Machine Learning Analysis of Consolidated Purchasing: A Case Study of Antiretroviral Medication 2019 Pricing Trends in Mexico

Análisis de aprendizaje automático de compras consolidadas: un estudio de caso sobre las tendencias de precios de medicamentos antirretrovirales en México en 2019

Blanca Iveth Mayorga Basurto
EGADE Business School, Tecnológico de Monterrey, Mexico

Galo Moncada Freire
Business School, Tecnológico de Monterrey, Mexico

Received: February 27, 2024.
Approved: May 15, 2024.

Abstract

This paper investigates trends in antiretroviral medication prices and their impact on public health in Mexico during 2019. Using three machine learning models developed in Python (logistic regression, random forest, and K-Nearest Neighbors or KNN), this study discerns increasing or decreasing patterns in antiretroviral (ARV) drug price fluctuations using a dataset comprising 15,220 observations of ARV drugs acquired between 2016 and 2019. Results indicate that random forests exhibited the highest precision in predicting price changes, followed by KNN and logistic regression. Significant factors affecting acquisition prices, such as drug type and duration of procurement strategy, were identified. In addition to analyzing price trends, the paper explores the budgetary considerations associated with these fluctuations, providing insights into the financial implications for healthcare systems and stakeholders. It is important to note that this paper focuses on a specific ARV pharmaceutical purchasing scheme. Moreover, the study emphasizes the creation of a unified and detailed medication price database, highlighting the significant effort invested in compiling complete and comprehensive information from various sources. This study's findings underscore the effectiveness of initiatives such as consolidated purchasing approaches and the integration of newer, cost-effective medications into treatment protocols. These initiatives have led to significant cost savings in antiretroviral medication procurement, contributing to improved access for individuals living with HIV/AIDS. Overall, the research highlights the importance of data-driven approaches and strategic planning in optimizing pharmaceutical purchasing processes and ensuring sustainable access to essential medications for public health interventions.

Keywords: *purchasing strategies, public health, Mexico, machine learning.*
JEL Classification: *I11, I18, H57.*

Resumen

Este trabajo examina las tendencias en los precios de medicamentos antirretrovirales y su impacto en la salud pública, centrándose en el contexto mexicano en 2019. Por medio de técnicas de aprendizaje automático, el estudio analiza las fluctuaciones en los precios de los medicamentos antirretrovirales, con base en un conjunto de datos que comprende 15,220 observaciones de medicamentos antirretrovirales (ARV) adquiridos entre 2016 y 2019, con énfasis particular en el año 2019. Los resultados indican que los "bosques aleatorios" demostraron la mayor precisión en la predicción de cambios de precios, seguidos por K-Nearest Neighbors *o KNN (algoritmo de k vecinos más próximos) y la regresión logística. Se identificaron factores significativos que afectan los precios de adquisición, como el tipo de medicamento y la duración de la estrategia de adquisición. Además, el estudio profundiza en las consideraciones presupuestarias, evaluando las implicaciones financieras de estas fluctuaciones de precios. Estos hallazgos destacan la efectividad de iniciativas como los enfoques de adquisición consolidada y la integración de medicamentos más nuevos y rentables en los protocolos de tratamiento, lo que conduce a ahorros significativos y mejor acceso para las personas que viven con VIH/sida. Es importante destacar que este artículo se centra en analizar un esquema específico de adquisición de medicamentos antirretrovirales. Los estudios futuros ampliarán este análisis para abarcar un espectro más amplio de esquemas de adquisición de medicamentos, proporcionando así una comprensión más completa de la dinámica de fijación de precios y sus implicaciones para la salud pública. Además, la investigación en curso perfeccionará la metodología e integrará fuentes de datos adicionales para mejorar la precisión y profundidad del análisis.*

Palabras clave: *ARV, tendencias de precios, compras consolidadas, México.*
Clasificación JEL: *I11, I18, H57.*

1. Introduction[1]

In April 2019[2], a citizen petition was submitted to the Secretaría de Hacienda y Crédito Público (SHCP, Ministry of Finance and Public Credit) to address an extremely urgent case related to the shortage of medications intended for the treatment of over 100,000 patients infected with the Human Immunodeficiency Virus (HIV) in public hospitals. Dr. Carlos Urzúa, then Minister of SHCP, pointed out that pharmaceutical purchasing had became complicated due to many HIV specialists' insistence on acquiring over thirty different HIV inhibitors, in order to perform various therapeutic combinations. These factors, coupled with restrictions imposed by antiretroviral (ARV) drug providers and associated cost overruns, posed a risk to the supply of ARV medications and the care of all HIV patients in the public sector (Urzúa, 2020).

In response to the situation, the SHCP summoned various multidisciplinary groups including medical experts, scientists, and administration and finance professionals with experience in consolidated purchasing. These groups met to analyze the problem's root causes and seek possible solutions, focusing primarily on the public health perspective. As a result of this analysis, a drastic change in the consolidated purchasing model was implemented in 2019. This change also involved the modification of the purchasing procedure, previously coordinated by the Mexican Social Security Institute (IMSS, Instituto Mexicano del Seguro Social).

In this context, interest in conducting a comprehensive analysis of this new approach to consolidated purchasing arose in order to assess its effectiveness. Efficient resource management in the healthcare sector is a crucial concern today, and one of the most important fields in this area is the procurement and distribution of ARV medications for HIV treatment. While it cannot eradicate HIV infection, "Antiretroviral

[1] This study was possible thanks to the support and collaboration of various government entities in Mexico. We especially thank the Secretaría de Hacienda y Crédito Público (Ministry of Finance and Public Credit), Secretaría de Salud (Health Ministry), CENSIDA (Centro Nacional para la Prevención y Control del VIH y el sida), IMSS (Mexican Social Security Institute), ISSSTE (Institute of Security and Social Services for State Workers), Pemex (Mexican Petroleum), and INAI (National Institute for Transparency, Access to Information, and Protection of Personal Data) for providing us with the necessary data to conduct this research. Their contribution was fundamental to the success and significance of the results obtained.

[2] In preparing this paper, we consulted and interviewed officials from the Mexican government, primarily from the SHCP, the Secretaría de Salud (SS), and the Centro Nacional para la Prevención y Control del VIH y el sida (CENSIDA).

therapy (ART) suppresses viral replication, increases $CD4^+$ cell-count, decreases AIDS-related mortality and morbidity and comorbidities, improves the quality of life of HIV-infected patients, and prevents sexual transmission of HIV" (Lozano & Domingo, 2011). Within this framework, there is a need to explore the results of a radical ARV procurement strategy implemented by the Mexican government, and how this strategy has affected both the economic aspect and the quality of care for patients.

This research takes shape as part of a collaborative effort that brings experts from various disciplines together, including finance, information technology, and ARV medication administration. The main objective is to assess whether the paradigm shift in the ARV medication procurement process—through the implementation of consolidated purchasing—has achieved the expected benefits both in terms of economic efficiency and improvement of healthcare provided to HIV patients.

Consolidated purchasing has been presented as a promising strategy to achieve significant economic savings and ensure the constant availability of essential medications. In this regard, this article thoroughly analyzes whether this approach has led to substantial economic savings and improved the quality of care for HIV patients at the same time.

A crucial aspect of the methodology is the application of advanced data analysis and machine learning techniques, which play a key role in evaluating this strategy's results. These techniques allow for a rigorous and comprehensive analysis of ARV medication procurement data, and the exploration of patterns and relationships that help better understand the consolidated purchasing strategy's underlying dynamics and assess its real impact.

The essential contribution of this research lies in three interrelated dimensions. First, it examines whether consolidated purchasing strategies for ARV medications have had a positive impact on the costs and financial sustainability of HIV treatment programs. Pattern and trend identification through machine learning techniques offers valuable information for informed decision-making regarding resource allocation.

Second, it addresses a gap in the literature by exploring how the implementation of advanced data analysis and machine learning techniques can optimize the financial administration of ARV procurement. These techniques provide a predictive insight into future trends in pharmaceutical purchasing, enabling efficient resource planning and allocation.

Lastly, this research highlights the importance of a multidisciplinary approach to addressing the complex challenges in ARV procurement management. Collaboration among experts in different fields provides a unique and enriching perspective in analyzing the challenges linked to the purchasing and efficient distribution of ARV medications.

2. Methodology

The method adopted is designed to comprehensively evaluate the efficiency and effectiveness of the ARV medication consolidated purchasing scheme in Mexico during 2019. The methodological approach includes the following steps:

- *Demand and Coverage Analysis*: It starts with a detailed analysis of statistical data sourced from international organizations and the Government of Mexico. (WHO, 2021a; WHO, 2021b). This analysis encompasses the demand for ART among individuals living with HIV/AIDS, internationally, in Latin America, and specifically in Mexico, including relevant information regarding ART coverage worldwide (33%), in Latin America (38%), and in Mexico, where 172,221 people have access to it. Of these, 99.2% receive it from the public sector and 0.8% from the private sector (SS, 2018; UNAIDS, 2020; UNAIDS, 2021).
- *Budgetary Analysis and Savings Evaluation:* Budgets are highlighted as a key factor in determining savings derived from variations in ARV medication prices. Fluctuations in average prices are quantified, which were around 55% in this study.
- *Medication Price Analysis:* We conducted a detailed analysis of ARV medication prices using the set of prices of all 15,220 medications procured by the National Center for HIV Prevention and Control of HIV and AIDS (Centro Nacional para la Prevención y Control del VIH y el sida, CENSIDA) from 2017 to 2020 as a basis. Additionally, we have a database of medications procured in fiscal years prior to 2019, when the remaining agencies and entities procured ARV medications outside the consolidated purchasing scheme. In this way, the results of different contracting schemes can be contrasted using these databases.
- *Analysis with Machine Learning Algorithms:* In order to project ARV medication prices, we employed machine learning algorithms in R and Python statistical environments (McKinney, 2013). We identified relevant data sets, including

variables such as purchasing agencies and entities, procured ARV regimens, and we classified all 15,220 data items used in the model construction. Additionally, we carried out analyses on the suppliers with allocations in 2017, 2018, and 2019, in response to the relevance identified through information requests to the Mexican government.

The methodology describes in detail the implementation of logistic regression, random forest, and K-Nearest Neighbors (KNN) models, including a theoretical explanation of the functioning of each. Research results are presented at the end of the development section, providing a thorough data-supported analysis of the consolidated purchasing scheme's effectiveness in 2019.

The adopted methodology seeks to ensure comprehensiveness and rigor in evaluating the ARV medication consolidated purchasing scheme. It relies on robust analytical approaches and machine learning algorithms to draw objective conclusions based on relevant empirical evidence.

3. Development

We conducted a comprehensive analysis to evaluate the effectiveness of the 2019 consolidated purchasing program, which ensures the supply of ARV medications to HIV patients.

We refer to the 2019 consolidated purchasing, which comprises the 2019-2020 fiscal cycle, from April 1, 2019, to March 31, 2020.

In order to ensure result accuracy and ensure that the decrease in ARV medication costs was not due to insufficient use of budgetary resources nor to a reduction in the quality of regimens provided by the Secretaría de Salud (SS, Health Ministry), we applied rigorous evaluation criteria:

- *Quality:* We examined improvements in HIV patient treatment.
- *Cost (budget):* We analyzed the reduction in ARV medication procurement prices.

The analysis was based on identifying and segmenting various relevant variables, including state, age, and agency. Table 1 shows the amounts of ARV medication procured, the quantity of ARV procured, and various key indicators at different times (see Table 1).

Table 1. ARV Medication Procured from 2017 to 2020

Cutoff Date	Amount of ARV Procured (in Mexican pesos, $MX)	Quantity of ARV Procured	Individuals on ART	Individuals with CD4 <200	Awarded Providers
31/03/17	537,812,507.78	339,566			7
30/06/17	1,276,375,861.82	608,514			3
30/09/17	488,984,118.03	232,862			6
31/12/17	960,073,004.33	485,409	87,026	13,101	8
31/03/18	393,950,549.44	251,783	89,315	12,813	8
30/06/18	985,625,837.20	446,229	91,194	13,259	3
30/09/18	815,559,854.53	438,252	93,037	13,762	8
31/12/18	871,150,963.21	467,876	95,732	14,004	8
31/03/19	350,896,989.49	221,428	97,245	3,498	8
30/06/19	811,635,301.73	521,787	98,765	3,498	14
30/09/19	685,147,693.28	537,324	99,531	6,921	14
31/12/19	630,004,095.86	425,418	100,409	11,273	15
31/03/20	553,478,872.57	346,809	102,288	2,147	15

Source: Compiled by the authors with data from CENSIDA (2017a), CENSIDA CENSIDA (2017b), CENSIDA (2017c), CENSIDA (2017d), CONASIDA (2018), CENSIDA (2018a), CENSIDA (2018b), CENSIDA (2018c), CENSIDA (2018d), CENSIDA (2019e), CENSIDA (2019f), CENSIDA (2019g), CENSIDA (2019h), CENSIDA (2019i), CENSIDA (2020a), CENSIDA (2021a) y CENSIDA (2021b).

The study relies on data collection and examination to provide a comprehensive understanding of the consolidated purchasing scheme's effectiveness in terms of its influence on treatment quality and cost reduction in the procurement of ARV medication for HIV patients. Segmentation based on key variables allows for a complete and enriching view of the results obtained, enabling a precise and detailed evaluation of the implemented initiative.

3.1. Demand and Coverage Analysis

The analysis focuses on the comprehensive evaluation of demand and coverage of ARV treatment regimens provided to HIV patients in Mexico. The aim is to understand the distribution, effectiveness, and characteristics of care programs in the public sector. The data used came from various official sources, including CENSIDA reports, which is the primary national agency for the prevention and control of HIV/AIDS in Mexico, as well as responses to information requests submitted to the Plataforma Nacional de Transparencia (PNT, National Transparency Platform) of the Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI, National Institute for Transparency, Access to Information, and Protection of Personal Data), Mexico's principal body responsible for ensuring transparency and access to public information.

According to the National Report on Monitoring Commitments and Expanded Goals to End AIDS, the distribution of ART in Mexico is largely concentrated in public sector institutions, with 99.2% of individuals receiving treatment in these entities, while 0.8% receive it in the private sector (SS, 2018).

Within the public sector, government institutions that provide ART to HIV patients include Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE, Institute of Security and Social Services for State Workers), IMSS, Petróleos Mexicanos (Pemex, Mexican Petroleum), Secretaría de la Defensa Nacional (Sedena, National Defense Ministry), Secretaría de Marina (Semar, Ministry of the Navy), and Secretaría de Salud (SS) through CENSIDA.

Regarding the specific patient distribution by institution and cutoff date, we can observe in Table 2 that the SS treats the largest number of patients, with a total of 98,100 patients under ART as of June 2019. Likewise, IMSS plays an important role, treating 55,818 patients as of December 2018, while Pemex treats 881 patients on the same date. However, it is worth noting that data provided by Sedena, Semar, and ISSSTE do not specify the number of patients treated (see Table 2).

Table 2. Patients Receiving ART

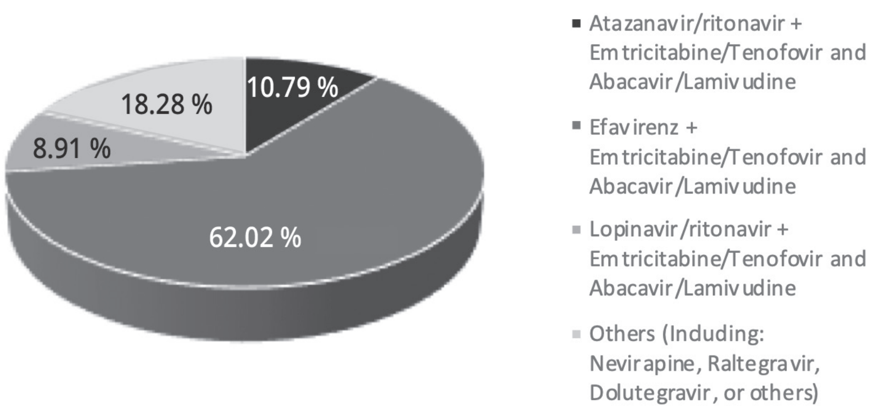
Institution	Number of Patients on ART	Cutoff Date
SS	98,100	June 10, 2019
IMSS	55,818	December 1, 2018
Pemex	881	June 25, 2019
Semar	No data available	No data available
Sedena	No data available	No data available
ISSSTE	No data available	No data available
Total	**154,799**	

Source: Compiled by the authors with data from information requests made to INAI in 2021.

3.2. ARV Treatment Regimens in Mexico

Mexico's ARV medicine market situation reflects concentration on first-line regimens, as revealed by the study "Antiretroviral Purchasing and Prescription Practices in Mexico: Constraints, Challenges and Opportunities" and shown in Figure 1 (Chaumont et al., 2015) (see Figure 1).

Figure 1. Antiretroviral Regimens in Mexico



Source: Compiled by the authors with data from information requests made to INAI in 2021 and with information from the basic framework and catalogue of medications issued by the General Health Council (Consejo de Salubridad General, 2017).

According to records up to June 2019 shown in Figure 1, regimens commonly used by government institutions include Efavirenz plus Emtricitabine/Tenofovir and Abacavir/Lamivudine, administered to 96,013 patients; Atazanavir/Ritonavir plus Emtricitabine/Tenofovir and Abacavir/Lamivudine, administered to 16,697 patients; Lopinavir/Ritonavir plus Emtricitabine/Tenofovir and Abacavir/Lamivudine, administered to 13,795 patients, and other regimens that incorporate agents such as Nevirapine, Raltegravir or others, administered to approximately 28,294 patients on ART (CENSIDA, 2019e).

Moreover, treatment regimens used adhere to CENSIDA's guidelines for administering ARV medications, which predominantly consist of groupings of Nucleoside Reverse Transcriptase Inhibitors (NRTIs) and/or Protease Inhibitors (PIs). These groupings vary, with Emtricitabine/Tenofovir plus a third agent such as Efavirenz, Atazanavir/Ritonavir, or Lopinavir/Ritonavir being the most common. Additional agents such as Nevirapine, Raltegravir, or Dolutegravir are also used, albeit in smaller proportions, particularly in cases of viral resistance.

Treatment regimen selection is guided by expert recommendations and customized to address the specific conditions and mutations present in each patient (CENSIDA, 2021c).

Table 3 provides a summary of the amount each of the regimens has been prescribed, commonly supplied by government agencies and entities in the Mexican Government before consolidated purchasing (see Table 3).

Table 3. Most Used ARV Treatment Regimens

ART Regimen	CENSIDA[3]	IMSS[4]	Pemex[5]	Total
Efavirenz + Emtricitabine/Tenofovir and Abacavir/Lamivudine	66,880	28,888	245	96,013
Others (Including: Nevirapine, Raltegravir, or Dolutegravir, among others)	21,166	6,938	190	28,294

[3] Response to request 1200800005619. Numbers as of June 10, 2019.
[4] Response to request 64101707619. Numbers as of December 2018.
[5] Response to request 1857200281519. Numbers as of June 25, 2019.

ART Regimen	CENSIDA[3]	IMSS[4]	Pemex[5]	Total
Atazanavir/Ritonavir + Emtricitabine/Tenofovir and Abacavir/Lamivudine	4,504	12,053	140	16,697
Lopinavir/Ritonavir + Emtricitabine/Tenofovir and Abacavir/Lamivudine	5,550	7,939	306	13,795
Total	**98,100**	**55,818**	**881**	**154,799**

Source: Compiled by the authors with data as of June 2019 from information requests made to INAI in 2021.

3.3 Budgetary Analysis and Savings Evaluation

This analysis focuses on examining the financial aspect of acquiring ARV medications for HIV/AIDS treatment in Mexico, using different sources of information for a comprehensive assessment.

We initially analyzed projected and published savings in 2019 by CENSIDA. During the 2019 to 2020 analysis period, we can observe an allocation of resources from the Catastrophic Expenses Fund of the Seguro Popular (SP, People's Insurance). CENSIDA obtained the authorization from the Technical Committee of the Fideicomiso del Sistema de Protección Social en Salud (Trust for the Social Protection System in Health), resulting in an allocation of 3,207,277,135.17 Mexican pesos. Of this amount, 2,820,476,482.87 were allocated for ARV medication procurement, while the remaining 386,800,652.30 were used to finance viral load, CD4, and genotype tests. This allocation highlights a significant component of estimated savings, totaling 1,373,101,058 Mexican pesos. According to this report, the estimated savings at that time would represent an approximately 55% reduction compared to purchases from the previous year, and they originate from the procurement of single-source and generic ARV medications (CENSIDA, 2019a).

Based on the response provided by CENSIDA, derived from information access request No. 0001200157620 submitted through the National Transparency Platform, the evaluation focuses on the purchasing periods spanning from 2015 to 2020, with a special emphasis on 2019 and 2020, as shown in Table 4 (see Table 4).

Table 4. Expended Budget on HIV/AIDS Care with Resources from the Fideicomiso del Sistema de Protección Social en Salud (Trust for the Social Protection System in Health), in Mexican pesos

Fiscal Year	Period[6]	ARV Medication	Laboratory Tests (viral load, CD4, and genotype)	Total
2015	2015-2016	2,525,542,291.89	348,666,843.00	2,874,209,134.89
2016	2016-2017	2,826,011,136.95	325,624,483.00	3,151,635,619.95
2017	2017-2018	3,097,064,812.92	329,155,024.44	3,426,219,837.36
2018	2018-2019	3,023,233,644.43	348,368,940.43	3,371,602,584.86
2019	2019-2020	2,471,152,477.44	363,581,350.00	2,834,733,827.44

Source: Compiled by the authors with data from information requests made to INAI in 2020.

In the 2018 fiscal year, spanning from April 1, 2018, to March 31, 2019, considerable resources were allocated to ARV medication procurement, reaching an expended budget of 3,023,233,644.43 Mexican pesos. However, a significant decrease is evident in the purchases corresponding to the 2019 fiscal year, with a reduction in the expended budget of 552,081,166.99, representing an 18.26% contraction compared to the previous year and marking it out as the lowest expended budget since 2015.

These numbers reflect the budget executed with resources from the Fideicomiso del Sistema de Protección Social en Salud (Trust for the Social Protection System in Health), although it is relevant to mention that the amounts do not include credit notes, which are incorporated into the estimated savings prior to the consolidation of ARV purchasing.

Specifically, the allocation of resources in different periods for ARV pharmaceutical purchasing and laboratory tests can be observed. Data illustrate an evolution in the expended budget over the years studied, with numbers ranging from 2,525,542,291.89 to 3,097,064,812.92 Mexican pesos for ARV medications and between 325,624,483.00 and 363,581,350.00 for laboratory tests, between 2015 and 2020.

The financial analysis provides a comprehensive insight into resource allocation and changes in the budget allocated for pharmaceutical purchasing.

[6] The periods cover from April 1, 2015, to March 31, 2020.

3.4 Medication Price Analysis

This study embarked on a comprehensive analysis of the numbers disclosed by CENSIDA regarding procured medications, focusing on the effects following the implementation of a renewed therapeutic approach, derived from a collaboration between SS and SHCP. We identified the medications that experienced the most significant reductions in unit prices in this context.

Table 5 shows the medications that reflected notable changes in their unit prices after the adoption of the new therapeutic approach (see Table 5).

It is important to note that this representation is a selection based on the numbers published by CENSIDA up to December 31, 2019. It is worth mentioning that data present in CompraNet, the official platform for government procurement registration, does not provide a standardized breakdown of the details of each medication procured.

This analysis is based on all medications reported by CENSIDA. Table 6 summarizes the data used, considering breakdowns by state, institute, and hospital, throughout each month from January 2017 to July 2020 (see Table 6). The results provide a holistic view of the evolution of ARV medication prices and procurement in Mexico (CENSIDA, 2021b).

In conclusion, the decrease in the budget allocated for purchasing ARV medications in 2019 compared to 2018 was largely offset by the estimated savings in their procurement, allowing for greater efficiency in the utilization of resources allocated for this purpose. Additionally, the implementation of the new therapeutic approach contributed to a reduction in the unit price of some ARV medications (Fernando & Pere, 2011; European Medicines Agency, 2018; NIH, n.d.).

Table 5. Price Comparison of Medications Procured in 2018 vs. Medications Procured with Consolidated Purchasing in 2019 (Mexican pesos)

ARV	Monthly Average Quantity Procured	2018 Unit Price (as of March 2019)	2018 Monthly Average	2019 Unit Price after Consolidated Purchasing	2019 Monthly Average	Decrease %
ABACAVIR, 300 mg: 60 tablets per package	1,477	MX$543	MX$801,849		MX$664,650	21%
ABACAVIR/LAMIVUDINE, 600/300 mg: 30 tablets per package	10,800	MX$990	MX$10,692,000	MX$282	MX$3,045,600	251%
DARUNAVIR, 400 mg: 60 tablets per package	3,500	MX$3,286	MX$11,500,545	MX$1,600	MX$5,600,000	105%
DARUNAVIR, 600 mg: 60 tablets per package	1,300	MX$4,481	MX$5,824,949	MX$1,834	MX$2,384,200	144%
EFAVIRENZ 600 mg, EMTRICITABINE 200 mg, TENOFOVIR disoproxil succinate 300.6 mg: 30 tablets per package	100,000	MX$1,300	MX$130,000,000	MX$800	MX$80,000,000	63%
EFVIRENZ, 600 mg: 3 0 tablets per package	11,900	MX$162	MX$1,927,800	MX$99	MX$1,178,100	64%
EMTRICITABINE-TENOFOVIR, 200/245 mg: 30 tablets per package	26,000	MX$2,061	MX$53,582,880	MX$710	MX$18,460,000	190%
LOPINAVIR/RITONAVIR, 200/50 mg: 120 tablets per package	900	MX$1,730	MX$1,557,000	MX$988	MX$888,750	75%
TENOFOVIR DISOPROXIL FUMARATE, 300 mg: 30 tablets per package	2,050	MX$2,000	MX$4,100,246	MX$1,040	MX$2,132,246	92%

Source: Compiled by the authors with data from CENSIDA, 2019e.

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.07

Table 6. Summary of Medications Used in the Development of Machine Learning Models

ARV	2016	2017	2018	2019	2020
ABACAVIR, 2 g Bottle 240 ml	578	578	578	578	577
ABACAVIR, 300 mg. 60 tablets	500	475	543	450	309
ABACAVIR/LAMIVUDINE, 600/300 mg. 30 tablets	1379	1379	990	282	399
ATAZANAVIR, 300 mg. 30 capsules	2765	2668	2641		
BICTEGRAVIR/EMTRICITABINE/TENOFOVIR ALAFENAMIDE, 50/200/25 mg. Box with 30 tablets				1720	1720
DARUNAVIR 150 mg. 240 tablets	4979	4979	4979	4979	4979
DARUNAVIR 400 mg. 60 tablets	3319	3286	3286	1600	1521
DARUNAVIR 600 mg. 60 tablets	4979	4979	4481	1834	2049
DARUNAVIR 75 mg. 480 tablets		4979	4979	4979	
DARUNAVIR/COBICISTAT, 800/150 mg. 30 tablets			2960	2915	2915
DIDANOSINE, 250 mg. 30 capsules	657	657			
DIDANOSINE, 400 mg. 30 capsules	1057	1057			
DOLUTEGRAVIR, 50 mg. 30 tablets	4077	4077	3335	3269	3000
DOLUTEGRAVIR/ABACAVIR/LAMIVUDINE, 50/600/300 mg. 30 tablets			4665	4135	3000
EFAVIRENZ, 600 mg. 30 tablets	259	221	162	99	85
EFAVIRENZ/EMTRICITABINE/TENOFOVIR DISOPROXIL, 600/200/245 30 tablets	2404	2332	2332	805	800
ELVITEGRAVIR/COBICISTAT/EMTRICITABIN A/ TENOFOVIR, 150/150/200/245 mg. 30 tablets			2000	2000	
EMTRICITABINE/TENOFOVIR ALAFENAMIDA, 200/10 gr. 30 tablets				1720	1720
EMTRICITABINE-TENOFOVIR, 200/245 mg. 30 tablets			2134	710	710
EMTRICITABINE, 200 mg. 30 capsules	490	343	601		
EMTRICITABINE/TENOFOVIR ALAFENAMIDA, 200/25 mg. 30 tablets					1720
EMTRICITABINE/TENOFOVIR DISOPROXIL FUMARATE, 200/245 mg. 30 coated tablets	2125	2061			

ARV	2016	2017	2018	2019	2020
ENFUVIRTIDE, INJECTABLE SOLUTION. 108 mg. 60 3 ml syringes, 60 1 ml syringes, and 180 alcohol wipes.	22450	22450	22608	20330	20330
ETRAVIRINE, 100 mg. 120 tablets	5429	5429	5429		
ETRAVIRINE, 200 mg. 60 tablets			5429	5429	5429
GLECAPREVIR/PIBRENTASVIR, 100/40 mg. 4 boxes, with 7 strips of 3 tablets each					68250
LAMIVUDINE, 150 mg. 60 tablets	584	394	394	365	335
LAMIVUDINE. 1 g per 100 ml. 240 ml and pipette	835	835	894	891	890
LAMIVUDINE/ZIDOVUDINE, 150/300 mg. 60 tablets	772	772	595	425	234
LOPIANVIR/RITONAVIR, 100/25 mg. 60 tablets	1100	1100	1100	1100	1100
LOPIANVIR/RITONAVIR, 200/50 mg. 120 tablets	2656	2063	988	1730	2010
LOPIANVIR/RITONAVIR, 8.0/2.0 g per 100 ml. Amber glass jar with 160 ml and dosing cup	1714	1714	1714	1714	1714
MARAVIROC 150 mg. 60 tablets	6622	6622	6622	6622	6612
MARAVIROC 300 mg. 60 tablets	6622	6622	6622	6622	6612
NEVIRAPINE, 1.0 g per 100 ml. 240 ml and dosing cup	333	328	313	313	313
NEVIRAPINE, 200 mg. 60 or 100 tablets	384	380	377		
RALTEGRAVIR, 400 mg. 60 tablets	5309	4247	3610	3395	3574
RITONAVIR, 100 mg. 30 tablets	348	348	348	348	348
SOFOSBUVIR-VELPATASVIR, 400/100 mg. 28 tablets.					49896
TENOFOVIR, 300 mg or 245 mg. 30 tablets	2000	2000	2000	1040	510
TIPRANAVIR, 250 mg. 120 capsules	3229	3229			
ZIDOVUDINE INJECTABLE SOLUTION, 200 mg. 5 vials (200 mg/20 ml)				995	993
ZIDOVUDINE ORAL SOLUTION, 1 g per 100 ml. 240 ml	515	500	680	485	460
LOPIANVIR/RITONAVIR, 200/50 mg. 60 tablets		1100		1100	
ETRAVIRINE, 200 mg. 120 tablets		5429	5429		

Source: Compiled by the authors with data from CENSIDA (2019e).

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.07

3.5 Analysis Using Machine Learning Algorithms

In this phase of the study, we proceeded with the creation and evaluation of machine learning models using logistic regression, random forest, and KNN techniques. The analysis procedure consisted of the stages shown in Figure 2 (see Figure 2).

Figure 2. Diagram of the Machine Learning Algorithm Analysis

Source: Compiled by the authors.

a) *Data collection and preparation:* We obtained data from various sources, including the National Transparency Platform and CENSIDA. Subsequently, we carried out data cleaning and transformation with the aim of using it in the analysis.

b) *Exploratory data analysis:* We employed various visualization techniques and descriptive statistics to better understand the data, as well as to detect patterns and relationships between variables.

c) *Variable selection and application of Machine Learning Models:* We conducted the identification of the most pertinent variables for the analysis, followed by the implementation of diverse machine learning models, such as logistic regression, random forest, and KNN. We employed these models to forecast potential decreases in ARV medication prices.

d) *Training and validation of each model:* We divided the data into training and validation sets and adjusted model parameters to achieve the best performance. We then evaluated the trained models using metrics such as accuracy, precision, and recall assessing their performance.

e) *Model comparison:* The performance of different algorithms is compared to determine which one provides the best results for the given dataset, predicting whether there was a variation in ARV medication price.

In this section, we elaborated in more detail on the key components of our data analysis and machine learning modeling. We addressed the independent and dependent variables used in the analysis of ARV medication procurement, as well as the machine learning methods applied, based on logistic regression, random forest, and KNN techniques.

3.5.1 Independent Variables (Medication Characteristics)

- ARV Type: We used this variable to categorize medications in terms of their commercial and market identification, allowing for the consideration of differences between generic, branded, or single-source medications.
- Short description: Brief description of the 45 different medications analyzed.
- Detailed description: Detailed description of the 45 different medications analyzed.
- Unit price: Amount in Mexican pesos paid per unit procured corresponding to the purchase year.
- Price from previous period: Amount in Mexican pesos paid per unit procured corresponding to the immediate previous purchase year.
- Quantity procured: Number of units procured corresponding to the purchase year.
- Total amount: The total amount in Mexican pesos paid for all units procured of this medication corresponding to the purchase year.
- Active components: This variable includes the chemical composition of the medication and its potential therapeutic effect, enhancing our understanding of the analyzed ARVs.
- Brief presentation: An indicator that includes a summarized presentation of the medication: capsule, tablet, solution, suspension, pill.
- Detailed presentation: An indicator that includes an extensive presentation of the medication: capsule, tablet, solution, suspension, pill.

- Dosage: A quantitative variable that plays a vital role in quantifying the amount of active ingredients in a treatment unit, including the unit of measurement.
- Content per package: Linked to treatment duration, this variable provides insights into medication availability and continuity.
- Supplier/Manufacturer: A factor that adds an important nuance by identifying the supply and production source of ARV medications, whether it be the name of the individual or legal entity from which the medication was purchased.
- Client: Government department or entity in Mexico that receives the purchased medications.
- Procurement Year: An indicator that contextualizes transactions over time, allowing for analysis of temporal patterns in medication purchasing.

We analyzed and selected relevant columns for the study exploratory data analysis (EDA), correlations, and feature importance with random forest.

3.5.2 Dependent Variable: Price Change

The dependent variable "price change" captures variations in ARV medication prices. This binary variable, encoded as "1" for price increase and "0" for decrease, allows us to analyze trends in procurement prices over time.

3.5.3 Data Analysis Methods and Machine Learning Modeling

3.5.3.1 Logistic regression

In this phase of the study, logistic regression was employed as one of the primary machine learning techniques for analyzing medication prices. Logistic regression remains a cornerstone in predictive modeling, particularly in domains where the outcome of interest is binary or categorical. Its flexibility, interpretability, and ability to handle complex relationships make it indispensable in the analysis of healthcare data, including pharmaceutical pricing dynamics Hastie, Tibshirani, and Friedman (2021a) and Hastie, Tibshirani, James, and Witten (2021b).

According to the logistic regression approach described by Hastie et al. (2021a) in their book *The Elements of Statistical Learning*, the procedure for analyzing medication prices using logistic regression is as follows:

- *Data preprocessing*. Missing values should be removed, and categorical variables encoded. Additionally, the dataset should be split into training and testing sets. We created a database containing the prices and characteristics of ARV medications procured by the Government of Mexico from 2015 to 2020. We selected relevant predictor variables that may affect medication prices, such as: medication type, short description, quantity procured, total amount, active components, summarized presentation, dosage, content per container, supplier/manufacturer, client, and procurement year. Finally, we partitioned the dataset into training and testing subsets, with an 80% training and 20% testing ratio.
- *Model training*. The logistic regression model is the training set. The goal is to find the optimal coefficient values that maximize the probability of the model correctly predicting each observation's class. The selected predictive variables are medication type, short description, quantity procured, total amount, active components, summarized presentation, dosage, content per container, supplier/manufacturer, client, and procurement year.
- *Model evaluation*. The model's accuracy is evaluated on the test data using measures such as the error rate and confusion matrix. Logistic regression modeled the probability of price change based on the independent variables.

Through the logistic function, we examine the relationships between these variables and their impact on the dependent variable.

We formulated a logistic regression model in Python with the aim of anticipating whether medication prices exhibit an increase or a decrease. To accomplish this, a function is employed to model $p(X)$, producing values that are exclusively binary, zero or one for all X values. In logistic regression, the logistic function is introduced within this context.

The logistic regression formula is expressed as:

$$log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Where $p(X)$ is the probability that $Y = 1$ given $X = x$, β_0 is the intercept term, and $\beta_1, \ldots, \beta_p$ are the coefficients of the predictor variables $X_1, \ldots, X_p$, also known as the regression coefficients.

The logistic regression formula is:

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}$$

This function converts the weighted sum of the input variables and coefficients into a probability value, which falls within the range of 0 and 1. If this probability surpasses the threshold of 0.5, the model indicates a prediction for the positive category; otherwise, it predicts the negative category.

In the analysis of medication prices, logistic regression offers a robust framework for modeling the likelihood of price changes based on various predictor variables. Its application in forecasting price fluctuations in pharmaceutical markets has yielded valuable insights for policymakers and healthcare administrators (Hosmer Jr et al., 2013).

With the increasing availability of healthcare data, logistic regression serves as a powerful tool for analyzing trends in medication pricing and procurement. Its ability to handle large datasets and incorporate diverse predictor variables makes it indispensable for understanding the drivers of price variations in the pharmaceutical sector (Agresti, 2015).

3.5.3.2 Random forest

The implementation of random forest allowed for a more sophisticated and accurate analysis. This technique combines multiple decision trees to identify subtle patterns and nonlinear relationships in the data.

Aligned with Hastie et al.'s (2021b) description of the random forest methodology, this approach involves the creation of decision trees using a set of training data. However, a distinct feature of this technique is that during the development of each tree, a random subset of predictors is chosen from the complete set of available predictors when evaluating a potential split. Among these selected predictors, only one is used for the actual split. Moreover, with each split, a new subset of predictors is randomly sampled, typically with the number of predictors being close to the square root of the total number of predictors ($m \approx \sqrt{p}$).

Random forest is recognized for its versatility and capability to handle complex datasets in various domains (Fernández-Delgado et al., 2014). In healthcare analytics, it has been extensively utilized for predicting clinical outcomes, disease diagnoses, and medication responses.

Opting for a smaller *m* value when constructing random forest proves advantageous, particularly in scenarios where there is substantial correlation among numerous predictors. This strategy aids in predicting the decrease in the price of ARV medications by effectively handling the complexities and interrelationships among various factors involved.

According to Marsland (2014), the random forest algorithm leverages ensemble learning by aggregating multiple decision trees, each trained on a bootstrap sample and a subset of features, thereby mitigating overfitting and enhancing generalization performance.

Moreover, in financial domains, random forest has proven effective for tasks such as credit risk assessment, stock market prediction, and fraud detection (Liaw & Wiener, 2002). Its ability to handle large volumes of data, nonlinear relationships, and high-dimensional feature spaces makes it a valuable tool for analyzing pharmaceutical pricing dynamics and forecasting changes in medication costs.

3.5.3.3 K-Nearest Neighbors Method

The K-nearest neighbors (KNN) method, as described by Hastie et al. (2021b), plays a crucial role in our predictive approach. This method relies on the idea that a new data point's category can be predicted by examining the categories of nearby data points. It assumes that data points with similar characteristics tend to belong to the same category, providing useful insights for classification and prediction tasks and improving our understanding of price changes.

Implementation of the KNN technique entails the following procedural steps:

- *Selection of K* value: A suitable value for *K*, denoting the number of nearest neighbors considered during prediction, is determined.
- *Distance computation*: Utilizing a Euclidean metric, the distance between the data point to be predicted and all other points in the training set is computed.
- *Neighbor identification*: The *K* nearest points to the data point under prediction are identified based on the calculated distances.
- *Category prediction*: Among these *K* neighbors, the category that appears most frequently is designated as the predicted category for the data point in question.

By following these steps, the KNN method facilitates informed predictions regarding price changes, thereby enhancing decision-making processes within our predictive framework.

KNN demonstrates its effectiveness particularly in situations where there is no clear functional relationship between the features and categories, as well as when the data exhibits nonlinear structures. However, it is important to consider that KNN may present significant computational consumption, especially in extensive datasets, due to the distance calculation required for all points in the training set.

In the context of this research, the KNN method was employed in conjunction with the two predictive approaches mentioned earlier: logistic regression and random forest, in order to explore its predictive potential and complement the comprehensive analysis of ARV medication prices, following the methodology proposed by Hastie, Tibshirani, and Friedman (2021a), and Hastie, Tibshirani, James, and Witten (2021b) and Badal and Sungkur (2023).

We separated data into training and testing sets with 80% of the data for training and 20% for testing. Prior to this separation, the data underwent preprocessing that normalized and standardized the variables, ensuring the consistency and accuracy of the results.

Given the significance of KNN in pharmaceutical price prediction, it is noteworthy to mention the seminal work by Cover and Hart (1967) which underscore KNN's versatility, efficacy, and simplicity in pharmaceutical price forecasting by leveraging similarities among medication attributes and purchase patterns.

4. Scope and Limitations

The scope of this study is based on the evaluation of consolidated purchasing of ARV medications in Mexico, carried out by the Federal Government through the SS. Specifically, the analysis focuses on the procurements made by CENSIDA with the purpose of meeting the therapeutic demands of HIV patients during the 2019 fiscal year. This interval spans from April 1, 2019, to March 31, 2020, covering the transactions aimed at addressing the medical needs of 98,100 patients undergoing ARV treatment. This patient cohort constitutes approximately 63.37% of the total population of individuals identified in this category. It is important to highlight that this analysis is exclusively limited to procurements made by the Federal Government, excluding any information regarding purchases made by private entities, which represent only 0.8% of the total population receiving ARV treatment in Mexico.

To better understand patterns in ARV medication procurement by the Mexican government in 2019, it has been decided to adopt an initial approach focused on the most representative medication within our dataset. This decision is based on the need to delve into the analysis of a specific medication to identify and understand the factors influencing its price and variability. The choice of this approach is justified by the wide gap observed between the minimum and maximum prices of the various medications procured. By focusing on a specific medication in this initial study, we aim to establish a solid foundation for understanding the determinants of price and procurement patterns before expanding our analysis to the entirety of ARV medications.

It is essential to recognize that this study faces certain intrinsic limitations due to the nature and source of the data used for the analysis:

- *Inconsistencies in information:* Discrepancies have been identified in the total expenditure amounts present in various official sources such as information access requests addressed to the INAI Institute, compared to the numbers published on the official CENSIDA website. Despite this inconsistency, the analysis was based on the data provided directly by the official CENSIDA portal.
- *Heterogeneous data source:* Since we relied upon data obtained from multiple official government sources and international organizations, there is a possibility of introducing a certain degree of bias due to inherent variations in the sources of information.
- *Absence of data in international statistics:* The lack of specific data for Mexico in statistics published by international organizations, such as the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO), led to the need to complement these statistics with information provided by Mexican government agencies responsible for the official data generation. Obtaining this information often involved submitting requests to the INAI.

Despite these limitations, however, we expect this study to provide a comprehensive and objective analysis of the consolidated purchasing of ARV medications in Mexico during the 2019 fiscal year, providing valuable insights for the evaluation of the effectiveness and efficiency of this procurement system in the realm of HIV patient treatment.

The Anáhuac Journal, **Vol. 24, núm. 1, 2024.**
https://doi.org/10.36105/theanahuacjour.2024v24n1.07

5. Results

This study aimed to evaluate the consolidated purchasing of ARV medications in Mexico in 2019. It primarily employed machine learning techniques to predict changes in drug prices and analyze their economic impact. Below, we present the results obtained from the analyses conducted.

In order to better understand ARV medication procurement patterns, we applied exploratory data analysis techniques. Table 7 shows a statistical summary of the numerical variables of the medications considered in this study (see Table 7), while Table 8 displays the statistics of the categorical variables (see Table 8).

The analysis reveals a variety in the characteristics and prices of the procured medications, suggesting a wide range of therapies and treatment approaches.

Table 7. Summary of Numerical Variables

Characteristic	Count	Mean	Std	Min	25%	50%	75%	Max	Unique Ratio
Year	1653	2019.40	0.49	2019	2019	2019	2020	2020	0.00
Purchase Year	1653	2019	0.00	2019	2019	2019	2019	2019	0.00
Month Number	1653	5.06	2.10	1	3	6	6.00	12.00	0.01
Unit Price	1653	2099.11	2621.30	99	578	1714	2915	20330	0.02
Quantity Procured	1653	1107.89	5992.21	1	18	75	467	104000	0.40
Total Amount	1653	1621455.51	8032435.54	99	20568	98462.83	613440	145750000	0.81
Previous Period Price	1365	2656.09	3055.69	162	680	2000.12	3610.07	22608.35	0.02

Source: Compiled by the authors.

Table 8. Summary of Categorical Variables

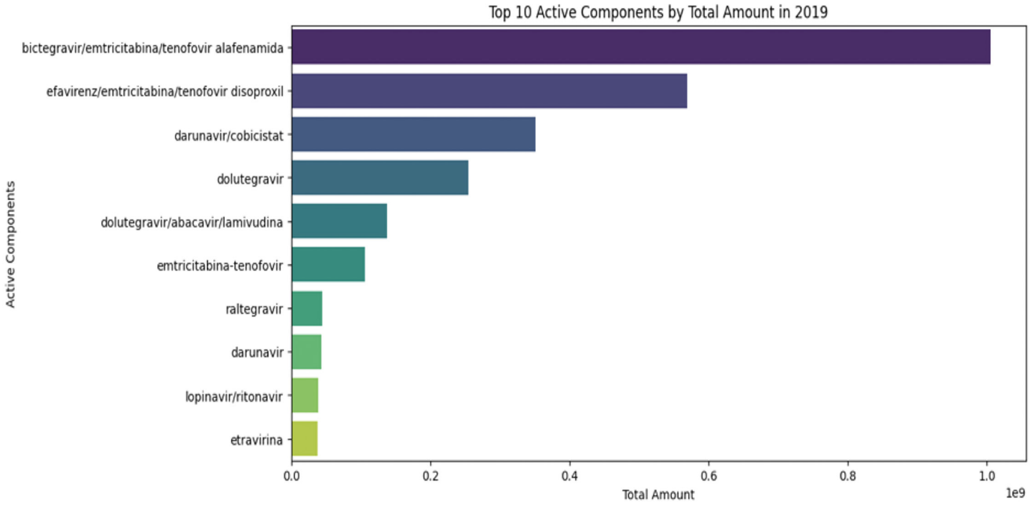
Characteristic	Count	Unique	Top	Freq.	Unique Ratio
Supplier	1653	15	Glaxosmithkline México, S.A. de C.V.	283	0.01
Client	1580	41	Mexico City	50	0.02
Presentation Brief	1653	4	Tablet	1301	0.00
Presentation Detailed	1653	9	Tablet	1213	0.01
Container Content	1653	13	30 tablets	589	0.01
Dosage	1653	25	50/200/25 mg	158	0.02
Active Components	1653	23	Lopinavir/Ritonavir	161	0.01
Short Description	1653	33	Bictegravir/Emtricitabina/Tenofovir Alafenamida	158	0.02
ARV Type	1642	2	Single source	890	0.00

Source: Compiled by the authors.

5.1 Results of the Medication Price Analysis

The comparative analysis of ARV medication prices revealed a downward trend during the period of consolidated purchasing. On average, we observed a 58.6% reduction in the prices of key medications, i.e. medications representing 80% of the budget, as shown in Figure 3. This significant decrease in prices confirms the effectiveness of the consolidated purchasing strategy in achieving substantial economic savings (see Figure 3).

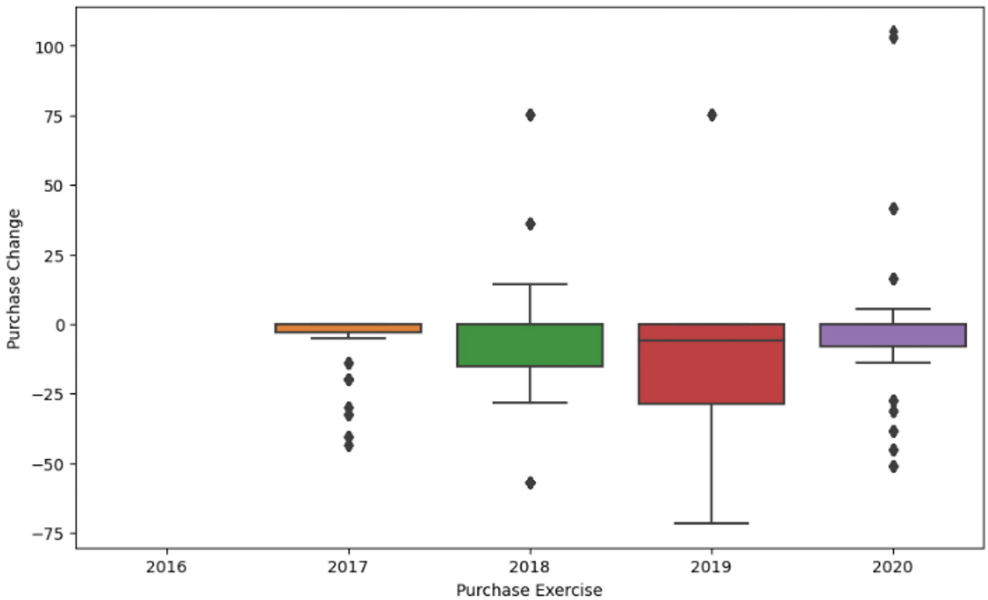
Figure 3. Top 10 Active Components by Total Amount in 2019



Source: Compiled by the authors.

Globally, considering all medications procured, the decrease in ARV medication prices for the 2019 procurement period is 26.1%, as shown in Figure 4 (see Figure 4).

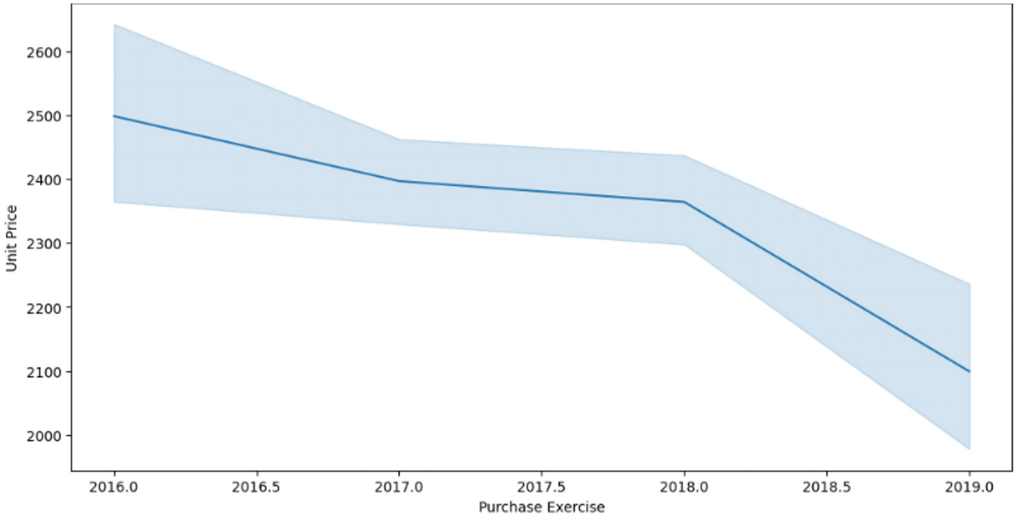
Figure 4. Percentage Change in Prices by Purchase Exercise



Source: Compiled by the authors.

According to numbers provided by the Mexican government, total economic savings during the analyzed period ranged from 1,070 million Mexican pesos to 1,559 million in total, considering laboratory tests, credit notes, and changes in ARV regimens, depending on the consulted source and specific criteria considered. This represents an average reduction of 55% of the budget allocated to CENSIDA for ARV medication procurement, amounting to 2,820,476,482.87 Mexican pesos. A subsequent study will analyze the database to contrast the results with these numbers; however, the detailed analysis of medications' unit prices during the 2019 procurement period reveals a clear downward trend, highlighting a marked decrease compared to previous periods, as shown in Figure 5. This result is evident in the visualization of temporal trends, where unit prices show a descending trajectory throughout different months of the year. This reduction in unit prices translates into a positive impact on total procurement costs, generating substantial savings for ARV medication procurement during the analyzed period (see Figure 5).

Figure 5. Temporal Price Trends



Source: Compiled by the authors.

5.2 Budget Analysis Results and Savings Evaluation

Information provided by CENSIDA highlights the allocation of resources from the Catastrophic Expenses Fund of the People's Insurance for the period from 2019 to 2020. Out of the 3,207,277,135.17 Mexican pesos in the authorized budget, 2,820,476,482.87 was allocated to ARV medication procurement. This report emphasized a crucial component of estimated savings, amounting to 1,373,101,058 Mexican pesos, representing approximately a 55% reduction compared to purchases from the previous year. These savings were primarily derived from the procurement of single-source and generic ARV medications (CENSIDA, 2019a; CENSIDA, 2019c).

The analysis also encompasses the evolution of the expended budget over the years analyzed, ranging from 2,525,542,291.89 Mexican pesos to 3,097,064,812.92 for ARV medications and from 325,624,483.00 to 363,581,350.00 Mexican pesos for laboratory tests, from 2015 to 2020. It is relevant to mention that the reported amounts do not include credit notes, which are incorporated into the estimated savings prior to the consolidation of ARV purchasing.

The budget analysis exhibited a downward trend in resource allocation for ARV medication procurement in Mexico. In the 2018 fiscal year, an expended budget of 3,023,233,644.43 Mexican pesos was reached. However, we observed a significant decrease in purchases corresponding to the 2019 fiscal year, with a reduction in the expended budget by 552,081,166.99, representing a 18.26% contraction compared to the previous year. This decrease marks the lowest budget since 2015.

It is important to highlight that the results of the analysis of our database align with the savings trend. This is due to the evident reduction in the unit price of several medications, which coincides with the overall cost decrease. However, it is crucial to highlight that this study did not assess the overall savings through the calculation of treatment costs for patients treated throughout the 2019 period. We will conduct this detailed analysis in a subsequent study.

5.3 Machine Learning Modeling and Price Change Prediction

For the three models generated in this analysis, we used a database composed of 15,220 observations of ARV medications procured between 2016 and 2019. The focus was on 2019 procurements, limited to medications with the active components Efavirenz, Emtricitabine, and Tenofovir. Each observation contained detailed information about various relevant attributes, such as medication type, quantity procured, total amount, active components, supplier/manufacturer, client, and procurement year. We selected attributes that could influence price changes.

5.3.1 Analysis with Logistic Regression

During training, we adjusted the model coefficients using the gradient descent algorithm. This involved iteratively updating the coefficients to minimize the logarithmic loss function, which measures the difference between the predicted probabilities and the actual outcomes. Feature selection techniques were employed to identify the most influential variables affecting price changes in ARV medications. Additionally, cross-validation was performed to assess the model's generalization ability and to mitigate overfitting.

We thoroughly evaluated the logistic regression model using data collected on ARV medications and their price changes. In addition to traditional evaluation metrics like precision, sensitivity, and specificity, supplementary techniques such as cross-validation and bootstrapping were utilized to validate model robustness. Calibration plots were generated to scrutinize the calibration of predicted probabilities across diverse thresholds.

We generated a confusion matrix, shown in Table 9, in order to analyze the correct and incorrect classification of the model's predictions. The confusion matrix offers a comprehensive overview of the outcomes, encompassing true positives, false positives, true negatives and false negatives (see Table 9).

Table 9. Confusion Matrix for the Logistic Regression Model

	No Decrease Prediction	**Decrease Prediction**
No Decrease	56	8
Decrease	6	64

Source: Compiled by the authors.

We calculated precision, sensitivity, and specificity metrics to assess the overall performance of the model in classification. We obtained the following metrics:

- Precision: 0.8750
- Sensitivity: 0.9143
- Specificity: 0.8750

Although the logistic regression model showed reasonable ability to predict price changes in ARV medications, it revealed certain limitations in its predictive capacity in some instances. Despite an overall accuracy of 89.29%, we identified areas for improvement in sensitivity and specificity.

To enhance the performance of the logistic regression model, we will consider incorporating additional data to enrich the model.

5.3.2 Analysis with Random Forest

Throughout the training phase, we made adjustments to fine-tune the model's settings using the Grid Search method. Parameters such as the number of trees in the forest, maximum tree depth, and minimum number of samples required to split a node were fine-tuned to enhance predictive accuracy. Feature importance analysis was conducted to identify significant predictors of ARV medication price changes.

We thoroughly evaluated the random forest model using standard evaluation metrics to understand its performance and predictive capability. In addition to traditional evaluation metrics, such as precision, sensitivity, and specificity, techniques such as out-of-bag error estimation and permutation feature importance were employed to evaluate the model's robustness and generalization ability. Moreover, visualization techniques, such as partial dependence plots and feature interaction analysis, were utilized to gain insights into the complex relationships between predictors and price changes.

The confusion matrix offers a comprehensive breakdown of true positives, correct positives, correct negatives, false positives, and false negatives, providing valuable insights into the model's performance, as shown in Table 10 (see Table 10).

Table 10. Confusion Matrix for the Random Forest Model

	No Decrease Prediction	**Decrease Prediction**
No Decrease	63	1
Decrease	3	67

Source: Compiled by the authors.

We assessed the random forest model's discriminative ability between positive and negative instances using the ROC curve. We quantified the quality of the model by calculating the AUC.

The precision of the random forest model, 98.53%, suggests that the model has a high level of confidence in identifying instances where prices are likely to increase. With a sensitivity that identifies 95.74% of actual price increases in ARV medications, this implies that the model effectively captures the majority of instances where prices are indeed observed to increase. The specificity, 98.53% of the time, the model accurately predicts instances where prices remain stable or decrease.

The overall accuracy of the random forest model indicates that the model's predictions, both positive and negative, are correct 97.32% of the time. It demonstrates the model's effectiveness in providing accurate forecasts of price changes in ARV medications.

To address any potential overfitting of the model, we will consider further optimization of hyperparameters and inclusion of more training data to improve the model's generalization. Additionally, feature selection techniques may be explored to identify the most relevant attributes contributing to the model's predictions.

5.3.3 K-Nearest Neighbors (KNN) Method

The KNN method, leveraging similarity between nearby data points, was employed to predict price changes of ARV medications based on drug characteristics. Prior to model training, dataset preprocessing was performed to ensure compatibility with the KNN algorithm. Techniques such as partitioning, normalization, and standardization of independent variables were employed.

For analysis purposes, we partitioned the dataset into training sets comprising 80% of the data and test sets comprising the remaining 20%.

For each data point in the test set, we calculated the Euclidean distance between that point and all points in the training set. The Euclidean distance is calculated using the formula:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Where *n* is the number of features, x_i and y_i are the values of feature *i* in the test point and training point respectively.

In our analysis, we chose a *K* value which signifies the number of nearest neighbors used to predict the label of a data point within the test set. We set the *K* value to 5. We determined the most common label among the *K* nearest neighbors and assigned that label to the test point. If the majority of neighbors experienced a price increase, the test point was labeled "1" (price increase), and if the majority experienced a decrease, it was labeled "0" (price decrease).

Once we assigned labels to all test points, we constructed the confusion matrix to evaluate the KNN model's performance on the test set. The confusion matrix compares the model's predictions against the actual outcomes, providing insight into the accuracy of the classifications.

The KNN model demonstrated a precision of 97.67%, a sensitivity of 94.33%, and a specificity of 97.67% in predicting price changes in ARV medications. The confusion matrix associated with these results is shown in Table 11 (see Table 11).

Table 11. Confusion Matrix Associated with KNN Analysis Results

	No Decrease Prediction	Decrease Prediction
No Decrease	58	2
Decrease	4	70

Source: Compiled by the authors.

Below is a comparative analysis of the logistic regression, random forest, and K-nearest neighbors (KNN) models in terms of precision, sensitivity, and specificity for predicting price changes in ARV.

Although the KNN model provided reasonable capability to predict price changes in ARV medications, it also revealed certain limitations in its predictive capacity in some instances. The precision results for each method are detailed in Table 12 (see Table 12).

Table 12. Precision Results for the Three Methods Used

Method	Precision	Sensitivity	Specificity	Accuracy
Logistic regression	87.50%	91.43%	87.50%	89.29%
Random forest	98.53%	95.74%	98.53%	97.32%
K-nearest neighbors	97.67%	94.33%	97.67%	96.30%

Source: Compiled by the authors.

The results show that the random forest demonstrated 98.53% precision in predicting price changes, followed by KNN and logistic regression. This suggests that the random forest is the most suitable method for analyzing and predicting price fluctuations in ARV medications.

Additionally, we observed that the models yielded consistent results with conventional data analysis. We identified influential factors in procurement prices, such as medication type, quantity procured, and duration of consolidated purchasing strategy. We detected complex interactions among these variables, highlighting the importance of considering multiple dimensions when analyzing ARV drug prices.

Specifically, the logistic regression model indicated that the duration of consolidated purchasing strategy had a significant effect on reducing procurement prices. On the other hand, the random forest model underscored the importance of medication type in price determination. The KNN model identified clustering patterns among different ARV medications, suggesting the existence of complex relationships between them in terms of prices and characteristics.

6. Discussion

The implementation of a new consolidated purchasing scheme by the Mexican government for ARV medication procurement has shown significant results. Data reveal that this approach succeeded in generating substantial savings in the cost of these medications, contributing to an optimization in public health investment. These results are consistent with government reports reporting savings of 1.7 billion pesos by implementing the new procurement scheme (Ramírez, 2019).

A key factor in these savings relates to the inclusion of the drug Bictegravir, which has demonstrated greater efficacy and a broader resistance profile for viruses, as well as fewer adverse effects (DHHS, 2021). This inclusion allowed the Mexican government to achieve a more efficient and simplified ARV regimen for over 36,000 patients, resulting in increased coverage and optimization of ARV treatment in the country. This is supported by statements from those involved in the procurement process, who highlight the improvement in care and the ability to effectively meet demand.

Substantial savings in purchases, representing nearly a 60% decrease compared to previous prices, enabled the Mexican government to carry out an update to the basic care package and make state-of-the-art medications available at reduced costs. Furthermore, the implementation of the consolidated purchasing scheme has set a precedent that aligns with international best practices and positions Mexico as a leader in medication provision globally.

When comparing the results obtained in this research with government reports, we can observe that savings in the purchasing of ARV medications were consistent with price decreases recorded during the period when the new procurement scheme was implemented. However, it is important to highlight that the cause of these savings has not yet been precisely determined. While the integration of the drug Bictegravir was a key factor, further research is needed to quantify its coverage and evaluate its impact compared to medications procured in periods prior to 2019.

We will make an attempt to include comparisons with previous studies related to savings in ARV medications and procurement policies.

7. Conclusions

It should be noted that, despite the positive results, this study faced limitations related to the nature of the data and information sources. Discrepancies in expenditure numbers and the heterogeneity of data sources may have introduced some degree of bias into the results.

Given the complexity and variability observed in the prices of ARV drugs procured by the Mexican government, the decision has been made to postpone the construction of predictive models for other medications in this initial paper. This decision is based on the need to focus our efforts on understanding in depth the factors affecting the most representative medication before generalizing our findings to the rest of the medications. Additionally, this approach will allow us to validate our models and methodologies in a more controlled context before applying them to a wider range of drugs. Once we have established a solid framework and identified key patterns in the most representative medication, we can expand our analysis to other ARV medications, allowing us to obtain a more comprehensive understanding of procurement patterns and price determinants in the context of pharmaceutical purchasing by the Mexican government.

In summary, this research provides a deep understanding of consolidated purchasing of ARV medications in Mexico. Machine learning models proved valuable in predicting price changes, with significant economic savings observed through this procurement approach.

However, it is imperative to acknowledge the constraints when interpreting the outcomes owing to variations in data and information origins. These insights can facilitate well-informed decision-making in ARV medication management and the enhancement of HIV patient care.

References

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. John Wiley & Sons.

Badal, Y.T. & Sungkur, R.K. Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and Information Technologies* 28, 3027–3057 (2023). https://doi.org/10.1007/s10639-022-11299-8

Chaumont, C. G. V., Bautista-Arredondo, S., José Calva, J., Isaac Bahena-González, R., Hitz Sánchez-Juárez, G., en Der, L., González de Araujo-Muriel, A., Magis-Rodríguez, C. & Hernández-Ávila, M.(2015). Antiretroviral purchasing and prescription practices in Mexico: constraints, challenges and opportunities. Salud Pública de M*éxico* 57 (2) 171–82. https://doi.org/10.21149/spm.v57s2.7606

Cover, T. & Hart, P. (1967). Nearest neighbor pattern classification*. IEEE Transactions on Information Theory,* 13(1), 21-27. https://doi.org/10.1109/TIT.1967.1053964

Department of Health and Human Services (DHHS). (2021). *Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV*. https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/adult-adolescent-arv/guidelines-adult-adolescent-arv.pdf

European Medicines Agency. (2018, June 22). *Biktarvy*. *Summary of Product Characteristics*. https://www.ema.europa.eu/en/medicines/human/EPAR/biktarvy

Fernández-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*(1), 3133–3181.

Fernando, L., & Pere, D. (2011). Tratamiento antirretroviral de la infección por el VIH. *Enfermedades Infecciosas y Microbiología Clínica*, *29*, 455–465. https://doi.org/10.1016/j.eimc.2011.02.009

Gobierno de México, Consejo de Salubridad General. (2017). Cuadro básico y catálogo de medicamentos. Issue 2017, 162–245. https://www.csg.gob.mx/descargas/pdf/priorizacion/cuadro-basico/med/catalogo/2017/EDICION_2017_MEDICAMENTOS-FINAL.pdf

Gobierno de México, Secretaría de Salud (SS). (2018). Informe nacional del monitoreo de compromisos y objetivos ampliados para poner fin al SIDA. (Informe GAM). https://www.gob.mx/censida/documentos/informe-nacional-del-monitoreo-de-compromisos-y-objetivos-ampliados-para-poner-fin-al-sida-informe-gam-mexico-2018?%20idiom=es

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017a). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas* *que Viven con VIH,* 3(1), 11-18. https://www.gob.mx/cms/uploads/attachment/file/226743/Boletin_Nal_CENSIDA_AT_IN_ene_mar_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017b). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 3(2), 9-16. https://www.gob.mx/cms/uploads/attachment/file/254117/Boletin_Nal_CENSIDA_AT_IN_abr_jun_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017c). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 3(3), 9-14. https://www.gob.mx/cms/uploads/attachment/file/299954/Boletin_Nal_CENSIDA_AT_IN_jul_sep_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2017d). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 3(4), 11-18. https://www.gob.mx/cms/uploads/attachment/file/332699/Boletin_Nal_CENSIDA_AT_IN_oct_dic_2017.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018a). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(1), 15-22. https://www.gob.mx/cms/uploads/attachment/file/332700/Boletin_Nal_CENSIDA_AT_IN_ene-mzo2018.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018b). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(2), 9-16. https://www.gob.mx/cms/uploads/attachment/file/370474/Boletin_Nal_CENSIDA_AT_IN_abr_jun_2018_1.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018c). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(3), 9-16. https://www.gob.mx/cms/uploads/attachment/file/420856/Boletin_Nal_CENSIDA_AT_IN_jul_sep_2018.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2018d). Personas con VIH en tratamiento antirretroviral en la Secretaría de Salud. *Boletín de Atención Integral de Personas que Viven con VIH,* 4(4), 5-12. https://www.gob.mx/cms/uploads/attachment/file/441126/Boletin_Nal_CENSIDA_AT_IN_oct_dic_2018.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019a, July 19). *Ahorros producto de la estrategia de adquisición de antirretrovirales 2019*. https://www.gob.mx/cms/uploads/attachment/file/478447/ahorros_medicamentos.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019b). Datos Abiertos de la Compra Consolidada de Medicamentos ARV 2019-2020. Gobierno de México.

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019c). *Los nuevos mecanismos de optimización y mejora de los tratamientos antirretrovirales en México*. *5*(3), 4–7.

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019d). *Vigilancia Epidemiológica de casos de VIH/SIDA en México. Registro Nacional de Casos de SIDA Actualización al Cierre de 2019*.

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019e, November 11). *Antirretrovirales*. 2019. http://www.censida_2015.salud.gob.mx/contenidos/censida/antirretrovirales.html

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019f). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(1), 15-22. https://www.gob.mx/cms/uploads/attachment/file/473713/Bolet_n_DAI_1ER_TRIM_ENE_MZO2019_VF1.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019g). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(2), 12-19. https://www.gob.mx/cms/uploads/attachment/file/500787/BOLET_N_DAI_2DO_TRIM_ABRIL_JUNIO2019_VF_2.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019h). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(3), 13-21. https://www.gob.mx/cms/uploads/attachment/file/525270/Bolet_n_DAI_3er_trim._Vol.5_N.3_Jul-Sept_VF2.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2019i). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 5(4), 16-22. https://www.gob.mx/cms/uploads/attachment/file/545616/Bolet_n_DAI_4to_trimestre_VF.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2020a). Personas en tratamiento antirretroviral (TAR). *Boletín de Atención Integral de Personas que Viven con VIH,* 6(1), 20-26. https://www.gob.mx/cms/uploads/attachment/file/563143/Bolet_n_DAI_1er_trim_2020_V090720_f.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2021a, January 20). *Adquisición de Antirretrovirales Diciembre 2018*. 2021. http://www.censida_2015.salud.gob.mx/descargas/transparencia/ARV_diciembre_18.pdf

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2021b, January 20). *Compra de Antirretrovirales CENSIDA 2019*. 2021. http://www.censida_2015.salud.gob.mx/contenidos/censida/antirretrovirales.html

Gobierno de México, Secretaría de Salud, Centro Nacional para el Control y la Prevención del VIH y el sida (Censida). (2021c, March 31). *Boletín de Atención Integral de Personas con VIH/Censida*. 2021. https://www.gob.mx/censida/es/articulos/boletin-de-diagnostico-y-tratamiento-antirretroviral-censida?idiom=es

Gobierno de México, Centro Nacional para la Prevención y el Control del VIH y el sida (CONASIDA). (2018). *Boletín del Grupo de Información Sectorial en VIH, sida e ITS*. SS, Semar, Sedena, IMSS, ISSSTE, Pemex.

Hastie, T., Tibshirani, R. & Friedman, J. (2021a). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hastie, T., Tibshirani, R., James, G. & Witten, D. (2021b). An introduction to statistical learning. *Springer Texts in Statistics* (Vol. 102). https://doi.org/10.1007/978-1-4614-7138-7

Hosmer Jr, D. W., Lemeshow, S. & Sturdivant, R. X. (2013). *Applied Logistic Regression.* John Wiley & Sons.

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2020). *GLOBAL HIV STATISTICS*. https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/strategic-information/hiv-data-and-statistics

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2021). *HIV treatment*. https://www.unaids.org/es/topic/treatment

Liaw, A. & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22. https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf

Lozano, F. & Domingo, P. (2011). Tratamiento antirretroviral de la infección por el VIHAntiretroviral therapy for HIV infection [Antiretroviral therapy for HIV infection]. *Enfermedades Infecciosas y Microbiología Clínica*, *29*(6), 455–465. doi.org/10.1016/j.eimc.2011.02.009

https://doi.org/10.36105/theanahuacjour.2024v24n1.07

Marsland, S. (2014). *Random Forests. In Machine Learning: An Algorithmic Perspective*. Chapman & Hall/CRC.

McKinney, W. (2013). Python for Data Analysis. O'Reilly Media, Inc.

National Institutes of Health (NIH). (n.d.). Aprobación de los medicamentos contra el VIH por la FDA. https://hivinfo.nih.gov/es/understanding-hiv/infographics/aprobacion-de-los-medicamentos-contra-el-vih-por-la-fda

Ramírez, M. (2019, July 23). Presume México su cambio de estrategia en VIH; #IAS2019. *El Economista*. https://www.eleconomista.com.mx/opinion/Presume-Mexico-su-cambio-de-estrategia-en-VIH-IAS2019-20190724-0007.html

Urzúa, C. (2020, January 20). El episodio de los antirretrovirales. *El Universal*.

World Health Organization (WHO). (2021a). *Global Price Reporting Mechanism*. https://apps.who.int/hiv/amds/price/hdd/

World Health Organization (WHO). (2021b). *HIV/AIDS*. https://www.who.int/health-topics/hiv-aids#tab=tab_1

About the authors

Blanca Iveth Mayorga Basurto is a professional in the field of information technology and data analysis, with a master's degree in Information Technology Management. She is currently in the process of completing her Doctorate in Financial Sciences at EGADE Business School, demonstrating a profound commitment to research and the application of advanced machine learning techniques in financial and public health contexts.

biveth@gmail.com
https://orcid.org/0000-0002-3643-3743?lang=en

Galo Moncada Freire works as a professor of finance and economics at the Tecnológico de Monterrey (ITESM) and at the Universidad Nacional Autónoma de México (UNAM). He holds a PhD in Economic Sciences from the Universidad Autónoma Metropolitana (UAM–Mexico), a master's in Business Economics from Tecnológico de Monterrey, and a degree in Economics from the Universidad Católica de Guayaquil. He also has consultation experience in administration, finance and economics and has worked in consultancy, banking and the commercial sector.

galomf@yahoo.com
https://orcid.org/0009-0007-4831-6599