Universidad Pablo de Olavide (España)
Revista de Métodos Cuantitativos para la Economía y la Empresa
número 39, 2025
ISSN: 1886-516X
DOI: 10.46661/rev.metodoscuant.econ.empresa.10760
Sección: Artículos
Recibido: 28-06-2024
Aceptado: 04-11-2024
Publicado: 26-06-2025
Páginas: 1-22

Revista de Métodos Cuantitativos para la Economía y la Empresa

# Clasificación laboral en México usando un enfoque de aprendizaje automático

## *Job classification in Mexico using a machine learning approach*

Luz Judith Rodríguez Esparza
Universidad Autónoma de Aguascalientes (México)
https://orcid.org/0000-0003-2241-1102
judithr19@gmail.com

Dolly Anabel Ortiz Lazcano
Universidad Autónoma de Aguascalientes (México)
https://orcid.org/0000-0003-3452-3291
dolly.ortiz@edu.uaa.mx

Mónica Fernanda Llamas Valle
Universidad Autónoma de Aguascalientes (México)
https://orcid.org/0009-0000-5365-1657
mf.maslle@gmail.com

**RESUMEN**

En este estudio, se aborda el desaliento laboral en México desde una perspectiva de modelación matemática. Se consideran dos condiciones de empleabilidad: desocupado y desalentado, y se caracteriza la clasificación de estos grupos utilizando modelos de aprendizaje automático y variables sociodemográficas, tales como nivel de instrucción, sexo, edad, estado conyugal, número de hijos, parentesco y ámbito de residencia. Considerando datos de la Encuesta Nacional de Ocupación y Empleo, la mayor precisión de clasificación de los algoritmos abordados la obtuvieron las redes neuronales y los bosques aleatorios. Estos modelos indicaron que las características principales que distinguen a los desalentados de los desempleados son: mujeres de 20-29 años, con educación media superior y superior, sin hijos, solteras y residentes en zonas urbanas. Lo más relevante es que, gracias a los resultados obtenidos con los modelos de aprendizaje automático, es posible no solo predecir con mayor precisión quiénes podrían caer en el desaliento laboral, sino también proponer políticas públicas más efectivas y focalizadas. Estas políticas pueden estar orientadas específicamente a los sectores identificados como más vulnerables, contribuyendo así a la disminución del desaliento laboral y a la mejora de la empleabilidad en el país.

**PALABRAS CLAVE**

Desánimo; aprendizaje automático; clasificación; México; ENOE.







***Cómo citar:*** Rodríguez Esparza, L. J., Ortiz Lazcano, D. A., & Llamas Valle, M. F. (2025). Clasificación laboral en México usando un enfoque de aprendizaje automático. Revista de Métodos Cuantitativos para la Economía y la Empresa, (39), 1–22. https://doi.org/10.46661/rev.metodoscuant.econ.empresa.10760

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



**ABSTRACT**

In this study, work discouragement in Mexico is addressed from a mathematical modeling perspective. Two conditions of employability are considered: unemployed and discouraged, and the classification of these groups is characterized using machine learning models and sociodemographic variables, such as educational level, sex, age, marital status, number of children, relationship, and area of residence. Considering data from the National Occupation and Employment Survey, the highest classification accuracy of the algorithms addressed was obtained by neural networks and random forests. These models indicated that the main features that distinguish the discouraged from the unemployed are women aged 20-29, with high school and higher education, without children, single, and residing in urban areas. The most relevant thing is that, thanks to the results obtained with the machine learning models, it is possible not only to predict with greater precision who could fall into work discouragement, but also, to propose more effective and focused public policies. These policies can be specifically aimed at the sectors identified as most vulnerable, thus contributing to the reduction of job discouragement and the improvement of employability in the country.

**KEYWORDS**

Discouragement; machine learning; classification; Mexico; ENOE.

Clasificación JEL: E24, C10, J21, J64.

MSC2010: 62-07, 62P25, 68T01

## 1. INTRODUCCIÓN

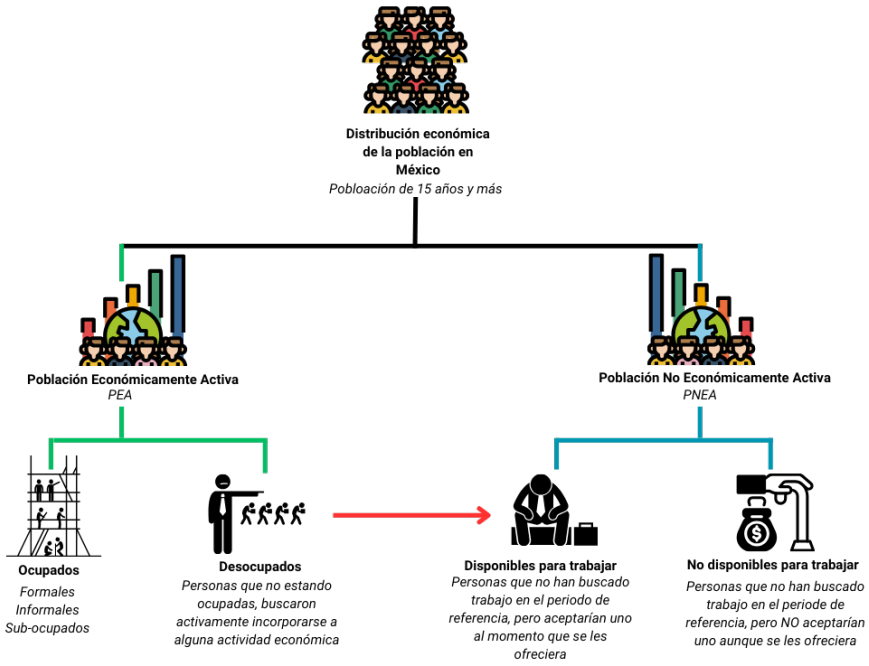
El tema de los trabajadores desalentados ha sido relevante durante décadas. Los primeros estudios que abordaron el Discouraged-worker effect (DWE) fueron los de Long (1953, 1958). Más recientemente, Martín-Román (2022) analizó un modelo que incorporó incertidumbre sobre los resultados de la búsqueda de empleo y los costos de transacción asociados a este proceso. Particularmente en México, este tema es de gran relevancia en el ámbito laboral y social, ya que refleja una problemática que afecta a una parte significativa de la fuerza laboral en el país, en particular a la población juvenil, entre 15 y 29 años (Murguía Salas et al., 2023).

La Figura 1 representa de manera esquemática la distribución económica de la población en México. En esta clasificación, la Población Económicamente Activa (PEA) se divide en personas ocupadas y desocupadas, donde las personas desocupadas son aquellas que buscan empleo activamente, comúnmente referidas como desempleadas. Por otra parte, la Población No Económicamente Activa (PNEA) se compone de personas que están disponibles y no disponibles para trabajar. Entre quienes están disponibles, se encuentran los trabajadores desalentados o desanimados, que son personas que, aunque están dispuestas a trabajar, han perdido la motivación para buscar empleo, en muchos casos debido a factores externos, y por tanto no realizan acciones concretas para encontrar trabajo (INEGI, 2023).

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



**Figura 1. Distribución económica de la población en México.**




Fuente: Elaboración propia.

La transición de una persona desocupada (desempleada) a desalentada (desanimada) es especialmente significativa, pues implica un cambio de la PEA a la PNEA. Este movimiento puede tener repercusiones en la medición de indicadores laborales, ya que la tasa de desempleo puede no reflejar completamente la situación real si no se toma en cuenta a este grupo (Heath, 2014).

En México, la situación laboral presenta grandes desafíos (Ortiz Lazcano y Rodríguez Esparza, 2023), y uno de los fenómenos preocupantes es la creciente tasa de trabajadores desalentados (Moy, 2020). Según Scotti (2015), las personas desalentadas han perdido la motivación para buscar empleo debido a diversas razones, como la falta de oportunidades laborales, la discriminación en el mercado laboral, la desesperanza de encontrar trabajo, entre otros factores. Estas personas, aunque estarían dispuestas a trabajar si se les presentara una oportunidad adecuada, han dejado de buscar activamente empleo, lo que las convierte en parte de la PNEA. Así pues, el desempleo desalentado es un fenómeno preocupante que puede afectar la participación laboral y el bienestar económico de la sociedad (Peón, 2021).

En México, existe la Encuesta Nacional de Ocupación y Empleo (ENOE) que es una importante fuente de información sobre el mercado laboral mexicano ("Encuesta Nacional de Ocupación y Empleo (ENOE)" (s.f.)). Esta encuesta, realizada por el Instituto Nacional de Estadística y Geografía (INEGI), proporciona datos mensuales y trimestrales sobre diversos aspectos relacionados con el empleo y la ocupación en México. La ENOE se inició en 2005 y es el resultado de la fusión de la Encuesta Nacional de Empleo Urbano (ENEU) y la Encuesta Nacional de Empleo (ENE), que durante más de dos décadas recopilaron información sobre la población ocupada y desocupada en el país. Esta consolidación permitió mejorar la captación y comprensión de las características del mercado laboral mexicano. La Encuesta Nacional de Ocupación y Empleo recopila información a través de dos cuestionarios: el Cuestionario Sociodemográfico y el Cuestionario de Ocupación y Empleo, básico y ampliado. Estos cuestionarios se han adaptado a los cambios en el mercado laboral mexicano, siguiendo un marco conceptual actualizado y alineado con estándares internacionales y la legislación nacional. Proporcionando información detallada so-

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



bre diversos aspectos del mercado laboral, como la fuerza de trabajo, la ocupación, la informalidad laboral, la subocupación, la desocupación, entre otros.

La literatura sobre estudios del desaliento laboral en México es limitada, y prácticamente todos ellos se han enfocado en una metodología en particular: los modelos logísticos multinomiales. Se recomienda al lector consultar la referencia Martín-Román (2022), que presenta los modelos econométricos, datos y resultados de los principales estudios que han abordado este tema de manera global.

Así, en este trabajo, mediante el uso de datos de la Encuesta Nacional de Ocupación y Empleo y la aplicación de herramientas estadísticas avanzadas, como los algoritmos de aprendizaje automático (en inglés Machine Learning), se analiza el fenómeno del desaliento en México con un nivel de detalle sin precedentes. Esta metodología identifica características clave de las personas afectadas, ofreciendo una comprensión más profunda de los factores subyacentes y patrones que contribuyen a esta población.

El aprendizaje automático es un conjunto de métodos o algoritmos que permiten a las computadoras aprender a partir de datos para realizar y mejorar predicciones a través de algoritmos computacionales diseñados para emular la inteligencia humana (El Naqa & Murphy, 2015; Molnar, 2020). Un algoritmo de aprendizaje automático es un proceso computacional que utiliza datos de entrada para lograr una tarea deseada sin ser programado literalmente para producir un resultado particular. El objetivo del aprendizaje automático es emular la forma en que los seres humanos aprenden a procesar señales sensoriales para lograr un objetivo. A este ideal se puede llegar principalmente a través de dos tipos de aprendizaje: el aprendizaje supervisado, que son algoritmos que al programarse permiten a la máquina aprender a distinguir características específicas de los objetos a evaluar, desde la observación de los datos de entrenamiento; y el aprendizaje no supervisado, en donde el aprendiz no cuenta con toda la información inicial, a diferencia del supervisado, pues en este caso el entrenamiento no asocia una configuración cinemática de entrada particular con un resultado específico.

Siendo que la transición de estar desocupado a desalentado es importante en el mercado laboral, los métodos de clasificación de aprendizaje automático (Casal et al., 2020) permiten segmentar y entender diferencias clave entre estos grupos, además de tener una comprensión más profunda y detallada de los factores que afectan el desaliento laboral. Así pues, el objetivo principal de este trabajo es caracterizar la condición de empleabilidad (desocupado y desalentado) a través de variables sociodemográficas (nivel de instrucción, sexo, grupo de edad, estado conyugal, número de hijos, posición en el hogar y ámbito de residencia) utilizando algoritmos de aprendizaje automático. Además, el uso de aprendizaje automático ofrece la posibilidad de anticipar tendencias y mejorar las predicciones sobre el desaliento laboral, lo cual representa un avance significativo en la comprensión de esta problemática en el mercado laboral y aporta una perspectiva innovadora frente a investigaciones tradicionales.

La falta de estudios específicos del mercado laboral limita la capacidad de los responsables de la toma de decisiones, líderes empresariales y formuladores de políticas para diseñar estrategias efectivas que aborden las causas subyacentes de la desmotivación laboral. Por lo tanto, es fundamental investigar y analizar en profundidad este fenómeno para poder desarrollar políticas y programas que promuevan un ambiente laboral más saludable, inclusivo y motivador en México.

Este artículo está organizado de la siguiente manera: en la Sección 2 se presentan algunos antecedentes encontrados en la literatura; en la Sección 3 se describen los modelos de aprendizaje automático utilizados para la clasificación. Luego, en la Sección 4 se muestran los resultados obtenidos, y, finalmente, se presentan las conclusiones en la Sección 5.

## 2. ANTECEDENTES

Los efectos sobre la oferta de mano de obra se examinan en los trabajos de Woytinsky (1940) y Long (1953). En esos estudios, se revisan conceptos clave del mercado laboral, como la población activa, los desempleados, los trabajadores por cuenta propia y la población inactiva,

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]

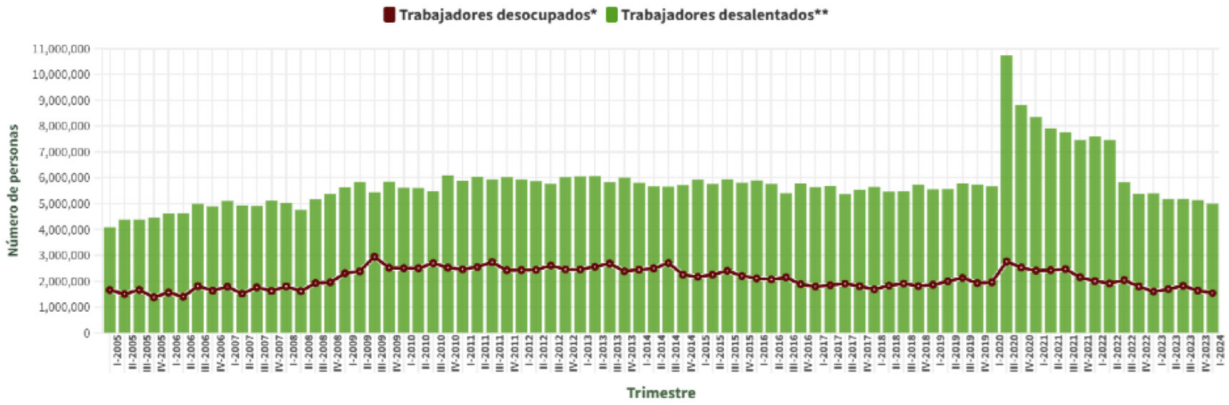


que incluye a estudiantes y amas de casa cuyo trabajo "no tiene ningún sentido económico" ya que no perciben salario. El término added-worker o trabajador adicional se utiliza para describir a los miembros de una familia, generalmente mujeres o adultos jóvenes, que ingresan al mercado laboral en respuesta a un aumento en la tasa de desempleo o la reducción de los ingresos del hogar. Esto sucede porque los hogares necesitan compensar la pérdida de ingresos de un miembro desempleado. Por su parte, Humphrey (1940) realiza una crítica a la teoría de Woytinsky sobre el desempleo, argumentando que la relación con su teoría del trabajador adicional requiere un enfoque redistributivo. Señala que es crucial considerar las características demográficas de los trabajadores, como la edad, el sexo, la raza, la calificación y la experiencia laboral, ya que estas diferencias pueden contribuir a la formación de mercados de trabajo precarizados debido a las variaciones salariales.

Long (1953) cuestiona los postulados keynesianos de su época, argumentando que estos no incluyen a los desempleados en la oferta laboral, lo cual resulta fundamental para dimensionar adecuadamente la fuerza de trabajo. En este sentido, es relevante considerar el valor económico de las personas no económicamente activas que están disponibles para trabajar, aunque no busquen empleo activamente. La Organización Internacional del Trabajo (ILO, por sus siglas en inglés International Labor Organization) clasifica a estos individuos, junto con los desempleados, como trabajadores potenciales (International Labour Organization, 2023), una categoría que no debería excluirse por completo de la oferta laboral y que ayudaría a comprender mejor el verdadero tamaño del mercado de trabajo.

En el ámbito específico del mercado laboral mexicano, la Figura 2 podría facilitar una comprensión más clara de la dinámica interconceptual entre la PEA y la PNEA en relación con los trabajadores desocupados y desalentados. En este contexto, el número de personas que reportan estar desempleadas representa consistentemente menos de la mitad de aquellas que han dejado de buscar empleo pero que estarían dispuestas a aceptar uno de inmediato. Esta disposición para trabajar, más allá de los argumentos de la teoría del trabajador adicional, refleja una dinámica relacionada más con los bajos salarios que con el desempleo.

**Figura 2. Número de personas desocupadas y desalentadas en México, por trimestre y por año.**




Nota: INEGI (2024) Encuesta Nacional de Ocupación y Empleo •
* Se refiere a la población que no trabajó siquiera una hora durante la semana de referencia de la encuesta, pero manifestó su disposición para hacerlo y realizó alguna actividad para obtener empleo.
** Población no ecómicamente activa (PNEA) que representa a individuos inactivos disponibles para trabajar, es decir, personas que buscaron trabajo en el pasado, pero desistió de hacerlo, o que no han buscado trabajo porque consideran que no tienen oportunidad alguna, aunque estarían dispuestos a ocupar un empleo inmediatamente.

Fuente: Elaboración propia.

Hasta ahora se ha estudiado ampliamente el desempleo en México y su relación con diferentes variables a través de distintos modelos (Barajas & Ibarra, 2016; Mariscal & Villarreal, 2017; Sánchez-Salgado & Alonso-Villarreal, 2018). Ruiz Nápoles & Ordaz Díaz (2011), por ejemplo, ana-

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



lizaron la evolución y las tendencias del empleo y del desempleo en México desde la aplicación de las reformas económicas iniciadas en los años ochenta. En esa referencia se muestra que no se cumplieron las expectativas de una mejora del desempeño laboral despertadas por las reformas económicas de esas décadas. Por otro lado, Hernández Pérez (2020) estimó el efecto que tiene el sexo, edad y nivel de instrucción sobre la tasa de desempleo mediante un análisis de panel. Sus resultados muestran que las características que se asocian a un mayor desempleo son: el sexo y el nivel de instrucción, con primaria incompleta y completa, mientras que las características que se asocian a un menor desempleo son: la edad de 45 años y más y el nivel de instrucción de secundaria completa, medio superior y superior. En la referencia de Arroyo-Martínez & Ortega-Ovalle (2020) encontraron que la inversión influyó en la tasa de desempleo en México de manera significativa, al igual que la educación; mientras que el salario tuvo un impacto poco significativo en la tasa de desempleo. Segovia Hernández (2021) realizó un análisis de series de tiempo de la tasa de desempleo trimestral para el periodo 1998-2020, con el objetivo de evaluar el impacto de eventos actuales sobre los futuros niveles de desempleo. Sus resultados sugirieron que, de mantenerse invariables los esquemas de contratación y despido ante los efectos de la crisis COVID-19, los niveles de desempleo se verían incrementados.

Recientemente, Maridueña-Larrea & Martín-Román (2024) examinaron la relación entre el desempleo y las tasas de actividad en seis países de América Latina, incluido México. Sus conclusiones señalan la existencia de rigideces que dificultan la mejora en el mercado laboral. Además, sus resultados muestran que no hay una relación de equilibrio a largo plazo en el modelo agregado para Brasil y México, lo que confirma la hipótesis de invariabilidad del desempleo en estas economías. Esto sugiere que las políticas laborales implementadas para responder a los cambios en la participación laboral en ambos países no han influido en sus tasas de desempleo a largo plazo.

Respecto a estudios previos del desaliento, Scotti (2015) trabajó con la hipótesis del desaliento como factor de exclusión del mercado laboral, analizando en qué medida y para quiénes es una expresión de esta exclusión o un fenómeno transitorio. Utilizando microdatos de la Encuesta Nacional de Ocupación y Empleo y un modelo logístico multinomial, observó que, para el segundo trimestre de 2012, el 10 % de la población estaba afectada por el desempleo abierto y el desaliento laboral. Además, encontró que los hombres tenían mayores tasas de participación laboral que las mujeres. De hecho, históricamente, las mujeres han tenido menor probabilidad de participar en el mercado laboral que los hombres en todo el mundo (International Labour Organization, 2024).

Por su parte, Escoto et al. (2017), en su investigación "Desempleo abierto y desalentado en tres mercados de trabajo latinoamericanos", analizaron la desocupación desde dos perspectivas: el desempleo abierto y el desaliento, para mostrar la desigualdad en el mercado laboral y la persistencia de dinámicas de exclusión laboral. El estudio abarcó tres países latinoamericanos: Costa Rica, México y Uruguay. Costa Rica fue seleccionado por su estancamiento en la lucha contra la pobreza y aumento de la desigualdad; México, por su constante desigualdad y baja institucionalidad laboral; y Uruguay, por su recuperación de los niveles de ocupación y fortalecimiento de la institucionalidad laboral. En ese trabajo, también se utilizó un modelo logístico multinomial. Más recientemente, Murguía Salas et al. (2023) analizaron el desaliento laboral en México considerando tres años: 2013, 2019 y 2022, también utilizando un modelo logístico multinomial, los factores sociodemográficos que los autores identificaron para el desaliento fueron: mujer, tener hijos, cuidar de otras personas, estudiar y tener mayores niveles de cualificación.

## 3. METODOLOGÍA

Consideremos el siguiente modelo:

**(1)**

$$Y = f(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$$

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



donde la variable dependiente *Y* indica la condición de empleabilidad de las personas, de tipo cualitativo y cuenta con dos categorías, si la persona está desocupada o desalentada (ver Figura 1). Las variables independientes s *X*i, i=1,⋯,7, son de tipo categóricas y se describen en la Tabla 1. Éstas son: nivel de instrucción (que llamaremos Educación), sexo, grupo de edad (que llamaremos Edad), estado conyugal (que llamaremos Conyugal), número de hijos (que llamaremos Hijos), posición en el hogar (que llamaremos Parentesco) y ámbito de residencia (que llamaremos Residencia). La función f depende de cada modelo de aprendizaje automático a utilizar.

**Tabla 1. Codificación de las variables para el análisis de clasificación.**

| Tipo | Variable | Categorías | Codificación |
|---|---|---|---|
| Dependiente | *Y*=Condición de empleabilidad | Desocupado<br>Desalentado | 1<br>2 |
| Independientes | $X_1$=Educación | Ninguno<br>Primeria incompleta<br>Primaria completa<br>Secundaria completa<br>Medio superior y superior<br>No especificado (NE) | 0<br>1<br>2<br>3<br>4<br>5 |
| | $X_2$ =Sexo | Hombre<br>Mujer | 1<br>2 |
| | $X_3$ =Edad | No aplica (NA)<br>15 a 19 años<br>20 a 29 años<br>30 a 39 años<br>40 a 49 años<br>50 a 59 años<br>60 años y más | 0<br>1<br>2<br>3<br>4<br>5<br>6 |
| | $X_4$ =Conyugal | NE<br>Unión libre<br>Separado<br>Divorciado<br>Viudo<br>Casado<br>Soltero | 0<br>1<br>2<br>3<br>4<br>5<br>6 |
| | $X_5$ =Hijos | NA- Hombres<br>Sin hijos<br>1 a 2 hijos<br>3 a 5 hijos<br>6 y más hijos<br>NE | 0<br>1<br>2<br>3<br>4<br>5 |
| | $X_6$ =Parentesco | Jefe de hogar<br>Cónyuge<br>Hijo/a<br>Sin parentesco<br>Otro parentesco | 0<br>1<br>2<br>3<br>4 |
| | $X_7$ =Residencia | Urbano<br>Rural | 1<br>2 |

Fuente: Elaboración propia.

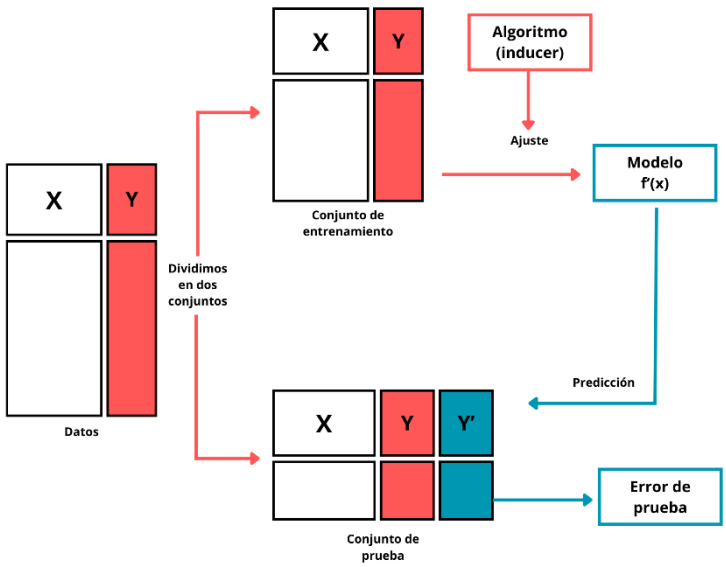
N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



Para llevar a cabo la clasificación de la condición de empleabilidad, se realiza un aprendizaje supervisado, pues ya se cuenta con la clasificación de la población dada por los microdatos de la Encuesta Nacional de Ocupación y Empleo, por lo que se busca un modelo que mejor clasifique a las personas de acuerdo con las variables sociodemográficas antes mencionadas.

En el aprendizaje automático el manejo de los datos es a través de dos conjuntos: el de entrenamiento y el de prueba (ver Figura 3). En el primero de ellos se realizan los ajustes con los algoritmos propuestos, luego, se evalúan (o predicen) en el conjunto de prueba para poder calcular el error de la predicción, entre menor sea este error indicará que el modelo ajusta mejor a los datos. En el caso de clasificación, este error de prueba se puede calcular a través de la matriz de confusión, que más adelante se aborda.

**Figura 3. Esquema general sobre el manejo de los datos con aprendizaje automático.**




Fuente: Elaboración propia.

Los algoritmos de aprendizaje automático propuestos en este trabajo para analizar los datos se describen a continuación.

## 3.1. Regresión logística

La regresión logística es un modelo estadístico utilizado para predecir la probabilidad de una variable dependiente binaria. Utiliza una función logística para modelar la variable dependiente, resultando en valores entre 0 y 1 que pueden interpretarse como probabilidades. La ecuación de la regresión logística es:

**(2)**

$$P(Y=1|\boldsymbol{x}) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}$$

donde $P(Y=1|\boldsymbol{x})$es la probabilidad de que $Y$ tome el valor de 1 dado $\boldsymbol{x}=(x_1,\ldots,x_n)$, $\beta_0$ es el parámetro de la intersección y $\beta_1,\cdots,\beta_n$ son los coeficientes de las variables independientes. Se recomienda la referencia Hosmer Jr, Lemeshow & Sturdivant (2013) para más información.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



## 3.2. Árboles de decisión (ver, por ejemplo, Loh (2011)).

Los árboles de decisión dividen el espacio de características en regiones homogéneas en cuanto a la variable objetivo. Cada nodo interno representa una prueba en una característica, cada rama representa el resultado de la prueba, y cada hoja representa una clase de etiqueta. La construcción del árbol se basa en la partición recursiva de los datos, seleccionando en cada paso la característica que mejor separa las clases objetivo. El criterio de división puede ser la entropía o el índice de Gini:

(3)

$$Entropía = -\sum_{i=1}^{k} p_i \log(p_i)$$

(4)

$$Índice\ de\ Gini = 1 - \sum_{i=1}^{k} p_i^2$$

donde $k$ es el número de clases y $pi$ es la proporción de observaciones de la clase $i$, para $i = 1,\cdots, k$. Se debe aplicar el mismo proceso a cada subconjunto de datos hasta que todas las instancias en un nodo pertenezcan a la misma clase, o bien, no queden características por dividir.

## 3.3. Bosques aleatorios (ver Loh (2011)).

Los bosques aleatorios son un conjunto de árboles de decisión entrenados de manera independiente utilizando diferentes subconjuntos del conjunto de datos. La clasificación final se obtiene mediante un voto mayoritario de las predicciones de todos los árboles. Este método mejora la precisión y reduce el sobreajuste. Para una nueva observación $\boldsymbol{x}$, la predicción $\hat{Y}$ está dada por

(5)

$$\hat{Y} = moda(\{T_b(\boldsymbol{x})\}_{b=1}^{B})$$

donde $T_b(\boldsymbol{x})$ es la predicción del $b$–ésimo árbol y $B$ es el número total de árboles.

Los bosques aleatorios combinan la salida de múltiples árboles de decisión para mejorar la precisión de la predicción y controlar el sobreajuste. La clave de su éxito es la combinación de bagging (técnica de remuestreo) y la selección aleatoria de características en cada nodo, lo que introduce diversidad en los árboles y mejora la generalización del modelo.

## 3.4. Clasificador bayesiano ingenuo (ver, por ejemplo, Sen et al. (2022)).

El clasificador bayesiano ingenuo (en inglés *Naive Bayes*) es un modelo probabilístico basado en el teorema de Bayes con la suposición "*naive*" (ingenua) de que todas las características (o atributos) son independientes entre sí dada la clase $C$. Para clasificar una nueva observación $\boldsymbol{x}=(x_1,\ldots,x_n)$ primero se calcula la probabilidad a posteriori para cada clase $C_k$:

(6)

$$P(C_k|\boldsymbol{x}) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(\boldsymbol{x})}$$

donde $P(C_k)$ es la probabilidad a priori de la clase $C_k$ y $P(x_i | C_k)$ es la probabilidad de observar $x_i$ dado que la clase es $C_k$. Como $P(\boldsymbol{x})$ es constante para todas las clases, entonces

**(7)**

$$P(C_k|\boldsymbol{x}) \propto P(C_k)\prod_{i=1}^{n} P(x_i|C_k).$$

Finalmente, la clase asignada a la observación $\boldsymbol{x}$ es la que maximiza la probabilidad a posteriori

**(8)**

$$\hat{C} = \arg max_{C_k} P(C_k|\boldsymbol{x}).$$

## 3.5. K–Vecinos más cercanos (ver El Naqa & Murphy (2015)).

El método *k*- vecinos más cercanos (en inglés k-nearest neighbors (k-NN)) es un algoritmo de clasificación no paramétrico basado en la similitud. Para clasificar una instancia, el algoritmo busca los *k* vecinos más cercanos en el espacio de características y asigna la clase más común entre esos vecinos. La distancia entre puntos se puede calcular mediante la distancia euclidiana:

**(9)**

$$d(\boldsymbol{x}, \boldsymbol{x}') = \sqrt{\sum_{i=1}^{n}(x_i - x_i')^2}$$

donde donde $\boldsymbol{x}=(x_1,\cdots,x_n)$ y $\boldsymbol{x}'=(x'_1,\cdots,x'_n)$ son dos puntos en un espacio *n*–dimensional. Así, para una nueva observación $\boldsymbol{x}$ se calcula la distancia dada en la ecuación (9) entre $\boldsymbol{x}$ y cada instancia $\boldsymbol{x}'$ del conjunto de datos de entrenamiento, seleccionando los *k* vecinos más cercanos a $\boldsymbol{x}$ en términos de la distancia calculada. Finalmente, la clase de $\boldsymbol{x}$ se determina por mayoría de votos entre los *k* vecinos más cercanos:

**(10)**

$$\hat{C} = argmax_C \prod_{i=1}^{k} 1(C_i = C)$$

donde $1(C_i = C)$ es la función indicadora que es 1 si la clase del *i*–ésimo vecino es *C* y 0 en caso contrario.

## 3.6. Redes Neuronales (ver, por ejemplo, Goodfellow, Bengio y Courville (2016)).

Las redes neuronales son modelos de aprendizaje inspirados en la estructura del cerebro humano. Son especialmente potentes para problemas de clasificación, principalemente cuando se trabaja con grandes conjuntos de datos y características complejas. Consisten en capas de neuronas artificiales (nodos) que procesan entradas y transmiten señales a través de sinapsis ponderadas a otras neuronas en capas posteriores. La salida de una neurona en la capa *l* es:

**(11)**

$$a_i' = f\left(\sum_{j=1}^{n} w_{ij}^{l} a_j^{l-1} + b_i^{l}\right)$$

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



donde $a_i^l$ es la activación de la neurona *i* en la capa *l*, $w_{ij}^l$ es el peso de la conexión entre la neurona *j* en la capa *l-1* y la neurona *i* en la capa *l*, $b_i^l$ es el sesgo, y *f* es la función de activación (como la sigmoide, ReLU (Rectified Linear Unit) y softmax). En particular, para clasificación binaria se utiliza la función de activación sigmoide que está dada por

**(12)**

$$\sigma(z) = \frac{1}{1+e^z} .$$

## 3.7. Máquinas de Soporte Vectorial (SVM por sus siglas en inglés Support vector machine) (ver, por ejemplo, El Naqa & Murphy (2015)).

Las SVM son modelos de clasificación que buscan el hiperplano que mejor separa las clases en el espacio de características, maximizando el margen entre las clases. Para problemas lineales, el hiperplano se define como:

**(13)**

$$w \cdot x + b = 0,$$

donde *w* es el vector de pesos y *b* es el sesgo. El margen se maximiza resolviendo el problema de optimización:

**(14)**

$$\min_{w,b} \frac{1}{2}||w||^2 \text{ sujeto a } y_i(w \cdot x_i + b) \geq 1, \forall i$$

donde $y_i$ es la etiqueta de clase de la *i*–ésima observación.

Para más detalles de los modelos se recomienda Probst et al. (2019) y Fernández-Delgado et al. (2019).

Luego de ajustar los modelos para cada año, se evalúa su eficiencia mediante las medidas de desempeño como la precisión, sensibilidad y especificidad, a partir de una matriz de confusión (Dinov, 2018). Ésta es una herramienta que permite visualizar el rendimiento de un modelo de clasificación al comparar las predicciones del modelo con los valores reales de las clases de la variable respuesta. La matriz organiza las predicciones en cuatro categorías: verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN). Estos valores se presentan en forma de tabla, como se muestra a continuación:

| | | Clase predicha | |
|---|---|---|---|
| | | **Positivo** | **Negativo** |
| Clase real | **Positivo** | *VP* | *FN* |
| | **Negativo** | *FP* | *VN* |

Al construir la matriz de confusión de un modelo, sus medidas de desempeño se definen de la siguiente manera. La precisión mide la proporción de todas las predicciones correctas, tanto de verdaderos positivos como de verdaderos negativos entre todas las predicciones del modelo, i.e.,

**(15)**

$$Precisión = \frac{VP+VN}{VP+FP+VN+FN}.$$

Por otro lado, se tiene la sensibilidad, también llamada tasa de verdaderos positivos o exhaustividad, que es una métrica que mide la proporción de verdaderos positivos identificados correctamente por el modelo respecto al total de casos positivos reales. Indica la capacidad del modelo para identificar correctamente los casos positivos. Se calcula mediante la fórmula

**(16)**

$$Sensibilidad = \frac{VP}{VP+FN}.$$

La especificidad es una métrica que mide la proporción de verdaderos negativos identificados correctamente por el modelo respecto al total de casos negativos reales. Indica la capacidad del modelo para identificar correctamente los casos negativos. Se calcula con la fórmula

**(17)**

$$Especificidad = \frac{VN}{VN+FP}.$$

Las métricas anteriores sirven para determinar los mejores modelos, con los cuales, se obtienen las categorías más importantes de cada variable, para finalmente caracterizar a la población desalentada. En este trabajo la clase real positivo es Desempleado y la case real negativo es Desalentado, por lo que es importante calcular, además de la precisión, la especificidad.

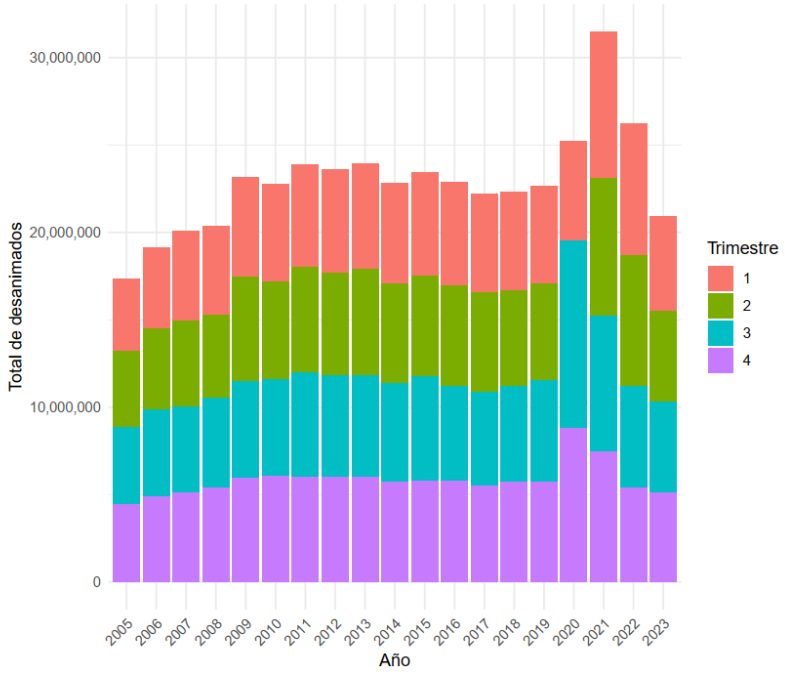
## 4. RESULTADOS

Para contextualizar el fenómeno del desaliento laboral en México, primero se obtuvo el total de personas desalentadas para el periodo comprendido entre 2005 y 2023 (los datos que se utilizaron en este trabajo se recuperaron de la Encuesta Nacional de Ocupación y Empleo el 13 de marzo de 2024). Estos resultados se presentan en la Figura 4. Se observa un aumento significativo en el año 2009, que se mantuvo estable hasta 2019. En 2020, hubo un ligero crecimiento, seguido de un gran repunte en 2021. Sin embargo, en los últimos años, el número de desalentados ha ido disminuyendo. En 2023, el total de personas desalentadas se redujo a niveles similares a los registrados en 2008.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



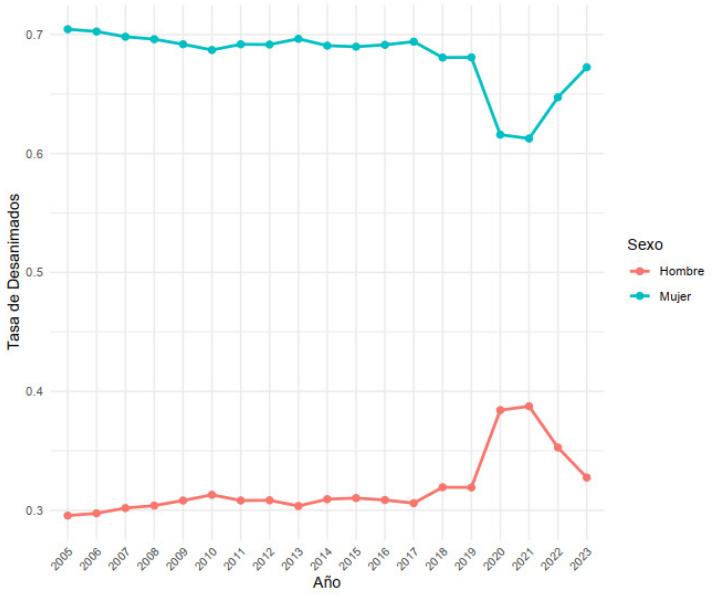
**Figura 4. Número total de personas desalentadas en México de 2005 a 2023.**




Fuente: Elaboración propia.

Particularmente, y con base en Scotti (2015), en la Figura 5 se presenta la proporción de desalentados por sexo, esto con el objetivo de visualizar cómo a las mujeres siempre les ha afectado más el desaliento laboral.

**Figura 5. Proporción de la tasa de desalentados de 2005 a 2023 por sexo.**




Fuente: Elaboración propia.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



Para el estudio de clasificación, se consideraron los datos de personas clasificadas como desocupadas (dentro de la PEA) y aquellas disponibles para trabajar (dentro de la PNEA). Se utilizaron los microdatos de la Encuesta Nacional de Ocupación y Empleo de cuatro años específicos: 2005 por ser el primer año en que se aplicó esta encuesta en México, 2009 debido a la crisis económica que vivió el país, el año 2020 por la crisis sanitaria debida al COVID-19, y el 2023, por ser el año más reciente de aplicación de la encuesta. En la Figura 6 se presentan las proporciones de estas personas de acuerdo con las variables independientes. Como se puede observar, en todas las categorías de todas las variables independientes, la proporción de personas desalentadas es mucho mayor a la de los desocupados.

**Figura 6. Proporción de desocupados y desalentados considerando la muestra general de todos los años 2005, 2009, 2020 y 2023.**


Fuente: Elaboración propia.

Los conjuntos de entrenamiento para cada año se construyeron con el 80 % del total de los datos, seleccionados de manera aleatoria, teniendo el 20 % restante como conjuntos de prueba. En la Tabla 2, se muestran el total de datos que tuvo cada conjunto para los años analizados.

**Tabla 2. Número de personas (desocupadas y desalentadas) usadas en el análisis de clasificación.**

| Datos | Año | | | |
|---|---|---|---|---|
| | 2005 | 2009 | 2020 | 2023 |
| Entrenamiento | 75,417 | 99,466 | 70,566 | 75,280 |
| Prueba | 18,854 | 24,865 | 17,641 | 18,819 |
| Total | 94,271 | 124,331 | 88,207 | 94,099 |

Fuente: Elaboración propia.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



Utilizando el paquete estadístico R, y las librerías rpart, randomForest, e1071, kknn y nnet se aplicaron los modelos de clasificación mencionados en la Sección 3 considerando los datos de entrenamiento cada año, para determinar si una persona estaba desocupada o bien desalentada, según las variables descritas en la Tabla 1. Una vez ajustados los modelos, se hizo la predicción de clasificación a los datos del conjunto de prueba y se compararon con sus valores reales (clasificación real), con lo que se obtuvo la precisión de cada algoritmo usando la fórmula (15). Los resultados de las precisiones se muestran en la Tabla 3.

**Tabla 3. Precisión de los modelos de clasificación para cada año**

| Modelo | Precisión | | | |
|---|---|---|---|---|
| | 2005 | 2009 | 2020 | 2023 |
| Regresión logística | 0.789328 | 0.769113 | 0.769740 | 0.785429 |
| Árboles de decisión | 0.787843 | 0.766096 | 0.774615 | 0.788299 |
| Bosques aleatorios | 0.797602 | 0.780856 | 0.775239 | 0.795632 |
| Naive-Bayes | 0.735918 | 0.729418 | 0.731081 | 0.741803 |
| k-NN | 0.764931 | 0.757973 | 0.735445 | 0.761464 |
| Redes neuronales | 0.800201 | 0.780977 | 0.776940 | 0.794144 |
| SVM | 0.788161 | 0.771043 | 0.769287 | 0.791912 |

Fuente: Elaboración propia.

Las precisiones obtenidas estuvieron entre el 72 % y el 80 %, este es el porcentaje de clasificación correcta que los algoritmos arrojaron tanto para los desempleados como para los desalentados. Para los años 2005, 2009 y 2020, el mejor modelo, según su precisión de predicción, fue redes neuronales y para el año 2023 fue árboles de decisión. Cabe señalar que, en todos años, la diferencia absoluta entre estos dos modelos fue de a lo más 1.7 %, es decir que ambos algoritmos generaron predicciones satisfactorias.

Se obtuvo además la sensibilidad (usando la fórmula (16)), para medir el porcentaje de la población desocupada que se clasificó correctamente, y la especificidad (usando la fórmula (17)) para el porcentaje de población desalentada clasificada correctamente. En la Figura 7 se muestran estas métricas considerando el año 2005. Los bloques sombreados en color verde para cada categoría representan la sensibilidad para la categoría desocupado y la especificidad para la categoría desalentado, éste último es el que particularmente nos interesa en este trabajo. Se observa que los desocupados se clasificaron correctamente entre el 32 % y el 56 % de los datos. Mientras que los desalentados, en general tuvieron clasificaciones bastante satisfactorias con todos los modelos, pues al menos el 79 % de los datos fue clasificado correctamente. La mejor clasificación de personas desalentadas se obtuvo a través de los árboles de decisión y bosques aleatorios, teniendo una especificidad de casi el 95 %, no muy alejados de éstos se encuentran las redes neuronales con el 93.6 % de clasificaciones correctas.

Dado que para el resto de los años (2009, 2020 y 2023), los resultados fueron muy similares, se omiten las gráficas.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]

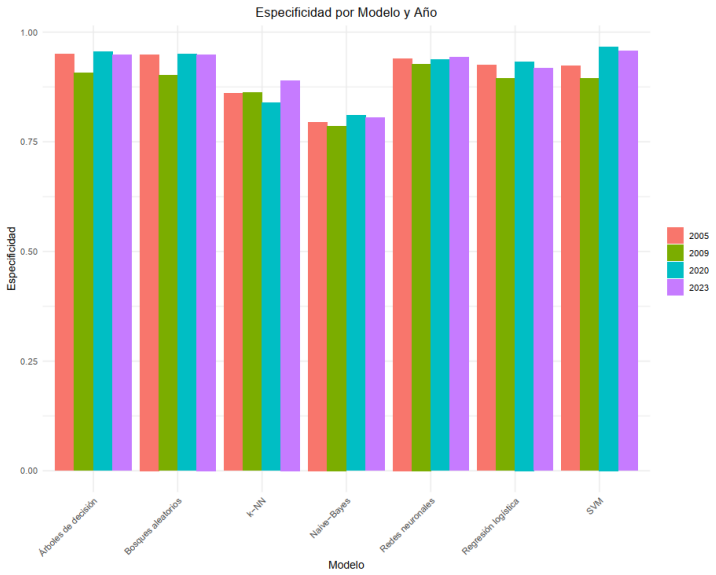


**Figura 7. Sensibilidad y especificidad (en porcentaje) de los modelos para el año 2005.**


Fuente: Elaboración propia.

En la Figura 8 se muestra la especificidad de todos los años y de todos los métodos, notando que los métodos k-NN y bayesiano ingenuo fueron los que menor especificidad obtuvieron. Otro patrón importante de notar es que, en el año 2009, generalmente se obtuvieron las menores especificidades. Es decir, los algoritmos de aprendizaje automático para ese año en particular no lograron identificar de manera precisa las características que distinguieron a las personas desalentadas de las desocupadas.

**Figura 8. Especificidad de los modelos para los años considerados.**




Fuente: Elaboración propia.

En la Tabla 4 se presentan las mayores especificidades de los cuatro años analizados, demostrando que la modelación a través de árboles de decisión, bosques aleatorios, redes neuronales y SVM ha sido muy eficaz en la clasificación de los desanimados.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



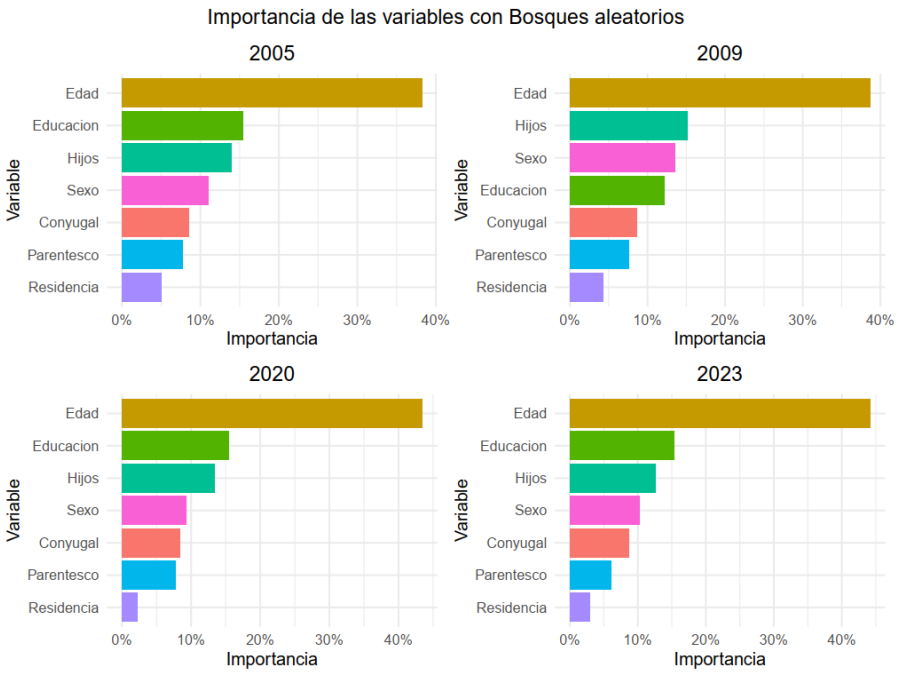
**Tabla 4. Especificidades de los mejores métodos de aprendizaje automático de los cuatro años analizados.**

| Año | Método | Especificidad (porcentaje) |
|---|---|---|
| 2005 | Árboles de decisión | 94.97 |
| | Bosques Aleatorios | 94.90 |
| 2009 | Redes neuronales | 92.35 |
| | Árboles de decisión | 90.71 |
| 2020 | SVM | 96.66 |
| | Árboles de decisión | 95.56 |
| 2023 | SVM | 95.73 |
| | Bosques Aleatorios | 94.89 |

Fuente: Elaboración propia.

Un algoritmo que obtuvo altas precisiones y alta especificidad fue el de bosques aleatorios. Con este modelo se logró obtener el porcentaje de importancia de las variables para cada año, como se muestra en la Figura 9. Es posible observar que, en todos los años, la edad es la variable con mayor peso y las que menos aportan son las variables conyugal, parentesco y residencia. En los años 2005, 2020 y 2023, la variable educación fue la segunda de mayor importancia, seguida del número de hijos y sexo.

**Figura 9. Importancia de las variables para cada año obtenida con el algoritmo de bosques aleatorios.**




Fuente: Elaboración propia.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



Generalmente, los algoritmos de aprendizaje automático no nos proveen las categorías de las variables que mayor aportan al modelado. Sin embargo, en el paquete estadístico R existe la librería lime que nos ayuda con la explicación de las variables para las redes neuronales, este algoritmo, al igual que bosques aleatorios, también obtuvo altas precisiones en la predicción.

El modelo de redes neuronales permitió obtener una cuantificación del peso que cada característica de las siete variables aportaba al modelo. En la Figura 10 se presentan estos pesos, empleando la codificación de la Tabla 1.

La Tabla 5 muestra la característica de cada variable que mayor peso obtuvo en el modelo de clasificación de redes neuronales.

**Tabla 5. Características importantes de las variables con redes neuronales por año.**

| Variable | 2005 | 2009 | 2020 | 2023 |
|---|---|---|---|---|
| Empleabilidad | Desanimado | Desanimado | Desanimado | Desanimado |
| Educación | Secundaria completa | Secundaria completa | Medio superior y superior | Medio superior y superior |
| Sexo | Mujer | Mujer | Mujer | Mujer |
| Grupo de edad | 20 a 29 años | 20 a 29 años | 20 a 29 años | 20 a 29 años |
| Estado conyugal | Soltero | Soltero | Soltero | Soltero |
| Número de hijos | Hombre, Sin hijos | Hombre, Sin hijos | Hombre, Sin hijos | Hombre, Sin hijos |
| Parentesco | Hijo | Hijo | Hijo | Hijo |
| Ámbito de residencia | Urbano | Urbano | Urbano | Urbano |

Fuente: Elaboración propia.

La información presentada en la Tabla 5 conjuntamente con la Figura 10 es uno de los aportes más significativos de este estudio. La edad es una variable muy importante para distinguir entre los grupos de desempleados y desalentados, particularmente de 20 a 29 años. Luego, se encuentra la variable educación, si bien, en años anteriores las personas que tenían secundaria completa era un factor importante, en los últimos años, específicamente, derivado de la pandemia por COVID-19, se ha visto que las personas con nivel de estudios de educación media superior y superior están resultando determinantes para la clasificación en cuestión. Obviamente el sexo, con la categoría mujer, también es relevante, concordando con el dato descriptivo de que desde 2005 hasta la fecha la tasa de mujeres desalentadas siempre ha sido mucho mayor que la de los hombres (ver Figura 5). Respecto a las otras características, resaltamos a las personas principalmente solteras, sin hijos, que ocupan la posición de hijos o hijas en sus hogares y que residen en zonas urbanas.

## 5. CONCLUSIONES

Esta investigación ha permitido analizar la influencia de variables sociodemográficas en el desaliento laboral de México considerando los años 2005, 2009, 2020 y 2023. A través de métodos de aprendizaje automático de clasificación, se identificaron patrones y tendencias significativas en la población trabajadora del país. Esta clasificación ayudó a segmentar y diferenciar entre los grupos de desocupados y desalentados, además de poder identificar perfiles de individuos con características particulares que afectan su estado laboral.

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



Aunque el tema de la clasificación de los desempleados junto con los desalentados ya había sido explorado previamente utilizando modelos logísticos, este trabajo representa un avance significativo. En esta investigación, se ha llevado a cabo una primera tentativa de modelar esta clasificación empleando modelos de aprendizaje automático, los cuales no solo superan en precisión y robustez a los modelos logísticos tradicionales, sino que también ofrecen un potencial considerablemente mayor para las predicciones futuras.

Considerando la precisión, los algoritmos de bosques aleatorios y redes neuronales demostraron una precisión destacada (aproximada del 80%), identificando a mujeres jóvenes entre los 20 y 29 años, solteras, con educación superior, sin hijos y residentes urbanos como el grupo con mayor propensión al desaliento laboral.

**Figura 10. Pesos de las características de las variables con redes neuronales en 2005, 2009, 2020 y 2023.**


Fuente: Elaboración propia.

En general, el haber considerado los años 2005, 2009, 2020 y 2023, como referentes económicos y sociales del país, no resultó relevante en este trabajo; si bien, hubo un incremento notorio en el total de desalentados por efecto de la pandemia, se esperaría que la modelación hubiese sido diferente en cada año, y no fue así. A través de los cuatro años, se pudo observar un patrón estable de características que ayudaron a caracterizar el grupo de los desalentados; el único cambio que es importante resaltar es el de nivel de instrucción a partir del 2020.

En este apartado, resultó notable observar que, hasta antes del año 2020, la población con estudios de secundaria completa mostraba un mayor índice de pertenencia al grupo de desalentados. Sin embargo, debido a la crisis sanitaria provocada por la COVID-19, este patrón

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



cambió significativamente, destacándose ahora el nivel de educación superior como el más representativo. Este cambio sugiere una proyección de aumento en la tasa de desaliento entre las personas con estudios de medio superior y superior en los próximos años.

A pesar de los avances logrados, es importante señalar las limitaciones inherentes al uso de algoritmos de aprendizaje automático, los cuales no siempre proporcionan una comprensión completa de las variables más importantes en la predicción del desaliento laboral. Aún así, el principal aporte de este estudio radica en la aplicación de técnicas de aprendizaje automático para analizar el desaliento laboral en México, lo cual no había sido abordado previamente en la literatura. Esto resalta la importancia de utilizar enfoques innovadores para comprender y abordar los desafíos laborales en el país.

Además, los resultados de este estudio proporcionan una base sólida para futuras investigaciones en el campo del desaliento laboral en México. Sería beneficioso llevar a cabo estudios longitudinales que utilicen los modelos aquí desarrollados para realizar predicciones sobre la evolución del desaliento laboral en el país en los próximos años. Estas predicciones podrían ser de gran utilidad para informar a las autoridades laborales y gubernamentales en la formulación de políticas y programas destinados a abordar este problema de manera proactiva y efectiva.

## AGRADECIMIENTOS

LJRE agradece a la Universidad Autónoma de Aguascalientes por su apoyo a través del proyecto PIM23-3. Un reconocimiento especial a los revisores, cuya atenta revisión y valiosas sugerencias permitieron fortalecer este artículo.

## CONTRIBUCIÓN DE LOS AUTORES

Conceptualización: DAOL, LJRE. Análisis de datos: LJRE, MFLV. Análisis formal: LJRE, DAOL, MFLV. Metodología: LJRE. Software: LJRE, MFLV. Validación: LJRE, DAOL, MFLV. Investigación: LJRE, DAOL, MFLV. Recursos: LJRE, DAOL. Preparación de escritura del artículo: LJRE, DAOL, MFLV. Escritura y edición: LJRE, DAOL, MFLV. Visualización: MFLV. Administración: LJRE. Todos los autores han leído y están de acuerdo en publicar esta versión del manuscrito.

## REFERENCIAS

Arroyo-Martínez, S., & Ortega-Ovalle, L. G. (2020). El impacto de los salarios en la tasa de desempleo en México del periodo 2000–2017 a través de un modelo estocástico. *Revista de investigación en ciencias contables y administrativas*, *6*(1), pp. 19–48. https://ideas.repec.org/a/msn/rijrnl/v6y2020i1p19-48.html

Barajas, A., & Ibarra, C. A. (2016). Modelos econométricos del desempleo en México: una revisión de la literatura. *Memorias del Congreso Nacional de Economía Política*.

Casal, R.F., Costa, J., & Oviedo, M. (2020). *Métodos de Aprendizaje estadístico*. https://rubenfcasal.github.io/aprendizaje_estadistico/aprendizaje_estadistico.pdf

Dinov, I. D. (2018). *Data science and predictive analytics*. Springer. https://doi.org/10.1007/978-3-319-72347-1

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, R. Li, & M. J. Murphy (Eds.), *Machine learning in radiation oncology: Theory and applications* (pp. 3-11). Springer International Publishing. https://doi.org/10.1007/978-3-319-18305-3_1

Encuesta nacional de ocupación y empleo (ENOE). (n.d.). *Instituto Nacional de Estadística y Geografía (INEGI)*. Retrieved February 17, 2024, https://www.inegi.org.mx/programas/enoe/

Escoto, A., Márquez, C., Prieto, V., Ochoa, S., & Reyes, P. (2017). Desempleo abierto y desalentado en tres mercados de trabajo latinoamericanos. *Población y mercados de trabajo en América Latina: temas emergentes*, 81–119. https://www.researchgate.net/profile/Ana-Escoto/publication/317389203_Desempleo_abierto_y_desalentado_en_tres_mercados_de_trabajo_latinoamericanos/links/60b169efa6fdcc1c66ebcc94/Desempleo-abierto-y-desalentado-en-tres-mercados-de-trabajo-latinoamericanos.pdf

Fernández-Delgado, M., Sirsat, M., Cernadas, E., Alawadi, S., Barro, S., & Febrero-Bande, M. (2019). An extensive experimental survey of regression methods. *Neural Networks*, *111*, pp. 11–34. https://doi.org/10.1016/j.neunet.2018.12.010

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press. https://www.deeplearningbook.org/

Heath, J. (2014). Unemployment in mexico revisited [Recuperado el 12 de septiembre de 2024]. https://jonathanheath.net/unemployment-in-mexico-revisited/

Hernández Pérez, J. (2020). "Desempleo en México por características sociodemográficas, 2005-2018". In: *Economía UNAM*, *17*(50), pp. 166–181. http://revistaeconomia.unam.mx/index.php/ecu/article/view/524

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wi- ley & Sons. https://doi.org/10.1002/9781118548387

Humphrey, D.D. (1940). Alleged "additional workers" in the measurement of unemployment. *Journal of Political Economy*, *48*(3), pp. 412–419. https://www.journals.uchicago.edu/doi/abs/10.1086/255613?journalCode=jpe

INEGI. (2023). Cómo se hace la ENOE: Métodos y procedimientos. *Encuesta Nacional de Ocupación y Empleo*. https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/ bvinegi/productos/nueva_estruc/702825190613.pdf

International Labour Organization. (2023). Resolution to amend the 19th icls resolution concerning statistics of work, employment and labour underutilization [Accessed: 2024-10-29 https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@stat/documents/normativeinstrument/wcms_230304.pdf

International Labour Organization. (2024). Estadísticas sobre las mujeres – ilostat [Accedido: 2024-06-27]. https://ilostat.ilo.org/es/topics/women/

Loh, W. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *1*(1), pp. 14–23. https://doi.org/10.1002/widm.8

Long, C. D. (1953). Impact of effective demand on the labor supply. *The American Economic Review*, *43*(2), pp. 458–467.

Long, C. D. (1958). *The labor force under changing income and employment*. NBER Books.

Maridueña-Larrea, Á., & Martín-Román, Á. (2024). The unemployment invariance hypothesis and the implications of added and discouraged worker effects in Latin America. *Latin American Economic Review*, *33*, pp. 1–25. https://doi.org/10.60758/laer.v33i.213

Mariscal, R., & Villarreal, E. (2017). Modelos econométricos del mercado laboral en México: un enfoque de series temporales. *Investigación Económica*, *76*(299), pp. 149–174.

Martín-Román, Á. L. (2022). Beyond the added-worker and the discouraged-worker effects: The entitled-worker effect. *Economic Modelling*, *110*, 105812. https://www.sciencedirect.com/science/article/pii/S026499932200058X

Molnar, C. (2020). Interpretable machine learning. https://christophm.github.io/interpretable-ml-book/index.html

Moy, V. (2020). Trabajadores desanimados y sin empleo. https://imco.org.mx/trabajadores-desanimados-y-sin-empleo/

Murguía Salas, V., Ronzón, Z. & Jardón A. (2023). Desaliento laboral. Una aproximación al análisis de la subutilización de la fuerza de trabajo juvenil de México. *Repositorio Institucional de la Universidad Autónoma del Estado de México*, pp. 145–162. http://hdl.handle.net/20.500.11799/140182

Ortiz Lazcano, D. A., & Rodríguez Esparza, L. J. (2023). Unemployment Vulnerability Index in Mexico: Effects of the covid-19 pandemic. *Economía, sociedad y territorio*, *23*(71), pp. 309–338. https://doi.org/10.22136/est20231862

N. 39, 2025 – ISSN: 1886-516X – DOI: 10.46661/rev.metodoscuant.econ.empresa.10760 – [Págs. 1-22]



Peón, F. V. (2021). Precariedad y desaliento laboral de los jóvenes en México. *Contraste Regional 9*(17), pp. 185–189. https://www.ciisder.mx/images/revista/contraste-regional-17/93_Precariedad_y_desaliento_laboral_de_los_jvenes_en_Mxico.pdf

Probst, P., Boulesteix, A., & Bischl, B. (2019). Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *Journal of Machine Learning Research 20*(53), pp. 1–32. http://jmlr.org/papers/v20/18-444.html.

Ruiz Nápoles, P., & Ordaz Díaz, J. L. (2011). Evolución reciente del empleo y el desempleo en México. *Economía UNAM 8*(23), pp. 91–105. https://www.scielo.org.mx/scielo.php?pid=S1665-952X2011000200005&script=sci_arttext

Sánchez-Salgado, R., & Alonso-Villarreal R. (2018). Análisis econométrico del desempleo en México: una aplicación de la metodología VAR. *Estudios Económicos 33*(1), pp. 27–56.

Scotti, M. C. M. (2015). Buscadores, desalentados y rechazados: Las dinámicas de inclusión y exclusión laboral enraizadas en la desocupación. *El Colegio de México*. https://repositorio.colmex.mx/concern/theses/w95050724?locale=es

Segovia Hernández, D. A. (2021). *Pronóstico del desempleo en México: aplicación de series de tiempo multivariadas*. Universidad Veracruzana. Facultad de Estadística e Informática. Región Xalapa. https://cdigital.uv.mx/server/api/core/bitstreams/0ae33d76-e58d-487e-a998-38851a467e45/content

Sen, J., Mehtab, S., Sen, R., Dutta, A., Kherwa, P., Ahmed, S., Berry, P., Khurana, S., Singh, S., Cadotte, D., Anderson, D., Ost, K., Akinbo, R., Daramola, O., & Lainjo, B. (2022, January). Machine learning: Algorithms, models, and applications. https://doi.org/10.48550/arXiv.2201.01943

Woytinsky, W. S. (1940). Additional workers on the labor market in depressions: A reply to Mr. Humphrey. *Journal of Political Economy*, *48*(5), pp. 735–739. https://doi.org/10.1086/255613