



**COLEGIO DE POSTGRADUADOS**

**INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN CIENCIAS AGRÍCOLAS**

**CAMPUS MONTECILLO**

**POSTGRADO EN SOCIOECONOMÍA, ESTADÍSTICA, E INFORMÁTICA**

**CÓMPUTO APLICADO**

**ALGORITMOS DE APRENDIZAJE AUTOMÁTICO  
PARA LA CLASIFICACIÓN DEL REZAGO SOCIAL A  
NIVEL MUNICIPAL EN MÉXICO**

**PABLO RODRIGO ÁVILA SOLÍS**

**T E S I S**

**PRESENTADA COMO REQUISITO PARCIAL  
PARA OBTENER EL GRADO DE:**

**MAESTRO EN CIENCIAS**

**MONTECILLO, TEXCOCO, ESTADO DE MÉXICO**

2021

La presente tesis titulada: **ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL REZAGO SOCIAL A NIVEL MUNICIPAL EN MÉXICO**, realizada por el alumno: **PABLO RODRIGO ÁVILA SOLÍS**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS  
SOCIOECONOMÍA ESTADÍSTICA E INFORMÁTICA  
CÓMPUTO APLICADO

CONSEJO PARTICULAR

CONSEJERO



---

**DR. JUAN MANUEL GONZÁLEZ CAMACHO**

ASESOR



---

**DR. PAULINO PÉREZ RODRÍGUEZ**

ASESOR



---

**DR. DELFINO VARGAS CHANES**

Montecillo, Texcoco, Estado de México, noviembre de 2021.

# ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DEL REZAGO SOCIAL A NIVEL MUNICIPAL EN MÉXICO

Pablo Rodrigo Ávila Solís, M.C.  
Colegio de Postgraduados, 2021

## RESUMEN

El grado de rezago social es una medida que permite estimar las carencias fundamentales de millones de personas en México, que además de asociarse con la pobreza, cuantifica un fenómeno distinto porque no toma en cuenta el ingreso. Este indicador se estima oficialmente, por el Consejo Nacional para la Evaluación de la Política de Desarrollo Social cada cinco años a nivel municipal; sin embargo, disponer de predicciones anuales es importante para mejorar el establecimiento de políticas de mitigación del rezago social. En este trabajo se propone un nuevo enfoque para predecir el grado de rezago social (alto, medio y bajo) a nivel municipal con base en algoritmos de inteligencia artificial (aprendizaje automático y aprendizaje profundo) e información anual de la estadística oficial de las unidades económicas de cada municipio. Con aprendizaje automático se implementaron, en el lenguaje de código abierto Python, tres clasificadores: regresión logística (LR), máquina de soporte vectorial (SVM), y bosques aleatorios (RF). Con aprendizaje profundo se implementaron tres arquitecturas de redes neuronales: red neuronal con escala nominal (DL-NOM); y dos redes neuronales con escala ordinal (DL1-ORD y DL2-ORD). Los modelos se entrenaron bajo diferentes escenarios y métodos de regularización, y se optimizaron con la técnica de validación cruzada y búsqueda de hiperparámetros por retícula. Para evaluar el desempeño de los modelos se calcularon las métricas precisión global de clasificación (*ACC*), *F1-macro*, y el error absoluto promedio (*MAE*). Los mejores resultados con aprendizaje automático supervisado se obtuvieron con el clasificador bosques aleatorios, con un valor de *F1-macro* 71.6% y *ACC* de 74.4%; así como valores *F1*, para las clases de rezago social alto, medio y, bajo de 59.6%, 73.0% y 82.2% respectivamente. Por otro lado, el mejor desempeño global para la predicción de las tres clases de rezago social se obtuvo con la arquitectura de redes neuronales DL1-ORD que presentó una *ACC* promedio de 76.2%, *MAE* promedio de 0.241 y *F1-macro* promedio de 73.8%; y valores *F1* para las clases objetivo de rezago social de 63.3% para alto, 73.8% para medio, y 84.6% para bajo. En este estudio se muestra la fuerte relación entre el grado de rezago social y las unidades económicas a nivel municipal. Asimismo, se ilustran en mapas la distribución geográfica del grado de rezago social en México. La metodología propuesta representa una valiosa herramienta para obtener estimaciones anuales del rezago social; en particular en periodos donde no existen estimaciones oficiales anuales.

**Palabras clave:** pobreza, redes neuronales, inteligencia artificial, clasificación supervisada.

# MACHINE LEARNING ALGORITHMS FOR THE CLASSIFICATION OF SOCIAL LAG AT MUNICIPAL LEVEL IN MEXICO

Pablo Rodrigo Ávila Solís, M.C.  
Colegio de Postgraduados, 2021

## ABSTRACT

The social lag index is a measure that allows estimating the primary deficiencies of millions of people in Mexico, which, in addition to being correlated with poverty, quantifies a different phenomenon because it does not take income into account. This indicator is officially estimated by the National Council for the Evaluation of Social Development Policy every five years at the municipal level. However, having annual predictions is important to improve the establishment of policies to mitigate the social lag. In this work, a new approach is proposed to predict the degree of social lag (high, medium and low) at the municipal level based on artificial intelligence algorithms (machine learning and deep learning) as well as annual information from the official statistics of economic units of each municipality. Using machine learning tools, three classifiers were implemented in the Python programming language: logistic regression (LR), support vector machine (SVM), and random forests (RF). With deep learning, three neural network architectures were implemented: neural network with nominal scale (DL-NOM); and two neural networks with ordinal scale (DL1-ORD and DL2-ORD). The models were trained under different scenarios and regularization methods, and were optimized with the cross-validation technique and hyperparameters grid-search. To evaluate the performance of the models, the global classification precision metrics (*ACC*), *F1-macro*, and the mean absolute error (*MAE*) were calculated. The best results with supervised machine learning algorithms were obtained with the random forests classifier, with a value of *F1-macro* 71.6% and *ACC* of 74.4%; as well as *F1* values, for the high, medium and low social lag classes of 59.6%, 73.0% and 82.2% respectively. On the other hand, the best global performance for the prediction of the three target classes of social lag was obtained with the DL1-ORD neural network architecture that presented an average *ACC* of 76.2%, average *MAE* of 0.241 and average *F1-macro* of 73.8%; and *F1* values for the target classes of social lag of 63.3% for high, 73.8% for medium, and 84.6% for low. This study shows the strong relationship between the degree of social lag and economic units at the municipal level. Also, the geographical distribution of the degree of social lag in Mexico is illustrated on maps. The proposed methodology represents a valuable tool to obtain annual estimates of social lag; particularly in periods where there are no official annual estimates.

**Keywords:** poverty, neural networks, artificial intelligence, supervised learning.

## **AGRADECIMIENTOS**

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo recibido para la elaboración de este trabajo durante la duración de mis estudios de maestría.

Al Colegio de Postgraduados, Campus Montecillo por abrirme las puertas y permitirme continuar con mi formación académica.

Al Posgrado de Socioeconómicas, Estadística e Informática – Computo Aplicado por el

Al Dr. Juan Manuel González Camacho por su tiempo, dedicación, valiosos consejos y toda su confianza, además de ser mi guía durante mis estudios de maestría y sobretodo en la realización de este trabajo.

A mis asesores Dr. Paulino Pérez Rodríguez y Dr. Delfino Vargas Chanes por su apoyo, comentarios puntuales y supervisión del desarrollo de esta tesis.

A mi esposa, Monserrat Marín Beltrán por ser mi compañera de vida; sin su apoyo, amor y comprensión este trabajo no hubiera sido posible.

A mi madre, María de los Ángeles Solís Acosta por el apoyo que siempre me ha brindado, así como su amor, cariño y confianza que me han permitido recorrer el camino para llegar hasta donde estoy.

A mi hija, Renata Mailén Ávila Marín por ser el motivo que día a día me impulsa a seguir adelante.

A toda mi familia, que siempre ha creído en mi y me han apoyado cuando lo he necesitado.

## **DEDICATORIA**

Este trabajo se lo dedico con mucho cariño a mi esposa, a mi hija y a mi madre que son mis pilares, mis fuentes de inspiración y mi motivación para nunca darme por vencido.

## CONTENIDO

RESUMEN .....	ii
ABSTRACT.....	iii
LISTA DE FIGURAS.....	viii
LISTA DE CUADROS.....	ix
INTRODUCCIÓN GENERAL .....	1
OBJETIVOS E HIPÓTESIS.....	3
Objetivo general.....	3
Objetivos específicos .....	3
Hipótesis.....	3
CAPÍTULO I PREDICCIÓN DEL REZAGO SOCIAL EN MÉXICO: UN ENFOQUE BASADO EN APRENDIZAJE AUTOMÁTICO A PARTIR DE DATOS DE UNIDADES ECONÓMICAS.....	4
1.1. RESUMEN.....	4
1.2. ABSTRACT.....	5
1.3. INTRODUCCIÓN .....	6
1.4. MATERIALES Y MÉTODOS .....	8
1.4.1. Colecta de datos.....	8
1.4.2. Procesamiento de la base de datos.....	8
1.4.3. Clasificadores de aprendizaje automático .....	9
1.4.3.1. Modelo de regresión logística.....	9
1.4.3.2. Máquina de soporte vectorial.....	10
1.4.3.3. Bosques aleatorios .....	11
1.4.4. Métricas de desempeño .....	12
1.4.5. Entrenamiento de los clasificadores .....	14
1.4.5.1. Escalamiento de variables.....	14
1.4.5.2. Selección de hiperparámetros .....	14
1.4.6. Evaluación del desempeño de los clasificadores.....	15
1.5. RESULTADOS Y DISCUSIÓN.....	16
1.5.1. Resultados de los clasificadores .....	16
1.5.2. Discusión.....	18

1.6. CONCLUSIONES .....	22
CAPÍTULO II EVALUACIÓN DE ARQUITECTURAS DE REDES NEURONALES DE APRENDIZAJE PROFUNDO PARA LA PREDICCIÓN ORDINAL DEL GRADO DE REZAGO SOCIAL EN MÉXICO .....	
2.1. RESUMEN.....	23
2.2. ABSTRACT.....	24
2.3. INTRODUCCIÓN .....	25
2.4. MATERIALES Y MÉTODOS .....	27
2.4.1. Conjuntos de datos .....	27
2.4.2. Hardware y software.....	28
2.4.3. Redes neuronales de aprendizaje profundo .....	29
2.4.3.1. Modelo de aprendizaje profundo con respuesta nominal .....	29
2.4.3.2. Modelos aprendizaje profundo con respuesta ordinal .....	30
2.4.4. Métricas de desempeño .....	31
2.4.5. Entrenamiento y selección de hiperparámetros.....	33
2.4.6. Comparación del desempeño de los clasificadores .....	34
2.5. RESULTADOS Y DISCUSIÓN.....	36
2.5.1. Resultados de entrenamiento.....	36
2.5.2. Resultados por clase objetivo .....	38
2.5.3. Comparación de los modelos ajustados.....	38
2.6. CONCLUSIONES .....	42
CÓDIGO DE PROGRAMACIÓN .....	43
CONCLUSIONES GENERALES.....	44
LITERATURA CITADA .....	45

## LISTA DE FIGURAS

### CAPÍTULO I

- Figura 1.1. (A) Curvas ROC y (B) curvas *Precision-Recall* del desempeño del clasificador bosques aleatorios (RF); P: precisión; S: sensibilidad; clases objetivo de rezago social (B), medio (M) y alto (A); TVP: tasa de verdaderos positivos; TFP: tasa de falsos positivos; AUC: área bajo la curva. .... 17
- Figura 1.2. Matriz de confusión normalizada para el modelo bosques aleatorios (RF), para las clases de rezago social bajo (B), medio (M), y alto (A). .... 18
- Figura 1.3. Comparación de las clases de rezago social para 2015; B: bajo; M: medio; A: alto; (A) propuestas por Vargas-Chanes y Valdés-Cruz versus (B) las predicciones realizadas con el clasificador bosques aleatorios. .... 19
- Figura 1.4. Distribución de las clases de rezago social; B: bajo; M: medio; A: alto; predichas con el clasificador bosques aleatorios, a partir de unidades económicas de 2016 (A) y 2017 (B), agregadas a nivel subsector. .... 19

### CAPÍTULO II

- Figura 2.1. Visualización de las clases de rezago social (B: bajo; M: medio; A: alto) aplicando la técnica de componentes principales, a partir de la matriz de covarianzas de los datos de entrada a nivel subsector. .... 28
- Figura 2.2. Arquitectura de una red neuronal artificial conectada hacia adelante con una capa oculta para la clasificación nominal de tres clases objetivo de rezago social. .... 29
- Figura 2.3. Entrenamiento por medio de parada anticipada para los modelos de DL-NOM, DL1-ORD y DL2-ORD. .... 36
- Figura 2.4. Comparación de resultados de la validación cruzada (CV), con 10 particiones, de las métricas de desempeño: *F1-macro*, precisión global (ACC) y error absoluto medio (MAE). .... 40
- Figura 2.5. Comparación de clases de rezago social; (A) propuestas por Vargas-Chanes y Valdés-Cruz (2019), y (B) predichas con el modelo de aprendizaje automático ordinal a nivel municipal de 2015. .... 41

## LISTA DE CUADROS

### CAPÍTULO I

Cuadro 1.1. Estructura del Sistema de Clasificación de América del Norte-México para identificar unidades económicas por actividad (INEGI, 2013).....	8
Cuadro 1.2. Matriz de confusión para el caso de un clasificador binario.....	12
Cuadro 1.3. Hiperparámetros e intervalos de valores de los clasificadores regresión logística (LR), máquina de soporte vectorial (SVM) y bosques aleatorios (RF). ....	15
Cuadro 1.4. Desempeño promedio de los clasificadores de regresión logística (LR), máquina de soporte vectorial (SVM) y bosques aleatorios (RF); con validación cruzada con 5 particiones. ....	16
Cuadro 1.5. Comparación de municipios por clase de rezago social; B: bajo; M: medio; A: alto; predichos con el modelo bosques aleatorios para 2016 y 2017 versus las clases de rezago de 2015 propuestas por Vargas-Chanes y Valdés-Cruz.....	20

### CAPÍTULO II

Cuadro 2.1. Matriz de confusión para el caso de un clasificador binario.....	32
Cuadro 2.2. Resultados del entrenamiento con regularización combinada: parada anticipada y L2, a través de validación cruzada con 5 particiones, para las métricas de desempeño <i>F1-macro</i> , error absoluto medio ( <i>MAE</i> ) y precisión global ( <i>ACC</i> ). ....	37
Cuadro 2. 3. Resultados por clase con regularización combinada: parada anticipada y L2, a través de validación cruzada con 10 particiones.....	38
Cuadro 2.4. P-valores de las pruebas de t por pares corregidas con método de Holm y pruebas de Wilcoxon, para las métricas de desempeño <i>F1-macro</i> , error absoluto medio ( <i>MAE</i> ) y precisión global ( <i>ACC</i> ).....	39

## **INTRODUCCIÓN GENERAL**

En este trabajo se presenta un enfoque metodológico para predecir el grado de rezago social a nivel municipal con base en dos paradigmas de aprendizaje de inteligencia artificial: aprendizaje automático y aprendizaje profundo; y el uso de información de la estadística oficial anual de las unidades económicas presentes en los municipios. Las clases o grados de rezago social se abordaron desde un enfoque nominal y un enfoque ordinal. La complejidad de los algoritmos de aprendizaje se incrementó con el objeto de mejorar la generalización de clasificadores.

El grado de rezago social determina la magnitud de cuatro indicadores sociodemográficos relacionados; que son el rezago educativo, el acceso a los servicios de salud, el acceso a los servicios básicos, y la calidad de la vivienda. La estimación periódica del grado de rezago social a nivel municipal proporciona un panorama de la distribución geográfica del rezago social que facilita la definición de políticas públicas para mitigar el rezago social alto, así como mejorar la gestión de los recursos federales a nivel de municipios.

En México, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) es el organismo responsable de realizar la estimación oficial quinquenal del grado de rezago social a diferentes niveles geográficos. Sin embargo, en los años donde no se dispone de la estimación oficial, no es posible conocer la situación de este fenómeno. Por esta razón, en este trabajo se propone una herramienta de predicción confiable con base en algoritmos de inteligencia artificial para cumplir con la asignación de recursos y apoyos federales a nivel municipal, en función del grado de rezago social, como se estipula en la Ley General de Desarrollo Social vigente de 2014 con base en la utilización de información de las estadísticas oficiales anuales de las unidades económicas activas en los municipios, como características de entrada de los modelos de aprendizaje. Estas unidades producen bienes y servicios, tanto en el sector público como en el privado, se colectan periódicamente por el Instituto Nacional de Estadística e Informática (INEGI), y se disponen públicamente por medio del Directorio Estadístico Nacional de Unidades Económicas (DENUE). Cabe destacar que, los escenarios de entrenamiento se realizaron de acuerdo con la categorización del tipo de actividad económica asociada a las unidades económicas pertenecientes al DENUE. El Sistema de Clasificación Industrial de América del Norte (SCIAN) estructura jerárquicamente las unidades económicas en: sectores, subsectores, ramas, subramas y clases.

En el capítulo I se presentan tres algoritmos con base en el aprendizaje automático supervisado para predecir la clase o grado de rezago social con una escala nominal. Estos son: Regresión logística (LR), máquina de soporte vectorial (SVM) y bosques aleatorios (RF). Para este propósito, se consideraron las tres clases de rezago social propuestas por Vargas-Chanes y Valdés-Cruz (2019): bajo, medio y alto. Se revisó el estado del arte sobre las aplicaciones de los algoritmos de aprendizaje automático y aprendizaje profundo para modelar variables relacionadas con la pobreza, y evaluar el desempeño con métricas ad hoc para clases no balanceadas (Alsharkawi *et al.*, 2021; Sani *et al.*, 2018). Igualmente, se consideró el trabajo de Powell *et al.* (2007), donde demostraron que la distribución de las unidades económicas influye en la caracterización sociodemográfica del área donde se encuentran. Asimismo, se presentan mapas de predicción con el modelo bosques aleatorios optimizado para describir la distribución geográfica en México, de los grados de rezago social a nivel municipal en 2016 y 2017.

En el capítulo II se evaluaron tres arquitecturas de redes neuronales de aprendizaje profundo para predicción del grado de rezago social: una con enfoque nominal y dos con enfoque ordinal. Estos tipos de arquitecturas se han empleado con éxito para predecir variables binarias asociadas con la pobreza de los hogares en Filipinas e Indonesia (Azcárraga y Setiono, 2018; Poerwanto y Fajriani, 2020). Se utilizó la información de unidades económicas del estudio previo descrito en el capítulo I, agrupadas a nivel subsector. Finalmente, se presentó un ejemplo de aplicación del clasificador que obtuvo los mejores resultados, para representar en un mapa la distribución geográfica a nivel municipal de los grados de rezago social en 2015.

Con el propósito de implementar modelos de aprendizaje automático y aprendizaje profundo, para predecir el grado de rezago social en México, se plantearon los siguientes objetivos e hipótesis:

## **OBJETIVOS E HIPÓTESIS**

### **Objetivo general**

Desarrollar y probar un nuevo enfoque de estimación del grado de rezago social a nivel municipal con base en los paradigmas de aprendizaje automático y aprendizaje profundo, para predecir tres clases o grados de rezago social en México (bajo, medio y alto); a partir de datos anuales de la estadística oficial de las unidades económicas activas en los municipios.

### **Objetivos específicos**

Implementar y evaluar el desempeño en predicción de tres clasificadores de aprendizaje automático: máquina de soporte vectorial, regresión logística y bosques aleatorios para predecir con enfoque nominal tres grados de rezago social a partir de datos de unidades económicas agregadas por municipio.

Implementar y evaluar el desempeño en predicción de tres arquitecturas de aprendizaje profundo, dos con enfoque ordinal y una con enfoque nominal, para mejorar la predicción del grado de rezago social en México a nivel municipal, a partir conteos de unidades económicas agrupadas por subsector.

Elaborar mapas de la distribución geográfica en México de las predicciones de los grados de rezago social a nivel municipal con la metodología propuesta.

### **Hipótesis**

Es posible estimar el grado de rezago social en México, por medio de algoritmos de aprendizaje automático y aprendizaje profundo, a partir de datos de unidades económicas activas a nivel municipal.

# CAPÍTULO I PREDICCIÓN DEL REZAGO SOCIAL EN MÉXICO: UN ENFOQUE BASADO EN APRENDIZAJE AUTOMÁTICO A PARTIR DE DATOS DE UNIDADES ECONÓMICAS.

## 1.1. RESUMEN

El rezago social en México se estima oficialmente a nivel municipal, cada 5 años, por el Consejo Nacional para la Evaluación de la Política de Desarrollo Social (CONEVAL). Sin embargo, es recomendable contar con predicciones anuales para definir políticas públicas apropiadas. En este trabajo se presenta un enfoque basado en aprendizaje automático para predecir las clases o grados de rezago social (alto, medio, bajo) a nivel municipal en México, a partir de la información de unidades económicas, proveniente del Directorio Estadístico Nacional de Unidades Económicas (DENUE) del año 2015. Se implementaron tres clasificadores de aprendizaje automático supervisado: regresión logística, máquina de soporte vectorial y bosques aleatorios. Con base en la información de conteos de las unidades económicas, en sus diferentes categorías, población y coordenadas geográficas de los municipios, se definieron los escenarios para el entrenamiento y prueba de los modelos propuestos. Los criterios para evaluar el desempeño de los clasificadores fueron el valor de *F1-macro*, la precisión global y el área bajo la curva ROC. Los resultados indicaron que el mejor desempeño global se obtuvo con el clasificador bosques aleatorios, con un valor de *F1-macro* 71.6% y una precisión global 74.4%; así como valores *F1*, para las clases de rezago social alto, medio y, bajo de 59.6%, 73.0% y 82.2% respectivamente. Estos resultados muestran una relación entre las clases de rezago social a nivel municipal y las tasas de categorías de unidades económicas agregadas a nivel subsector, de acuerdo con el Sistema de Clasificación Industrial de América del Norte (SCIAN). Las predicciones de rezago social con las clases utilizadas se contrastaron con datos reportados en la literatura de 2015.

**Palabras clave:** pobreza, clasificación supervisada, árboles de decisión, bosques aleatorios.

## 1.2. ABSTRACT

Social lag in Mexico is officially estimated at the municipal level, every 5 years, by the National Council for the Evaluation of Social Development Policy (CONEVAL). However, it is highly desirable to have annual predictions to define proper public policies. This paper presents an approach based on machine learning algorithms to predict the classes or degrees of social lag (high, medium, low) at the municipal level in Mexico, based on information of economic units, from the National Statistical Directory of Economic Units. (DENUE) of the year 2015. Three supervised machine learning classifiers were implemented: logistic regression, support vector machines and random forests. Based on the information of the contents of the economic units, in their different categories, population and geographical coordinates of the municipalities, the scenarios for the training and testing of the proposed models were defined. The criteria to evaluate the performance of the classifiers were the F1-macro value, the accuracy and the area under the ROC curve. The results indicated that the best global performance was obtained with the random forests classifier, with a F1-macro value of 71.6% and an accuracy of 74.4%; as well as F1 values, for the high, medium and low social lag classes of 59.6%, 73.0% and 82.2% respectively. These results show a relationship among the classes of social lag at the municipal level and the rates of categories of economic units aggregated at the subsector level, according to the North American Industrial Classification System (SCIAN). The predictions of social lag with the labels used were contrasted with data reported in the 2015 literature.

**Keywords:** poverty, supervised learning, decision trees, random forests.

### 1.3. INTRODUCCIÓN

En México, de acuerdo con la Ley General de Desarrollo Social vigente de 2004, los recursos y apoyos federales se distribuyen a nivel de municipios y se asignan en función del grado de rezago social. Por ello es importante conocer periódicamente esta variable para identificar zonas de alto rezago. El rezago social, a diferencia de la pobreza, es un indicador socioeconómico de carencias sociales. CONEVAL (2016) reportó el agrupamiento del rezago social en cinco clases: muy bajo, bajo, medio, alto y muy alto, con base en un análisis de componentes principales, la técnica de estratificación propuesta por Dallenius y Hodges (1959) e información de la encuesta intercensal 2015. Esta encuesta engloba 11 variables sociodemográficas relativas a educación, salud, servicios básicos, activos en el hogar y calidad de la vivienda.

La estimación del rezago social se realiza cada cinco años de manera oficial, debido a la disponibilidad de las variables socioeconómicas. Vargas-Chanes y Valdés-Cruz (2019) propusieron un método alternativo para estimar el rezago social a nivel municipal, definieron tres clases (bajo, medio y alto), con base en un análisis de clases latentes y datos de la encuesta intercensal 2015 reportada por CONEVAL (2016), para dar mayor soporte estadístico a la estimación de rezago social a nivel municipal.

En la literatura se reportan métodos indirectos para estimar indicadores socioeconómicos relacionados con la pobreza. Engstrom *et al.* (2017) realizaron una propuesta con un modelo de regresión múltiple con penalización L1 para estimar las tasas de pobreza y consumo promedio de 1291 municipios en Sri Lanka a partir de imágenes satelitales. Estos autores encontraron que las características obtenidas de las imágenes explican alrededor de 60% de la variación de la variable respuesta estudiada. Alsharkawi *et al.* (2021) aplicaron 16 modelos de aprendizaje supervisado, con base en diferentes encuestas nacionales de hogares, para clasificar la situación de pobreza de hogares en Jordania. Entre los cuales destacan regresión logística, potenciación del gradiente, bosques aleatorios y máquina de soporte vectorial. Estos autores reportaron que el clasificador potenciación del gradiente, obtuvo 81% de desempeño con la métrica F1 con un conjunto de datos desbalanceado. Sani *et al.* (2018) utilizaron tres clasificadores para categorizar el ingreso mensual de los hogares de tres estados en Malasia: k-vecinos más cercanos, árboles de decisión y Bayesiano empírico. El clasificador árboles de decisión fue el más significativo para detectar si los hogares estaban dentro del 40% más pobre o fuera de éste, con clases desbalanceadas.

Powell *et al.* (2007) estudiaron la disponibilidad de unidades económicas en Estados Unidos; tales como tiendas de alimentos agregadas por código postal. Estos autores aplicaron el análisis de regresión multivariado para asociar las características de la ubicación de vecindarios, a partir de datos censales del año 2000; e indicaron que los supermercados y tiendas de abarrotes se localizan en vecindarios con ingresos bajos. Esto muestra la posibilidad de calcular indicadores socioeconómicos de una región geográfica a partir de unidades económicas.

El objetivo de este capítulo es implementar tres modelos aprendizaje automático: regresión logística, máquina de soporte vectorial y bosques aleatorios para predecir tres clases de rezago social (bajo, medio y alto) a nivel municipal, a partir de datos de las unidades económicas de 2015, población y coordenadas geográficas de los municipios. Así como, comparar estas predicciones con las obtenidas para 2015 por Vargas-Chanes y Valdés-Cruz (2019) para realizar, con base en el enfoque propuesto, predicciones subsecuentes de rezago social en 2016 y 2017.

## 1.4. MATERIALES Y MÉTODOS

### 1.4.1. Colecta de datos

Para realizar el estudio se obtuvieron datos de las unidades económicas (UE), aproximadamente cinco millones de registros, procedentes de la quinta edición del Directorio Estadístico Nacional de Unidades Económicas - DENUE (INEGI, 2015). INEGI realiza el DENUE desde 2010 y a partir de 2015 se ha actualizado al menos una vez al año. Las UE se jerarquizaron conforme al Sistema de Clasificación Industrial de América del Norte (SCIAN). INEGI (2013) describió esta clasificación en cinco niveles: sector, subsector, rama, subrama y clase (Cuadro 1.1). Un sector consta de subsectores, un subsector se divide en ramas, una rama deriva en subramas y una subrama se descompone en clases. Cada UE tiene los siguientes atributos: identificador único, código de actividad SCIAN, clave de entidad y clave de municipio. La localización geográfica y la población municipal se obtuvo del marco geoestadístico nacional (INEGI, 2014).

**Cuadro 1.1. Estructura del Sistema de Clasificación de América del Norte-México para identificar unidades económicas por actividad (INEGI, 2013).**

Dígitos por actividad	Nivel de agregación	Categorías
2	Sector	20
3	Subsector	94
4	Rama	303
5	Subrama	614
6	Clase	1059

Como clases objetivo se usaron las tipologías de rezago social a nivel municipal, bajo (B), medio (M) y alto (A) consideradas en la base de datos descrita por Vargas-Chanes y Valdés-Cruz (2019). En las fuentes de información se utilizó la clave de municipio y se identificaron 2457 municipios; de los cuales 971 (39.51%) corresponden a la clase B, 1064 (43.30%) a la clase M, y 422 (17.17%) a la clase A; por lo tanto, se contó con clases objetivo no balanceadas.

### 1.4.2. Procesamiento de la base de datos

La recolección, el procesamiento y la integración de la base de datos se realizó en el lenguaje de programación Python, v. 3.8. Los modelos de aprendizaje automático se implementaron en la plataforma Scikit-learn v. 0.23.2 (Pedregosa *et al.*, 2011). Para ejecutar los procesos se utilizó un

equipo de cómputo MacBook Air 2017 con sistema operativo macOS 10.14 Mojave de 64 bits, procesador Intel Core i5 a 1.8 GHz y 8 GB de memoria RAM.

Con los datos de las UE de 2015 y el marco geoestadístico nacional de 2014, se calculó para cada nivel de agregación y cada municipio, la tasa de categorías de UE municipales ( $TUE$ ) definida como:

$$TUE_{c,m} = \frac{UE_{c,m}}{P_m} \times 1000,$$

donde  $TUE_{c,m}$  es la tasa de UE para la  $c$ -ésima categoría (del nivel  $k$  de agregación del SCIAN) y  $m$ -ésimo municipio, por mil habitantes;  $c = 1, 2, \dots, nk$ ;  $nk = (20, 94, 303, 614, 1059)$ ;  $m = 1, 2, \dots, 2457$ ;  $UE_{c,m}$  es el conteo de UE para  $c$  y  $m$  específicos, en cada nivel  $k$  y;  $P_m$  es la población del municipio.

Cinco escenarios de entrada de datos se definieron: sector, subsector, rama, subrama y clase, de acuerdo con las cinco jerarquías definidas por el SCIAN. Las variables  $TUE$  se calcularon para las categorías económicas definidas en cada escenario y se agregaron dos variables geográficas, latitud y longitud de cada municipio. De este modo, el primer escenario de entrada de datos consistió en 22 variables de entrada ( $20 TUE + 2$ ) para el nivel sector, 96 para subsector, 305 para rama, 616 para subrama; y 1061 variables para clase.

### 1.4.3. Clasificadores de aprendizaje automático

Tres modelos de aprendizaje automático se implementaron, regresión logística (LR, logistic regression); máquina de soporte vectorial (SVM, support vector machine); y bosques aleatorios (RF, random forests) para los diferentes escenarios de entrada de datos.

#### 1.4.3.1. Modelo de regresión logística

LR es un clasificador binario lineal, que se basa en la teoría estadística de los modelos lineales generalizados y utiliza una función de liga logit (McCullagh y Nelder, 1989). La función de liga logit se define como:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right),$$

donde  $p$  es la probabilidad del evento de interés.

LR se modela como la combinación lineal de los parámetros y las características de entrada, esto es:

$$\text{logit}(p^{(i)}) = \sum_{j=0}^m w_j x_j^{(i)} = z^{(i)},$$

donde  $w_j$  es el  $j$ -ésimo parámetro y  $x_j^{(i)}$  es la  $j$ -ésima característica de la  $i$ -ésima muestra del conjunto de datos de entrenamiento.

En este trabajo se consideró el modelo LR con un término de regularización L2 (James *et al.*, 2013) para evitar el sobreajuste del modelo con los datos de entrenamiento; el cual consiste en optimizar la función de pérdida  $J(\mathbf{w})_{LR}$  expresada por:

$$J(\mathbf{w})_{LR} = \sum_{i=1}^n \left[ -y^{(i)} \log(\phi(z^{(i)})) - (1 - y^{(i)}) \log(1 - \phi(z^{(i)})) \right] + \frac{\lambda}{2} \sum_{j=1}^m w_j^2,$$

donde  $\phi(z) = \frac{1}{1+e^{-z}}$  es la función de activación sigmoide;  $w_j$  el  $j$ -ésimo parámetro, nótese que  $w_0$  no se penaliza;  $y^{(i)}$  es la clase objetivo de la  $i$ -ésima muestra, que considera los valores 0 y 1;  $\lambda$  es el hiperparámetro de regularización L2;  $n$  es el número de muestras u observaciones; y  $m$  es el número de parámetros. Este modelo se generaliza para clases múltiples, por medio de los enfoques de clasificación múltiple: uno *versus* el resto y uno *versus* uno, combinando clasificadores binarios.

Otro enfoque para la clasificación clases múltiples con este modelo se aborda por medio de la función *softmax*; la cual se define como,

$$p(y = k|z) = \phi(z) = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}},$$

donde  $K$  representa el número de clases objetivo;  $k$  es una clase, en particular, para calcular la función *softmax*; y  $z_k$  es la combinación lineal de una clase, en particular, con su propio vector de pesos. *Softmax* obtiene las probabilidades de pertenencia para cada clase y la probabilidad máxima se utiliza para realizar una predicción. La función de costo utilizada para RL es la función de entropía cruzada, la cual se define como,

$$J(\mathbf{w}) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(p(y = k|z)),$$

donde  $y_k^{(i)}$  es la probabilidad de que la muestra  $i$  pertenezca a la clase  $k$ ;  $m$  es el número de muestras; y  $K$  el número de clases objetivo. En particular, si  $K = 2$ , se obtiene la función de costo del modelo RL binario.

#### 1.4.3.2. Máquina de soporte vectorial

El clasificador SVM consiste en maximizar el margen del hiperplano que separa dos clases. El margen se define como la distancia entre el hiperplano de separación y las muestras de

entrenamiento más cercanas a éste. SVM se utilizó bajo el enfoque de clasificación multiclase uno *versus* el resto, para las tres clases de rezago social consideradas en este estudio. SVM con regularización L2 para clasificación margen-flexible, consistió en minimizar la función de pérdida  $J(\mathbf{w})_{SVM}$  (Wang *et al.*, 2006), la cual se expresa por,

$$J(\mathbf{w})_{SVM} = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{2} \sum_{i=1}^m \xi^{(i)2},$$

$$t^{(i)}(\mathbf{w}^T \mathbf{x}^{(i)} + b) \geq 1 - \xi^{(i)},$$

donde  $\mathbf{w}$  es el vector de pesos o parámetros;  $b$  es el término de sesgo;  $t^{(i)}$  es la clase objetivo de la  $i$ -ésima muestra, que considera los valores  $-1$  y  $1$ ;  $C$  es un hiperparámetro que controla el ancho del margen del hiperplano y, por ende, del término de regularización L2; y  $\xi$  es una variable de holgura que introdujeron Cortes y Vapnik (1995), para tratar con datos no linealmente separables.

En una clasificación binaria se tienen dos casos  $w_0 + \mathbf{w}^T \mathbf{x}^{(i)} \geq 1 - \xi^{(i)}$  para  $t^{(i)} = 1$  y  $w_0 + \mathbf{w}^T \mathbf{x}^{(i)} \leq -1 + \xi^{(i)}$  para  $t^{(i)} = -1$ , donde  $\mathbf{x}^{(i)}$  es la  $i$ -ésima muestra del conjunto de datos de entrenamiento.

El uso de funciones kernel (Bishop, 2006) facilita la solución de problemas de clasificación no lineal, por medio de combinaciones lineales de las variables o características de entrada proyectadas en un espacio de dimensión más alta. Los kernels utilizados son lineal,  $\kappa(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \mathbf{x}^{(i)T} \mathbf{x}^{(j)}$ ; función de base radial (rbf),  $\kappa(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = e^{-\gamma \|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|^2}$ ; y sigmoide,  $\kappa(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \tanh(\gamma \cdot \mathbf{x}^{(i)T} \mathbf{x}^{(j)} + r)$ ; donde  $\gamma$  y  $r$  son hiperparámetros que se optimizan por medio de una retícula con intervalos de valores discretos.

#### 1.4.3.3. Bosques aleatorios

El clasificador RF propuesto por Breiman (2001) es un método de ensamble que utiliza combinaciones de un número determinado de árboles de decisión. Estos se toman de una muestra aleatoria con remplazo y son entrenados a partir de un número determinado de características de entrada tomadas de una muestra sin remplazo. Después de que se generan los árboles de decisión se utiliza el voto mayoritario para predecir una clase. Este enfoque busca minimizar la entropía o la impureza de Gini de los árboles de decisión del modelo. La medida de entropía  $I_E(q)$  para las clases no vacías se define como,

$$I_E(q) = - \sum_{k=1}^c p(k|q) \log_2(p(k|q)),$$

donde  $p(k|q)$  es la proporción de observaciones de un nodo  $q$ , que pertenecen a la clase  $k$ , de forma que  $I_E(q) = 0$ , si todos los ejemplos del nodo pertenecen a la misma clase; y  $I_E(q) = 1$  si las observaciones se distribuyen en las clases de forma uniforme.

El criterio de impureza de Gini  $I_G(q)$ , se define como,

$$I_G(q) = -\sum_{k=1}^K p(k|q)(1 - p(k|q)) = 1 - \sum_{k=1}^K p(k|q)^2,$$

donde  $I_G(q)$  alcanza su valor máximo cuando las muestras tienen una distribución uniforme dentro de las clases (Raschka y Mirjalili, 2019).

Los hiperparámetros del modelo RF que se optimizan por medio de una búsqueda por retícula son el número de estimadores (árboles de decisión) del modelo ( $NE$ ); el criterio de calidad para dividir los nodos ( $CR$ ); la profundidad máxima de los árboles de decisión ( $MP$ ) que controla la complejidad del modelo; y el número máximo de características o variables de entrada ( $MC$ ) que se eligen al azar para dividir los nodos.

#### 1.4.4. Métricas de desempeño

Para evaluar el desempeño de los clasificadores se utilizaron las siguientes métricas; precisión global ( $ACC$ , accuracy), precisión ( $P$ ), sensibilidad ( $S$ ), valor  $F1$ , valor  $F1-macro$ , y áreas bajo la curva ( $AUC$ , area under the curve) de la curva ROC y la curva  $P-S$ ; las cuales se originan a partir de la matriz de confusión.

La matriz de confusión describe las predicciones de clase (representadas por columnas) *versus* las clases verdaderas (representadas por hileras). Para el caso binario (Cuadro 1.2) esta matriz reporta el número de verdaderos positivos ( $VP$ ), observaciones positivas, predichas como positivas; el número de falsos positivos ( $FP$ ), observaciones negativas predichas como positivas; el número de falsos negativos ( $FN$ ), observaciones positivas predichas como negativas; y el número de verdaderos negativos ( $VN$ ), observaciones negativas predichas como negativas.

**Cuadro 1.2. Matriz de confusión para el caso de un clasificador binario.**

		Clases predichas	
		Positivo	Negativo
Clases verdaderas	Positivo	$VP$ : verdadero positivo	$FN$ : falso negativo
	Negativo	$FP$ : falso positivo	$VN$ : verdaero negativo

De esta manera,  $ACC$  se define por,

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

$P$  es la fracción de  $VP$  con respecto al total de positivos predichos y se calcula por,

$$P = \frac{VP}{VP+FP}.$$

$S$  es la fracción de  $VP$  con respecto al total de positivos observados y se define como,

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

$F1$  representa la media armónica de  $P$  y  $S$ , esto es:

$$F1 = \frac{2P \times S}{P+S}.$$

donde  $F1$  toma valores entre 0 y 1; 0 indica un ajuste pobre y 1 un ajuste perfecto. La métrica  $F1$  favorece a los clasificadores con  $P$  y  $S$  similares (Géron, 2019).

$F1$ -macro, en el caso de un problema multiclase, se calcula como el promedio aritmético de las medidas  $F1$  obtenidas para cada clase:

$$F1\text{-macro} = \frac{1}{K} \sum_i^N F1_i,$$

donde  $F1_i$  es el valor de  $F1$  de la  $i$ -ésima clase objetivo; y  $K$  es el número total de clases. Esta métrica es apropiada cuando se tienen clases desbalanceadas; cada clase tiene el mismo peso y se hace énfasis en las clases con menor frecuencia (Lipton *et al.*, 2014).

La curva ROC asocia la tasa de  $VP$  en el eje vertical *versus* la tasa de  $FP$  en el eje horizontal. Para comparar el desempeño de los clasificadores se utiliza  $AUC$ , que varía entre 0 y 1, valores cercanos a 1 indican un buen desempeño del clasificador (Fawcett, 2006). En el caso de una clasificación binaria se tiene una curva ROC por modelo, y en el caso multiclase se tiene una curva ROC para cada clase del modelo. La curva  $P$ - $S$  representa la gráfica de valores de  $S$  en el eje horizontal *versus* valores de  $P$  en el eje vertical. El valor  $AUC$  de la curva  $P$ - $S$  toma valores entre 0 y 1, valores próximos a 1 indican un buen desempeño del modelo. En el caso binario, cuando los datos no están balanceados, la curva  $P$ - $S$  es más apropiada para comparar el desempeño de los clasificadores (Saito y Rehmsmeier, 2017).

## 1.4.5. Entrenamiento de los clasificadores

### 1.4.5.1. Escalamiento de variables

El conjunto de datos (o características) de entrada se estandarizó antes de realizar el entrenamiento de los clasificadores LR, SVM y RF, con base en la expresión,

$$x'_j = \frac{x_j - \bar{x}_j}{s_j},$$

donde  $x'_j$  es un vector de observaciones de la  $j$ -ésima característica de entrada, del conjunto de entrenamiento de  $n$  observaciones;  $\bar{x}_j$  es la media muestral de  $x'_j$ ; y  $s_j$  es la desviación estándar muestral.

Lo anterior permite mejorar el desempeño de los algoritmos de optimización. Cabe resaltar que los parámetros de escala  $\bar{x}_j$  y  $s_j$  para cada característica  $j$ , se aplicaron posteriormente para transformar las características del conjunto de prueba y evaluar el desempeño de los clasificadores.

### 1.4.5.2. Selección de hiperparámetros

La selección de hiperparámetros óptimos para cada clasificador se realizó con una búsqueda en retícula y validación cruzada (CV) con  $k = 10$  particiones. Una partición aleatoria se utilizó con 1965 observaciones, 80 % del conjunto de datos de entrada. Para cada clasificador, se seleccionó la combinación de hiperparámetros que obtuvo el valor promedio máximo del criterio *F1-macro*, de las combinaciones posibles en cada retícula de hiperparámetros.

Los hiperparámetros e intervalos de valores analizados para cada clasificador se describen en el Cuadro 1.3. En particular, LR se entrenó con el inverso del hiperparámetro de regularización L2, con dos estrategias de clasificación: uno *versus* el resto y multinomial; y se optimizó con el algoritmo L-BFGS-B (Morales y Nocedal, 2011).

**Cuadro 1.3. Hiperparámetros e intervalos de valores de los clasificadores regresión logística (LR), máquina de soporte vectorial (SVM) y bosques aleatorios (RF).**

Modelo	Hiperparámetro	Rango
LR	$\lambda$	0.001, 0.01, 0.1, 1, 10, 100, 1000 y 10000
	método	uno contra el resto y multinomial
SVM	Kernels	Lineal, sigmoide y rbf
	C	0.0001, 0.001, 0.01, 0.1, 1, 10, 100 y 1000
	$\gamma$	0.0001, 0.001, 0.01, 0.1, 1, 10, 100 y 1000 (para kernels sigmoide y rbf)
	r	0, 1, 2,3, 4, y 5 (para el kernel sigmoide)
RF	<i>NE</i>	100, 150, 200, 250, 300, 350 y 400
	<i>CR</i>	Entropía y Gini
	<i>MC</i>	Raíz cuadrada y logaritmo base 2
	<i>MP</i>	10, 15, 20, 25 y 30

*NE*: número de estimadores; *CR*: criterio de división; *MC*: número máximo de características; *MP* profundidad máxima. Los intervalos de valores se seleccionaron en forma heurística y experimental.

#### 1.4.6. Evaluación del desempeño de los clasificadores

La evaluación del desempeño final de los clasificadores LR, SVM y RF se realizó con los conjuntos de datos completos (2457 observaciones) y con el conjunto de variables o características de entrada retenido. Una CV se realizó con  $k = 5$  particiones aleatorias, estratificadas por las tres clases de rezago social, para obtener las métricas de desempeño promedio. Los mismos escenarios de entrada de datos (categorías del SCIAN) se utilizaron y particiones aleatorias de los datos para evaluar la capacidad predictiva de los clasificadores y evitar sesgos en la comparación de los modelos.

## 1.5. RESULTADOS Y DISCUSIÓN

### 1.5.1. Resultados de los clasificadores

Los escenarios de entrada de datos e hiperparámetros optimizados de los clasificadores fueron: LR (a nivel subsector,  $\lambda = 10$ ); SVM (a nivel subsector, kernel sigmoide,  $C = 1000$ ,  $r = 0$  y  $\gamma = 0.0001$ ); y RF (a nivel subsector, entropía  $CR$ ,  $MP = 20$ ,  $MC = \sqrt{96}$  y  $NE = 250$ ). De acuerdo con los hiperparámetros de regularización L2 optimizados, LR requirió mayor regularización que SVM. De los cinco escenarios de entrada de datos analizados, el nivel de agregación subsector del SCIAN, obtuvo el mejor desempeño en predicción en los tres modelos.

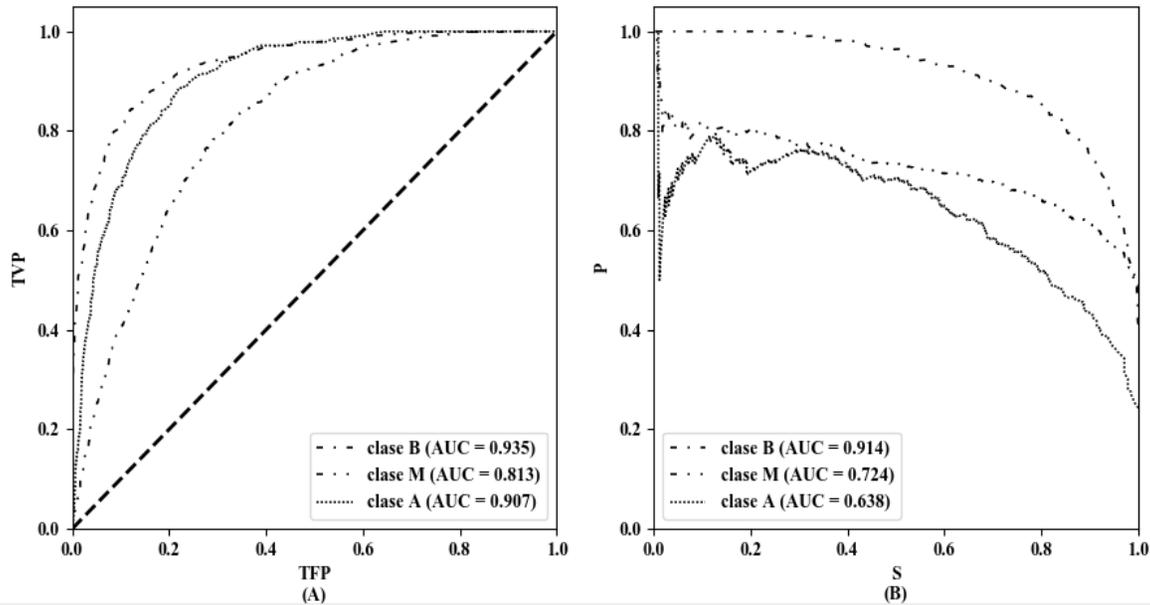
RF fue superior a LR y SVM, con  $F1$ -macro y  $ACC$ . RF fue superior a LR y SVM, para  $P$  en clases B y A; para  $S$  en clase M y para  $F1$  en clases B y M. LR fue superior a SVM y RF, para  $P$  en clase M y para  $S$  en clase B. SVM fue superior a LR y RF, para  $S$  y  $F1$  en clase A. Además, se observó que los tres modelos de clasificación se desempeñaron mejor para identificar la clase B (Cuadro 1.4). Para la clase de rezago social A, RF minimiza  $FP$  ( $P = 0.668$ ), es decir, RF predijo la clase A con mejor precisión. Por su parte, SVM minimiza  $FN$  ( $S = 0.616$ ), es decir, SVM detectó la mayor cantidad de clases A verdaderas.

**Cuadro 1.4. Desempeño promedio de los clasificadores de regresión logística (LR), máquina de soporte vectorial (SVM) y bosques aleatorios (RF); con validación cruzada con 5 particiones.**

Modelo	$F1$ -macro	$ACC$	Clase	$P$	$S$	$F1$
LR	0.709 +/- 0.028	0.732 +/- 0.016	B	0.839	0.786	0.812
			M	0.687	0.737	0.711
			A	0.617	0.595	0.606
SVM	0.708 +/- 0.024	0.728 +/- 0.016	B	0.836	0.777	0.805
			M	0.685	0.727	0.706
			A	0.612	0.616	0.614
RF	0.716 +/- 0.021	0.744 +/- 0.010	B	0.864	0.785	0.822
			M	0.679	0.789	0.730
			A	0.668	0.538	0.596

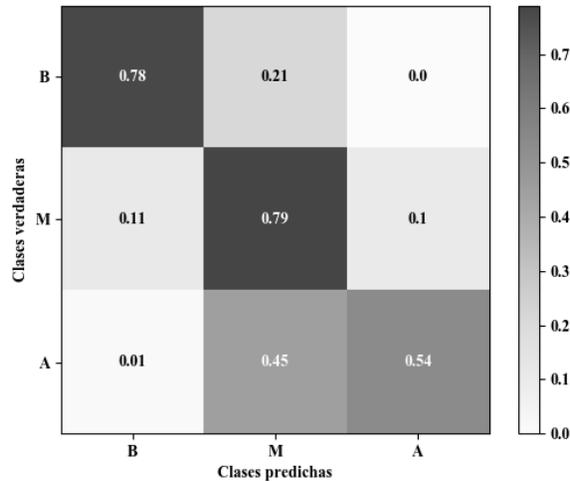
$ACC$ : precisión global; clases objetivo de rezago social: bajo (B), medio (M) y alto (A);  $P$ : precisión;  $S$ : sensibilidad.

Las curvas ROC y curvas  $P$ - $S$  para cada clase objetivo del clasificador RF se describen en la Figura 1.1. El valor máximo de  $AUC$  se relacionó con la clase B y el menor relacionó con la clase M, lo cual contradice los valores de  $F1$  para cada clase. La visión optimista de la curva ROC se atribuye a la distribución no uniforme de las tres clases de rezago social, para evaluar el desempeño de los clasificadores en este tipo de situaciones, existen alternativas más robustas; como la curva ROC concentrada, la curva de costos y la curva  $P$ - $S$ ; esta última opción se considera la más adecuada (Saito y Rehmsmeier, 2015).



**Figura 1.1. (A) Curvas ROC y (B) curvas  $P$ - $S$  del desempeño del clasificador bosques aleatorios (RF); P: precisión; S: sensibilidad; clases objetivo de rezago social (B), medio (M) y alto (A); TVP: tasa de verdaderos positivos; TFP: tasa de falsos positivos; AUC: área bajo la curva.**

Otra forma de medir el comportamiento del modelo RF fue a través de la matriz de confusión normalizada, que describe la proporción de clases verdaderas dentro cada clase predicha. (Figura 1.2). Además, los resultados de la matriz de confusión normalizada muestran los porcentajes de observaciones que pertenecen correctamente a la clase predicha; las cuales son: 78%, 79% y 54%, para las clases B, M y A, respectivamente. Las clases M y B tuvieron mejor predicción que A; además, muestras de la clase M se confunden con la clase A, que son clases contiguas.

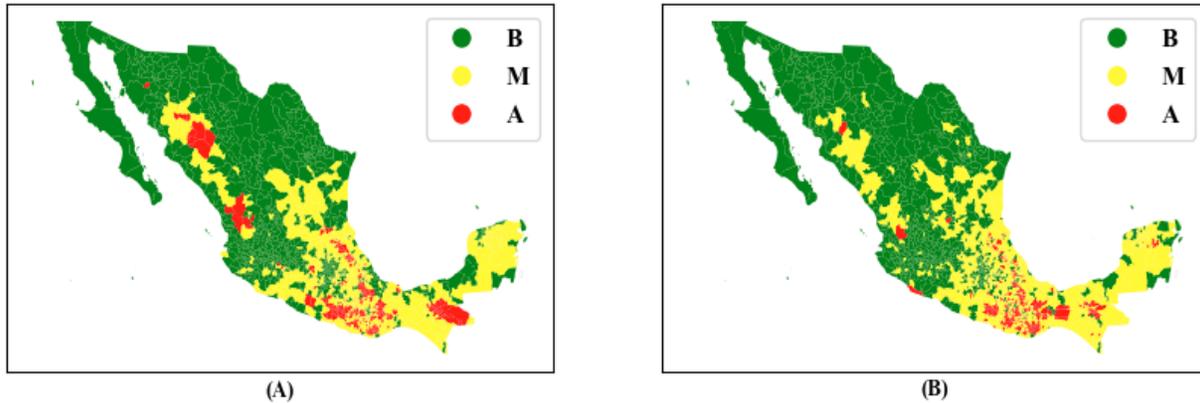


**Figura 1.2. Matriz de confusión normalizada para el modelo bosques aleatorios (RF), para las clases de rezago social bajo (B), medio (M), y alto (A).**

### 1.5.2. Discusión

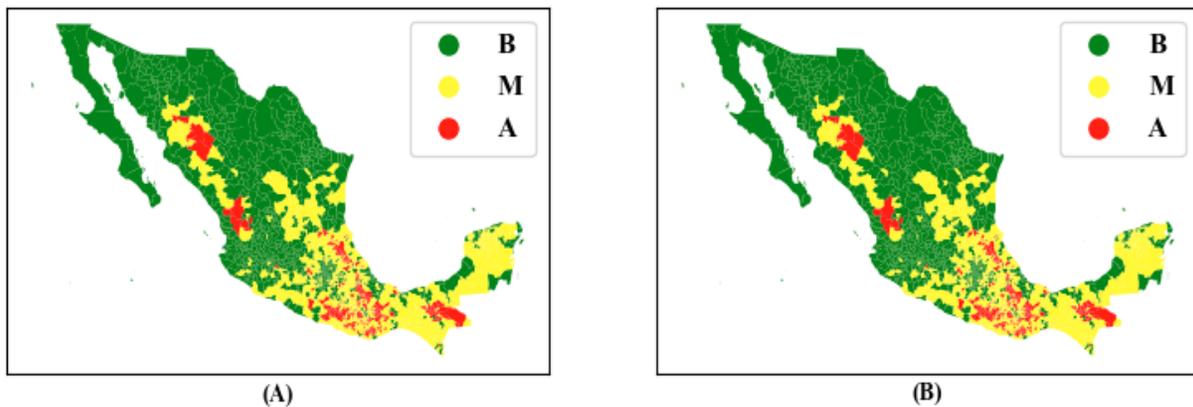
De acuerdo con las predicciones de las clases de rezago social, con el modelo RF a partir de UE agregadas a nivel subsector, se realizó una comparación geográfica a nivel municipal con las clases de rezago reportadas por Vargas-Chanes y Valdés-Cruz (2019).

En los mapas de comparación de rezago social se observó que la distribución a nivel municipal de la clase B se localizó en su mayor parte en el norte de México y en ambos mapas las distribuciones fueron muy similares (Figura 1.3). La distribución de la clase M fue muy similar en el centro y sureste, y en el norte en algunos municipios, el modelo los clasificó como clase B o como clase A. La distribución de la clase A fue muy similar en el centro y sureste de México; sin embargo, en el norte, RF clasificó algunos municipios de clase A como clase M.



**Figura 1.3. Comparación de las clases de rezago social para 2015; B: bajo; M: medio; A: alto; (A) propuestas por Vargas-Chanes y Valdés-Cruz versus (B) las predicciones realizadas con el clasificador bosques aleatorios.**

Finalmente, se utilizó el modelo RF para realizar predicciones a partir UE de los datos del DENUE de octubre de 2016 y noviembre de 2017 (Figura 1.4), donde se obtuvo una distribución de clases de rezago social similar a la reportada por Vargas-Chanes y Valdés-Cruz (2019).



**Figura 1.4. Distribución de las clases de rezago social; B: bajo; M: medio; A: alto; predichas con el clasificador bosques aleatorios, a partir de unidades económicas de 2016 (A) y 2017 (B), agregadas a nivel subsector.**

La aplicación de modelos de aprendizaje automático para predecir el rezago social en México propuesto en este capítulo es un enfoque novedoso y viable, sobre todo, por el tipo de características o variables de entrada. Los resultados mostraron (Cuadro 1.5) que de 2015 a 2016, hubo una disminución en la proporción de municipios con rezago social alto de 17.2 a 15.5 %, esto es 42 municipios pasaron de un grado de rezago social alto a medio; de igual manera, no se observa variación en los municipios con rezago social bajo. Por otra parte, las predicciones para los años 2016 y 2017, mostraron que los municipios con rezago social alto no variaron, y cinco municipios con rezago social medio en 2016, pasaron a un grado de rezago social bajo. La mayoría de los municipios no tuvieron un cambio en su grado de rezago social predicho, pese a utilizar datos de distintos años del DENUÉ como entradas del modelo RF.

**Cuadro 1.5. Comparación de municipios por clase de rezago social; B: bajo; M: medio; A: alto; predichos con el modelo bosques aleatorios para 2016 y 2017 versus las clases de rezago de 2015 propuestas por Vargas-Chanes y Valdés-Cruz.**

Municipios por clase de rezago social			
Año	B	M	A
2015	971 (39.5%)	1064 (43.3%)	422 (17.2%)
2016	971 (39.5%)	1106 (45.0%)	380 (15.5%)
2017	976 (39.7%)	1101 (44.8%)	380 (15.5%)

Los municipios que cambiaron positivamente su nivel de rezago social de 2016 a 2017, de acuerdo con el modelo RF, son: Eduardo Neri, Guerrero; Comonfort, Guanajuato; El Marqués, Querétaro; Tapalpa, Jalisco; y Perote, Veracruz. Los municipios Eduardo Neri, Guerrero; Tapalpa, Jalisco; y Comonfort, Guanajuato; tuvieron ligeros incrementos en UE para 2017, principalmente en aquellas relacionadas al comercio al por menor. El municipio de El Marqués, Querétaro, aumentó significativamente el registro de las UE relacionadas a la construcción y a la industria, lo cual detecta el modelo y predice el cambio. Por último, el municipio de Perote, Veracruz, tiene cambios mixtos en las diferentes categorías de rezago social a nivel subsector, por lo que requiere de un mayor análisis para explicar el motivo de cambio en la predicción de su rezago social.

Así, la metodología propuesta para predecir el grado de rezago social a partir de la información de UE, obtuvo resultados similares a las metodologías tradicionales, que resaltan la poca variación que tienen los niveles de rezago a nivel municipal. En este contexto, la capacidad de los tomadores de decisiones responsables de mitigar el rezago social en México es limitada, en particular en las zonas de rezago social alto.

En este capítulo no se consideró la naturaleza ordinal de los grados de rezago social para optimizar el desempeño de los clasificadores de aprendizaje automático. Como se ha mostrado en otros estudios (Pérez-Rodríguez *et al.*, 2020; Cao *et al.*, 2020), el desempeño de los clasificadores tiende a mejorar, bajo el supuesto de una jerarquía equidistante entre las clases objetivo bajo estudio. Asimismo, es posible enfocar el desempeño de los clasificadores con base en una categoría de rezago social de interés, esto permitiría mejorar la clasificación para una clase específica; por ejemplo, si se emplea la métrica AUC de las curvas *P-S* como criterio de desempeño.

## 1.6. CONCLUSIONES

Con base en los resultados obtenidos en este capítulo, se pudo probar que la predicción de las clases o grados de rezago social; alto, medio y bajo; se puede llevar a cabo por medio de algoritmos de aprendizaje automático supervisado a partir de datos del Directorio Nacional de Unidades Económicas, agregadas a nivel subsector.

Los modelos de regresión logística, máquina de soporte vectorial y bosques aleatorios alcanzaron un buen desempeño en predicción. En particular, el clasificador bosques aleatorios fue superior a los modelos de regresión logística y máquina de soporte vectorial con una precisión global promedio de 74.4% y *F1-macro* promedio de 71.6%.

La distribución geográfica de las clases de rezago social resultó muy similar a la distribución del rezago social considerado. El desempeño en predicción de los clasificadores implementados mostró que existe correlación entre las tasas de categorías de unidades económicas y localización geográfica de los municipios, con las clases de rezago social. Las predicciones del rezago social para 2016 y 2017 mostraron que es posible estimar el grado de rezago social de manera indirecta, cuando no se dispone del cálculo oficial.

## CAPÍTULO II EVALUACIÓN DE ARQUITECTURAS DE REDES NEURONALES DE APRENDIZAJE PROFUNDO PARA LA PREDICCIÓN ORDINAL DEL GRADO DE REZAGO SOCIAL EN MÉXICO

### 2.1. RESUMEN

El rezago social es una condición que afecta a millones de personas que se asocia positivamente con la pobreza, por esta razón es importante medir este fenómeno con regularidad para mitigar su impacto. La toma de decisiones estratégicas y el monitoreo continuo del rezago social requieren contar con procedimientos y herramientas que faciliten la estimación precisa y oportuna del grado de rezago social en México. En el presente trabajo se implementaron, en el lenguaje de código abierto Python, tres arquitecturas de redes neuronales de aprendizaje profundo; una con escala nominal (DL-NOM) y dos con escala ordinal (DL1-ORD y DL2-ORD); para predecir el grado de rezago social en México, a partir de datos de conteos de unidades económicas provenientes de estadísticas oficiales a nivel municipal. Estos clasificadores se entrenaron con métodos de regularización combinada y se optimizaron por medio de un procedimiento de validación cruzada y búsqueda de hiperparámetros por retícula. El clasificador DL1-ORD obtuvo mejor desempeño que los clasificadores DL-NOM y DL2-ORD; sin embargo, las diferencias no fueron estadísticamente significativas, con base en las pruebas de t por pares, y de Wilcoxon. DL1-ORD obtuvo una precisión global de clasificación promedio de 76.2%, error absoluto medio de 0.241 y valor de *F1-macro* promedio de 73.8%. La metodología propuesta para la predicción ordinal del grado de rezago social en México, por medio de redes neuronales de aprendizaje profundo entrenadas a partir de datos anuales de unidades económicas, es una valiosa herramienta de estimación alterna a los métodos que se basan en datos quinquenales o decenales.

**Palabras clave:** pobreza, inteligencia artificial, optimización estocástica, aprendizaje supervisado.

## 2.2. ABSTRACT

Social lag is a condition that affects millions of people and is positively correlated with poverty, for this reason it is important to measure this phenomenon regularly to mitigate its impact. Strategic decision-making and continuous monitoring of social lag require procedures and tools that facilitate accurate and timely estimation of the degree of social lag in Mexico. In the present work, three deep learning neural network architectures were implemented in the open-source language Python; one with nominal scale (DL-NOM) and two with ordinal scale (DL1-ORD and DL2-ORD); to predict the degree of social lag in Mexico, based on data from counts of economic units from official statistics at the municipal level. These classifiers were trained with combined regularization methods and were optimized by the cross-validation method and hyperparameter grid search. The DL1-ORD classifier performed better than the DL-NOM and DL2-ORD classifiers; however, the differences were not statistically significant, based on the paired t-tests and Wilcoxon non-parametric test. DL1-ORD obtained an average accuracy of 76.2%, mean absolute error of 0.241 and F1-macro value of 73.8%. The proposed methodology for ordinal prediction of the degree of social lag in Mexico, using deep learning neural networks trained from annual data from economic units, is a valuable alternative estimation tool to methods that are based on every five-year or ten-year data.

**Keywords:** poverty, artificial intelligence, stochastic optimization, supervised learning.

### 2.3. INTRODUCCIÓN

El grado de rezago social es una medida que integra cuatro indicadores de carencias sociales fundamentales: el rezago educativo, el acceso a los servicios de salud, el acceso a los servicios básicos, y la calidad de la vivienda. Esta medida permite a los tomadores de decisiones conocer la situación actual en la que se encuentran los Estados, municipios y localidades en México. El cálculo o estimación confiable del grado de rezago social a nivel municipal facilita la definición de políticas públicas de mitigación del rezago social y la transferencia de recursos federales a las zonas con mayor grado de rezago. En México, el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) es responsable de realizar la estimación oficial del grado de rezago social cada cinco años con base en encuestas intercensales, así como información de los censos de población y vivienda más recientes. Sin embargo, durante los años intermedios, donde no se dispone de la estimación oficial, no existe un referente anual de la distribución de los grados de rezago social a nivel municipal.

Las unidades económicas son entidades activas, tanto públicas como privadas, productoras de bienes y servicios con fines mercantiles o no. La divulgación de esta información económica ocurre con mayor frecuencia que la publicación oficial del rezago social, la cual se reporta inclusive hasta dos veces por año. Esta información se recaba por el Instituto Nacional de Estadística e Informática (INEGI), y se ofrece al público a través del Directorio Estadístico Nacional de Unidades Económicas (DENUE). Las actividades económicas relativas a las unidades económicas contenidas en el DENUE se categorizan de acuerdo con el Sistema de Clasificación Industrial de América del Norte (SCIAN), cuya estructura jerárquica se compone de sectores, subsectores, ramas, subramas y clases.

Las redes neuronales representan una estrategia novedosa para la predicción de variables socioeconómicas. Azcarraga y Setiono (2018), utilizaron estos modelos para caracterizar la pobreza de los hogares en Filipinas con base en un clasificador binario entrenado con el algoritmo de propagación hacia atrás, a partir de datos no balanceados de perfiles sociodemográficos. Ellos concluyeron que es posible identificar correctamente hasta 7 de cada 10 hogares pobres. Poerwanto y Fajriani (2020), aplicaron un modelo de redes neuronales con una sola capa oculta para predecir niveles de pobreza en Indonesia. Utilizaron tres predictores y validación cruzada para optimizar la red de clasificación binaria de pobreza, obteniendo una precisión global de clasificación correcta de 95%.

Los algoritmos de aprendizaje profundo como las redes neuronales convolucionales, se han aplicado para la predicción de la pobreza. Babenko *et al.* (2017) aplicaron diferentes arquitecturas de redes neuronales convolucionales para predecir niveles de pobreza; con base en imágenes satelitales de municipios de México y resultados de la Encuesta Nacional de Ingresos y Gastos de los Hogares de 2014; de esta manera llegaron a la conclusión de que es posible estimar la pobreza con esta técnica, no obstante, es necesario comprender el proceso de entrenamiento que se genera. Estos métodos requieren mayor capacidad de procesamiento computacional para su entrenamiento. En la literatura también se reportan algoritmos de aprendizaje automático supervisado alternativos para la clasificación de variables relacionadas con la pobreza, como la máquina de soporte vectorial (Naviamos y Niguidula, 2020) y los bosques aleatorios (Mohamud y Gerek 2019).

El objetivo de esta investigación fue implementar y evaluar el desempeño de tres arquitecturas de redes neuronales de aprendizaje profundo (dos ordinales y una nominal) para mejorar la metodología propuesta por Ávila-Solís *et al.* (2021), para predecir el grado de rezago social en México (alto, medio y bajo) a nivel municipal, a partir de datos anuales de unidades económicas agrupadas por actividad económica a nivel subsector; así como, comparar los mapas de predicción de la distribución geográfica del grado de rezago social a nivel municipal en México.

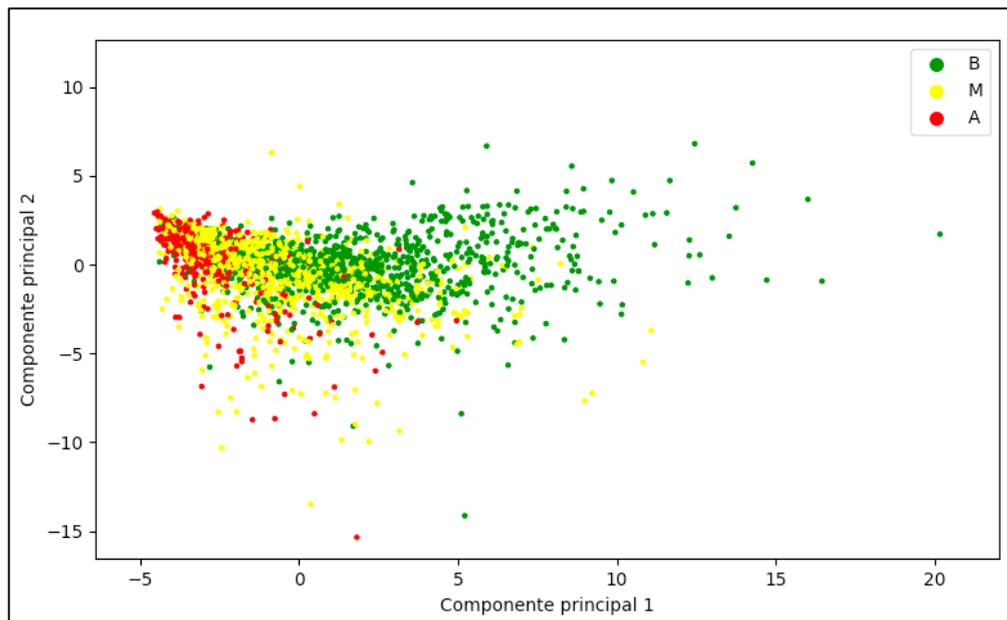
## 2.4. MATERIALES Y MÉTODOS

### 2.4.1. Conjuntos de datos

Para predecir el grado de rezago social de 2457 municipios de México se emplearon datos de México en 2015 que se obtuvieron por medio de un análisis longitudinal por medio de clases latentes y reportados por Vargas-Chanes y Valdés-Cruz (2019). En este estudio se consideraron tres grados o clases de rezago social: alto (A), medio (M) y bajo (B); las cuales, conservan alta correlación con la estimación pública de los grados de rezago social que reporta CONEVAL (2016). Estos autores señalan que la clasificación del rezago social que proponen tiene mayor soporte estadístico. Dada la naturaleza del fenómeno de rezago social, se contó con clases objetivo desbalanceadas: 422 municipios en clase A, 1064 en la clase M y 971 en la clase B.

La información de las unidades económicas del DENU se obtuvo para el mismo año que las clases objetivo de rezago social (INEGI, 2015). Los datos se agruparon por municipio con base en conteos de las 94 categorías de actividades económicas a nivel subsector del SCIAN, con este nivel de agrupamiento, los clasificadores alcanzan mejor desempeño (Ávila-Solís *et al.*, 2021). Como método de ingeniería de variables de entrada, se crearon tasas de unidades económicas (TUE) para cada categoría de actividad económica; los conteos de unidades económicas se dividen por categoría entre la población del municipio correspondiente, y el resultado se multiplica por mil. Las variables TUE, latitud y longitud de cada municipio (INEGI, 2014), se utilizaron como características de entrada o variables predictoras de los modelos (96 variables). Estas variables se estandarizaron previo a la etapa de entrenamiento para mejorar la estabilidad numérica de los procedimientos de optimización de los modelos de aprendizaje.

Se trató de evitar trabajar con todas las variables y prevenir algún efecto de multicolinealidad entre ellas, aplicando la técnica de reducción de dimensionalidad de componentes principales a partir de la matriz de covarianza, la cual no resultó útil para dicho objetivo por consiguiente se emplearon las características completas para ajustar los modelos. Aunque con este método fue posible llevar a cabo una visualización de las clases objetivo de rezago social en dos dimensiones, con los datos de entrada escalados a nivel subsector (Figura 2.1). La proporción de varianza explicada para los dos primeros componentes fue de 13.5% y 5.3%, respectivamente, es decir estos explicaron en conjunto el 18.8% de la variabilidad de los datos.



**Figura 2.1. Visualización de las clases de rezago social (B: bajo; M: medio; A: alto) aplicando la técnica de componentes principales, a partir de la matriz de covarianza de los datos de entrada a nivel subsector.**

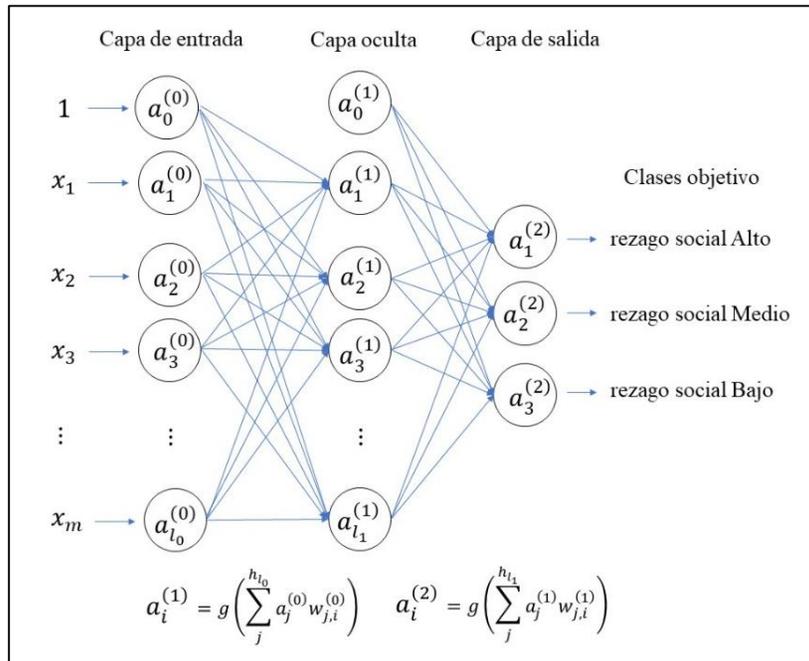
#### 2.4.2. Hardware y software

Las diferentes arquitecturas de redes neuronales de aprendizaje profundo se implementaron con el lenguaje de código abierto Python en su versión 3.7, mediante las bibliotecas Keras y Tensorflow, versión 2.5. Las particiones aleatorias de los conjuntos de entrenamiento y prueba, y el cálculo de las métricas de desempeño de los clasificadores se realizaron con el paquete Scikit-learn 0.24.1 (Pedregosa et al, 2011). El proceso de modelación se ejecutó en una MacBook Air con sistema operativo macOS Mojave con 8 GB de RAM y procesador Intel Core i5 de 1.8 GHz.

### 2.4.3. Redes neuronales de aprendizaje profundo

#### 2.4.3.1. Modelo de aprendizaje profundo con respuesta nominal

El modelo de aprendizaje profundo utilizado para la predicción del grado de rezago social, mediante un enfoque nominal, es una red neuronal artificial conectada hacia adelante (DL-NOM), formada por una capa de entrada, una o más capas ocultas y una capa de salida (Chollet, 2018). Cada capa se compone por una o más neuronas (unidades de activación), denotados como  $a_i^{(l)}$ , donde  $i$  representa el índice de la unidad de activación y  $l$  la capa correspondiente. En la capa de entrada, el índice  $i$  varía, tal que  $0 \leq i \leq m$ , donde  $m$  es el número de características de entrada y la unidad con índice  $i = 0$ , representa el sesgo de la capa; en las capas ocultas o intermedias de la red, el número de unidades de activación es un hiperparámetro que se ajusta durante la etapa de entrenamiento del modelo y en la capa de salida el número de unidades de activación, en el caso de un problema de clasificación, depende del número de clases objetivo  $K$  (Figura 2.2).



**Figura 2.2. Arquitectura de una red neuronal artificial conectada hacia adelante con una capa oculta para la clasificación nominal de tres clases objetivo de rezago social.**

Cuando se aplica una función de activación  $g(\cdot)$  a la combinación lineal  $z_i^{(l)} = \sum_j^{h_l} a_j^{(l)} w_{j,i}^{(l)}$  se obtiene la respuesta de la  $i$ -ésima unidad de activación de la siguiente capa ( $l + 1$ ), esto se denomina un proceso de propagación hacia adelante. Los pesos o parámetros  $w_{j,i}^{(l)}$  se almacenan

en  $l - 1$  matrices  $W$ , las unidades de activación en la capa de entrada se definen por el número de características de entrada del modelo. Los modelos de red neuronal se entrenan, generalmente, con el algoritmo de propagación de los errores hacia atrás. En la capa de salida del modelo DL-NOM las respuestas se transforman en clases objetivo por medio de la función *softmax* o función exponencial normalizada,  $s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}}$ . Las funciones de activación utilizadas en las capas ocultas son: la función sigmoide,  $g(x) = \frac{1}{1+e^{-x}}$ ; la función tangente hiperbólica,  $g(x) = \frac{e^{-x}-e^{-x}}{e^x+e^{-x}}$ ; y la unidad lineal rectificadora (ReLU),  $g(x) = \max(0, x)$ .

#### 2.4.3.2. Modelos aprendizaje profundo con respuesta ordinal

Dado que las clases objetivo (o grados) de rezago social guardan un orden entre ellas;  $A > M > B$ ; en este estudio se implementaron dos arquitecturas de aprendizaje profundo con escala ordinal en las respuestas. propusieron un método para agregar una escala ordinal a las clases objetivo de una red neuronal conectada hacia adelante (DL1-ORD) (Cheng *et al.* 2008). El cual consiste en modificar la codificación de las clases objetivo y la función de activación de la capa de salida. Sea  $D = \{\mathbf{x}, y\}$  un conjunto de datos que consiste en  $n$  muestras y  $m$  características de entrada, donde  $\mathbf{x}$  es un vector  $m$ -dimensional, e  $y$  es la clase ordinal correspondiente de un conjunto finito discreto de tamaño  $K$ . El vector de codificación propuesto para las clases  $y$  es  $t = (t_1, \dots, t_i, \dots, t_K)$ , donde  $t_i$  es igual a 1 si la categoría ordinal pertenece a la clase objetivo o a una menor, y 0 de otro modo, para  $1 \leq i \leq K$ . Además, la capa de salida de la red cuenta con una función de activación sigmoide con  $K$  nodos. Para realizar predicciones con este método, se revisa en orden el vector de salida resultante, junto con un umbral específico (generalmente 0.5), cuando se encuentra el primer elemento  $t_i$  que es menor a dicho umbral, se asigna la clase  $i - 1$ , o la clase  $K$  en el caso de que no haya elementos menores al umbral. De acuerdo con lo anterior, la codificación para las clases objetivo utilizadas en la red neuronal DL1-ORD, son para la clase de rezago social alto  $[1, 0, 0]$ , para clase de rezago social medio  $[1, 1, 0]$  y para la clase de rezago social bajo  $[1, 1, 1]$ .

Otro enfoque utilizado para tomar en cuenta las clases objetivo de rezago social como clases ordinales, es mediante una arquitectura de aprendizaje profundo rango consistente *logit* (DL2-ORD), propuesta por Cao *et al.* (2020), que garantiza la consistencia de las predicciones para clasificación. Bajo este enfoque, dado el conjunto de datos  $D$ , los intervalos ordenados  $r_K$  de las

clases objetivo  $y_i$  se extienden en  $K - 1$  clases binarias  $y^{(1)}, \dots, y^{(K-1)}$ . De tal suerte que  $y^{(K-1)} \in \{0,1\}$ , indica si  $y_i$  excede el intervalo  $r_K$ , 1, ó no, 0. Con base en las clases objetivo modificadas se entrenan  $K - 1$  clasificadores binarios en la capa de salida. Así, dado un vector de entrada  $\mathbf{x}$ , la predicción se obtiene mediante  $h(\mathbf{x}) = r_q$ , y el índice del intervalo  $q$  es asignado por:

$$q = 1 + \sum_{j=1}^{K-1} f_j(\mathbf{x}_i),$$

donde  $h(\mathbf{x})$  es una función de aprendizaje; y  $f_j(\mathbf{x}_i) \in \{0,1\}$  es la predicción del  $j$ -ésimo clasificador en la capa de salida.

Durante el entrenamiento de DL2-ORD, se minimiza la función de costo de entropía cruzada promedio de los  $K - 1$  clasificadores con un umbral de 0.5 (similar al umbral de DL1-ORD), que garantiza que las probabilidades estimadas sean decrecientes. Con el enfoque de DL2-ORD se utilizó el mismo orden de las clases objetivo que se empleó con DL1-ORD para las clases objetivo de rezago social, de forma que se asignó 0 a la clase A (alto), 1 a la clase M (medio) y 2 a la clase B (bajo).

#### 2.4.4. Métricas de desempeño

Las métricas para medir el desempeño de los clasificadores DL-NOM, DL1-ORD y DL2-ORD fueron la precisión global de clasificación correcta (*Acc*, accuracy), el error absoluto promedio (*MAE*, mean absolute error), y el valor *F1-macro*. *Acc* es una medida apropiada cuando se tienen clases objetivo, distribuidas uniformemente; y *F1-macro* resulta más apropiada cuando las clases no están perfectamente balanceadas, es decir, se cuenta con más ejemplos de alguna de las clases objetivo, como en el caso de las clases de rezago social. Ambas métricas, tienen su origen en la matriz de confusión, la cual compara las predicciones de clase, columnas, contra las clases verdaderas, hileras.

En el caso binario (Cuadro 2.1) la matriz de confusión presenta los verdaderos positivos ( $VP$ ), los falsos positivos ( $FP$ ), los falsos negativos ( $FN$ ), y los verdaderos negativos ( $VN$ ), definiendo a partir de estos valores la precisión global como:

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

**Cuadro 2.1. Matriz de confusión para el caso de un clasificador binario.**

		Clases verdaderas	
		Positivo	Negativo
Clases predichas	Positivo	$VP$	$FP$
	Negativo	$FN$	$VN$

$VP$ : verdadero positivo;  $FN$ : falso negativo;  $FP$ : falso positivo;  $VN$ : verdadero negativo.

Otras métricas que se obtienen a partir de la matriz de confusión son la precisión ( $P$ ) y la sensibilidad ( $S$ ); que representan la fracción de positivos predichos, que en realidad son positivos; y la fracción de observaciones positivas detectadas por el clasificador, respectivamente, definidas como  $P = \frac{VP}{VP+FP}$ , y  $S = \frac{VP}{VP+FN}$ .

La medida  $F1$  es la media armónica de  $P$  y  $S$ , es decir  $F1 = \frac{2P \times S}{P+S}$ , que toma valores entre 0 (ajuste mediocre) y 1 (ajuste perfecto). La medida  $F1$  favorece a los modelos con  $P$  y  $S$  similares (Géron, 2019). En el caso de clasificación múltiple, se define la medida  $F1$ -macro como:

$$F1\text{-macro} = \frac{1}{K} \sum_i^K F1_i,$$

donde  $F1_i$  es el valor de  $F1$  de la  $i$ -ésima clase objetivo y  $K$  es el número de clases. En situaciones con clases desbalanceadas, el valor  $F1$  – macro resulta apropiado, ya que da a cada clase el mismo peso (Chai *et al.*, 2002).

Métricas de desempeño para variables continuas se utilizan en escenarios de optimización de variables ordinales. Tal es el caso de la métrica  $MAE$  que se define por:

$$MAE = \frac{\sum_{i=1}^n e_i}{n} = \frac{\sum_{i=1}^n |y_i - h(\mathbf{x}_i)|}{n},$$

donde  $y_i$  representa la  $i$ -ésima clase objetivo o valor verdadero;  $h(\mathbf{x}_i)$  es una función que aproxima el valor de  $y_i$ , a partir de  $\mathbf{x}_i$ ;  $\mathbf{x}_i$  es un vector de variables para la  $i$ -ésima observación; y  $e_i$  es el  $i$ -ésimo error absoluto entre  $y_i$  y  $h(\mathbf{x}_i)$ .

#### 2.4.5. Entrenamiento y selección de hiperparámetros

En la etapa de entrenamiento se utilizó validación cruzada (CV) con 5 particiones y búsqueda por retícula para realizar la selección de hiperparámetros con 80% del conjunto original de datos y en la etapa de prueba en predicción se utilizó CV con 10 particiones para la validación de los mejores modelos con 100% del conjunto original de datos. La CV consiste en crear  $k$  subconjuntos disjuntos del mismo tamaño a partir de un conjunto original de datos. En cada iteración  $k - 1$  subconjuntos se utilizan como datos de entrenamiento y el subconjunto restante como validación o prueba. Posteriormente se realiza el cálculo de las métricas de desempeño deseadas; valores comunes de  $k$  son 5 ó 10. Esta técnica se utilizó en la selección de hiperparámetros óptimos y en la evaluación final del modelo seleccionado.

El proceso de entrenamiento de una red neuronal es iterativo y experimental; para realizar el ajuste de los pesos óptimos de los clasificadores DL-NOM, DL1-ORD y DL2-ORD por medio de los optimizadores de sus funciones de costo se requiere determinar los hiperparámetros de cada modelo. Para ello, se definió un procedimiento de búsqueda por retícula que consiste en definir intervalos discretos de los hiperparámetros de interés y generar un conjunto de combinaciones de valores para encontrar la mejor combinación de hiperparámetros que produzca el mejor desempeño del modelo con base en una métrica definida, por ejemplo, *Acc*. Con este procedimiento se buscaron los valores óptimos de los siguientes hiperparámetros: probaron de una a tres capas ocultas, con tres funciones de activación en cada capa (*sigmoide*, *tanh* y *ReLU*); y de 4 a 256 unidades de activación en cada capa oculta. Tanto las funciones como las unidades de activación fueron consistentes en cada capa oculta. La inicialización de los pesos de cada capa se basó en la técnica de Glorot (Glorot y Bengio, 2010), este esquema trata de uniformizar las varianzas de las entradas y salidas de una capa, que da como resultado mayor velocidad de convergencia durante el entrenamiento de los pesos. Las funciones de costo que se utilizaron son: *MAE* para los modelos DL-NOM y DL1-ORD, y entropía cruzada para DL2-ORD.

Para evitar problemas de sobreajuste se utilizaron estrategias de regularización; la regularización busca reducir el sobreajuste de un modelo, por medio de un balance entre el sesgo y la varianza del modelo. En este estudio se utilizó la regularización L2, que consiste en agregar a la función de costo un término cuadrático; esto es,  $J(\mathbf{W}; \mathbf{X}, \mathbf{y}) + \lambda \|\mathbf{W}\|^2$ ; donde  $\lambda$  es el parámetro

de regularización, y  $\|\cdot\|$  es la norma L2 de la matriz de pesos o parámetros  $\mathbf{W}$ . Otra técnica de regularización es la parada anticipada del entrenamiento (*early stopping*), en la cuál se observa el desempeño del modelo en el conjunto de entrenamiento y de validación al mismo tiempo, para determinar el número óptimo de pasadas donde el modelo generaliza de mejor manera, de acuerdo con alguna métrica de desempeño; ya que a partir de este punto su desempeño se degrada.

Para controlar el posible sobreajuste de los modelos empleados en este trabajo, se añadió doble regularización a la búsqueda de hiperparámetros, con distintos valores para la regularización L2 de la función costo dentro de cada capa (0.0001, 0.001, 0.01, 0, 1), así como el método de parada anticipada. Todos los modelos se optimizaron con el método Adam (Kingma y Ba, 2015), el cual es una variante del gradiente estocástico descendiente, simple y computacionalmente eficiente, con pocos requisitos de memoria. Este método calcula las tasas de aprendizaje adaptativo individuales para diferentes parámetros a partir de estimaciones del primer y segundo momento de los gradientes.

Finalmente, para realizar predicciones, se emplearon datos completos para entrenar el modelo con los parámetros óptimos, con lo que se obtienen los pesos finales de la red neuronal que se utiliza junto con datos de entrada del DENUÉ de años posteriores para predecir el nivel de rezago social de un municipio específico.

#### 2.4.6. Comparación del desempeño de los clasificadores

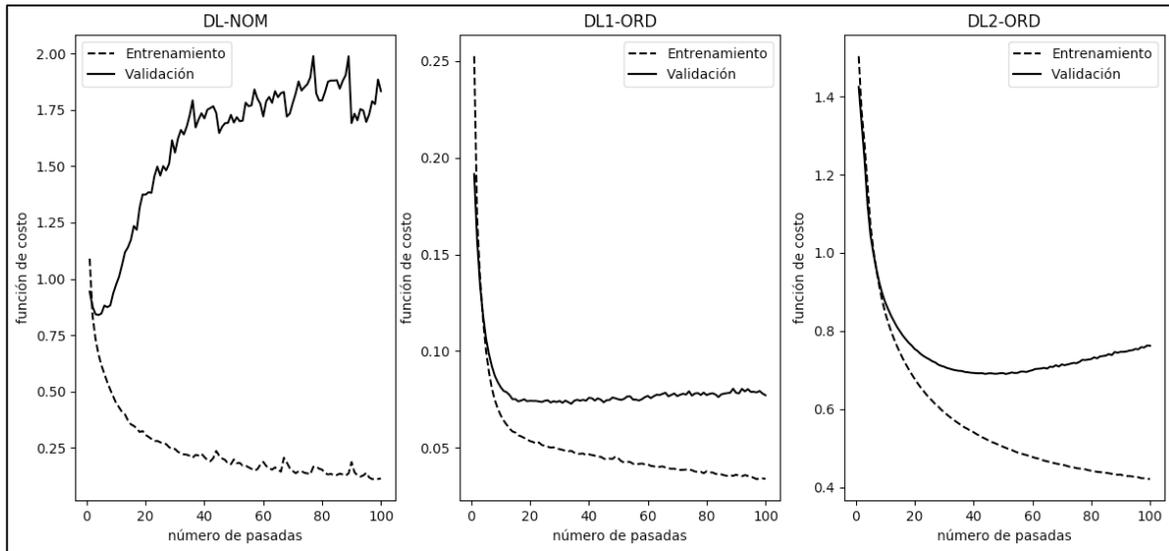
Una vez que se contó con los modelos optimizados, se aplicaron pruebas de hipótesis a las métricas de desempeño utilizadas para comparar si estas fueron significativamente diferentes, y la técnica de visualización del diagrama de caja y bigotes, para fines de comparación. Un procedimiento análogo al propuesto por Sani *et al.* (2018) se utilizó para realizar la comparación de los modelos de clasificación de pobreza con CV, se utilizaron las pruebas  $t$  por pares de las métricas  $F1 - macro$ ,  $ACC$  y  $MAE$ . Las comparaciones múltiples se corrigieron por medio del método de Holm, que es un método menos conservador que el de Bonferroni (James *et al.*, 2021). Además, se utilizó la desviación estándar agrupada, debido al tamaño de muestra que se tiene, aunque, de acuerdo con Japkowicz y Mohak (2011), estas pruebas tradicionales no son lo suficientemente robustas para comparar los algoritmos de aprendizaje. Una alternativa no paramétrica a las pruebas  $t$  por pares, es la prueba de rangos de Wilcoxon, misma que se aplicó

con corrección para comparaciones múltiples, el nivel de significancia  $\alpha$  se dividió entre el número total de comparaciones; este procedimiento es análogo al aplicado por Chang y de Rijke. (2018) para comparar el desempeño por pares de los algoritmos bajo la métrica *MAE* con regresión ordinal.

## 2.5. RESULTADOS Y DISCUSIÓN

### 2.5.1. Resultados de entrenamiento

El efecto de la regularización por parada anticipada, para las funciones de costo respectivas de cada modelo, con CV de 10 particiones, mostró que el número de ciclos (*epochs*) óptimos son, en promedio: 5, 15 y 36; para los modelos DL-NOM, DL1-ORD y DL2-ORD; respectivamente. (Figura 2.3).



**Figura 2.3. Entrenamiento por medio de parada anticipada para los modelos de DL-NOM, DL1-ORD y DL2-ORD.**

Las métricas de desempeño que se obtuvieron durante el entrenamiento, para los mejores parámetros e hiperparámetros, permitieron realizar la comparación con las obtenidas en la CV final para la selección del mejor modelo (Cuadro 2.2). Se observó que el modelo que más sobreajuste tuvo fue DL-NOM. Los modelos DL-NOM y DL1-ORD incrementaron la variabilidad de las métricas de desempeño, al comparar las obtenidas durante el entrenamiento con las de prueba. En este sentido, el modelo que mejor se comportó, una vez ajustado, fue DL2-ORD, pues no presentó sobreajuste y su variabilidad es similar a la obtenida durante el entrenamiento del modelo.

**Cuadro 2.2. Resultados del entrenamiento con regularización combinada: parada anticipada y L2, a través de validación cruzada con 5 particiones, para las métricas de desempeño *F1-macro*, error absoluto medio (*MAE*) y precisión global (*ACC*).**

Modelo	<i>F1-macro</i>	<i>MAE</i>	<i>ACC</i>
DL-NOM	0.732+/-0.014	0.2514+/-0.015	0.7547+/-0.013
DL1-ORD	0.736+/-0.022	0.2473+/-0.021	0.7567+/-0.018
DL2-ORD	0.716+/-0.027	0.2737+/-0.028	0.7323+/-0.028

Los hiperparámetros optimizados para DL1-ORD fueron: tres capas ocultas con 24 unidades, regularización L2 de 0.001 y función de activación *tanh*. DL2-ORD alcanzó los mejores resultados durante el entrenamiento con dos capas ocultas, con 4 unidades en cada capa, regularización L2 de 0.001 y función de activación ReLU. Por último, DL-NOM se optimizó con tres capas ocultas, con 24 unidades en cada capa, regularización L2: 0.001 y función de activación tangente hiperbólica. DL1-ORD y DL2-NOM compartieron los hiperparámetros ajustados, mientras que los tres algoritmos compartieron el valor de regularización óptimo.

### 2.5.2. Resultados por clase objetivo

Los resultados obtenidos por clase, conforme a la métrica *F1*, mostraron que el modelo DL1-ORD alcanzó los mejores resultados para las clases de rezago B y M, y el modelo DL2-ORD el mejor resultado para la clase de rezago A (Cuadro 2.3).

**Cuadro 2. 3. Resultados por clase con regularización combinada: parada anticipada y L2, a través de validación cruzada con 10 particiones.**

Modelo	<i>F1-macro</i>	<i>MAE</i>	<i>ACC</i>	Clase	<i>P</i>	<i>S</i>	<i>F1</i>
DL-NOM	0.718 +/-0.036	0.262 +/-0.033	0.744 +/-0.034	A	0.642	0.586	0.613
				M	0.699	0.748	0.723
				B	0.846	0.808	0.826
DL1-ORD	0.738 +/-0.032	0.241 +/-0.035	0.762 +/-0.032	A	0.656	0.611	0.633
				M	0.717	0.761	0.738
				B	0.864	0.829	0.846
DL2-ORD	0.716 +/-0.026	0.275 +/-0.027	0.732 +/-0.027	A	0.617	0.682	0.648
				M	0.723	0.648	0.683
				B	0.796	0.846	0.820

*MAE*: error absoluto medio; *ACC*: precisión global; *P*: precisión; *S*: especificidad.

En general, los resultados de las redes neuronales artificiales, son mejores a los alcanzados por el modelo de bosques aleatorios reportado por Ávila-Solís *et al.* (2021), para *F1-macro* comparado con DL-NOM y *ACC* comparado con DL-NOM y DL1-ORD. De acuerdo a la validación final, el modelo DL1-ORD contó con los mejores resultados de las métricas de desempeño.

### 2.5.3. Comparación de los modelos ajustados

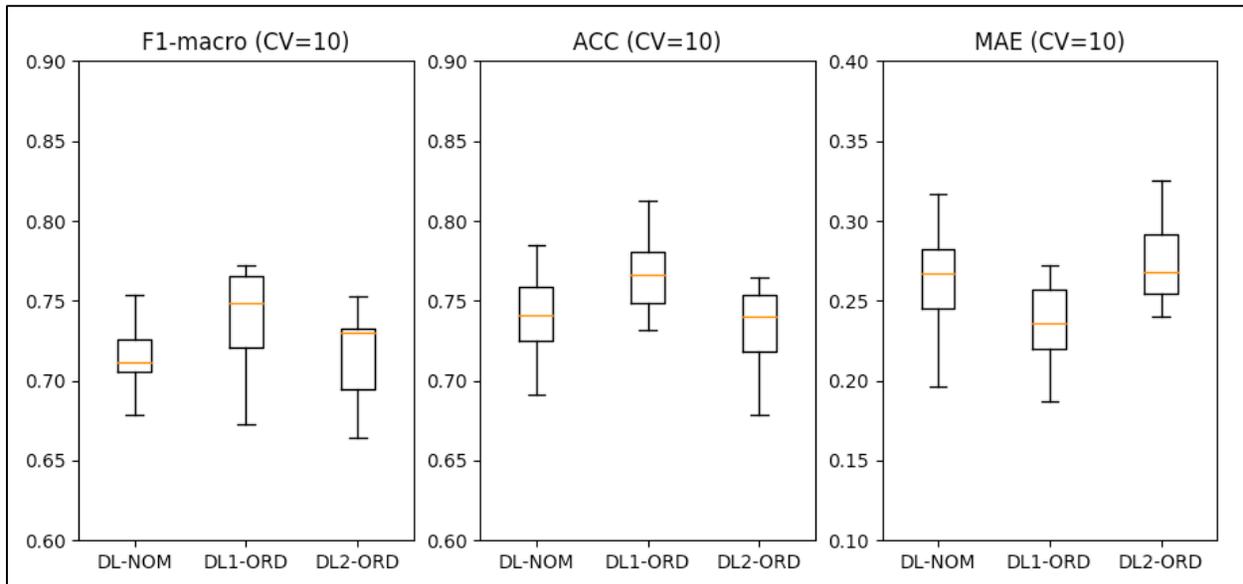
Los p-valores observados de las pruebas *t* por pares, corregidas mediante el método de Holm, para las métricas de desempeño *F1 – macro*, *MAE* y *ACC*, resultaron no significativos, es decir, se consideró que no existen diferencias en el enfoque de los clasificadores para niveles aceptables

del nivel de significancia  $\alpha = 0.05$  (Cuadro 2.4). Los p-valores de las pruebas Wilcoxon por pares resultaron significativos para las tres métricas de desempeño entre DL1-ORD y DL2-ORD, ya que se requerían valores menores a 0.016 para rechazar la igualdad de medias con un  $\alpha$  de 0.05, ya que se tuvieron 3 comparaciones por pares.

**Cuadro 2.4. P-valores de las pruebas de t por pares corregidas con método de Holm y pruebas de Wilcoxon, para las métricas de desempeño *F1-macro*, error absoluto medio (*MAE*) y precisión global (*ACC*).**

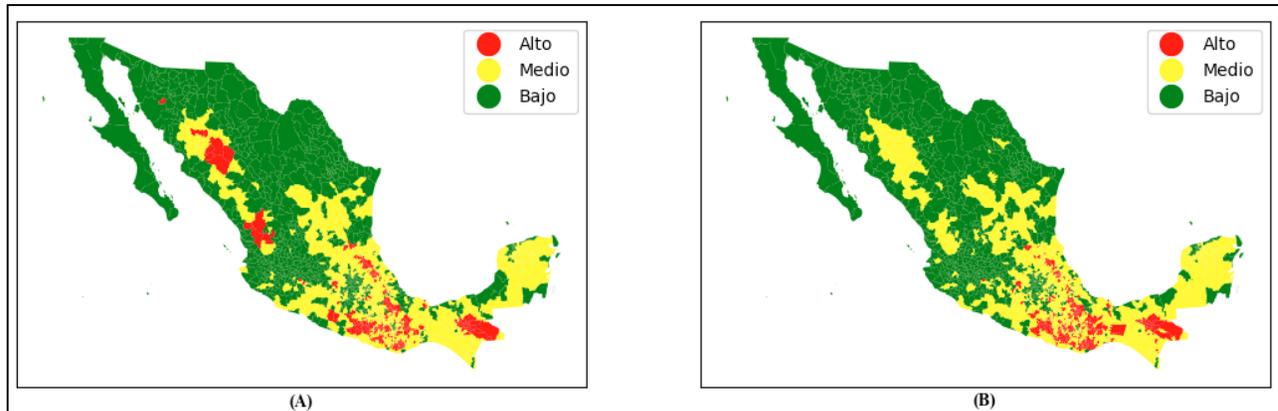
Modelo <i>versus</i> Modelo	<i>F1-macro</i> (Holm)	<i>MAE</i> (Holm)	<i>ACC</i> (Holm)	<i>F1-macro</i> (Wilcoxon)	<i>MAE</i> (Wilcoxon)	<i>ACC</i> (Wilcoxon)
DL-NOM DL1-ORD	0.440	0.362	0.471	0.275	0.492	0.375
DL-NOM DL2-ORD	0.877	0.394	0.471	0.769	0.374	0.374
DL1-ORD DL2-ORD	0.440	0.101	0.152	0.014	0.007	0.007

La comparación visual, por medio del gráfico de caja, de las métricas de desempeño que fueron evaluadas, permitió observar que la métrica *ACC* presentó una distribución más simétrica que los observados para la métrica *F1 – macro* y *MAE* (Figura 2.4). Aunque las pruebas de significancia, para comparar el desempeño de los modelos ordinales con los de la red neuronal artificial sin enfoque ordinal, no arrojaron diferencias estadísticas, las mejores prácticas invitan a emplear modelos diseñados para el tipo de variable de respuesta que se pretende analizar.



**Figura 2.4.** Comparación de resultados de la validación cruzada (CV), con 10 particiones, de las métricas de desempeño: *F1-macro*, precisión global (*ACC*) y error absoluto medio (*MAE*).

Finalmente, se mapeó y comparó la clasificación del modelo DL1-ORD, mediante predicción por validación cruzada, contra las clases verdaderas para 2015 (Figura 2.5). Donde se observó que, el modelo DL1-ORD, presentó errores de clasificación para las clases de rezago alto en la región noroeste del país. Mientras que, para la zona de la península de Yucatán, donde prevalece el rezago social medio, presentó una buena aproximación a los datos originales.



**Figura 2.5. Comparación de clases de rezago social; (A) propuestas por Vargas-Chanes y Valdés-Cruz (2019), y (B) predichas con el modelo de aprendizaje automático ordinal a nivel municipal de 2015**

En la literatura, se reportan otras estrategias para modelar las clases objetivo en escala ordinal; por ejemplo, mediante métodos de regularización Bayesianos, probado en datos genómicos e implementado en el lenguaje de programación R (Pérez-Rodríguez *et al.*, 2020). Sin embargo, por motivos de su implementación y disponibilidad en el ecosistema de programación para el desarrollo de este trabajo, se optó por los primeros dos enfoques ordinales descritos con anterioridad.

## 2.6. CONCLUSIONES

Con base en los objetivos planteados, las arquitecturas de redes neuronales implementadas ayudaron a predecir el rezago social, ya que los mejores resultados promedio de prueba fueron una precisión global de 76.2%, error absoluto medio de 0.241 y *F1-macro* de 73.8%, alcanzados a través de regularización combinada; parada anticipada y L2. Por su parte, la comparación de los resultados con validación cruzada resultó no significativos por medio de una prueba paramétrica, prueba de *t* por pares, y una no paramétrica, Wilcoxon. Por lo tanto, no fue estadísticamente distinto abordar la predicción de las clases objetivo de rezago social con modelos que incluyeron un enfoque ordinal *versus* un enfoque nominal; con base en características de entrada provenientes de conteos de unidades económicas a nivel subsector para el año 2015. Sin embargo, de acuerdo a las visualizaciones presentadas y la naturaleza intrínseca del fenómeno de rezago social se presume adecuado considerar la ordinalidad de las clases objetivo para una modelación íntegra.

## CÓDIGO DE PROGRAMACIÓN

Los códigos de programación en Python 3, con el fin de replicar el contenido de este trabajo, se encuentran alojados, para uso libre, en el siguiente repositorio de GitHub:

[https://github.com/Roderich25/mac/tree/master/python\\_thesis](https://github.com/Roderich25/mac/tree/master/python_thesis).

Inicialmente, se generan los archivos que funcionan como entrada para los escenarios de entrenamiento descritos en el capítulo I, con el *script* 01\_inegi\_selenium.py; se consiguen los nombres de los archivos para cada estado y edición del DENUÉ de interés. Una vez que se obtienen los nombres de los archivos, se procede a descargarlos en formato CSV de la página oficial de descargas del INEGI (<https://www.inegi.org.mx/app/descarga/default.html>), esto se puede realizar tanto de forma síncrona con el código 02\_inegi\_denue.py, como de forma asíncrona con el código 02\_inegi\_denue\_async.py. Por medio del *script* 03\_denue\_dataframe.py; se procesan los archivos para almacenarlos en una base de datos tipo SQLite3. Posteriormente, con el *script* 04\_rezago\_dataframe.py, se agrupan los conteos de unidades económicas conforme a su categoría de actividad económica, por municipio en formato CSV, para obtener archivos con prefijo “denue\_wide\_” seguidos del número de dígitos manejados en la nomenclatura del SCIAN (2 para sector; 3 para subsector; 4 para rama; 5 para clase; y 6 para subclase).

En los *scripts* 05\_logistic\_regression.py, 06\_svm.py y 07\_random\_forest.py; se ajustan los modelos expuestos en el capítulo I, con base en los archivos generados en el *script* 04\_rezago\_dataframe.py. Mientras que, en los *scripts* 08\_dl\_nom.py, 09\_dl1\_ord.py y 10\_dl2\_ord; se muestran los códigos para entrenar los algoritmos especificados en el capítulo II.

En cada archivo, se detallan las bibliotecas y semillas de números aleatorios empleadas, siendo prominentes: Scikit-Learn para aprendizaje automático (Pedregosa *et al.*, 2011) en el capítulo I, y la biblioteca Keras; que es una API de alto nivel de Tensorflow, especializada en redes neuronales artificiales; en el capítulo II.

Cualquier duda o referencia de este trabajo, puede ser remitida al autor a través del correo electrónico: [rodrigoavilasolis@gmail.com](mailto:rodrigoavilasolis@gmail.com).

## CONCLUSIONES GENERALES

Con base en los objetivos planteados y los resultados obtenidos en los capítulos I y II, se muestra que es factible emplear los paradigmas de aprendizaje automático y aprendizaje profundo para realizar la predicción anual de los grados de rezago social en México a partir de datos de la estadística oficial de las unidades económicas activas en cada municipio para 2015; metodología que representa una nueva propuesta de estimación indirecta del rezago social a nivel municipal en México. Los modelos ajustados pueden ser utilizados para obtener estimaciones del grado de rezago social, para años subsecuentes, donde no se cuenta con estadísticas oficiales, pero se tienen datos actualizados de las unidades económicas de los municipios a través del Directorio Nacional de Unidades Económicas.

El clasificador bosques aleatorios, optimizado por medio de validación cruzada, se desempeñó mejor que el clasificador lineal de regresión logística y el clasificador no lineal máquina de soporte vectorial. Bosques aleatorios alcanzó una precisión global promedio de 74.4% y un valor de *F1-macro* promedio de 71.6%. Por su parte, una arquitectura de redes neuronales artificiales con enfoque ordinal y regularización combinada logró una precisión global promedio de 76.2% y un valor de *F1-macro* promedio de 73.8%, por lo que se considera ideal tomar en cuenta las clases objetivo de rezago social, con la naturaleza ordinal que guardan entre ellas.

Por último, de acuerdo con las visualizaciones que se mostraron en ambos capítulos, las predicciones de las clases de rezago social se distribuyeron geográficamente de forma similar a la distribución de las clases de rezago social consideradas como punto de referencia, y se comportaron prácticamente invariantes para los años posteriores donde no se contó con puntos de referencia.

## LITERATURA CITADA

- Alsharkawi A., M. Al-Fetyani, M. Dawas, H. Saadeh, and M. Alyaman. 2021. Poverty classification using machine learning: the case of Jordan. *Sustainability*. 13: 1-16. DOI: <https://doi.org/10.3390/su13031412>
- Ávila-Solís P. R., J. M. González-Camacho, P. Pérez-Rodríguez y D. Vargas-Chanes. 2021. Predicción del rezago social en México: un enfoque basado en aprendizaje automático a partir de datos de unidades económicas. *Agrociencia*, (en revisión).
- Azcárraga, A., and R. Setiono. 2018. Neural network rule extraction for gaining insight into the characteristics of poverty. *Neural Comput & Applic* 30, 2795–2806.
- Babenko, B., J. Hersh, D. Newhouse, A. Ramakrishnan, and T. Swartz. 2017. Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico. *Proceedings from NIPS 2017: Neural Information Processing Systems Workshop on Machine Learning for the Developing World*. Long Beach, CA.
- Bishop, C. M. 2006. *Pattern Recognition and machine learning*. Springer. pp: 291-294.
- Breiman L. 2001. Random Forests. *Mach. Learn.* 45: 5–32.
- Cao W., Mirjalili V., and Raschka S. 2020. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognit. Lett.*, 140: 325-331.
- Chai, K. M. A., H. T. Ng, and H. L. Chieu. 2002. Bayesian online classifiers for text classification and filtering. In: *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*. pp: 97-104.
- Chang L., and M. de Rijke. 2018. Incremental sparse Bayesian ordinal regression. *Neural Networks*. Volume 106: 294-302. DOI: <https://doi.org/10.1016/j.neunet.2018.07.015>.
- Cheng J., Z. Wang, and G. Pollastri. 2008. A neural network approach to ordinal regression *IEEE International Joint Conference on Neural Networks, IEEE World Congress on Computational Intelligence 2008*, pp. 1279-1284.
- Chollet, F. 2018. *Deep Learning with Python*. Manning Publications. pp: 27-60.
- CONEVAL (Consejo Nacional para la Evaluación de la Política de Desarrollo Social). 2016. Índice de Rezago Social 2015. Presentación de Resultados. [https://www.coneval.org.mx/Medicion/Documents/Indice\\_Rezago\\_Social\\_2015/Nota\\_Rezago\\_Social\\_2015\\_vf.pdf](https://www.coneval.org.mx/Medicion/Documents/Indice_Rezago_Social_2015/Nota_Rezago_Social_2015_vf.pdf) (Consulta: enero 2021).
- Cortes C., and V. Vapnik. 1995. Support-vector networks. *Mach Learn.* 20: 273–297.
- Dalenius T., and J. Hodges. 1959. Minimum variance stratification. *J. Am. Stat. Assoc.* 54: 88-101.

- Engstrom, R., J. Hersh, and D. Newhouse. 2017. Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being. World Bank Policy Research Working Paper No. 8284, Available at SSRN: <https://ssrn.com/abstract=3090770>
- Fawcett T. 2006 An introduction to ROC analysis. *Pattern Recognit. Lett.* 27: 861-874.
- Géron, A. 2019. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. 2a ed. O'Reilly Media. Sebastopol, CA. pp: 60-66.
- Glorot, X. and Y. Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research 9:249-256.
- INEGI (Instituto Nacional de Estadística y Geografía). 2013. Sistema de Clasificación. Industrial de América del Norte, México SCIAN 2013. [https://www.inegi.org.mx/contenidos/productos//prod\\_serv/contenidos/espanol/bvinegi/productos/clasificadores/SCIAN/SCIAN\\_2013/702825051693.pdf](https://www.inegi.org.mx/contenidos/productos//prod_serv/contenidos/espanol/bvinegi/productos/clasificadores/SCIAN/SCIAN_2013/702825051693.pdf) (Consulta: enero 2021).
- INEGI (Instituto Nacional de Estadística y Geografía). 2014. Marco geoestadístico 2014 versión 6.2 (DENUE). <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825004386> (Consulta: enero 2021).
- INEGI (Instituto Nacional de Estadística y Geografía). 2015. Directorio Nacional de Unidades Económicas DENUE. <https://www.inegi.org.mx/app/mapa/denue/default.aspx>
- James G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An introduction to statistical learning: with applications in R. Springer. pp: 215-219.
- Japkowicz, N. and S. Mohak. 2011. Evaluating Learning Algorithms: A Classification Perspective. pp 248-251.
- Kingma, D. and J. Ba. 2015. Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015).
- Lipton Z.C., C. Elkan, and B. Naryanaswamy. 2014. Optimal Thresholding of Classifiers to Maximize F1 Measure. In: Calders T., Esposito F., Hüllermeier E., Meo R. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2014. Lecture Notes in Computer Science, vol 8725. Springer, Berlin, Heidelberg. DOI: [https://doi.org/10.1007/978-3-662-44851-9\\_15](https://doi.org/10.1007/978-3-662-44851-9_15)
- McCullagh, P. and J. A. Nelder. 1989. Generalized Linear Models, volume 37 of Monographs on Statistics and Applied Probability. 2a Ed. Chapman and Hall. London, UK. 532 p.
- Mohamud J. H., and O. N. Gerek. 2019. Poverty level characterization via feature selection and machine learning. In: 27th Signal Processing and Communications Applications Conference, SIU 2019.

- Morales J. L. and J. Nocedal. 2011. Remark on "algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization". *ACM Trans Math Software*. 38: 1-4. DOI: <https://doi.org/10.1145/2049662.2049669>
- Naviamos, M. P., and J. D. Niguidula. 2020. A study on determining household poverty status: SVM based classification model. In: *ACM International Conference Proceeding Series*. pp: 79-84.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. 2011. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12: 2825-2830.
- Pérez-Rodríguez P, S. Flores-Galarza, H. Vaquera-Huerta, D. H. Del Valle-Paniagua, O. A. Montesinos-López, J. Crossa J. 2020. Genome-based prediction of Bayesian linear and non-linear regression models for ordinal data. *Plant Genome*. 13: 1-13. DOI: [10.1002/tpg2.20021](https://doi.org/10.1002/tpg2.20021)
- Poerwanto, B., and F. Fajriani. 2020. Resilient Backpropagation Neural Network on Prediction of Poverty Levels in South Sulawesi. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 11-18.
- Powell L. M., S. Slater, D. Mirtcheva, Y. Bao, and F. J. Chaloupka. 2007. Food store availability and neighborhood characteristics in the United States. *Prev. Med.* 44: 189-195.
- Raschka S. and V. Mirjalili. 2019. *Python Machine Learning*. 3a Ed. Packt Publishing Ltd. Birmingham, UK. pp: 90-96.
- Saito T., and M. Rehmsmeier 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 10(3). DOI: <https://doi.org/10.1371/journal.pone.0118432>
- Saito T., and M. Rehmsmeier. 2017. Precrec: fast and accurate precision–recall and ROC curve calculations in R. *Bioinformatics*. 33: 145-147.
- Sani N. S., M. A. Rahman, A. A. Bakar, S. Sahran, and H. M. Sarim. 2018. Machine learning approach for bottom 40 percent households (B40) poverty classification. *Int. J. Adv. Sci. Eng. Inf. Technol.* 8: 1698-1705.
- Stapor, K. 2017. Evaluation of classifiers: current methods and future research directions. *FedCSIS*.
- Vargas-Chanes D., and S. Valdés-Cruz. 2019. A longitudinal study of social lag: regional inequalities of growth in Mexico 2000 to 2015. *J. Chin. Sociol.* 6: 1-18. DOI: <https://doi.org/10.1186/s40711-019-0100-6>
- Wang L., J. Zhu, and H. Zou. 2006. The doubly regularized support vector machine. *Statistica Sinica*. 16: 589-615.