de diversos indicadores educativos para identificar y difundir información sobre los avances y desafíos del sistema educativo en Jalisco, presentándolos de manera organizada y accesible a la sociedad. Estos datos fueron obtenidos de fuentes oficiales como la SEP, la DGPPyEE, la SEJ, MEJORED, el INEE, la CEMEJ, MIDE Jalisco e INEGI, además de la Dirección de Evaluación Educativa de la SEJ, la Dirección General de Análisis y Diagnóstico del Aprovechamiento Educativo y el College Board.

Es importante destacar, en primer lugar, que los indicadores educativos fueron seleccionados en función de la accesibilidad y disponibilidad de la información en las fuentes mencionadas anteriormente. Por otro lado, las cifras reportadas son las más actualizadas disponibles al momento de la fecha de cierre de esta investigación. El micrositio resultante de este estudio será actualizado anualmente para dar seguimiento a las cifras históricas. Además, cabe señalar que no se encontró información desagregada por municipios o regiones de Jalisco para todos los indicadores, lo que resultó en la exclusión de estos datos en las secciones correspondientes del micrositio.

A continuación, se presentan los 11 indicadores seleccionados, organizados en cuatro grandes categorías:

**Categoría: Aprendizaje:**

- Repetición
- Reprobación
- Logros de aprendizaje

**Categoría: Acceso y trayectoria:**

- Abandono escolar
- Analfabetismo
- Cobertura
- Eficiencia terminal
- Grado promedio de escolaridad
- Nivel de escolaridad
- Tasa de terminación

**Categoría: Socioemocional:**

- Habilidades socioemocionales

El análisis se llevó a cabo en dos niveles. El primero es el análisis descriptivo, al mostrar cómo está la educación y especificar las propiedades y tendencias importantes a través de la lectura de los datos, resultado de los indicadores (García, 2018; Secretaría de Educación Pública [SEP], 2021), en el que se utilizó el análisis estadístico básico para responder a la pregunta: ¿cómo estamos? El segundo nivel es el análisis interpretativo, que busca responder a la pregunta: ¿por qué estamos así? Esto se logró identificando hechos y problemáticas sociales y de política pública (el micrositio presenta una línea de tiempo que sirvió como guía para este análisis).

Para el análisis descriptivo, el criterio principal utilizado fue el comportamiento estadístico histórico de los indicadores educativos, además de la identificación de avances y retrocesos en relación con las tendencias deseadas. En este sentido, se calculó la diferencia porcentual entre los indicadores del ciclo escolar más reciente con datos oficiales y los datos correspondientes a los cinco ciclos escolares anteriores (o, en su defecto, entre los últimos dos censos del INEGI). También se evaluó la tendencia del indicador, la variación entre fuentes oficiales y el dato más actual (ya sea estimado o preliminar) en comparación con el último dato oficial (no estimado ni preliminar).

En cuanto al análisis interpretativo, se identificó la relación entre los comportamientos de los indicadores educativos y diversos eventos, sucesos y problemáticas históricas dentro del contexto de Jalisco. Esto incluye acuerdos gubernamentales, cierres y aperturas de escuelas y otras estrategias derivadas de la pandemia, entre otros factores. Preguntas como ¿a qué se puede atribuir la obtención de estos resultados?, ¿qué los ocasiona? y ¿cómo se relacionan con otros indicadores?, guiaron este análisis.

Las etapas del proyecto de investigación fueron las siguientes:

1. Diseño de la investigación: establecimiento del objetivo, metodología, planteamiento del problema, marco de referencia y justificación, así como la evaluación del proyecto.
2. Recolección de información: definición de las categorías e indicadores a investigar en esta etapa, seguida de la obtención de información de fuentes oficiales.
3. Sistematización y análisis de la información: los datos recolectados fueron analizados desde un enfoque descriptivo e interpretativo, y los contenidos del micrositio fueron redactados.
4. Diseño del micrositio: se diseñó la interfaz y la estructura general de las páginas del micrositio, y se insertaron los contenidos correspondientes.
5. Difusión de resultados: se desarrolló una estrategia de comunicación y vinculación para presentar los resultados de la investigación.

Cabe destacar que, posteriormente a las etapas descritas, se planearon dos etapas a realizarse en el futuro con la finalidad



de robustecer el proyecto. La primera implica analizar e incluir en el microsítio información sobre indicadores de infraestructura, organización escolar y personal educativo. La segunda fase tiene como objetivo agregar un enfoque cualitativo al análisis, abordando las causas y consecuencias de estas problemáticas educativas, así como la percepción y testimonios de la ciudadanía.

## 5. Resultados

Para este trabajo de investigación, se presentan los resultados generales de los 11 indicadores educativos, desagregados por nivel educativo, en términos de si han avanzado o retrocedido según su tendencia deseable en los últimos cinco ciclos escolares.

Es relevante señalar que los resultados completos están disponibles en el microsítio [www.diagnosticoeducativojalisco.org](http://www.diagnosticoeducativojalisco.org), que es el producto resultante de esta investigación. En esta plataforma, cada indicador se presenta con: su definición, datos destacados, tablas, gráficos y descripciones de fuentes oficiales, interpretación de datos, explicación de variaciones de fuentes (en caso de tener), y desglose de resultados por municipio y/o región (si hay disponibilidad). Además, se incluye una línea de tiempo sobre eventos relacionados con la política educativa en Jalisco durante los últimos años, un glosario de términos educativos y las fórmulas de los indicadores, así como una herramienta de consulta interactiva para acceder a los indicadores educativos desagregados por región y municipio de Jalisco.

A continuación, se presentan los resultados por indicador educativo reportados por la DGPPyEE, entre ellos, repetición, cobertura, reprobación, abandono escolar, eficiencia terminal y tasa de terminación; por el INEGI, como analfabetismo, grado promedio de escolaridad y nivel de escolaridad; así como los resultados tanto de las evaluaciones PLANEA de los últimos dos años de levantamiento, como del Diagnóstico Socioemocional hecho por la SEJ.

Según cifras de la DGPPyEE (P. y E. E. [DGPPyEE] Dirección General de Planeación, 2022a, 2022b), se comparan en seguida los resultados del ciclo escolar 2017-2018 con el último ciclo escolar reportado con cifras oficiales (ni preliminares ni estimadas) para cada indicador, y se presenta su tendencia deseable.

Indicador	Tendencia deseable	Resultados por nivel educativo			
	Conviene que:	Preescolar	Primaria	Secundaria	Media Superior
Comparativo entre ciclo escolar 2017-2018 con el 2021-2022					
Repetición	Disminuya	No aplica	De 1.4% a 0.7% ▼	De 0.6% a 0.6%	De 4.9% a 49.3% ▽
Cobertura	Aumente	De 72.8% a 64.6% △	De 104.4% a 105.3% ▲	De 93.5% a 93.5%	De 73.7% a 71.6% △
Comparativo entre ciclo escolar 2017-2018 con el 2020-2021					
Reprobación	Disminuya	No aplica	De 1.2% a 0.1% ▼	De 3.5% a 0.6% ▼	De 14.8% a 18.7% ▽
Abandono escolar	Disminuya	No aplica	De 0.5% a -0.4% ▼	De 4.9% a 2.1% ▼	De 18.5% a 1.1% ▼
Eficiencia terminal	Aumente	No aplica	De 96.4% a 96.9% ▲	De 84.6% a 90.9% ▲	De 63.9% a 56.2% △
Tasa de terminación	Aumente	No aplica	De 100.7% a 101.2% ▲	De 85.3% a 88.7% ▲	De 59.8% a 59.1% △

Tabla 1. Tendencia de indicadores educativos reportados por la DGPPyEE.

Fuente: elaboración propia a partir de (P. y E. E. [DGPPyEE] Dirección General de Planeación, 2022b, 2022a).

Simbología: ▲Tendencia deseable ascendente, el resultado sí incrementó. △Tendencia deseable ascendente, el resultado disminuyó. ▼Tendencia deseable descendente, el resultado sí disminuyó. ▽Tendencia deseable descendente, el resultado incrementó.

Nota: Las celdas con la etiqueta "No aplica" indican que la DGPPyEE no presenta resultados de los indicadores correspondientes para el nivel de preescolar.

En la tabla anterior, se puede apreciar que los resultados de educación media superior muestran principalmente una tendencia no deseable, exceptuando la cifra de abandono escolar. En cambio, los niveles de primaria y secundaria permanecen estables o con una tendencia favorable en la mayoría de los casos. En cuanto al indicador único de preescolar, la comparación de los resultados sugiere una tendencia desfavorable.

Por otra parte, a continuación, se comparan los resultados de los siguientes indicadores educativos reportados por el INEGI en los últimos censos que recopilaron información (2010, 2020), de acuerdo con su tendencia deseable.

Indicador	Tendencia deseable	Resultados por año de censo	
	Conviene que:	2010	2020
Analfabetismo	Disminuya	4.4%	2.9%▼
Grado promedio de escolaridad	Aumente	8.8	9.9▲

Tabla 2. Tendencia de indicadores educativos reportados por el INEGI. Analfabetismo y grado promedio de escolaridad.

Fuente: elaboración propia a partir de (Instituto Nacional de Estadística y Geografía [INEGI], 2010, 2020b).

Simbología: ▲Tendencia deseable ascendente, el resultado sí incrementó. △Tendencia deseable ascendente, el resultado disminuyó. ▼Tendencia deseable descendente, el resultado sí disminuyó. ▽Tendencia deseable descendente, el resultado incrementó.

Nota: El indicador de Grado promedio de escolaridad se refiere a los grados promedios aprobados a partir de primero de primaria (Instituto Nacional de Estadística y Geografía [INEGI], n.d.), lo que significa que seis grados corresponden a primaria completa, nueve grados a secundaria completa y 12 grados a educación media superior completa.

Como se observa en la tabla, los porcentajes de la población en condición de analfabetismo en Jalisco disminuyeron entre el 2010 y el 2020. Esto quiere decir que se redujo 1.5 puntos porcentuales la población de 15 años y más que no son capaces de leer ni escribir una breve y sencilla exposición de hechos relativos a su vida cotidiana. En cuanto al grado promedio de escolaridad, se muestra que la tendencia es ascendente, lo cual es deseable para este indicador porque refleja un incremento en el promedio de grados escolares aprobados por la población de 15 años y más. Así, entre el 2010 y el 2020 el incremento fue de 1.1 grados en total.

Tendencia deseable					
Conviene que:	Aumente en niveles educativos altos, como media superior y superior.				
Resultados por año de censo					
Año	Sin escolaridad	Básica	Media Superior	Superior	No especificado
2010	5.2%	58.0%	18.5%	17.3%	1.0%
2020	3.4%▼	50.3%▼	23.7%▲	22.3%▲	0.2%▼

Tabla 3. Tendencia de indicadores educativos reportados por el INEGI. Nivel de escolaridad.

Fuente: elaboración propia a partir de (Instituto Nacional de Estadística y Geografía [INEGI], 2010, 2020b).

Simbología: ▲Tendencia deseable ascendente, el resultado sí incrementó. △Tendencia deseable ascendente, el resultado disminuyó. ▼Tendencia deseable descendente, el resultado sí disminuyó. ▽Tendencia deseable descendente, el resultado incrementó.

Aunado a la tabla anterior, se observa que el porcentaje de población no escolarizada en Jalisco tuvo una disminución de 1.8 puntos porcentuales, mientras que el porcentaje de personas que solo cuentan con educación básica disminuyó 7.7 puntos porcentuales. Esto es deseable ya que refleja que cada vez hay más personas que asisten a las escuelas y aprueban niveles educativos, incrementando la asistencia en la escolaridad media y superior. En relación con lo anterior, el porcentaje de personas que asisten a educación media superior y educación superior tuvo un incremento de 5.2 y 5 puntos porcentuales, respectivamente, lo cual es deseable.

En cuanto a los logros de aprendizaje, se puede comparar los resultados de la prueba PLANEA 2018 a nivel nacional y PLANEA 2021 a nivel estatal. A continuación, se presenta el porcentaje de estudiantes por nivel de logro de aprendizaje, donde el Nivel I es insuficiente, Nivel II es básico, Nivel III es satisfactorio y Nivel IV es sobresaliente.

Áreas	Año	Nivel I	Nivel II	Nivel III	Nivel IV
Lenguaje y comunicación	2018	45.0%	34.5%	16.6%	4.0%
	2021	41.3%▼	26.4%▼	24.1%▲	8.3%▲
Matemáticas	2018	53.3%	18.7%	17.2%	11.7%
	2021	56.9%▽	15.7%▼	15.3%△	12.1%▲

Tabla 4. Estudiantes de sexto de primaria en Jalisco por logro de aprendizaje en PLANEA.

Fuente: elaboración propia a partir de (Instituto Nacional para la Evaluación de la Educación [INEE], 2018; Secretaría de Educación Jalisco [SEJ], 2021).

Simbología: ▲Tendencia deseable ascendente, el resultado sí incrementó. △Tendencia deseable ascendente, el resultado disminuyó. ▼Tendencia deseable descendente, el resultado sí disminuyó. ▽Tendencia deseable descendente, el resultado incrementó.

Al comparar los resultados en el área de Lenguaje y comunicación entre PLANEA 2018 y PLANEA 2021, se observa una tendencia positiva, ya que se redujo el porcentaje de estudiantes en los *niveles I y II*, mientras que aumentó en los *niveles III y IV*. En Matemáticas, destaca positivamente el aumento en el *Nivel IV*, aunque el aumento del *Nivel I* no es deseable.

Por otro lado, en relación con las habilidades socioemocionales, la SEJ implementó el Diagnóstico Socioemocional en 2020. Según los resultados, 35.5% de los estudiantes de primaria alta, 38.9% de secundaria y 43.5% de educación media superior tuvieron dificultades para controlar sus emociones durante la pandemia. En cuanto a comportamiento agresivo, 16.1%, 17.7% y 15.7% de estudiantes en esos niveles educativos reportaron haber actuado de manera agresiva. Además, el 47.6% de estudiantes de secundaria y 55.0% de educación media superior afirmaron que su estabilidad emocional empeoró. En contraste, los estudiantes de primaria dieron un promedio de calificación de 6.7 en su respuesta a la pregunta de qué tan bien se sentían emocionalmente (Secretaría de Educación Jalisco [SEJ], 2021).

Varios de los resultados de estos indicadores pueden interpretarse a partir de las repercusiones de la pandemia por COVID-19, como el cierre de escuelas y la transición a la educación a distancia. Algunos estudiantes experimentaron una comunicación irregular (Comisión Estatal para la Mejora Continua en Jalisco [CEMEJ], 2021) y, debido a condiciones socioeconómicas, algunos tuvieron que apoyar económicamente a sus familias y trabajar. Muchos estudiantes carecían de los recursos necesarios para continuar sus estudios, entre otros elementos (Instituto Nacional de Estadística y Geografía [INEGI], 2020a). Esto podría estar relacionado con la disminución en logros de aprendizaje, los impactos socioemocionales y una reducción en la cobertura educativa en los niveles preescolar y educación media superior.

Además, es relevante recordar que la SEP emitió los Acuerdos números 16/06/21 y 11/06/22, estableciendo que las calificaciones de estudiantes de educación primaria y secundaria no podrían ser inferiores a 6, lo que significa que no podrían reprobado (ACUERDO Número 16/06/21 Por El Que Se Regulan Las Acciones Especificas y Extraordinarias Relativas a La Conclusión Del Ciclo Escolar 2020-2021, En Beneficio de Los Educandos de Preescolar, Primaria y Secundaria Ante El Periodo de Contingencia Sanitaria Generada Por El Virus SARS-CoV2 (COVID-19), 2021; ACUERDO Número 11/06/22 Por El Que Se Regulan Las Acciones Especificas y Extraordinarias Para La Conclusión Del Ciclo Escolar 2021-2022 y El Inicio Del Ciclo Escolar 2022-2023., 2022). Esto puede explicar por qué los resultados de reprobación, repetición y abandono escolar disminuyen, mientras que los de eficiencia terminal y tasa de terminación aumentan en primaria y secundaria. Sin embargo, en educación media superior, donde no se establecieron especificaciones de acreditación especiales, los indicadores muestran lo contrario. Solo los resultados de abandono escolar de educación media superior generan preguntas, ya que mejoran, lo cual podría deberse al método de cálculo empleado.

## 6. Discusión y conclusiones

Como se muestra en las tablas 1, 2 y 3 del apartado anterior, se destaca que la mayoría de los indicadores (reprobación, abandono escolar, eficiencia terminal y tasa de terminación) presenta una tendencia favorable en los niveles educativos de primaria y secundaria, así como en aquellos que abarcan a la población en general, como los indicadores reportados por el INEGI: analfabetismo, grado promedio de escolaridad y nivel de escolaridad. Sin embargo, es crucial dirigir una atención especial hacia los niveles de preescolar y media superior, ya que en la mayoría de los indicadores de estos niveles se observa una tendencia no deseable.

Además, es importante señalar que la comprensión del impacto generado por la pandemia resulta imprecisa debido a que las últimas cifras oficiales (no preliminares ni estimadas) se refieren a ciclos escolares en los cuales los estudiantes no estuvieron presentes de manera presencial en su totalidad.

Esta situación resalta la necesidad de contar con mecanismos ágiles, sólidos y oportunos para la recolección, análisis y difusión de información a través del uso e innovación de herramientas digitales, así como del fomento de una cultura de ética, responsabilidad y rendición de cuentas para eficientar, consolidar y fortalecer los procesos de transparencia de información. Esto permitiría que las cifras preliminares y estimadas fueran reportadas oficialmente por las instancias gubernamentales correspondientes y que contáramos con información actualizada. Asimismo, es fundamental seguir evaluando y sistematizando datos de indicadores que, a través del tiempo, ofrezcan una comprensión completa del contexto y estado actual de la educación, garantizando así las condiciones básicas para este derecho.

En particular, se vuelve necesario hacer hincapié en la relevancia de diseñar indicadores que se adapten a las necesidades contextuales, como la inclusión de indicadores que den seguimiento a las condiciones socioemocionales, de inclusión, equidad, etc. Estos son esenciales para el seguimiento de temas fundamentales, dado que el bienestar integral de los estudiantes influye de manera significativa en sus logros de aprendizaje. Además, es esencial contar con instrumentos de evaluación que permitan comparaciones sólidas y consistentes a lo largo del tiempo.

Con relación a los indicadores restantes, es esencial mantener un acceso público y accesible a los resultados con rastreabilidad histórica. Esto permitirá un seguimiento continuo de los avances y desafíos en la educación, tanto para

la ciudadanía como para la toma de decisiones basadas en evidencia. Esta información contribuirá a la implementación de acciones integrales que mejoren los aprendizajes, la permanencia estudiantil, el bienestar socioemocional y la eficiencia terminal, con el objetivo de maximizar el logro de aprendizajes de NNA y permitirles desarrollarse plenamente.

Otro punto de importancia radica en la variabilidad encontrada al analizar los datos presentados por instituciones oficiales. Se han identificado inconsistencias en las cifras reportadas en y entre algunas de estas fuentes, donde los resultados de algunos indicadores para los mismos ciclos escolares y niveles educativos difieren. Por ejemplo, el porcentaje de eficiencia terminal en educación media superior para el ciclo 2018-2019 es: 47.5% según la publicación de 2019 de la SEP y 67.3% según la publicación de 2020, también de la SEP (P. y E. [DGPPyEE] Dirección General de Planeación, 2019, 2020). Esto subraya la importancia de que las instancias correspondientes revisen la metodología utilizada para generar información de estos indicadores y establezcan acuerdos en la medición y reporte de resultados. Además, se recomienda mantener la información pública, ya que en ocasiones las fuentes originales de datos han dejado de estar disponibles o han modificado sus plataformas electrónicas, dificultando su consulta.

El micrositio resultante de esta investigación, además de proporcionar una visión histórica del estado y progreso de diversos indicadores educativos, y de ofrecer una herramienta interactiva para la consulta de datos municipales y regionales, brinda certeza en cuanto al registro de datos oficiales. Esto asegura la preservación de la información recopilada, independientemente de los cambios de administración. Y coloca la mejora de la educación en el centro de la agenda pública en diversos sectores y ámbitos sociales.

En última instancia, es necesario reconocer el esfuerzo realizado por diversas instituciones para generar, analizar y publicar información sobre diversos indicadores educativos. Esto permite que la sociedad en general comprenda el estado de la educación y evalúe en qué medida el sistema educativo, tanto estatal como nacional, ha implementado políticas que promueven mejoras, estancamiento o retroceso en la educación. Por lo tanto, se considera que la publicación oportuna, transparente y en formatos accesibles fortalece la democracia y la participación ciudadana.

La información educativa debe adherirse a los principios de transparencia, lo que significa que debe ser accesible, comprensible y clara para la ciudadanía. Además, debe ser de alta calidad, recopilada, calculada y analizada con base en sólidos lineamientos técnicos, confiables y de alto nivel. También debe ser relevante, de modo que resulte útil para diversos usuarios (sociedad, comunidad escolar, investigadores, académicos, empresarios, tomadores de decisiones, etc.) para construir opiniones, participar y tomar decisiones informadas. Por último, debe ser oportuna, es decir, obtenida y divulgada de manera puntual y actualizada.

## 7. Referencias

Arias Pérez, J. E. & Aristizábal Botero, C. A. (2011). El dato, la información, el conocimiento y su productividad en empresas del sector público de Medellín. *Semestre Económico*, 14(28), 95–109. <https://revistas.udem.edu.co/index.php/economico/article/view/370/329>

Ley del Sistema Nacional de Información Estadística y Geográfica, Pub. L. No. 18-12–2015, Diario Oficial de la Federación (2015). [https://sc.inegi.org.mx/repositorioNormateca/Lmj\\_SNIEG.pdf](https://sc.inegi.org.mx/repositorioNormateca/Lmj_SNIEG.pdf)

Ley General de Educación, Pub. L. No. 30-09–2019, Diario Oficial de la Federación (2019). <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGE.pdf>

Comisión Estatal para la Mejora Continua en Jalisco [CEMEJ]. (2021). Retos educativos en Jalisco para el logro y permanencia en pandemia. Documento ejecutivo. <https://recrea.jalisco.gob.mx/cemej/wp-content/uploads/sites/37/2023/07/Retos-educativos-en-Jalisco-documento-ejecutivo.pdf>

Comisión Nacional para la Mejora Continua de la Educación [MEJORED]. (2022). Indicadores nacionales de la mejora continua de la educación en México. Cifras del ciclo escolar 2020-2021. Principales hallazgos. <https://www.mejoredu.gob.mx/images/publicaciones/principales-hallazgos-22.pdf>

Dirección General de Planeación, P. y E. [DGPPyEE]. (2019). Principales Cifras del Sistema Educativo Nacional 2018-2019. Secretaría de Educación Pública [SEP]. [https://www.planeacion.sep.gob.mx/Doc/estadistica\\_e\\_indicadores/principales\\_cifras/principales\\_cifras\\_2018\\_2019\\_bolsillo.pdf](https://www.planeacion.sep.gob.mx/Doc/estadistica_e_indicadores/principales_cifras/principales_cifras_2018_2019_bolsillo.pdf)

Dirección General de Planeación, P. y E. [DGPPyEE]. (2020). Principales Cifras del Sistema Educativo Nacional 2019-2020. Secretaría de Educación Pública [SEP]. [https://www.planeacion.sep.gob.mx/Doc/estadistica\\_e\\_indicadores/principales\\_cifras/principales\\_cifras\\_2019\\_2020\\_bolsillo.pdf](https://www.planeacion.sep.gob.mx/Doc/estadistica_e_indicadores/principales_cifras/principales_cifras_2019_2020_bolsillo.pdf)

Dirección General de Planeación, P. y E. [DGPPyEE]. (2022a). Información Estadística e Indicadores Educativos. Gobierno de México. <https://www.planeacion.sep.gob.mx/estadisticaeindicadores.aspx>

Dirección General de Planeación, P. y E. E. [DGPPyEE]. (2022b). Sistema Interactivo de Consulta de Estadística Educativa. <https://www.planeacion.sep.gob.mx/principalescifras/>

García, M. (2018). Análisis descriptivo: qué es y las capacidades que necesitas. Tecnológico de Monterrey. <https://blog.maestriasydiplomados.tec.mx/analisis-descriptivo-que-es-y-las-capacidades-que-necesitas>

Instituto Mexicano para la Competitividad [IMCO]. (2022). ¿La educación es prioridad para los estados? Acciones para mitigar el impacto de la pandemia. [https://imco.org.mx/wp-content/uploads/2022/08/Acciones-estatales-educativas-para-mitigar-el-impacto-de-la-pandemia\\_IMCO.pdf](https://imco.org.mx/wp-content/uploads/2022/08/Acciones-estatales-educativas-para-mitigar-el-impacto-de-la-pandemia_IMCO.pdf)

Instituto Nacional de Estadística y Geografía [INEGI]. (n.d.). Escolaridad. Retrieved January 15, 2024, from <https://cuentame.inegi.org.mx/poblacion/escolaridad.aspx?tema=P>

Instituto Nacional de Estadística y Geografía [INEGI]. (2010). Censo de Población y Vivienda 2010. <https://www.inegi.org.mx/programas/ccpv/2010/#Tabulados>

Instituto Nacional de Estadística y Geografía [INEGI]. (2020a). Encuesta para la Medición del Impacto COVID-19 en la Educación (ECOVID-ED). [https://www.inegi.org.mx/contenidos/investigacion/ecovid/2020/doc/ecovid\\_ed\\_2020\\_presentacion\\_resultados.pdf](https://www.inegi.org.mx/contenidos/investigacion/ecovid/2020/doc/ecovid_ed_2020_presentacion_resultados.pdf)

Instituto Nacional de Estadística y Geografía [INEGI]. (2020b). Censo de Población y Vivienda 2020. <https://www.inegi.org.mx/programas/ccpv/2020/#Tabulados>

Instituto Nacional de Estadística y Geografía [INEGI]. (2021). Características educativas de la población. [https://www.inegi.org.mx/temas/educacion/#Informacion\\_general](https://www.inegi.org.mx/temas/educacion/#Informacion_general)

Instituto Nacional para la Evaluación de la Educación [INEE]. (2015). Programa para la Evaluación Internacional de los Estudiantes. Instituto Nacional Para La Evaluación de La Educación [INEE]. . <https://www.inee.edu.mx/evaluaciones/pisa/>

Instituto Nacional para la Evaluación de la Educación [INEE]. (2017a). Panorama Educativo de México. Indicadores del Sistema Educativo Nacional 2017. Educación básica y media superior. <https://www.inee.edu.mx/wp-content/uploads/2018/12/P1B116.pdf>

Instituto Nacional para la Evaluación de la Educación [INEE]. (2017b). Evaluación de Condiciones Básicas para la Enseñanza y el Aprendizaje. Instituto Nacional Para La Evaluación de La Educación [INEE]. <https://www.inee.edu.mx/evaluaciones/ecea/#:~:text=La%20Evaluaci%C3%B3n%20de%20las%20Condiciones,para%20su%20operaci%C3%B3n%20y%20funcionamiento.>

Instituto Nacional para la Evaluación de la Educación [INEE]. (2018). Planea: Resultados nacionales 2018. 6o de primaria. [https://www.inee.edu.mx/images/stories/2018/planea/PLANEA06\\_Rueda\\_de\\_prensa\\_27nov2018.pdf](https://www.inee.edu.mx/images/stories/2018/planea/PLANEA06_Rueda_de_prensa_27nov2018.pdf)

Instituto Nacional para la Evaluación de la Educación [INEE]. (2019). La educación obligatoria en México. Informe 2019. <https://www.inee.edu.mx/wp-content/uploads/2019/04/P11245.pdf>

Marrero García, M. M. (2013). La importancia de la información para la toma de decisiones. REMIJ, 14(2), 1–3. <https://remij.sld.cu/index.php/remij/article/view/68/143>

Mexicanos Primero. (2018). ICRE Índice de Cumplimiento de la Responsabilidad Educativa. Estatal 2018. [https://www.mexicanosprimero.org/wp-content/uploads/2022/07/icre\\_2018.pdf](https://www.mexicanosprimero.org/wp-content/uploads/2022/07/icre_2018.pdf)

Miranda López, F., Mendieta Melgar, G. M., Rivera Navarro, H. O., Othón Rivera Navarro, H., Islas Dossetti, J. M., Cabrero Iriberry, I. & Vázquez Viquez, Ma. E. (2020). Experiencias de las comunidades educativas durante la contingencia sanitaria por covid-19. Educación básica. Informe ejecutivo 2020 Coordinación general. <https://editorial.mejoredu.gob.mx/ResumenEjecutivo-experiencias.pdf>

Monitoreo de Indicadores del Desarrollo [MIDE] Jalisco. (2016). ¿Qué es MIDE? Monitoreo de Indicadores del Desarrollo de Jalisco. <https://mide.jalisco.gob.mx/mide/panelCiudadano/inicio>

Monitoreo de Indicadores del Desarrollo [MIDE] Jalisco. (2022). Búsqueda. <https://mide.jalisco.gob.mx/mide/panelCiudadano/>

# Control de acceso biométrico a espacios públicos con base en tecnologías emergentes

María Esmeralda Arreola Marín, Mariela Chávez Marcial & José Iraic Alcantar Al-cantar  
Tecnológico Nacional de México / 2ITS de Ciudad Hidalgo, México  
marreola@cdhidalgo.tecnm.mx, mchavez@cdhidalgo.tecnm.mx, jiraic@cdhidalgo.tecnm.mx

**Abstract** The purpose of this research is to generate knowledge about a biometric system. Through this system, we sought to control access to both public and private institutions. Efficient technology was integrated to allow accurate identification of people without the need to enter passwords, as well as the realization of quick access actions, including temperature measurement. In addition, the security of the technological equipment, which is difficult to breach, stands out. This automated system aims to provide secure and efficient access to both the parking lot and the main entrance of the Instituto Tecnológico Superior de Cd. Hidalgo (ITSCH), where both workers and students are registered. This translates into accurate, fast, secure, hygienic and personalized access. Compared to other biometric systems, facial recognition has proven to be an accurate identification technology for companies, whether public or private. Its algorithmic method of automatic programming, responsible for verifying and recognizing a person's identity from their facial features, provides greater security, control and accuracy in facial data, which is unique to each individual. This research focused on a qualitative approach to evaluate the fulfillment of the operational functionalities of the facial recognition system. Prototyping was developed in order to measure the fulfillment of the objectives. To carry out this study, the TensorFlow library for facial recognition and the Python programming language were used, using the prototype model. The result obtained is a biometric facial recognition system capable of registering the entrances and exits of students and employees of the university.

**Keywords:** *biometrics, facial recognition, TensorFlow, CNN, microcontrollers.*

## 1. Introducción

Tras la pandemia, muchas instituciones públicas y privadas actualizaron sus sistemas de acceso, incorporando tecnología eficiente para minimizar el contacto físico y eliminar la manipulación directa por parte del usuario. Se buscó evitar que los antiguos registros en máquinas de tickets, bitácoras o tarjetas de identificación se convirtieran en posibles focos de contagio.

Los beneficios de la implementación de sistemas biométricos de acceso son notables. Incluyen una mayor precisión en la identificación, la promoción de la higiene, la eliminación de la necesidad de ingresar contraseñas y la agilización de acciones como la medición de la temperatura y el registro, incluso cuando se utiliza cubrebocas. Además, estos sistemas son altamente seguros y difíciles de vulnerar.

Este sistema automatizado garantiza un acceso seguro y eficiente tanto al estacionamiento como a la entrada principal de la institución, en este caso, del Instituto Tecnológico Superior de Cd. Hidalgo (ITSCH), donde se encuentran registrados tanto estudiantes como trabajadores. Esto se traduce en un acceso preciso, rápido, seguro, higiénico y personalizado. Si una persona ajena a la institución desea ingresar, deberá ponerse en contacto con los responsables para obtener la autorización necesaria, lo que garantiza tanto la seguridad como el control del tráfico de visitantes.

En comparación con otros sistemas biométricos, el reconocimiento facial se ha convertido en una tecnología de identificación precisa para diversas empresas, ya sean públicas o privadas. Su método algorítmico de programación automática se encarga de verificar y reconocer la identidad de una persona a partir de sus rasgos faciales, proporcionando un alto nivel de seguridad, control y precisión en los datos faciales, que son únicos para cada individuo.

El desarrollo está basado en una combinación de tecnologías de internet de las cosas (IoT) con inteligencia artificial, de tal forma que desde un dispositivo en la entrada se permita enviar una imagen a un servidor, que, mediante una red neuronal, identifique a la persona.

Esta investigación, además de ofrecer precisión, calidad y eficiencia en su servicio, puede resultar útil tanto para empresas privadas como para instituciones de diferentes sectores, ya que puede adaptarse a las necesidades específicas de cada uno. Esto lo convierte en una exploración con un potencial significativo para la integración de tecnología vanguardista.

Para el desarrollo de este sistema biométrico, se empleó el diseño de un prototipo siguiendo el "Modelo de Prototipos", también conocido como modelo de desarrollo evolutivo, que se utiliza principalmente en proyectos de desarrollo de software. El proceso comienza con la definición de los objetivos, tanto generales como específicos para el software. Posteriormente, se identifican los requerimientos y las áreas del proyecto que requieren una mayor conceptualización. Los procesos de prototipado están estrechamente relacionados con la mejora continua y siguen el ciclo de Deming, que

se enfoca en un proceso iterativo diseñado para planificar, implementar, medir y ajustar. La figura 1 muestra la representación de la arquitectura del prototipo.

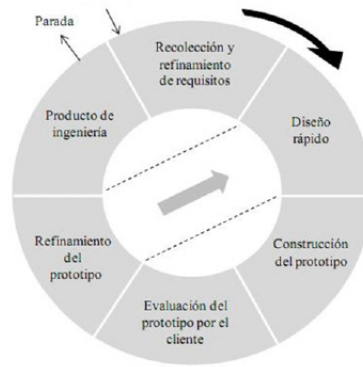


Figura 1. Arquitectura de prototipo

El proceso comienza con la definición de los requerimientos y las variables a explorar, identificando los elementos necesarios para el diseño del prototipo. Al concebir el prototipo, se analizan aspectos como el tipo de prototipo más adecuado, y se evalúan factores como el diseño, el montaje, la ergonomía, los materiales, las formas y las dimensiones, entre otros. Se procede al testeo, donde se presenta el prototipo a los usuarios, se observa y registra su rendimiento y se les permite aportar sus ideas. Una vez completado este proceso, se examinan los resultados y los aprendizajes adquiridos, lo que permite perfeccionar el concepto de solución.

## 2. Marco Teórico

El desarrollo de la investigación implica el uso de redes neuronales, para el proceso de clasificación de imágenes en este caso específico redes neuronales convolucionales.

Las redes neuronales convolucionales son un subconjunto de aprendizaje automático y están en el centro de los algoritmos de aprendizaje profundo. Estas estructuras se componen de capas de nodos, englobando una capa de entrada, una o más capas ocultas y, al fin, una capa de salida. Cada nodo se encuentra interconectado con otros, cargando consigo un peso y un umbral específico. Si el rendimiento de cualquier nodo individual supera el umbral prescrito, este se activa, transmitiendo así datos a la subsiguiente capa de la red; de lo contrario, se omite el envío de información.

Las redes neuronales convolucionales, abreviadas como ConvNets o CNN, hallan su máximo esplendor en tareas de clasificación y visión artificial. Destacan entre las demás variantes neuronales por su notoria aptitud para procesar entradas de señales de imagen, voz o sonido.

Las CNN orquestan el discernimiento de imágenes y las arduas labores de visión artificial, un nicho de la inteligencia artificial (IA) que permite a sistemas y dispositivos extraer conocimiento sustancial de imágenes digitales, vídeos y otros registros visuales. En la actualidad, existen diversa forma de llevar a cabo la programación de estas, en ese sentido TensorFlow es una famosa plataforma de código abierto que permite facilitar esto, generando así un modelo neuronal capaz de realizar el reconocimiento de imágenes.

Este software, fruto del ingenio de Google, proporciona una amplia gama de herramientas y recursos para diseñar, entrenar y evaluar modelos de aprendizaje profundo. Esto incluye capas predefinidas para construir redes neuronales, optimizadores para ajustar los parámetros del modelo, y herramientas para visualizar y analizar el proceso de entrenamiento y los resultados. Además de la biblioteca principal, cuenta con una API para JavaScript y otra para dispositivos IoT y móviles.

Por otro lado, para la captura de fotos, se empleó una cámara conectada a un microcontrolador en cual fue programado bajo la plataforma Arduino.

Arduino, su concepción resulta de la amalgama de tres elementos esenciales:

- Una placa de hardware de código abierto, con diversos modelos de placas de desarrollo con microcontroladores.
- Software gratuito y multiplataforma, que incluye un entorno de desarrollo (IDE) para escribir, compilar y cargar código en el microcontrolador.
- Un lenguaje de programación de acceso libre, basado en C y enriquecido con múltiples bibliotecas que simplifican la programación de microcontroladores.

Un **microcontrolador** se puede considerar como un ordenador en un único chip. La palabra micro indica que el dispositivo es pequeño, y controlador indica que el dispositivo se puede usar en aplicaciones de control, esto son capaces de trabajar con sensores y actuadores.

Un **sensor** es un dispositivo capaz de detectar magnitudes físicas o químicas, llamadas variables de instrumentación, y transformarlas en variables eléctricas. Por ejemplo, las cámaras.

Los **actuadores** son los componentes que interactúan con el mundo físico externo a la placa. Los actuadores transforman la señal eléctrica en energía que actúa directamente con el mundo, por ejemplo, los motores.

La biometría, se trata de una ciencia dedicada al análisis estadístico de las características biológicas. Se emplea en el contexto de evaluación de rasgos humanos con fines de seguridad. En ese sentido, el modelo CNN se encarga de almacenar los rasgos de la cara, relacionándolos a una etiqueta o identificador, en este caso se usó el número de control de cada alumno o trabajador, el cual a su vez está almacenado en una base de datos relacional, con toda la información adicional, donde se puede obtener datos como nombre, y si es activo actualmente.

A continuación, se presenta el modelo relacional, que representa la relación entre las tablas contenidas en la base de datos y la interacción, conexión y flujo de datos de cada tabla. Este modelo proporciona una visión unificada de los datos y un entendimiento claro de la función del sistema propuesto. Se puede consultar la figura 2 misma que representa el modelo relacional.

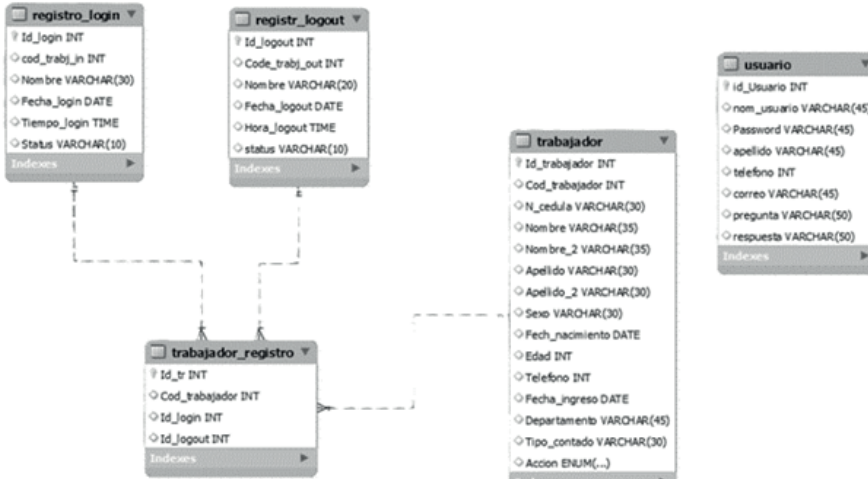


Figura 2. Modelo relacional de la DB del sistema biométrico

### 3. Métodos

Esta indagación se enfocó en la aplicabilidad de la tecnología de reconocimiento facial para resolver un desafío que en el ITSCH se presentó, siguiendo un enfoque cualitativo para definir el cumplimiento de las funcionalidades operativas del sistema propuesto y determinar qué tan efectivo es el reconocimiento de rostros utilizando una CNN con una cámara genérica a baja resolución. De esta forma, la estructura de sistema general es la siguiente, ver figura 3.

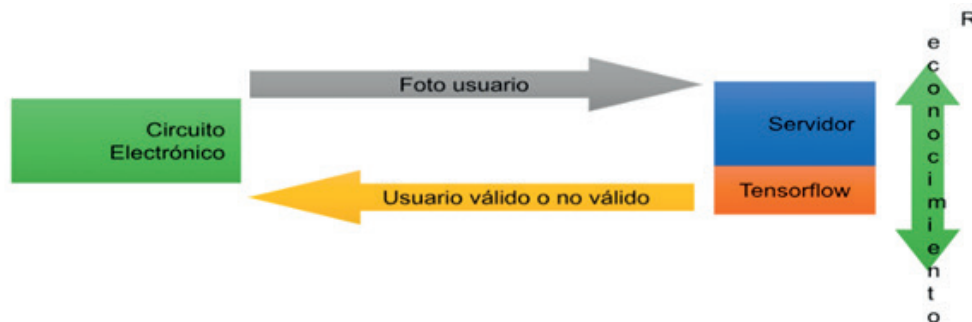


Figura 3. Esquema funcional

1. Un circuito electrónico en la entrada toma una captura de imagen y la envía a través de la red a un servidor, mediante protocolo HTTP.
2. El servidor recibe la imagen y la somete al modelo de red neuronal previamente entrenado con las fotos y a quién corresponde cada foto, de manera que puede determinar quién es.
3. El servidor envía la respuesta, ya sea indicando si el usuario, alumno o trabajador en este caso, es válido o en caso contrario indicar si no se reconoce.
4. El circuito electrónico recibe la respuesta y en base a ella decidirá si permite la entrada, activando el servomotor.

La recolección de requisitos se realizó mediante encuestas al personal y estudiantes que utilizan los sistemas de entrada y salida actuales en la institución, así como a través de observación directa e indirecta del proceso de control de accesos y análisis de documentos, lo que proporcionó una base sólida para el desarrollo. En este sentido, se determinó que el prototipo del circuito electrónico utilizara un microcontrolador con capacidad wifi, conectado a los sensores y actuadores ya mencionados, ver figura 4.



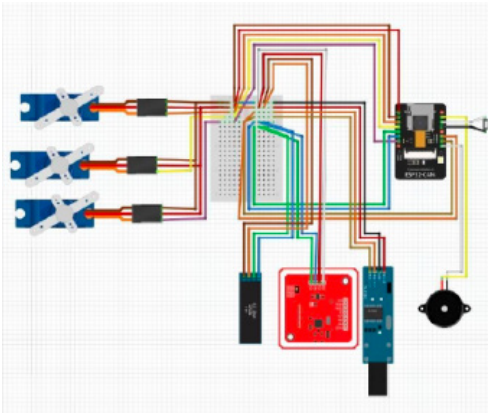


Figura 4. Módulo de sensor Arduino

**Microcontrolador:** Aunque la programación del microcontrolador, se empleó el software y el lenguaje de Arduino, ninguna placa Arduino fue necesaria, ya que se optó por el ESP3, el cual, además de ofrecer prestaciones avanzadas, se ajustaba perfectamente al propósito del prototipo, destacando la capacidad de comunicación inalámbrica WiFi, necesaria para la comunicación con el servidor. La placa elegida es ESP32-Cam, que incluye conexión a la cámara OV2640, como se aprecia en la figura 5.



Figura 5. Captura de imagen

**Sensores:**

1. PN532 NFC. Es un circuito integrado desarrollado por NXP Semiconductors que habilita la tecnología NFC para la comunicación inalámbrica entre dispositivos compatibles, a través de radiofrecuencia. Ofrece modos de operación versátiles para aplicaciones como control de acceso mediante tarjetas RFID, pudiendo leer y escribir chip de tarjetas que soporten esta tecnología. En el prototipo, se integró como medida adicional de acceso.
2. Cámara OV2640. Es un sensor de imagen CMOS (Complementary Metal-Oxide-Semiconductor) fabricado por OmniVision Technologies. Este sensor está diseñado para capturar imágenes fijas y video en aplicaciones electrónicas. Presenta una resolución de 2 megapíxeles. Integrada en la placa Esp32 Cam y en el prototipo fue utilizada para tomar la captura de imágenes desde el microcontrolador.
3. PIR. Se trata de un sensor de infrarrojos, que permite determinar la presencia de alguna persona, en este caso usada para saber si se toma la foto.
4. Botón. Un simple botón que al presionar permite el paso de una señal, se utilizó como alternativa para que un guardia pueda permitir el paso a personas visitantes o externas que no puedan ser reconocidas por el sistema.

**Actuadores:** Son los componentes que interactúan con el mundo físico externo a la placa. Los actuadores transforman la señal eléctrica en energía que actúa directamente con el mundo. Los actuadores utilizados son:

1. Servomotores MG90. Es un actuador de tamaño compacto y bajo costo. Suele tener una estructura interna que incluye un motor eléctrico, una caja de engranajes y un circuito de control que permite recibir señales de control, mediante pulsos digitales, para establecer su posición. Puede rotar en un rango específico de grados (generalmente entre 0 y 180 grados) en respuesta a las señales de control enviadas por un microcontrolador u otro dispositivo. Aunque en el circuito se usaron tres, para simular las puertas de acceso principal solo se usó el primero, ya que los otros dos implican tener una cámara para cada uno.
2. Oled 0.91". Una pantalla OLED con un tamaño de 0.91 pulgadas. Las pantallas OLED (Organic Light Emitting Diode) son un tipo de tecnología de visualización que utiliza materiales orgánicos para emitir luz y crear imágenes. Son conocidas por ofrecer colores vibrantes, negros profundos y un alto contraste en comparación con otras tecnologías de pantallas. El objetivo de usar esta pequeña pantalla es mostrar mensajes sencillos del estado de funcionamiento del microcontrolador.
3. Buzzer o zumbador. Es un pequeño altavoz que es capaz de vibrar si se le aplica un pulso eléctrico (una onda

digital o cuadrada) emitiendo un sonido constante a cierta frecuencia, utilizado para enviar mensajes sonoros a los usuarios.

La fase de modelado permitió crear un diseño preliminar del sistema, estableciendo así los cimientos para el diseño final. Luego, se fabricó un prototipo en 3D (figura 6) y se comenzaron las pruebas para entrenar el algoritmo, avanzando hacia la fase de desarrollo y retroalimentación del producto final.

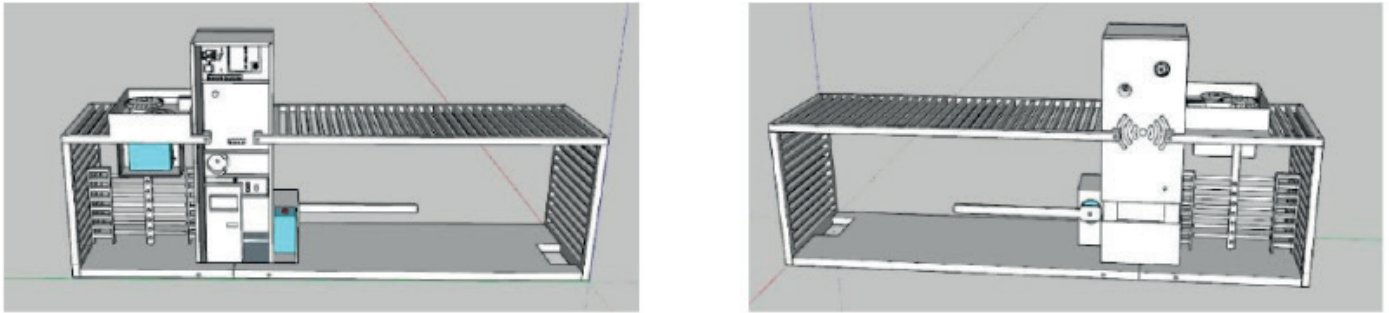


Figura 6. Maquetación de entrada

Este prototipo permite determinar cuándo tomar una foto y enviar al servidor, esperando la respuesta para establecer si se apertura la puerta o no, bajo la siguiente lógica:

```

Si (verificarPresencia() == Verdadero)
    foto = tomarFoto()
    enviarFotoAlServidor(foto)
    respuesta = esperarRespuesta()

    Si (respuesta == "valida") {
        activarServo()
    }
    Sino
        emitirSeñalDeError()
    Fin si
Fin si
    
```

Las principales funciones que se desarrollaron en esta exploración fueron: el desarrollo de un prototipo funcional de la entrada principal a la Institución y un sistema de reconocimiento usando redes neuronales. El modelo entidad-relación, representa la relación entre entidades con sus respectivos atributos; este determina los datos que se manejarán en el sistema. Cada una de ellas guarda información importante para la correcta funcionalidad del sistema. Además, un algoritmo que, mediante técnicas de comparación, permite determinar el reconocimiento facial de las personas, figura 7

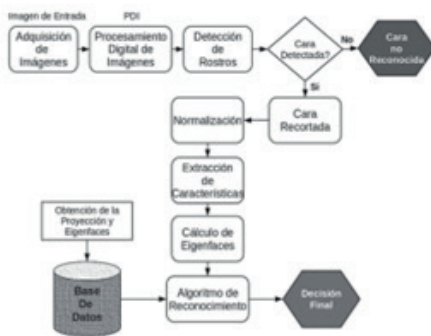


Figura 7. Algoritmo de detección de rostros del sistema de reconocimiento facial

Para este algoritmo de reconocimiento se implementó una CNN en TensorFlow. Desafortunadamente la API para dispositivos IoT, TensorFlow Lite, no se implementó directamente en el microcontrolador debido a que las capacidades para almacenar y procesar son insuficientes para este caso, ya que se pretende aplicar para alrededor de 2000 personas entre Alumnos y Trabajadores de la institución, por lo que, se utilizó la biblioteca principal mediante el lenguaje de programación JavaScript y Python para implementar el modelo un servidor.

Para realizar el entrenamiento del modelo de CNN, se utilizaron fotos de prueba de un grupo de estudiantes todas estas debidamente etiquetadas con su número de control a una resolución de 512x512 pixeles, para posteriormente mandarlas al servidor y este realice el entrenamiento de la red neuronal, esto mediante una interfaz web, ver figura 8.



Figura 8. Esquema de entrenamiento

Finalmente, se obtuvo un sistema de reconocimiento facial basado en la automatización de entradas y salidas de los estudiantes y personal del ITSCH. El producto resultante cuenta con seis funciones principales: registrar características faciales, identificar y reconocer a un estudiante y/o trabajador ya registrado en el sistema, identificar los rostros no registrados, registrar la hora de entradas y salidas, llevar control de registro de datos de un trabajador (administrativo o docente) y generar reportes (figura 9).

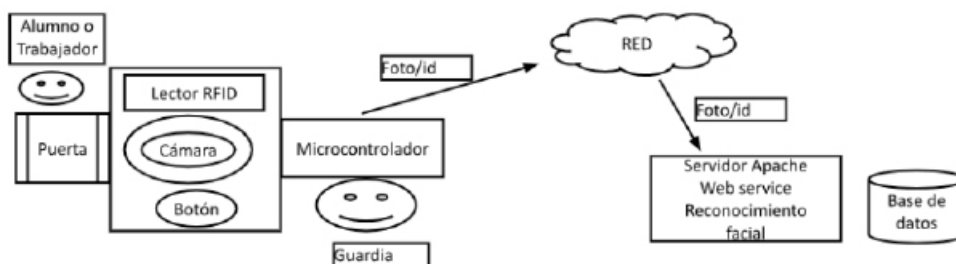


Figura 9. Proceso biométrico final

Los resultados experimentales fueron satisfactorios, las funcionalidades propuestas fueron ejecutadas de manera correcta. En la tabla 1 se aprecian las evaluaciones resultantes.

Funcionalidades	Cumplida	No cumplida
Registro de características faciales	X	
Identificación de rostros	X	
Registro de Entradas y salidas	X	
Detección de rostros no registrados	X	
Control de Registro	X	
Reportes	X	

Tabla 1. Cumplimiento de funcionalidades del sistema de reconocimiento facial

#### 4. Resultados

El sistema será capaz de tener un control en las entradas y salidas de estacionamientos públicos y privados. El equipo de trabajo realizó lo siguiente:

Maquetación. El equipo de trabajo realizó el prototipo en 3d, se imprimió para cumplir el objetivo del proyecto y se implementó la instalación de sensores que tienen la capacidad de controlar los procesos que llevarán a cabo el prototipo. Se realizaron pruebas, figura 10 y 11.

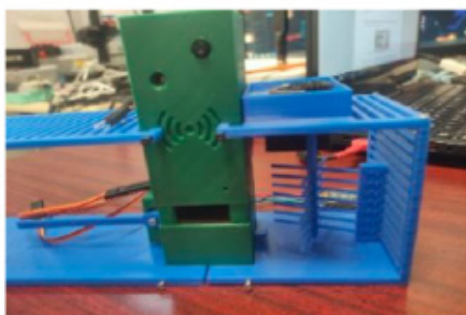


Figura 10. Instalando focos led

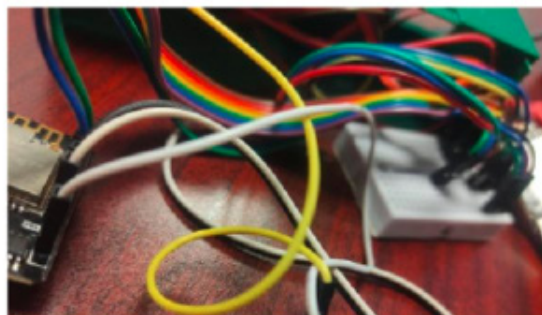


Figura 11. Instalación de cámara y conexión

Programación. Se desarrolló la programación del Arduino, los servomotores y del código para el sistema biométrico. En este último se realizó el entrenamiento del modelo de Red Neuronal Convolutacional para las pruebas bajo los siguientes parámetros:

- Se utilizó un modelo Secuencial
- Fotos de 33 alumnos a una resolución de 512 \* 512 píxeles, ver figura 12
- En este caso cada alumno corresponde a una clase, relacionada a su número de control
- Tres capas convolucionales (de 32, 64 y 128 con activación "relu")
- Una capa de salida con activación "softmax"
- Para el optimizador se utilizó "Adam" y función de pérdida "Sparse Categorical Cross Entropy"



Figura 12. Resultado de entrenamiento: precisión y pérdida.

Se realizó un entrenamiento durante 1000 épocas donde se puede apreciar que la mayor precisión se logra a partir de la época 750 con un 90% (línea azul) mientras que la precisión en la validación fue más baja, de un 50% aproximadamente (línea naranja), como se puede ver en el gráfico figura 13.

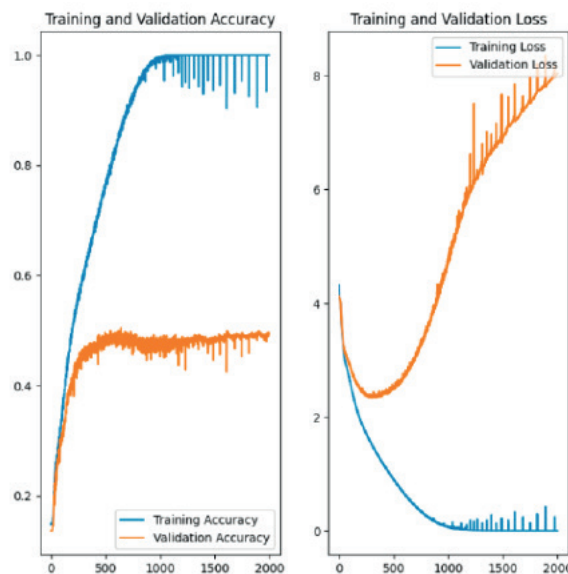


Figura 13. Resultado de entrenamiento: precisión y pérdida.

Finalmente, se logró que la red neuronal pueda devolver el número de control reconocido según el rostro o foto validada con una precisión promedio de 50%.

## 5. Discusión

La automatización de accesos a estacionamientos y las entradas principales a instalaciones se vislumbra como un tema de gran relevancia en las próximas décadas. El desarrollo de estos sistemas no solo contribuye a la seguridad, sino que también promueve el ahorro energético y mejora la comodidad. Los resultados experimentales de este proyecto demuestran que su desarrollo e implementación podrían abordar los desafíos que enfrentan las personas que son parte del ITSCH.

Un ejemplo significativo en esta línea proviene del trabajo de Calles Carrasco & Becerra Arévalo (2019), quienes crearon un sistema informático de reconocimiento facial para el registro y control de asistencia de los socios de la cooperativa de taxis y camionetas Puyo. Este sistema permite mantener un registro preciso de las asistencias de los socios de la cooperativa y emplea el algoritmo de Viola-Jones, la librería TensorFlow de Python y SQL Server como gestor de base de datos.

En una línea similar, Wang & Siddique (2020) proponen un sistema de reconocimiento facial que utiliza el reconocedor facial de histograma de patrones binarios locales montado en tecnología de drones. Los autores sostienen que este sistema puede identificar a las personas con una precisión cercana al 89.1%. Un antecedente adicional es el trabajo de Portilla Jiménez (2018), quien desarrolló un sistema de reconocimiento facial para la detección e identificación de intrusos. Utilizó la metodología de prototipo, el algoritmo de método comparativo, PhpMyAdmin como gestor de base de datos (XAMPP), Python y la librería OpenCV.

Es innegable que el campo del reconocimiento facial ha experimentado un avance significativo, con múltiples algoritmos y metodologías propuestas para sistemas basados en esta tecnología. Sin embargo, en el desarrollo de este sistema particular, se optó por Python como lenguaje de programación y se incorporó la librería TensorFlow, que desempeña una función crucial en el entrenamiento de la visión por computadora. Además, se empleó MySQL como administrador de la base de datos del sistema, encargado de almacenar los datos de todos los usuarios registrados. Para la interfaz del sistema, se recurrió a Tkinter, una herramienta multiplataforma que se incluye con Python y se considera un estándar en el diseño de interfaces gráficas de usuario. Esto garantiza un funcionamiento eficiente y de alta calidad, con la posibilidad de ser implementado tan pronto como las autoridades de la institución den su aprobación.

## 6. Conclusiones

El sistema propuesto se basó en el concepto de sistema biométrico para el reconocimiento facial, utilizando un reconocedor facial incorporado en la librería TensorFlow. Este sistema fue entrenado para gestionar las entradas y salidas de estudiantes y empleados del ITSCH.

Aunque las funcionalidades se cumplieron satisfactoriamente, es importante considerar que los experimentos se llevaron a cabo en un entorno controlado y presentaron limitaciones en términos de la calidad de las cámaras y la iluminación. El éxito de la implementación depende en gran medida de aspectos como la iluminación, que permite que el sistema reconozca con precisión los rasgos faciales, y de la calidad de la cámara digital, ya que menos estática y ruido en las imágenes aumenta la precisión del sistema. El sistema propuesto representa una mejora beneficiosa para el control de registro de entradas y salidas de estudiantes y trabajadores de la institución. Como plan a futuro, se busca mejorar la precisión del sistema, esto aumentando el número de imágenes para el entrenamiento, de manera que el modelo pueda generalizar más características, además, se tiene la intención de combinarlo con un modelo ya entrenado, aplicando el aprendizaje por transferencia.

La implementación de este proyecto aportará beneficios tanto a empresas públicas como privadas que requieran un control eficiente de las entradas y salidas, así como a aquellas que buscan mejorar el acceso a sus instalaciones. El Instituto Tecnológico de Ciudad Hidalgo, en particular, ha considerado esta propuesta, teniendo en cuenta los aspectos de seguridad, precisión, higiene y rapidez. De esta manera, el tecnológico se posiciona a la vanguardia de la tecnología, automatizando incluso los procesos más simples.

## 7. Referencias

A.W. Ahmad, (2011). "Implementation of Zigbee-GSM based Home Security Monitoring and Remote Control System".

Circuits and Systems (MWSCAS). IEEE 54th International Midwest Symposium on.

Alexandre, Luis, (2010). Tecnología NFC, Ecuador, Monografía.

Analog Devices, (2014). "MP36 Data Sheet": [www.analog.com](http://www.analog.com)

Andrejevic, M., & Selwyn, N. (2020). Facial recognition technology in schools: critical questions and concerns.

Learning, Media and Technology, 45(2), 115-128. <https://doi.org/10.1080/17439884.2020.1686014>

Banzi, Massimo (2009). Getting Started with Arduino (en inglés) (1ª edición). Make Books. p. 128.

Calles Carrasco, M. F., & Becerra Arévalo, N. P. (2019). Sistema informático de reconocimiento facial para el registro y control de asistencia de los socios de la cooperativa de taxis y camionetas puyo, 8(5).

Garduño Santana, M. A., Díaz-Sánchez, L. E., Tabarez Paz, I., & Romero Huertas, M. (2017). Estado del arte en reconocimiento facial. Research in Computing Science, 140(1), 19-27.

Granado, Emanuel, (2011). Sistemas con Radiofrecuencia, Venezuela.

Microchip Technology, (2014). "DSPIC30FXX Data Sheet High-Performance, 16-bit Microcontrollers". Chandler, Arizona: <http://www.microchip.com>

Mikroelektronika, (2014). "GSM Click Manual": [www.mikroe.com](http://www.mikroe.com)

Monroy, F. (2018). Estacionamientos automatizados. Disponible en: <https://multiplojp.wordpress.com/2018/07/26/estacionamientos-automatizados/>

Noble, J. (2009). Programming Interactivity: A Designer's Guide to Processing, Arduino, and openFrameworks (1ª edición). O'Reilly Media. p. 768. ISBN 0596154143.

Panasonic, (2014). "DN6851 Data Sheet". <http://www.semicon.panasonic.co.jp/>

Portilla Jimenez, J. J. (2018). Análisis y Diseño de un Sistema de Reconocimiento Facial aplicando Machine Learning para detectar e identificar intrusos. En Universidad de Guayaquil. <http://repositorio.ug.edu.ec/handle/redug/32954>

Ruiz, José Manuel, (2007). Manual de Programación Arduino, México.

Uber facial recognition. (2016). <https://www.techinasia.com/uber-china-facial-recognition>

Wadham, Rachel, (2003). "Radio Frequency Identification". Library Mosaics.

Wang, L., & Siddique, A. A. (2020). Facial recognition system using TensorFlow face recognizer for anti-theft and surveillance application based on drone technology. Measurement and Control, 53(7-8), 1070-1077. <https://doi.org/10.1177/0020294020932344>

Zumba Gamboa, J. P., & León Arreaga, C. A. (2018). Evolución de las Metodologías y Modelos utilizados en el Desarrollo de Software. INNOVA Research Journal, 3(10).

# Revisión de literatura: análisis de sentimientos en X para medir la aceptación de políticas públicas

Rodríguez Pinto Cristian David, Briseño Martínez Ramón Alejandro, Morales-Montelongo José Guadalupe

Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, Maestría en Tecnologías de Información, México, ORCID: 0000-0002-4173-7151; Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, Doctorado en Tecnologías de Información, México, ORCID: 0009-0009-6561-0807; Universidad de Guadalajara, Centro Universitario de Tonalá, Departamento de Ciencias de la Información y Desarrollos Tecnológicos, México, ORCID: 0000-0002-7546-3689

cristian.rodriguez7672@alumnos.udg.mx, alejandro.bmartinez@alumnos.udg.mx ,  
jose.gpe.morales@academicos.udg.mx

**Resumen.** La evaluación de políticas públicas es fundamental para garantizar la gobernabilidad de un país. La expansión del uso de internet ha permitido que las personas puedan expresar opiniones sobre asuntos públicos constantemente. La red social X, antes llamada Twitter, está posicionada globalmente como el principal sitio de internet para compartir y comentar noticias y asuntos públicos; continuamente se producen grandes volúmenes de datos susceptibles a ser analizados. Una de las técnicas más utilizadas para estudiar grandes volúmenes de opiniones expresadas en X es el análisis de sentimientos. Este artículo revisa la literatura existente sobre estudios que implementan análisis de sentimientos a información extraída de X como herramienta para conocer la opinión pública respecto a políticas públicas implementadas en diferentes geografías.

**Palabras clave.** *Análisis de sentimientos, X, políticas públicas, opinión pública, lexicón.*

## 1. Introducción

Se entiende como política pública toda acción gubernamental implementada para atender necesidades de la población o resolver algún problema de esta (Bueno Suárez, C., & Osuna Llana, J. L. (2013). En países como Reino Unido y Australia, recientemente han surgido debates respecto a la desvinculación entre los resultados de las evaluaciones tradicionales de políticas públicas, basadas generalmente en encuestas y sondeos, y el grueso de la población (Williams, K., & Lewis, J. M., 2021). La desvinculación de la sociedad hacia estos estudios abre espacio para reflexionar sobre la utilización de metodologías experimentales que, con rigor científico, aprovechen las nuevas tecnologías de información para recolectar y procesar datos y, finalmente, difundir los resultados.

La e-participación es definida como el uso de tecnologías de información como canal para que la ciudadanía comunique al gobierno sus necesidades (Macintosh, A., 2008). Sitios de redes sociales como X funcionan como espacios ideales para la e-participación debido a que ofrecen una experiencia de usuario que propicia la conversación de la ciudadanía sobre los temas que considera trascendentales, entre ellos las políticas públicas (Simonofski, A., Fink, J., & Burnay, C., 2021). Sin embargo, se advierte que la ciudadanía ejerce su e-participación generalmente de manera desestructurada, emitiendo grandes cantidades de mensajes de manera espontánea y descoordinada, cuya interpretación es un reto para las autoridades debido a su volumen. La gran cantidad de datos cualitativos y cuantitativos generados constantemente en X es materia de estudio sobre fenómenos sociales, como demuestran múltiples investigaciones alrededor del mundo (Chan, C., Zeng, J., & Schäfer, M. S., 2022; Freire Castello, N., 2019; García, J. G., Henríquez-Coronel, P., Ponce, J. P., & Herrera, J., 2017; Vukmirovic, M., Raspopovic Milic, M., & Jovic, J., 2022 y Pedroza, A., 2019).

La información de X como objeto de investigación se puede conseguir utilizando herramientas tecnológicas que permiten obtener y procesar grandes conjuntos de datos. El análisis de sentimientos ha sido implementado para múltiples fines como investigaciones de preferencias de consumo en estudios de mercado (Chauhan, P., Sharma, N. & Sikka, G., 2021); para conocer las necesidades de los usuarios de aplicaciones y mejorar su experiencia (Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M., 2023); para identificar violencia digital de género (Pedroza, A., 2019); e incluso para intentar predecir resultados electorales (Yili Wang, Jiaxuan Guo, Chengshen Yuan y Baozhu Li., 2022 y Singh, N.K., Tomar, D.S. & Sangaiah, A.K., 2018).

A inicios del siglo XXI surgieron investigaciones para interpretar las opiniones de usuarios en múltiples sitios de internet, hasta que en 2008 Twitter, hoy X, se posicionó como la principal plataforma de microblogging en el mundo y concentró los estudios de análisis de sentimientos debido a que "la conversación ofrece una gran cantidad de información relevante para varios temas y puede facilitar la comprensión de los sentimientos de las personas" (Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M., 2023). Las bases de datos académicas Elsevier's ScienceDirect, IEEE Xplore Digital Library, Springer Link, ACM Digital Library y Wiley Online Library

registraron un incremento promedio anual de investigaciones que aplican análisis de sentimientos de 34% entre los años 2008 y 2022 (Al-Otaibi, S. T., & Al-Rasheed, A. A., 2022).

El presente artículo es una revisión de literatura de los artículos obtenidos resultado de la búsqueda de los términos “sentiment analysis” y “policy acceptance” en la base de datos académica Web of Science que comprendió el periodo del año 2010 al 2023. Se obtuvieron 55 resultados, de los cuales se abordan los 25 que implementaron el análisis de sentimientos para orientar decisiones gubernamentales o de políticas públicas. El objetivo de esta revisión es explorar el potencial del análisis de sentimientos como herramienta para evaluar la popularidad de políticas públicas, una alternativa frente a encuestas y sondeos.

## 2. Revisión de literatura

El análisis de sentimientos es el proceso para definir si la opinión de una persona, o red de personas, sobre un tema es positiva, neutral o negativa. Su objeto de estudio es el mensaje, generalmente en formato escrito, y las unidades de análisis son las palabras que se utilizan (Serena Y. Kim, et al., 2021).

El análisis de sentimientos se aplica a distintos niveles: palabras individuales, oraciones o documentos completos. La etiqueta del sentimiento predominante se asigna teniendo en consideración las connotaciones de las palabras. Por ejemplo, la palabra “bueno” se vincula con una connotación positiva, mientras que la palabra “terrible” se vincula con una connotación negativa (Chih-yuan Li, et al., 2022). Existen enfoques que incluso asignan un valor numérico para intentar capturar la intensidad de un sentimiento. Por ejemplo, se utilizan escalas de -5 a 5 donde -5 representa un sentimiento extremadamente negativo, 5 un sentimiento extremadamente positivo y 0 un sentimiento neutral (Pedroza, A., 2019). Los enfoques de análisis de sentimientos son tres: el de lexicón, el de aprendizaje automático y el híbrido.

La presente revisión de literatura adopta la taxonomía de estudios de análisis de sentimientos propuesta por Wang (2022) que tiene múltiples alternativas de clasificación de polaridad, como se muestra en la Figura 1, y analiza los artículos encontrados conforme a la misma. Se adjunta la Tabla 1 con los hallazgos.

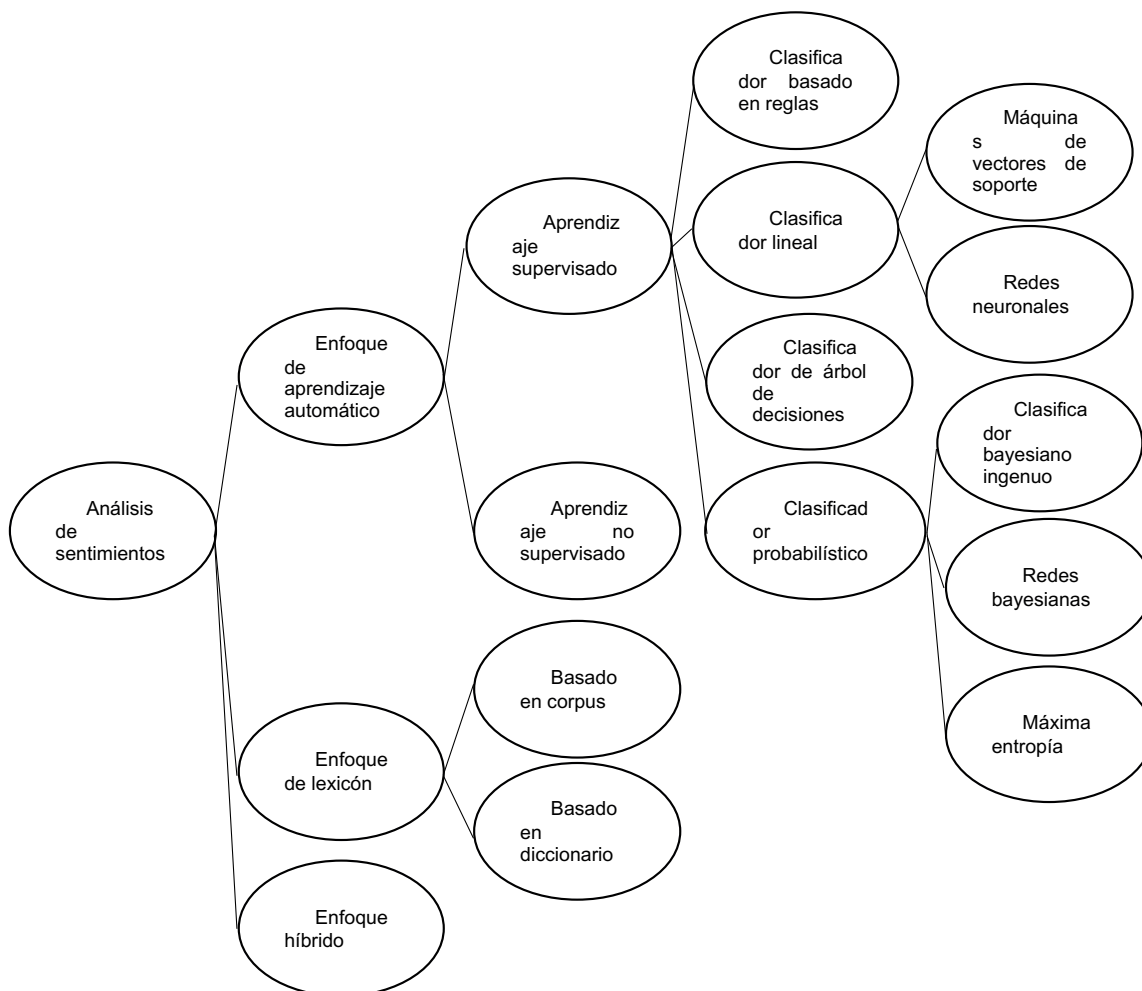


Figura 1. Taxonomía del análisis de sentimientos [14]



Referencia	Enfoque de la taxonomía	Lenguaje	Hallazgos	Particularidades
Acerbi A, Burns J, Cabuk U, Kryczka J, Trapp B, Valletta JJ, Mesoudi A. (2023)	Enfoque de lexicón basado en diccionario (paquete VADER de R).	R. API de Twitter.	Con una valencia de -1 a 1, la puntuación media y mediana de la emoción sobre el tema en cuestión fue, en ambos casos, negativa: -0.40 y -0.65, respectivamente. Los datos completos, incluyendo los retweets, están fuertemente influenciados por el tweet más retuiteado con una puntuación de emoción de -0.65.	Se discute la pertinencia de considerar los retweets, o solamente los tweets únicos, en el estudio.
Al-Otaibi, S. T., & Al-Rasheed, A. A. (2022)	Revisión de literatura	-	La precisión del análisis de sentimientos depende del conjunto de datos y el algoritmo de clasificación, por lo tanto, enfatiza su limpieza y legibilidad.	Se observa que la mayoría de los estudios de análisis de sentimientos se realizan en idioma inglés, y que se refieren principalmente a productos y servicios. También se encontraron estudios respecto a eventos críticos: desastres, actividades sociales y extremismo.
Cano-Marin, E., Mora-Cantalops, M., & Sánchez-Alonso, S., (2023)	Revisión de literatura.	-	El uso de la Inteligencia Artificial para el procesamiento del lenguaje natural está incrementando en los análisis de sentimientos. Los estudios de análisis de sentimientos para predecir fenómenos futuros son una línea de investigación con oportunidades.	El estudio se centra en el uso de la información obtenida de Twitter como una oportunidad para negocios.
E. del Valle & L. de la Fuente. (2023)	Revisión de literatura.	-	El estudio identifica como principales retos para los estudios de análisis de sentimientos en español la detección de ironía, la detección de conversaciones falsas producidas por bots, la localización geográfica de los tweets, las normas de la API de Twitter.	Destaca el uso del modelo de lenguaje pre entrenado de Google BERT como herramienta innovadora.
Fernandez, G.; et al. (2022)	Enfoque de lexicón basado en diccionario (NRC Emoticon Lexicon).	Python. API de Twitter.	La respuesta emocional a la pandemia de COVID-19 y las decisiones tomadas por las autoridades para evitar contagios fueron, en su mayoría, negativas, pero diferentes en distintas ciudades de Italia. Hay una relación entre eventos noticiosos importantes y la emoción de las personas a lo largo del tiempo de estudio.	Los formuladores de políticas públicas pueden utilizar los resultados para cuantificar el nivel de aceptación positiva o negativa de sus decisiones e identificar políticas públicas diferentes según la geografía.

Chih-yuan Li, et al. (2022)	Enfoque híbrido: lexicón (VADER, Sentiment Analyzer y TextBlob) y clasificador basado en reglas.	Python. API de Twitter.	Se propuso un sistema de monitoreo de las preocupaciones de los usuarios de X durante la pandemia de COVID-19. El experimento demostró que “ventiladores” y “mascarillas” fueron los temas que produjeron menos preocupación, mientras que “cierre de negocios” provocó más preocupación. La preocupación fue diferenciada en los distintos estados que conforman Estados Unidos, país en el que se concentró el estudio.	Propone una fórmula para medir el nivel de preocupación diario de las personas respecto a un tema. Pone a disposición un dataset de más de 16 millones de Tweets.
Majid Rahardi, et al. (2022)	Enfoque de aprendizaje automático, supervisado, con clasificador lineal de máquina de vectores de soporte.	Python. API de Twitter.	Los resultados de clasificación utilizando máquina de vectores de soporte tiene un resultado de 92% de precisión, mejor que el obtenido con kernel RBF, seguido por núcleo lineal y núcleo polinomial con 90% y núcleo sigmoidal con 89%.	La evaluación del modelo de máquina de vectores de soporte se presentó en una tabla de matriz de confusión que consta de clase de predicción y clase real. La comparación de cada núcleo.
Aijaz Ahmad Reshi, et al. (2022)	Enfoque híbrido: lexicón TextBlob, VADER y AFINN) y clasificador lineal basado en redes neuronales (LSTM-GRNN).	Python. API de Twitter.	Los modelos de aprendizaje automático basados en redes neuronales tienen mejor rendimiento que otros. Respecto a los modelos basados en lexicón, TextBlob fue el que tuvo mejor precisión.	El estudio sugiere que investigaciones futuras desarrollen un sistema de monitoreo de tweets en tiempo real y advierte que las noticias falsas inciden en los resultados de los análisis de sentimientos.
Vukmirovic, M., Raspopovic Milic, M., & Jovic, J. (2022)	Enfoque de lexicón basado en diccionario (VADER).	Python. API de Twitter.	Twitter puede ser considerado como herramienta para obtener datos de la opinión de los usuarios respecto a espacios públicos como calles y las intervenciones que deben hacer las autoridades.	Reconoce que el análisis de sentimientos basado en el lexicón VADER no detecta ironía, lo que causa una clasificación imprecisa del sentimiento.
K. D. S. Brito, R. L. C. S. Filho and P. J. L. Adeodato. (2021)	Revisión de literatura.	-	De la revisión de literatura, se encontró que 22 estudios que aplicaron análisis de sentimientos lograron predecir al menos tres elecciones, demostrando que son utilizables en próximas campañas electorales.	Los estudios que implementan análisis de sentimientos a información obtenida de redes sociales son, en general, poco replicados, por lo tanto sus resultados no son comparados.

Chauhan, P., Sharma, N. & Sikka, G. (2021)	Revisión de literatura.	-	La mayoría de los estudios que utilizan análisis de sentimientos para intentar predecir resultados electorales utilizan datos recabados de la API de Twitter, la mayoría de ellos han predecido de manera exitosa, lo que demuestra el potencial de estos estudios. Los principales retos son la precisión de dichos estudios, identificar los tweets producidos por cuentas falsas para alterar la conversación real e identificar el sarcasmo.	Compara los estudios de análisis de sentimientos con los estudios volumétricos y los de redes.
Hu T, et al. (2021)	Enfoque de lexicón basado en diccionario (VADER y National Research Council Canada Lexicon			
(NRCLex)) y clasificador basado en reglas (Latent Dirichlet Allocation).	Python. API de Twitter.	Los hallazgos demuestran una conexión entre la exposición a noticias o eventos importantes y cómo esto produce efectos en los sentimientos de la población. Además, los procesos psicológicos emocionales perceptivos y cognitivos producen efectos en la manifestación de sentimientos cuando se exponen a eventos noticiosos.	Se destaca el uso de información extraída de Twitter como fuente alternativa de datos que ayudan a comprender las opiniones de la ciudadanía hacia diferentes vacunas COVID-19 en Estados Unidos.	
Serena Y. Kim, et al. (2021)	Enfoque de aprendizaje automático, supervisado, con clasificador lineal de máquina de vectores de soporte.	Python. API de Twitter.	Se consiguió 80.2% de precisión utilizando RoBERTa, un modelo de lenguaje pre entrenado desarrollado por Facebook IA definiendo las clases positivo, neutro y negativo.	Compara la relación entre el sentimiento público hacia la energía solar, la política energética y el mercado.

M. Rodríguez-Ibáñez, F. -J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez. (2021)	Enfoque híbrido: lexicón basado en diccionario (AFINN, JAEN, Linguakit y ABU, todos en español) y clasificador basado en reglas.	Twitty: herramienta desarrollada por los autores.	Se encontró variación entre los sentimientos alusivos a partidos y actores políticos: los candidatos de partidos separatistas de España tienen un mayor índice de negatividad que los candidatos no separatistas.	Se propone el Índice de Positividad Agregada Compuesta (CAPI). Se identificó que la traducción de lexicones de inglés a español no garantiza precisión, por lo que se sugiere poner atención a dicho paso.
Sujon M, Dai, F. (2021)	Enfoque de lexicón basado en diccionario (SentiStrength y Linguistic			
Inquiry and Word Count (LIWC)).	Se firmó un acuerdo con Twitter para descargar 210 mil datasets referentes al transporte público de la ciudad de Washington.	La mayoría de los usuarios de Twitter que hablan sobre el transporte público en la ciudad de Washington lo consideran un tema importante; su principal sentimiento hacia la actuación de las autoridades es neutral.	Los autores argumentan que el incremento de la frecuencia de tweets respecto a un tema específico puede considerarse incremento de la atención del público sobre el tema en cuestión. La información analizada está puesta a disposición de investigadores, bajo solicitud.	
Adams-Cohen, N. J. (2020)	Enfoque de aprendizaje automático, supervisado, con clasificador basado en reglas.	Python. API de Twitter.	Combinar técnicas de análisis de sentimientos con datos de redes sociales brinda una nueva perspectiva en el análisis de la opinión pública.	Se observa que la cantidad de información en Twitter sobre un tema específico crece exponencialmente cuando dicho tema se vuelve de interés mediático, en este caso una decisión de la Suprema Corte de Estados Unidos.
Aseem Kinra, et al. (2020)	Enfoque híbrido: lexicón basado en diccionario y clasificador basado en reglas (Latent Dirichlet Allocation y algoritmo Fruchterman-Reingold).	Python. API de Twitter.	Los datos extraídos de Twitter son útiles para generar conocimiento sobre la opinión pública. Se puede modelar un algoritmo para identificar las palabras clave de un corpus. Se puede implementar análisis de sentimientos para conocer la postura de una persona respecto a un tema de interés público.	El estudio aplica análisis de sentimientos a tweets y artículos en periódicos para conocer las preocupaciones del público hacia políticas públicas.

<p>Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. (2020)</p>	<p>Enfoque de aprendizaje automático, no supervisado utilizando Latent Dirichlet Allocation.</p>	<p>Python. API de Twitter.</p>	<p>El COVID-19 afectó significativamente las condiciones psicológicas de las personas. El análisis de sentimientos sobre COVID-19 contribuye a la comprensión sobre la dinámica de los sentimientos y las preocupaciones de la gente a lo largo del tiempo. Esta información puede ser utilizada por las autoridades de salud para priorizar la salud mental y psicosocial durante pandemias como esta. Los patrones y emociones identificados podrían usarse para guiar programas de intervención específicos o políticas públicas.</p>	<p>Los autores afirman que los análisis de sentimientos superan las limitaciones de las técnicas de ciencias sociales como encuestas o entrevistas a pequeña escala.</p>
<p>Etter, M., Colleoni, E., Illia, L., Meggiorin, K., &amp; D'Eugenio, A. (2018)</p>	<p>Enfoque de aprendizaje automático, supervisado, con clasificador basado en reglas.</p>	<p>Python. API de Twitter.</p>	<p>A diferencia de los medios de comunicación tradicionales, la percepción de Italy Bank en Twitter prevalece en sentimiento negativo a lo largo de un año, con un promedio mensual de -0.29.</p>	<p>El estudio de caso revela cómo el análisis de cuatro fuentes diferentes -Twitter, encuesta, calificadora y periódicos- tienen juicios diferentes. La fuente de información más detallada y continúa para conocer la percepción de un banco a través del tiempo fue Twitter.</p>
<p>Gohil S, Vuik S, Darzi A. (2018)</p>	<p>Revisión de literatura.</p>	<p>-</p>	<p>Ninguna de las metodologías de análisis de sentimientos ha sido replicada en múltiples casos para medir su precisión.</p>	<p>El estudio sugiere desarrollar una única herramienta para análisis de sentimientos entrenada primero utilizando un corpus específico. En el caso de análisis de sentimientos sobre temas de salud se sugiere desarrollar un lexicón sobre el tema.</p>
<p>Ray, P., Chakrabarti, A., Ganguli, B., &amp; Das, P. K. (2018)</p>	<p>Enfoque híbrido: lexicón basado en diccionario (SentiWordnet); clasificador probabilístico bayesiano ingenuo y clasificador lineal basado en máquinas de vectores de soporte.</p>	<p>R. API de Twitter.</p>	<p>Se concluye que la población rural de India manifiesta en Twitter preocupaciones diferentes sobre la política pública de desmonetización respecto a las preocupaciones expresadas por la población urbana y los grupos académicos y políticos identificados.</p>	<p>La metodología puede ser replicable en eventos sociopolíticos y opiniones de clientes sobre servicios públicos que provoquen opiniones expresadas en Twitter.</p>

<p>Singh, N.K.,          Tomar, D.S. &amp;          Sangaiah, A. K.          (2018)</p>	<p>Revisión de          literatura.</p>	<p>-</p>	<p>Manejar modismos, símbolos,          palabras mal escritas y sarcasmo          es el desafío de los análisis de          sentimientos.</p>	<p>Las publicaciones          en varios idiomas          con tratamiento          geográfico en redes          sociales incrementan          la complejidad de          hacer análisis de          sentimientos con          niveles aceptables          de precisión y          coherencia.</p>
<p>Bain, J., Chaban,          N. (2017)</p>	<p>Enfoque de          aprendizaje          automático,          supervisado,          con clasificador          basado en          reglas.</p>	<p>Python. API de          Twitter.</p>	<p>Se encontró un alto nivel de          neutralidad en los tweets sobre la          Semana Europea de la Energía          Renovable. Se observa una          imagen positiva de la Unión          Europea como actor estratégico          involucrado en temas de energía          renovable.</p>	<p>“Semana Europea de          Energía Renovable”          está emergiendo          como narrativa          dentro y fuera de la          Unión Europea. En          este caso se utilizó          Twitter como espacio          para esparcir la          narrativa. Futuros          estudios deben          considerar análisis          de sentimientos para          identificar narrativas          sobre un tema          específico en otras          redes sociales.</p>
<p>Aghababaei S,          Makrehchi. M          (2016)</p>	<p>Enfoque de          lexicón basado          en diccionario.</p>	<p>Python. API de          Twitter, utilizando          muestras          aleatorias          al pasado          con enfoque          Coupling From          The Past          (CFTP).</p>	<p>En el desarrollo de un modelo          de predicción de ocurrencia de          crímenes basado en la minería          de textos de Twitter en un área          geográfica, el clasificador predice          dónde incrementará o reducirá          los crímenes con base en los          expedientes criminales, pero el          análisis de sentimiento basado          en lexicón ofrece una predicción          insignificante.</p>	<p>Si bien no se          encontró predicción          significativa de          delitos en un área          geográfica basada          en el análisis de          sentimientos en          Twitter, se sugiere          retomar los estudios          de análisis de          contenidos y cruzar          los datos con otras          variables para          demostrar correlación          con otros incidentes.</p>

Collins, Craig, et al. (2013)	Enfoque híbrido: lexicón basado en diccionario (SentiStrength) y clasificador basado en reglas.	Python. API de Twitter.	Los usuarios que padecen el tráfico suelen tener más expresiones negativas que positivas. Los resultados pueden ayudar a las autoridades de tránsito a mejorar la gestión del tráfico y, por lo tanto, la experiencia de los usuarios. Futuras investigaciones deben estudiar el impacto de las opiniones en las redes sociales sobre las decisiones de los gobernantes.	Se destaca que el costo de la recolección de datos fue mínimo, que los datos pueden ser recabados en tiempo real, que se puede identificar con claridad el sentimiento de un usuario y que las necesidades expresadas por los usuarios del transporte público pueden ser atendidas. El estudio se realizó en lenguaje inglés, por lo que personas que se expresen en otros idiomas sobre el mismo tema fueron excluidos.
-------------------------------	---	-------------------------	--	--

Tabla 1. Hallazgos de la revisión de literatura

El enfoque de lexicón toma como referencia un conjunto de palabras a las que se les asigna un valor sentimental, lo que permite establecer su polaridad o puntuación. El lexicón basado en corpus toma en cuenta la totalidad del texto para después asignar etiquetas con polaridades según el contexto del mensaje, esto implica que las palabras obtienen su polaridad a través del análisis contextual. Por su parte, el lexicón basado en diccionario establece una lista de palabras a las que se les asigna una carga de polaridad o puntuación que toma como referencia la definición general de cada palabra, sin tomar en cuenta el contexto en el que es empleada (Singh, N.K., Tomar, D.S. & Sangaiah, A.K., 2018).

El enfoque de aprendizaje automático implica utilizar un algoritmo que realice la tarea de predicción. Se divide en dos tipos: el supervisado y el no supervisado, este último mide la presencia de palabras o frases a las que, en la etapa final se les asigna una carga de valor (Al-Otaibi, S. T., & Al-Rasheed, A. A., 2022). El enfoque de aprendizaje automático de tipo supervisado implica la construcción de un clasificador capaz de predecir la polaridad de textos. El clasificador se construye a partir de un conjunto de datos que contenga una cantidad representativa de textos con una polaridad previamente etiquetada por humanos. El dataset sirve para entrenar el clasificador. Los clasificadores del enfoque de aprendizaje automático de tipo supervisado pueden ser de cuatro formas: basado en reglas, de árbol de decisiones, lineal y probabilístico (Yili Wang, Jiaxuan Guo, Chengshen Yuan y Baozhu Li., 2022).

La forma del clasificador basado en reglas implica identificar características del texto que puedan indicar una polaridad y después tomar en cuenta esas características para establecer reglas de asignación de etiquetas de polaridad. Se utiliza la condición “si-entonces”. Por ejemplo: si una frase incluye la palabra “feliz”, la oración se etiqueta como positiva, o si una frase contiene la oración “triste”, la oración se etiqueta como negativa. La forma del clasificador de árbol de decisiones establece condiciones para identificar características del texto llamadas nodos. Cada nodo identifica un atributo del texto y decide si lo transfiere a otro nodo. Esta serie de nodos emula las partes de un árbol: raíz, tronco, ramas y hojas, siendo estas últimas el nodo en el que se etiqueta la polaridad del texto, en caso de que cumpla con las características requeridas en los nodos previos. La diferencia entre el clasificador basado en reglas y el de árbol de decisiones es que el primero es más directo, pero el segundo puede realizar decisiones más complejas (Serena Y. Kim, et al., 2021).

La forma de clasificador lineal separa las características del texto en distintas clases. Puede ser de dos modelos: máquinas de vectores de soporte -SVM, por sus siglas en inglés- o redes neuronales. La primera requiere un conjunto de datos, o dataset, previamente etiquetado para posteriormente construir un algoritmo que genere una matriz de frecuencia de palabras y vectorice la Frecuencia de Término-Frecuencia Inversa de Documento -TF-IDF, por sus siglas en inglés- en un hiperplano. La idea de las SVM es clasificar el texto considerando cuántas veces aparece una palabra en el texto, cuántas veces aparece la misma palabra en la colección de textos y la carga sentimental de la palabra, lo que posicionará cada texto analizado en un punto del hiperplano (Mansoor, Muvazima & Gurumurthy, Kirthika & U, Anantharam & Prasad, V., 2020). Por su parte, la forma de clasificador de redes neuronales requiere un dataset previamente etiquetado para construir un algoritmo capaz de establecer relaciones entre los textos conocidos y asignar

etiquetas a nuevos textos. La característica que diferencia al clasificador de redes neuronales es que el algoritmo minimiza progresivamente sus errores al calibrar las polaridades de los textos conforme analiza más mensajes. Este entrenamiento de identificación de patrones emula el funcionamiento del cerebro humano. Las redes neuronales se consideran deep learning -aprendizaje profundo- porque trabajan con capas ocultas que les permiten realizar complejos procesamientos de datos (Al-Otaibi, S. T., & Al-Rasheed, A. A., 2022).

La forma de clasificador probabilístico consiste en utilizar algoritmos con modelos matemáticos de probabilidad para determinar la polaridad de un texto. Se divide en tres: el clasificador bayesiano se basa en el Teorema de Bayes para calcular la polaridad de un texto de acuerdo al conjunto de sus características, a diferencia del clasificador bayesiano ingenuo que asume que las características del texto son independientes entre sí. Finalmente, el clasificador máxima entropía modela relaciones complejas entre las características de los textos para asignar etiquetas de polaridad basadas en la probabilidad más uniforme (M. Rodríguez-Ibáñez, F. -J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez., 2021).

### 3. Metodologías

Existen diferentes vías de acceso a los datos de X. La oficial es la interfaz de programación de aplicaciones -API, por sus siglas en inglés-, que permite descargar una cantidad de datos limitada a su perfil de usuario (M. Rodríguez-Ibáñez, F. -J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez., 2021 y Aseem Kinra, et al., 2020). Por otro lado, se pueden descargar datos haciendo scraping, técnica de uso de software para hacer peticiones sistemáticas a sitios web por medio de sus respuestas HTML (Yili Wang, Jiakuan Guo, Chengshen Yuan y Baozhu Li., 2022). Los softwares utilizados para la recolección de información son principalmente Python (Serena Y. Kim, et al. 2021 y Chih-yuan Li, et al., 2022) y R (Cano-Marin, E., Mora-Cantalops, M., & Sánchez-Alonso, S., 2023 y Yili Wang, Jiakuan Guo, Chengshen Yuan y Baozhu Li., 2022). Una vez que se tiene el dataset se procede a la limpieza de datos que consiste en la eliminación de las palabras que no aportan carga sentimental como pronombres, preposiciones o artículos; también se borran signos de puntuación, caracteres especiales y se homogenizan todas las palabras a letras minúsculas. Este último paso se conoce como tokenización y agiliza las etapas posteriores del análisis (Aijaz Ahmad Reshi, et al., 2022).

En el caso del enfoque de lexicón, los estudios utilizan principalmente los lexicones en inglés Valence Aware Dictionary and Sentiment Reasoner -VADER- (Pedroza, A. (2019) y Affective Norms for English Words -AFINN- (Aijaz Ahmad Reshi, et al., 2022), el primero tiene 7,500 palabras con una valencia de -1 extremadamente negativo a 1 extremadamente positivo y el segundo tiene 2,476 palabras con valencia de -5 extremadamente negativo a 5 extremadamente positivo, ambos consideran el 0 como neutral. También se encontraron lexicones traducidos al español como AFINN, JAEN, Linguakit y SBU (M. Rodríguez-Ibáñez, F. -J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez., 2021). Un estudio realizado con enfoque híbrido comparó el nivel de precisión de etiquetas de un lexicón que incluyó VADER y AFINN con un procesador de redes neuronales y encontró que la precisión era del 95%, mayor en el modelo de redes neuronales pre entrenado con la paquetería TextBlob de Python (Aijaz Ahmad Reshi, et al., 2022).

Otro estudio implementó un kernel de función de base radial -RBF, por sus siglas en inglés- para transformar datos en un espacio de características de mayor dimensión y encontró que se alcanzó una precisión del 92% en la precisión de etiquetas, por encima de los kernels lineal, polifónico y sigmoidal (Majid Rahardi, et al., 2022). Se destacó el desempeño del modelo SVM al procesar grandes cantidades de información.

Una revisión de literatura que comparó el desempeño del modelo de redes neuronales BERT con clasificadores basados en lexicón y aprendizaje automático indicó que el modelo de aprendizaje Representaciones de Codificadores Bidireccionales de Transformers -BERT, por sus siglas en inglés- alcanza niveles de precisión con un promedio de 85% y destaca que “desde 2020, una tercera parte de los estudios que implementan aprendizaje automático utilizan este modelo, lo que da una idea de su éxito” (E. del Valle & L. de la Fuente., 2023).

Del modelo de árbol de decisiones se señala que tiene un “uso mucho menos extendido” debido a la complejidad de su entrenamiento, para lo que se propone entrenarlo con aprendizaje no supervisado que identifica el grado de polaridad mediante indicadores subjetivos generados a partir del léxico de sentimientos (Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M., 2023 y Yili Wang, Jiakuan Guo, Chengshen Yuan y Baozhu Li., 2022). Comparado con otros modelos, su nivel de precisión está por debajo de las SVM (Al-Otaibi, S. T., & Al-Rasheed, A. A., 2022).

Sobre el clasificador bayesiano ingenuo, se afirma que alcanza la misma precisión que el clasificador SVM, pero con un rendimiento más lento, además de que su rigidez causa problemas para clasificar modismos (Singh, N.K., Tomar, D.S. & Sangaiah, A.K., 2018). En otro estudio el clasificador bayesiano ingenuo produjo predicciones menos precisas en comparación con el árbol de decisiones (Nur Siyam, Omar Alqaryouti & Sherief Abdallah., 2019).



## 4. Temas

Los primeros análisis de sentimientos en X tuvieron la intención de predecir los resultados de elecciones presidenciales en distintos países con diferentes tiempos previos al día de la votación (Yili Wang, Jiaxuan Guo, Chengshen Yuan y Baozhu Li., 2022). Por otra parte, revisiones de literatura concluyen que si bien es posible predecir los resultados de una elección basándose en las opiniones que las personas expresan en X, las metodologías implementadas hasta el momento son casuísticas y no alcanzan la replicabilidad esperable (K. D. S. Brito, R. L. C. S. Filho and P. J. L. Adeodato., 2021).

Múltiples estudios que implementan análisis de sentimientos y aprovechan los datos de X se concentran en la opinión de los usuarios respecto a productos como hoteles (M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh and O. Qawasmeh., 2019 y C. Tsai, K. Chen, Y. Hu and W. Chen., 2020), restaurantes (L. Rafael, C. Pessutto, D. S. Vargas and V. P. Moreira., 2020), series de Netflix (Acerbi A, Burns J, Cabuk U, Kryczka J, Trapp B, Valletta JJ, Mesoudi A., 2023), libros y sitios web (H. Saif, Y. He and M. Fernandez. 2016 y R. Xia, F. Xu, J. Yu, Q. Yong, and E. Cambria., 2016).

El análisis de sentimientos también se ha implementado para estudiar la aceptación de políticas públicas o decisiones gubernamentales como la satisfacción de los usuarios de un sistema de transporte público (Collins, Craig, et al., 2013), el impulso al uso de energías renovables (Bain, J., Chaban, N., 2017), la reducción de emisiones de carbono (Serena Y. Kim, et al., 2021), la desmonetización de la economía de un país (Ray, P., Chakrabarti, A., Ganguli, B., & Das, P. K., 2018) y los permisos de circulación para vehículos autónomos (Aseem Kinra, et al., 2020). En Estados Unidos, también se estudió la discusión de la corte de ese país sobre derechos de las personas homosexuales (Adams-Cohen, N. J., 2020). Para el ámbito corporativo, se analizaron las expresiones de los trabajadores hacia su empresa para medir su nivel de satisfacción laboral (Etter, M., Colleoni, E., Illia, L., Meggiorin, K., & D'Eugenio, A., 2018). En lenguaje español se creó un modelo para detectar el lenguaje de odio (E. del Valle & L. de la Fuente., 2023).

Si bien ya existían estudios sobre recomendaciones médicas publicadas en X (Gohil S, Vuik S, Darzi A., 2018), a partir del 2020 la mayoría de los análisis de sentimientos en dicha red social se han enfocado en la pandemia de COVID-19. En un primer momento, se concentraron en estudiar las preocupaciones de la población angloparlante respecto a la propagación del virus (Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T., 2020 y Mansoor, Muvazima & Gurumurthy, Kirthika & U, Anantharam & Prasad, V., 2020) y el colapso de los sistemas de salud en países como Italia (Fernandez, G.; et al., 2022). Posteriormente, se analizó la aceptación de las políticas públicas de combate a la propagación del virus y se encontró que medidas como el uso de mascarillas era aceptado, mientras que el desabasto de ventiladores para respiración artificial provocaba expresiones de preocupación (Chih-yuan Li, et al., 2022).

Conforme avanzó la pandemia, surgieron nuevos temas objeto de estudio, como la confianza en las vacunas contra el COVID-19 en Estados Unidos (Hu T, et al., 2021) y en Indonesia (Majid Rahardi, et al., 2022); el nivel de rechazo a las campañas de vacunación y las tendencias de aceptación de los agentes biológicos (Aijaz Ahmad Reshi, et al., 2022). Estudios enfocados a la predicción de venta de productos sostienen que los análisis de sentimientos con información de X alcanzan niveles de precisión que brindan información útil respecto al comportamiento del tráfico en zonas urbanas (Sujon M, Dai, F., 2021).

## 5. Discusión

El principal cuestionamiento a los estudios de análisis de sentimientos es la ineficacia de las computadoras para descifrar mensajes sarcásticos (Sujon M, Dai, F., 2021). Se advierte que el discurso político está plagado de sarcasmo, doble sentido y rumores, lo que puede complicar su análisis literal. Además, si bien X es una fuente importante de datos, está plagado de "mensajes manipulados propagados por empresas que crean perfiles falsos para mejorar la imagen de personajes o instituciones" (Etter, M., Colleoni, E., Illia, L., Meggiorin, K., & D'Eugenio, A., 2018).

Sobre los análisis de sentimientos con enfoque de lexicón existen cuestionamientos respecto a sus omisiones de los matices contextuales de las palabras y está demostrado que enfoques como la máquina de vectores de soporte lo supera en precisión.

## 6. Conclusiones

Los estudios que utilizan el enfoque de lexicón, como los que emplean los diccionarios VADER y NRC, se caracterizan por asignar valores emocionales a palabras específicas, lo que permite una evaluación rápida y directa de la polaridad de los textos. Estos estudios son útiles para comprender las respuestas emocionales a eventos específicos, como la pandemia de COVID-19, y suelen ser replicables en diferentes idiomas. Sin embargo, se enfrentan a desafíos en la detección de ironía y la falta de contextualización, lo que puede afectar la precisión en la interpretación de sentimientos.

En contraste, los trabajos de aprendizaje automático, ya sea con enfoque de aprendizaje supervisado o no supervisado, se caracterizan por su capacidad para aprender patrones complejos y relaciones contextuales en los textos. Ya sea utilizando clasificadores lineales de máquinas de vectores de soporte, redes neuronales o métodos probabilísticos, estos estudios ofrecen un enfoque más sofisticado y adaptable. Su versatilidad permite que sean empleados en variedad de temas, desde la predicción de elecciones hasta la evaluación de la popularidad de instituciones financieras. Sin embargo, su éxito depende, en gran medida, de conjuntos de datos representativos y limpios.

De la revisión de literatura se concluye que los clasificadores más precisos son los de máquina de vectores de soporte, redes neuronales y bayesianos ingenuos, pero los análisis con lexicones tienen una implementación más ágil. La mayoría de los estudios encontrados que implementan la herramienta de análisis de sentimientos se realizan en idioma inglés. Si bien existen estudios que lograron altos niveles de precisión para etiquetar sentimientos o para emular los resultados de encuestas o votaciones, las metodologías son adaptadas a cada caso y difícilmente replicables.

La frecuencia de expresiones negativas en X es mayor que la de expresiones positivas (Acerbi A, Burns J, Cabuk U, Kryczka J, Trapp B, Valletta JJ, Mesoudi A., 2023), lo que respalda la vigencia de los datos objeto de estudio.

La revisión aquí planteada puede dar pie a nuevas investigaciones que implementen análisis de sentimientos en lenguajes como el español, para lo que es necesario desarrollar lexicones que tomen en cuenta las complejidades de la sociedad objeto de estudio y aprovechen la multitud de elementos analizables en internet, no solo los textos, sino también los audios y el vídeo, cuya producción y presencia crece considerablemente a diario.

## 7. Investigaciones futuras

Los recientes cambios en la política de X condicionan la obtención de datos mediante su API a pagar cuotas y establecen límites en la cantidad de datos descargables (Calma, J., 2023, 31 de mayo), lo que dificulta la replicabilidad de los estudios referidos en este trabajo. Una alternativa para realizar análisis de sentimientos es aplicar scraping, técnica de descarga automatizada de datos de una página web (Arias, et. al., 2022). La biblioteca Selenium de Python es una opción, ya que permite automatizar la descarga de cientos de tweets sobre un tema en un periodo de tiempo determinado y después exportarlos en formato .csv. La limpieza del conjunto de datos y la tokenización de las palabras puede hacerse utilizando las paqueterías dplyr, tidyr, stringr y janitor en R, un software eficiente para el manejo de grandes volúmenes de información. De esta manera, cada tweet quedará reducido a un conjunto de sustantivos y adjetivos en minúsculas, unidad de análisis.

Se sugiere hacer la clasificación de tweets utilizando el lexicón AFINN-111 traducida al español, el cual tiene 2,477 palabras, cada una con una puntuación asignada que va del -5 extremadamente negativo al 5 extremadamente positivo y considera al 0 como neutral (Pedroza, A., 2019 y M. Rodríguez-Ibáñez, F. -J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez., 2021). Una vez establecida la carga sentimental de cada unidad de análisis, será posible analizar longitudinalmente los datos calculando las medidas de tendencia central de la información y generar visualizaciones de la aceptación o rechazo de los usuarios de X hacia un tema o conjunto de temas, en este caso políticas públicas.

## 8. Referencias

Adams-Cohen, N. J. (2020). Policy Change and Public Opinion: Measuring Shifting Political Sentiment With Social Media Data. *American Politics Research*, 48(5), 612-621. <https://doi.org/10.1177/1532673X20920263>

Acerbi A, Burns J, Cabuk U, Kryczka J, Trapp B, Valletta JJ, Mesoudi A. (2023). Sentiment analysis of the Twitter response to Netflix's *Our Planet* documentary. *Conserv Biol.* 2023 Aug;37(4):e14060. doi: 10.1111/cobi.14060. Epub 2023 Apr 10. PMID: 36661052.

Aghababaei S, Makrehchi M (2016). Mining social media content for crime prediction. In: *IEEE/WIC/ACM International conference on web intelligence (WI)*. IEEE, pp 526–531

Al-Otaibi, S. T., & Al-Rasheed, A. A. (2022). A Review and Comparative Analysis of Sentiment Analysis Techniques. *Informatica*, 46(6). <https://doi.org/10.31449/inf.v46i6.3991>

- M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh and O. Qawasmeh. (2019). Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic, and semantic features. *Information Processing and Management*, vol. 56, no. 2, pp. 308-319, 2019. <https://doi.org/10.1016/j.ipm.2018.01.006>
- Arias, et. al. (2022). Analyzing Spanish-Language Public Sentiment in the Context of a Pandemic and Social Unrest: The Panama Case. *Int. J. Environ. Res. Public Health* 2022, 19,10328. <https://doi.org/10.3390/>
- Bain, J., Chaban, N. (2017). An emerging EU strategic narrative? Twitter communication during the EU's sustainable energy week. *Comp Eur Polit* 15, 135–155 (2017). <https://doi.org/10.1057/cep.2016.17>
- K. D. S. Brito, R. L. C. S. Filho and P. J. L. Adeodato. (2021). "A Systematic Review of Predicting Elections Based on Social Media Data: Research Challenges and Future Directions," in *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 819-843, Aug. 2021, DOI: 10.1109/TCSS.2021.3063660.
- Bueno Suárez, C., & Osuna Llana, J. L. (2013). Reflexiones epistemológicas y metodológicas para la evaluación de políticas públicas. *Andamios, Revista de Investigación Social*, 10(21), 95-117. <https://doi.org/10.29092/uacm.v10i21.300>
- Calma, J. (2023, 31 de mayo). Twitter just closed the book on academic research. *The Verge*. <https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research>
- Cano-Marin, E., Mora-Cantalops, M., & Sánchez-Alonso, S. (2023). Twitter as a predictive system: A systematic literature review. *Journal of Business Research*, 157, 113561. <https://doi.org/10.1016/j.jbusres.2022.113561>
- Chauhan, P., Sharma, n. & Sikka, G. (2021). The emergence of social data and sentiment analysis in election prediction. *Journal Ambient Intell Human Comput* 12, 2601-2627 (2021). <https://doi.org/10.1007/s12652-020-02423-y>
- Chih-yuan Li, et al. (2022). Public Health Policy Monitoring through Public Perceptions: A Case of COVID-19 Tweet Analysis. *Information* 2022, 13, 543. <https://doi.org/10.3390/info13110543>
- Collins, Craig, et al. (2013). A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data. *Journal of Public Transportation*, 16 (2): 21-45. DOI: <http://doi.org/10.5038/2375-0901.16.2.2>
- E. del Valle & L. de la Fuente. (2023). Sentiment analysis methods for politics and hate speech contents in Spanish language: a systematic review, in *IEEE Latin America Transactions*, vol. 21, no. 3, pp. 408-418, March 2023, doi:10.1109/TLA.2023.10068844
- Etter, M., Colleoni, E., Illia, L., Meggiorin, K., & D'Eugenio, A. (2018). Measuring Organizational Legitimacy in Social Media: Assessing Citizens' Judgments With Sentiment Analysis. *Business & Society*, 57(1), 60–97. <https://doi.org/10.1177/0007650316683926>
- Fernandez, G.; et al. (2022). Social Network Analysis of COVID-19 Sentiments: 10 Metropolitan Cities in Italy. *Int. J. Environ. Res. Public Health* 2022, 19, 7720. <https://doi.org/10.3390/ijerph19137720>
- Freire Castello, N. (2019). Por qué es twitter el territorio político digital. *Polis*, 15(2), 41-76. <http://doi.org/10.24275/uam/izt/dcsh/polis/2019v15n2/Freire>
- García, J. G., Henríquez-Coronel, P., Ponce, J. P., & Herrera, J. (2017). Analítica de Twitter para el estudio de las emociones primarias durante el terremoto de México 2017. *RISTI – Iberian Journal of Information Systems and Technologies*. 19 (abril) 479 – 492. ISSN: 1646-9895
- Gohil S, Vuik S, Darzi A. (2018). Sentiment Analysis of Health Care Tweets: Review of the Methods Used. *JMIR Public Health Surveill*. 2018 Apr 23;4(2):e43. doi: 10.2196/publichealth.5789. PMID: 29685871; PMCID: PMC5938573
- Hu T, et al. (2021). Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. *J Med Internet Res*. 2021 Sep 10;23(9):e30854. doi: 10.2196/30854. PMID: 34346888; PMCID: PMC8437406
- Aseem Kinra, et al. (2020). Examining the potential of textual big data analytics for public policy decision making: A case study with driverless cars in Denmark. *Transport Policy* <https://doi.org/10.1016/j.tranpol.2020.05.026>
- Macintosh, A. (2008). E-Democracy and E-Participation Research in Europe. En H. Chen, L. Brandt, V. Gregg, R. Traunmüller, S. Dawes, E. Hovy, A. Macintosh, & C. A. Larson (Eds.), *Digital Government* (Vol. 17, pp. 85-102). Springer US. [https://doi.org/10.1007/978-0-387-71611-4\\_5](https://doi.org/10.1007/978-0-387-71611-4_5)
- Majid Rahardi, et al. (2022). Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 6, 2022. <http://dx.doi.org/10.14569/IJACSA.2022.0130665>
- Mansoor, Muvazima & Gurumurthy, Kirthika & U, Anantharam & Prasad, V. (2020). Global Sentiment Analysis Of COVID-19 Tweets Over Time. <https://doi.org/10.48550/arXiv.2010.14234>
- Nur Siyam, Omar Alqaryouti & Sherief Abdallah. (2019). Mining government tweets to identify and predict citizens' engagement.

Technology in Society (2019), doi: <https://doi.org/10.1016/j.techsoc.2019.101211>

Pedroza, A. (2019). Detección de agresiones electrónicas en redes sociales mediante minería de texto. El caso de ataques contra mujeres en Facebook. Tesis de maestría en tecnologías de información, Universidad de Guadalajara. <https://hdl.handle.net/20.500.12104/80084>

Ray, P., Chakrabarti, A., Ganguli, B., & Das, P. K. (2018). Demonetization and its aftermath: An analysis based on twitter sentiments. *Sadhana*, 43(11), 186. <https://doi.org/10.1007/s12046-018-0949-0>

Serena Y. Kim, et al. (2021). Public Sentiment toward Solar Energy: Opinion Mining of Twitter Using a Transformer-Based Language Model. *Sustainability* 2021, 13, 2673. <https://doi.org/10.3390/su13052673>

Singh, N.K., Tomar, D.S. & Sangaiah, A.K. (2018). Sentiment analysis: a review and comparative analysis over social media. *J Ambient Intell Human Comput* 11, 97–117 (2020). <https://doi.org/10.1007/s12652-018-0862-8>

Simonofski, A., Fink, J., & Burnay, C. (2021). Supporting policy-making with social media and e-participation platforms data: A policy analytics framework. *Government Information Quarterly*, 38(3), 101590. <https://doi.org/10.1016/j.giq.2021.101590>

Sujon M, Dai, F (2021). Social media mining for understanding traffic safety culture in Washington state using twitter data. *Journal of Computing in Civil Engineering* 35(1):04020059

C. Tsai, K. Chen, Y. Hu and W. Chen. (2020). Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management*, Vols. 80, 104122, 2020. <https://doi.org/10.1016/j.tourman.2020.104122>

Vukmirovic, M., Raspopovic Milic, M., & Jovic, J. (2022). Twitter Data Mining to Map Pedestrian Experience of Open Spaces. *Applied Sciences*, 12(9), 4143. <https://doi.org/10.3390/app12094143>

Williams, K., & Lewis, J. M. (2021). Understanding, measuring, and encouraging public policy research impact. *Australian Journal of Public Administration*, 80(3), 554-564. <https://doi.org/10.1111/1467-8500.12506>

Yili Wang, Jiaxuan Guo, Chengshen Yuan y Baozhu Li. (2022). Sentiment Analysis of Twitter Data. *Appl. Sci.* 2022, 12, 11775. <https://doi.org/10.3390/app122211775>

L. Rafael, C. Pessutto, D. S. Vargas and V. P. Moreira. (2020). Multilingual aspect clustering for sentiment analysis, *Knowledge-Based Systems*, Vols. 192, 105339, 2020. <https://doi.org/10.1016/j.knosys.2019.105339>

Aijaz Ahmad Reshi, et al. (2022). COVID-19 Vaccination Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset. *Healthcare* 2022, 10, 411. <https://doi.org/10.3390/healthcare10030411>

M. Rodríguez-Ibáñez, F. -J. Gimeno-Blanes, P. M. Cuenca-Jiménez, C. Soguero-Ruiz and J. L. Rojo-Álvarez. (2021). "Sentiment Analysis of Political Tweets From the 2019 Spanish Elections," in *IEEE Access*, vol. 9, pp. 101847-101862, 2021, DOI: 10.1109/ACCESS.2021.3097492

Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, 223, 119862. <https://doi.org/10.1016/j.eswa.2023.119862>

H. Saif, Y. He and M. Fernandez. (2016). Contextual semantics for sentiment analysis of twitter, *Information Processing and Management*, vol. 52, no. 1, p. 5–19, 2016. <https://doi.org/10.1016/j.ipm.2015.01.005>

R. Xia, F. Xu, J. Yu, Q. Yong, and E. Cambria. (2016). Polarity shift detection, elimination and ensemble a three-stage model for document-level sentiment analysis. *Information Processing and Management*, vol. 52, no. 1, p. 36–45, 2016. <https://doi.org/10.1016/j.ipm.2015.04.003>

Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLoS ONE* 15(9): e0239441. <https://doi.org/10.1371/journal.pone.0239441>

# Aplicación Web para identificar personas a partir de la similitud semántica en descripciones físicas textuales utilizando técnicas de Procesamiento de Lenguaje Natural

Fabián Vera, Ramon A. Briseño , Guadalupe Alonso-Contreras

Universidad de Pamplona, Pamplona, Colombia, Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, Doctorado en Tecnologías de Información, México, Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA) Instituto Politécnico Nacional CDMX, México\

fabian.vera@unipamplona.edu.com, alejandro.bmartinez@alumnos.udg.mx, galonsoc1700@alumno.ipn.mx

**Resumen.** En Jalisco, uno de los estados con mayor número de desapariciones en el país (Ibarra, 2022), persiste el desafío de identificar a Personas Fallecidas Sin Identificar (PFSI). Por lo que, el proyecto RENIA (Red neuronal de investigación y análisis para la búsqueda e identificación de personas) busca la iniciativa de abordar este problema por medio de Inteligencia Artificial. El problema al que nos enfrentamos es una tarea compleja debido al alto número de desapariciones en el estado, por lo tanto, el trabajo propuesto muestra una solución basada en la implementación de una aplicación web con uso de técnicas de Procesamiento de Lenguaje Natural (PLN) para la comparación de descripciones textuales de datos de personas desaparecidas, con el fin de colaborar con las autoridades gubernamentales y facilitar el proceso de búsqueda e identificación de personas en el estado de Jalisco. El modelo Universal Sentence Encoder implementado en el aplicativo web, como parte del proyecto RENIA, alcanzó una efectividad del 90% al encontrar registros buscados dentro de las 100 mejores similitudes de cada búsqueda en la base de datos de PFSI (IJCF-PFSI, s.f.), que cuenta con 4,278 registros. La capacidad que presenta para localizar registros específicos dentro de la extensa base de datos de PFSI aporta una eficacia considerable, agilizando el proceso de identificación y brindando así una herramienta eficiente para aumentar las posibilidades de encontrar a personas reportadas como desaparecidas.

**Palabras clave.** *Procesamiento de Lenguaje Natural (PLN), Identificación de personas, Inteligencia Artificial (IA), Análisis Semántico, Aprendizaje automático, Embeddings.*

## 1. Introducción

A raíz de diversas situaciones como conflictos armados, desastres naturales, migraciones, entre otros, se presentan cientos de miles de personas desaparecidas a nivel mundial. Según la doctora en antropología Isabel Beltrán Gil (adondevanlosdesaparecidos, 2022), las desapariciones en México están relacionadas con la impunidad y el incumplimiento de políticas públicas para abordar el problema. Esto se debe a la falta de capacidad y recursos de las instituciones gubernamentales en los procesos de búsqueda e identificación, lo que agrava la crisis forense en el país. La falta de registros precisos y una base de datos actualizada por parte del gobierno mexicano, como resalta Consuelo Morales Elizondo (Morales, 2015), activista y fundadora de Ciudadanos en Apoyo a los Derechos Humanos, dificulta tener información estadística precisa sobre el problema.

Por otro lado, según la investigación de Josué Ibarra, estudiante de Estudios Políticos y Gobierno en la UDG (Ibarra, 2022), entre 2000 y 2022, se han reportado alrededor de 12,949 desapariciones en Jalisco, siendo los últimos años los más críticos con 7,912 casos. Por lo cual, la zona metropolitana de Guadalajara registra un 17% de todas las desapariciones en México.

Además, se destaca el conflicto entre las desapariciones forzadas y el incidente de los tráileres de la muerte en Jalisco, ya que, durante el gobierno de Aristóteles Sandoval, se utilizaron dos tráileres para transportar los 322 cadáveres que superaban la capacidad de las morgues (Vargas, 2020).

A través del proyecto RENIA, se propone abordar el problema de identificación de personas en condición de desaparición en Jalisco mediante la Inteligencia Artificial y el Procesamiento de Lenguaje Natural, con el objetivo de realizar búsquedas e identificar la similitud semántica entre descripciones textuales de las personas desaparecidas mediante un aplicativo web. El Procesamiento de Lenguaje Natural se encargará de la capacidad de las computadoras para manejar e interpretar el lenguaje humano y realizar diversas tareas como la traducción o el análisis de textos (Berryhill, 2019), como la similitud semántica, definida como una medida del grado en el que dos fragmentos de texto tienen el mismo significado.

Este trabajo toma como base de datos las fichas de personas fallecidas sin identificar (PFSI) del SEMEFO Jalisco (IJCF-PFSI, s.f.), las cuales son contrastadas con el significado semántico de descripciones de características físicas introducidas por el usuario en la aplicación web propuesta. Se implementa el modelo pre entrenado de PLN Universal Sentence Encoder (USE) para calcular la similitud entre el texto ingresado por el usuario y los registros de personas desaparecidas en la base de datos recolectada. Asimismo, la aplicación web demostró que esta búsqueda por medio

PLN puede ser una alternativa eficiente para usuarios sin conocimientos técnicos de IA, ya que ofrecerá una interfaz clara y sin filtros de búsqueda.

## 2. Planteamiento del problema

De acuerdo con Prieto Sferrazza Taibi (Sferrazza, 2021), la desaparición forzada de personas es considerada una de las más atroces violaciones a los derechos humanos, generando consecuencias perjudiciales tanto para la persona desaparecida, sus familiares y la sociedad en su totalidad. Por lo que, puede conceptualizarse como el arresto, detención, secuestro o cualquier forma de privación de la libertad por parte de agentes del Estado. Esta acción se acompaña de la negativa a reconocer la privación de libertad, ocultamiento de la suerte o paradero de la persona desaparecida, sustrayéndola a la protección de la ley.

Las víctimas de desapariciones son conscientes de que sus familias desconocen su paradero, reduciendo las posibilidades de recibir ayuda. Al quedar separados de la sociedad, se encuentran privadas de todos sus derechos y a merced de sus aprehensores, siendo el caso. Los familiares y amigos de estas personas pasan por una angustia similar, enfrentando la incertidumbre acerca del destino de la víctima. Vivir en constante preocupación por la causa de la desaparición es una angustia mental, ignorando si la víctima vive aún y de ser así, dónde se encuentra recluida, en qué condiciones y cuál es su estado de salud (“La desaparición forzada en México: una mirada desde los organismos del sistema de naciones unidas”, 2019).

Cepeda y Leetoy en el artículo ‘De víctimas a expertas: estrategias de agencia cívica para la identificación de desaparecidos en México’ (Cepeda, 2021), destacan la crisis de violencia sin precedentes en México. Jalisco es uno de los más afectados, con registros que indican 7,045 personas no localizadas y 2,238 desaparecidas desde 1995 hasta abril de 2020.

El colectivo de mujeres Por Amor a Ellxs (Facebook, s.f.), en la Zona Metropolitana de Guadalajara, ha implementado estrategias en la solución de la crisis de desapariciones. Utilizan redes sociales, especialmente Facebook, para amplificar la información que recolectan, extendiendo su alcance a otras familias en la misma situación. En su página comparten información sobre qué hacer en caso de la desaparición de un ser querido, resaltando la difusión de nombres, descripciones físicas y señas particulares de personas no reclamadas en las morgues. Desde su inicio en 2016, su labor ha contribuido a la identificación de más de 100 personas. La página se enriqueció con la participación de profesionales que asistieron a las instalaciones del SEMEFO para examinar fotografías de cuerpos no reclamados, con el objetivo de proporcionarles identidad y buscar a sus desaparecidos.

Además, el periódico El País en un video documental del año 2018, “Las familias de Jalisco en busca de sus desaparecidos” (Ibarra, 2022), presenta testimonios de diferentes personas buscando a sus desaparecidos. Consuelo Velázquez, entre ellos, menciona cómo le negaron tomar fotografías en el SEMEFO para difundirlas entre colectivos de búsqueda e identificación en Jalisco. Destaca la ausencia de difusión adecuada, lo que llevó a la necesidad de recurrir a descripciones manuscritas para compartir información entre colectivos y permitir que las personas sin identificar fueran reconocidas por sus familiares. Esta falta de herramientas tecnológicas de uso masivo para la difusión de información sensible dificulta aún más este proceso.

Para la propuesta de aplicativo web de este trabajo se toma como base de datos el Registro PFSI de SEMEFO Jalisco (IJCF-PFSI, s.f.), ya que posee un campo de búsqueda en su sitio web. Sin embargo, dentro de su buscador, se presenta una limitación ya que los usuarios que buscan personas desaparecidas desconocen las palabras exactas utilizadas por las autoridades para describir los cuerpos de personas fallecidas sin identificar.

## 3. Trabajos relacionados

Se ha revisado la literatura con el objetivo de identificar trabajos donde se produzcan aplicaciones de PLN enfocadas en la clasificación de textos y similitud semántica. Por ejemplo, se identifica que en (Merayo, 2019) se aplica PLN para identificar contenido inapropiado en texto, utilizando diversas técnicas como machine learning y deep learning, destacando máquinas de soporte vectorial. Esto demuestra la aplicabilidad de la técnica a redes sociales y plataformas como YouTube para identificar textos no regulados. Por otro lado, en (Hu, 2021) se aborda la clasificación de género de los nombres con modelos de aprendizaje automático basados en caracteres. Se utilizan algoritmos como Class Scaled Logistic Regression, Deep Neural Network, Long Short-Term Memory, Char-BERT y Name embedding, mostrando un enfoque interesante en la identificación.

En (Li, 2023) se emplean redes de gráficas convolucionales para hacer una clasificación del texto en una fuente heterogénea. El uso de grafos representa las conexiones que hay entre los pesos en la clasificación de los textos. Hemos encontrado que el esquema de encriptación de búsqueda semántica difusa (FSSE), descrito en (Liu, 2020), admite la búsqueda de múltiples palabras clave en datos encriptados en la nube. Permitiendo la búsqueda difusa y expansiones semánticas, utilizando huellas dactilares de palabras clave y distancia de Hamming para mejorar la precisión. De manera similar, en (Nemshaev, 2021) se discute la posibilidad de utilizar la búsqueda semántica para

nombrar a un experto en sistemas de automatización, describiendo características de la búsqueda por etiquetas, la ontología del dominio y el algoritmo de búsqueda semántica.

En (Jeong, 2022) se observa el enfoque hacia el análisis de conjunto de datos y el Procesamiento de Lenguaje Natural para la tarea de pregunta-respuesta basada en diálogos. Aborda las limitaciones de los modelos de lenguaje pre-entrenados que no suelen considerar el razonamiento de sentido común, por lo que se presenta el modelo Diálogo-QA con Razonamiento de Sentido Común (DQACR), el cual aprovecha la búsqueda semántica y el aprendizaje continuo para mejorar el razonamiento de sentido común y pérdida de información.

En cuanto al problema de Desambiguación del Sentido de Palabras (WSD en sus siglas en inglés) se aborda en (Nodehi, 2022) mediante técnicas de PLN y algoritmos metaheurísticos. La propuesta implica el uso de una red neuronal basada en grafos de WordNet para generar incrustaciones de palabras y sentidos simultáneos, mejorando la precisión mediante un método para agrupar y mapear sentidos en el grafo.

Por otro lado, considerando trabajos relacionados directamente con encontrar la similitud semántica de dos textos, se encontró que en (Balaha, 2021) se lleva a cabo la calificación automática de exámenes, comparando la similitud entre las respuestas de los estudiantes y las respuestas de referencia. Este trabajo comparó el rendimiento de varios modelos como BERT, Glove, FastText, Word2Vec y USE, donde éste último mostró mejor exactitud. En (Mahajan, 2020) se propone un ejercicio de identificación de grados de similitud entre registros de visitas periódicas de chequeos de salud a un hospital por parte de una persona. El trabajo destaca que las fichas de visitas suelen ser muy similares, pero es importante resaltar cuáles de ellas carecen de similitud en las descripciones textuales para identificar cambios en la salud del paciente. El experimento se llevó a cabo con el modelo pre entrenado de PLN BERT.

Adicionalmente, en (Sheth, 2021) se utiliza una aplicación con Universal Sentence Encoder para encontrar similitud en las descripciones de las características del recurso humano de una organización con las descripciones de capacitaciones. De esta forma, se busca identificar a las personas adecuadas para recibir entrenamiento en áreas específicas. En (Vowinckel, 2023), el contexto de las solicitudes de patentes, es fundamental identificar el estado de la técnica relevante y emplear estos documentos para evaluar la novedad e inventiva de la invención reivindicada. En este propósito, se utiliza el modelo BERT, que permite generar representaciones vectoriales de texto, facilitando el uso de similitud vectorial como indicador de la similitud semántica entre textos.

Finalmente, con el objetivo de conocer más sobre trabajos relacionados con identificar personas a partir de similitudes semánticas, se tiene que en (Zhou, 2023) se presenta una propuesta para encontrar imágenes de peatones utilizando PLN, buscando integrar características textuales de imagen y abordar la falta de información semántica en las características de las personas y analizando sus propiedades específicas en las imágenes y completándose con detalles semánticos. Se presenta una red de complemento de información que en (Frikha, 2021) se expone como un método de búsqueda de personas desaparecidas en áreas públicas, basado en atributos semánticos en sistemas de vigilancia. Utiliza clasificadores de atributos semánticos profundos basados en redes neuronales convolucionales para aprender y reconocer características en condiciones no controladas. En (Martin-Rodilla, 2019) se menciona la aplicación de técnicas de PLN en el análisis de informes forenses en el contexto de desapariciones forzadas durante la dictadura militar brasileña. Propone un sistema de extracción de información que identifica entidades nombradas y terminología clasificatoria e indexadora, asistiendo a los investigadores en la búsqueda de patrones en los informes de autopsia.

## **4. Propuesta**

La propuesta se basa en una aplicación web que utiliza el modelo USE de procesamiento de lenguaje natural, el cual codifica textos en vectores de alta dimensión que se pueden utilizar para calcular su similitud semántica. Con dicho modelo, un usuario internauta puede realizar una búsqueda en la aplicación web de una persona desaparecida escribiendo una descripción textual de características físicas, a lo que la aplicación responderá con las fichas de personas desaparecidas (provenientes del formulario PFSI para esta aplicación) ordena de mayor a menor coincidencia semántica.

Para la implementación de la aplicación web se llevaron a cabo 3 pasos fundamentales: primero, la recolección de información y generación de la base de datos; segundo, pruebas de desempeño del modelo USE en similitud semántica de textos con la base de datos recopilada; tercero, el desarrollo de la aplicación web para búsquedas textuales con similitudes semánticas en la base de datos recolectada. Mismos pasos que se describen a continuación.

### **4.1 Base de datos**

Como ya se mencionó anteriormente, la base de datos del aplicativo web se tomó del formulario PFSI (IJCF-PFSI, s.f.) que dispone SEMEFO Jalisco para la identificación de personas fallecidas sin identificar. El formulario PFSI descrito, cuenta con 9 campos: Id, Fecha ingreso, Sexo, Probable nombre, Edad, Tatuajes, Indumentarias, Señas particulares

y Delegación IJCF, los cuales pueden contener, o no, información de los cuerpos. Tatuajes, Indumentarias y Señas particulares son campos que tienen descripciones textuales. Con la finalidad de utilizar la similitud semántica y no una búsqueda por medio de filtros, por cada registro todos los campos, excepto el campo Id, se concatenaron en uno solo llamado descripción y se guardaron en una base de datos MySQL. Por consiguiente, la base de datos del aplicativo cuenta con 4,278 registros; cada registro sólo tiene dos campos: el Id del registro y el campo Descripción, el cual será utilizado para analizar las similitudes semánticas con las búsquedas ingresadas por el usuario en el aplicativo web.

## 4.2 Arquitectura del modelo USE

Antes de explicar la arquitectura del modelo USE, es importante definir primero algunos conceptos clave como:

**Tokenización:** En procesamiento del lenguaje natural, se basa en la idea de convertir una secuencia de texto en partes más pequeñas (“What is Tokenization? Types, Use Cases, Implementation”, s.f.).

**Incrustación:** se refiere a una representación vectorial de un objeto, como una palabra, un documento o una imagen. Esta representación captura características y relaciones importantes del objeto en un espacio de menor dimensión, permitiendo que las máquinas los manipulen y analicen de manera más eficiente.

La arquitectura del codificador universal de frases (USE) incluye un modelo de codificación formulado como una red de promediación profunda (DAN). A continuación, se explica paso a paso la arquitectura del modelo utilizado para el desarrollo del prototipo:

### 4.2.1 Codificador de red de promediado profundo (DAN):

El codificador de red de promediado profundo (DAN), un componente del codificador Universal Sentence Encoder (USE) (Cer, 2018), está diseñado para generar incrustaciones de frases de forma eficiente. La arquitectura se explica paso a paso a continuación:

1. Incrustaciones de entrada: El codificador DAN toma las incrustaciones de entrada para palabras y bi-gramas y las promedia juntas (Cer, 2018).
2. Red neuronal: Las incrustaciones promediadas se introducen en una red neuronal profunda (DNN).
3. Incrustación de frases: La DNN procesa las incrustaciones promediadas para producir incrustaciones de frases de 512 dimensiones.
4. Entrenamiento e implementación: El codificador DAN se entrena de forma similar al codificador basado en transformadores y se implementa en TensorFlow (Cer, 2018).
5. Eficiencia y compensaciones: El codificador DAN está optimizado para una inferencia eficiente, ofreciendo una precisión ligeramente reducida en comparación con el codificador basado en transformador (Cer, 2018). Este compromiso entre precisión y recursos computacionales lo hace adecuado para diversas tareas de procesamiento del lenguaje natural y aplicaciones de aprendizaje por transferencia.

Otro de los puntos fundamentales de la arquitectura son las tareas de transferencia en el contexto del codificador universal de frases (USE), ya que implican la aplicación de incrustaciones de frases pre-entrenadas a diversas tareas de procesamiento del lenguaje natural, ofreciendo una visión del impacto de la selección e integración de modelos en el rendimiento de las tareas.

Esta arquitectura proporciona un enfoque flexible y eficiente para generar incrustaciones de frases, atendiendo a las diversas necesidades de las tareas de procesamiento del lenguaje natural.

## 4.3 Pruebas de desempeño del modelo USE

El modelo USE está entrenado y optimizado para procesar texto que consiste en más de una palabra, como oraciones, frases o párrafos cortos. Su entrenamiento abarca diversas fuentes de datos y tareas, con el objetivo de adaptarse dinámicamente a la comprensión del lenguaje natural. La entrada al modelo es un texto de longitud variable, y la salida es un vector de 512 dimensiones. Se aplica el modelo de referencia STS (Semantic Textual Similarity) para medir la similitud semántica, evaluado mediante el benchmark STS. Este benchmark proporciona una medida de cómo las puntuaciones de similitud, calculadas mediante incrustaciones de oraciones, se relacionan con los juicios humanos.

Por otro lado, se emplea la correlación de Pearson (una prueba que mide la relación estadística entre dos variables continuas) para evaluar la calidad de las puntuaciones de similitud generadas por la máquina en comparación con las evaluaciones humanas (adondevanlosdesaparecidos, 2022). Este codificador se distingue de otros modelos de incrustación de nivel de palabra, porque se entrenó en una serie de tareas de predicción de lenguaje natural que requieren modelar el significado de secuencias de palabras individuales (adondevanlosdesaparecidos, 2022).

Las pruebas de rendimiento del modelo USE implican analizar qué tan bien el modelo relaciona una descripción textual



con su registro similar escrito con palabras diferentes. Para llevar a cabo estas pruebas, se seleccionaron 40 registros de manera aleatoria, con la condición de que tuvieran más de 30 palabras (garantizando así que tuvieran una descripción textual). Después, esos 40 registros se parafrasearon tres veces con palabras distintas y con diferentes cantidades de palabras, utilizando la plataforma de IA Chat GPT (ChatGPT, s.f.). Se solicitó el parafraseo con una cantidad similar de palabras, el parafraseo con el 30% menos de palabras y el parafraseo con el 30% más de palabras. Posteriormente, se compararon las descripciones parafraseadas una a una, evaluando la similitud semántica con el modelo USE y los 4,278 registros de la base de datos. Al identificar cada registro parafraseado con su ID, se pudo determinar si el modelo lo clasificó entre los 100 IDs de registros con mayor similitud mediante el coeficiente de correlación de Pearson. El hecho de que el modelo detecte similitud de un registro parafraseado dentro de los 100 IDs de registros con mejor similitud se consideró como un acierto del modelo. De las 40 pruebas por cada uno de los tres tipos de parafraseo, se obtuvo el porcentaje de aciertos del modelo. El porcentaje de aciertos del modelo ayudó a estimar cuánto más fácil resulta realizar una búsqueda con la aplicación web en comparación con hacerlo manualmente en la base de datos cuando no se conocen las palabras exactas con las que se registró el cuerpo de una persona fallecida sin identificar en SEMEFO.

#### 4.4 Desarrollo de la aplicación web

Diagrama de casos de uso

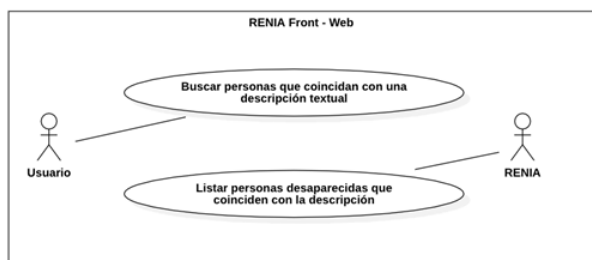


Figura 1. Diagrama de casos de uso del aplicativo web

El desarrollo del aplicativo se dividió en Front-End y Back-End, los cuales conforman módulos de una arquitectura orientada a servicios (siendo el cliente el Front-end y el servidor el Back-End) como se muestra en la Figura 1 y Figura 2, respectivamente. El Back-End implementa una API escrita en lenguaje Python, que proporciona un end-point donde se calcula la similitud semántica de la búsqueda del usuario con los registros de la base de datos a través del modelo USE, devolviendo los registros ordenados de mayor a menor similitud semántica. Por otro lado, el Front-End está diseñado con el framework React para crear interfaces de usuario. La Figura 3 presenta la interfaz gráfica de búsqueda y obtención de resultados del aplicativo web. En el repositorio <https://github.com/alexAgnus/RENIAtexto.git> (Briseño, s.f.) se puede descargar el código de la implementación para pruebas y mejoras del aplicativo.

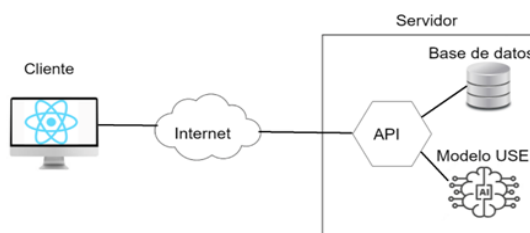


Figura 2. Arquitectura cliente-servidor de la aplicación web del proyecto RENIA.



Figura 3. Interfaz gráfica de la aplicación web en el apartado de búsqueda de personas desaparecidas a través de similitud semántica de textos.

## 5. Resultados

Con el objetivo de parafrasear los 40 registros seleccionados de manera aleatoria y asegurando que tuvieran más de 30 palabras, se proporcionaron las siguientes instrucciones a Chat GPT: *reescribe las siguientes oraciones con diferentes palabras, manteniendo el significado semántico; reescribe las siguientes oraciones con diferentes palabras manteniendo el significado semántico y disminuyendo el número de palabras en un 30%; y reescribe las siguientes oraciones con diferentes palabras manteniendo el significado semántico y aumentando el número de palabras en un 30%. Dichas instrucciones generaron tres bases de datos de registros parafraseados: una con la misma cantidad de palabras, otra con menos palabras y una tercera con más palabras, respectivamente.*

De las 120 pruebas totales realizadas, el modelo acertó en 95 ocasiones, con un 79% de efectividad al encontrar el registro buscado de manera parafraseada en los primeros 100 registros con mayor similitud. En las 40 pruebas de búsqueda con parafraseo y una cantidad similar de palabras distintas, el modelo acertó en 32 ocasiones con un 80% de efectividad. En las búsquedas con parafraseo disminuyendo la cantidad de palabras, el modelo acertó en 36 ocasiones (90% de efectividad). Finalmente, en las búsquedas con parafraseo aumentando el número de palabras, el modelo acertó en 27 ocasiones (67.5% de efectividad) como se muestra en la Tabla 1.

Tipo de búsqueda	Aciertos de 40 búsquedas	Porcentaje de efectividad
Parafraseo con la misma cantidad de palabras distintas	32	80%
Parafraseo con el 30% menos de palabras	36	90%
Parafraseo con el 30% más de palabras	27	67.5%

Tabla 1. Resultados de las pruebas de desempeño del modelo USE.

## 6. Discusión y conclusiones

Las pruebas de PLN utilizando similitud semántica de textos realizados en este trabajo, utilizando el modelo USE y la base de datos de personas desaparecidas sin identificar (PFSI) de SEMEFO Jalisco, demostraron ser una herramienta potente en la identificación de personas a través de descripciones textuales de rasgos físicos. También se observó que al reescribir un registro con palabras distintas y reducir en un 30% la cantidad de palabras para su posterior búsqueda de similitud semántica, se obtiene una mayor efectividad en las coincidencias que mantener o aumentar el número de palabras del registro original. Esto podría deberse a que, al parafrasear con palabras distintas en menor cantidad, el registro conserva las palabras clave con más peso en la interpretación del significado semántico. Por otro lado, el aumentar palabras que no aporten sentido semántico podría sesgar el significado del texto a buscar.

El mejor resultado que puede alcanzar la aplicación web es que en un 90% de las búsquedas de personas, donde se utilicen en su mayoría palabras clave en las descripciones, la persona buscada se encuentre dentro de los primeros 100 registros devueltos. Considerando que la base de datos de esta aplicación contiene 4,278 registros, explorar sólo

los primeros 100 en 9 de cada 10 búsquedas agiliza el proceso de identificación de personas.

Posiblemente, la mejor forma de identificar personas por descripciones textuales sea combinar técnicas tradicionales para comparación de similitud sintáctica con técnicas de comparación en el significado semántico, con el objetivo de mejorar la exactitud en las búsquedas.

## 7. Trabajo futuro

Como trabajo futuro, se plantea realizar una comparación entre el modelo USE utilizado en la presente propuesta y las ventajas que brindan las relaciones semánticas presentes en una base de datos basada en grafos, como NEO4J. Se busca discutir las fortalezas y debilidades de ambos enfoques. Además, se buscará mejorar el rendimiento de la aplicación mediante la creación de un modelo híbrido, que incorpore diferentes paradigmas de similitud textual, tales como el sintáctico y el orientado a grafos.

## 8. Referencias

- adondevanlosdesaparecidos. (23 de junio de 2022). Desaparición de personas: un paradigma del crimen perfecto. A dónde van los desaparecidos. Recuperado de: <https://adondevanlosdesaparecidos.org/2022/06/22/desaparicion-de-personas-un-paradigma-del-crimen-perfecto/>
- Balaha, H. M., & Saafan, M. M. (2021). Automatic Exam Correction Framework (AECF) for the MCQs, Essays, and Equations Matching. IEEE Access, 9, 32368-32389. Recuperado de: <https://doi.org/10.1109/ACCESS.2021.3060940>
- Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (noviembre de 2019). Hello, World: Artificial intelligence and its use in the public sector. OECD. Recuperado de: <https://doi.org/10.1787/726fd39d-en>
- Briseño, R. A. (s.f.). Alexagnus/Reniatexto. GitHub. Recuperado el 4 de noviembre de 2023, de <https://github.com/alexAgnus/RENIAtexto>
- Cepeda, A., & Leetoy, S. (enero de 2021). De víctimas a expertas: estrategias de agencia cívica para la identificación de desaparecidos en México. Íconos - Revista de Ciencias Sociales, (69). Recuperado de: <https://doi.org/10.17141/iconos.69.2021.4197>
- Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., St. John, R., ... & Sung, Y.-H. (2018). Universal Sentence Encoder. arXiv preprint arXiv:1803.11175.
- ChatGPT. (s.f.). Recuperado el 31 de agosto de 2023, de <https://chat.openai.com>
- Facebook. (s.f.). Recuperado el 31 de agosto de 2023, de <https://www.facebook.com/PorAmorAEIIXS>
- Frikha, M., Fendri, E., & Hammami, M. (enero de 2021). Deep Semantic Attributes for People Search. Procedia Computer Science, 192, 90-99. Recuperado de: <https://doi.org/10.1016/j.procs.2021.08.010>
- Hu, Y., Hu, C., Tran, T., Kasturi, T., Joseph, E., & Gillingham, M. (7 de febrero de 2021). What's in a Name? -- Gender Classification of Names with Character Based Machine Learning Models. arXiv.org. Recuperado de: <https://arxiv.org/abs/2102.03692v1>
- Ibarra, J. (5 de mayo de 2022). Jalisco registra números rojos en desaparición de personas. Zona Docs. Recuperado de: <https://www.zonadocs.mx/2022/05/05/jalisco-registra-numeros-rojos-en-desaparicion-de-personas/>
- IJCF-PFSI. (s.f.). Recuperado el 29 de agosto de 2023, de [https://cienciasforenses.jalisco.gob.mx/registro\\_pfsi.php](https://cienciasforenses.jalisco.gob.mx/registro_pfsi.php)
- Jeong, S., Oh, D., Park, K., & Lim, H. (enero de 2022). Considering Commonsense in Solving QA: Reading Comprehension with Semantic Search and Continual Learning. Applied Sciences, 12(9). Recuperado de: <https://doi.org/10.3390/app12094099>
- LA DESAPARICIÓN FORZADA EN MÉXICO: UNA MIRADA DESDE LOS ORGANISMOS DEL SISTEMA DE NACIONES UNIDAS. (2019). Recuperado de [https://www.cndh.org.mx/sites/default/files/documentos/2019-09/lib\\_DesaparicionForzadaMexicoUnaMirada.pdf](https://www.cndh.org.mx/sites/default/files/documentos/2019-09/lib_DesaparicionForzadaMexicoUnaMirada.pdf)
- Li, H., Yan, Y., Wang, S., Liu, J., & Cui, Y. (marzo de 2023). Text classification on heterogeneous information network via enhanced GCN and knowledge. Neural Computing and Applications, 35(20), 14911-14927. Recuperado de: <https://doi.org/10.1007/s00521-023-08494-0>

- Liu, G., Yang, G., Bai, S., Zhou, Q., & Dai, H. (2020). FSSE: An Effective Fuzzy Semantic Searchable Encryption Scheme Over Encrypted Cloud Data. *IEEE Access*, 8, 71893-71906. Recuperado de: <https://doi.org/10.1109/ACCESS.2020.2966367>
- Mahajan, D., et al. (noviembre de 2020). Identification of Semantically Similar Sentences in Clinical Notes: Iterative Intermediate Training Using Multi-Task Learning. *JMIR Medical Informatics*, 8(11), e22508. Recuperado de: <https://doi.org/10.2196/22508>
- Martin-Rodilla, P., Hattori, M. L., & Gonzalez-Perez, C. (julio de 2019). Assisting Forensic Identification through Unsupervised Information Extraction of Free Text Autopsy Reports: The Disappearances Cases during the Brazilian Military Dictatorship. *Information*, 10(7). Recuperado de: <https://doi.org/10.3390/info10070231>
- Merayo-Alba, S., Fidalgo, E., González-Castro, V., Alaiz-Rodríguez, R., & Velasco-Mata, J. (2019). Use of Natural Language Processing to Identify Inappropriate Content in Text. En *Hybrid Artificial Intelligent Systems* (pp. 254-263). Springer International Publishing. Recuperado de: [https://doi.org/10.1007/978-3-030-29859-3\\_22](https://doi.org/10.1007/978-3-030-29859-3_22)
- Morales, E. C. (2015). La desaparición: un problema que impacta a las personas vulnerables.
- Nemshaev, S., Barykin, L., & Dadteev, K. (enero de 2021). Selection of experts for scientific and technical expertise based on semantic search. *Procedia Computer Science*, 190, 643-646. Recuperado de: <https://doi.org/10.1016/j.procs.2021.06.102>
- Nodehi, A. K., & Charkari, N. M. (septiembre de 2022). A metaheuristic with a neural surrogate function for Word Sense Disambiguation. *Machine Learning with Applications*, 9, 100369. Recuperado de: <https://doi.org/10.1016/j.mlwa.2022.100369>
- Sferrazza Taibi, P. (julio de 2021). La búsqueda de personas desaparecidas: derecho humano de las víctimas y obligación internacional del Estado. *Estudios constitucionales*, 19(1), 265-308. Recuperado de: <https://doi.org/10.4067/S0718-52002021000100265>
- Sheth, D., Gupta, A. R., & D'Mello, L. (noviembre de 2021). Using Universal Sentence Encoder for Semantic Search of Employee Data. En *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)* (pp. 1-4). Recuperado de: <https://doi.org/10.1109/ICCICA52458.2021.9697114>
- Vargas Fara González, D. F., Tapia, F., Gallardo, A., & Samantha. (11 de noviembre de 2020). Jalisco: La verdad de los "tráileres de la muerte". *Crisis Forense*. Recuperado de: <https://quintoelab.org/crisisforense/jalisco-la-verdad-de-los-trailerres-de-la-muerte/>
- Vowinckel, K., & Hähnke, V. D. (junio de 2023). SEARCHFORMER: Semantic patent embeddings by siamese transformers for prior art search. *World Patent Information*, 73, 102192. Recuperado de: <https://doi.org/10.1016/j.wpi.2023.102192>
- What is Tokenization? Types, Use Cases, Implementation. (s.f.). Recuperado el 18 de diciembre de 2023, de <https://www.datacamp.com/blog/what-is-tokenization>
- Zhou, H., Li, F., Tian, X., & Huang, Y. (2023). Feature semantic alignment and information supplement for Text-based person search. *Frontiers in Physics*, 11. Recuperado de: <https://www.frontiersin.org/articles/10.3389/fphy.2023.1192412>



[www.iieg.gob.mx](http://www.iieg.gob.mx)

Dirección de Información de Gobierno, Seguridad Pública e Impartición de Justicia