

Universidad Autónoma De Nuevo
León

Facultad de Ciencias Físico
Matemáticas

Un análisis
sociodemográfico de
Nuevo León

Ricardo Luna Escobedo

Introducción

Hay muchos temas que son de mucho interés hoy en día: salud, cambio climático, educación, natalidad, mortalidad, transición política, etc. Pero de todos, el que más me ha causado interés es uno en particular: **Salario**.

El salario está definido como una retribución económica que un trabajador recibe a cambio por la prestación de sus servicios laborales. Se trata de la cantidad de dinero que el trabajador recibe de su empleador, generalmente de forma periódica, a cambio del trabajo que realiza.

El salario tiene un origen que proviene de la necesidad. En tiempos antiguos, recordemos que no existía una moneda cambio; en su lugar la gente recurría al trueque. Este método cambio se fue volviendo obsoleto con el pasar de los años. Ya que si una persona quería conseguir ganado, necesitaba conseguir otro artículo en particular, y podía pasar toda una vida entera buscándolo sin tener éxito. Es por ello por lo que luego se introdujo el pago con semillas, sales, animales... hasta llegar a la formalidad de la moneda.

He oído que mucha gente siempre ha dicho comentarios como los siguientes: mientras más estudies más dinero ganas, mientras más trabajes más dinero ganas, el salario mensual depende del género, el salario mensual depende, etc. Pero como podemos ver, todo lo anterior es pura opinión. No podemos realizar inferencias solamente en base a opiniones, si queremos afirmar alguna hipótesis, o teoría, necesitamos recurrir a una herramienta construida en base a la lógica y a la exactitud; es por ello por lo que existen las **matemáticas**.

Afortunadamente, hoy en día contamos con increíbles avances matemáticos y con nuevas tecnologías que nos permiten realizar cálculos de manera más rápida y eficaz, entre ellos: **Aprendizaje automático**. Ésta es una rama de la inteligencia artificial que se centra en desarrollar sistemas capaces de aprender y mejorar automáticamente a partir de la experiencia sin ser programados explícitamente para ello. En lugar de seguir reglas predefinidas, los algoritmos de aprendizaje automático identifican patrones en los datos para hacer predicciones, clasificaciones o tomar decisiones.

Y si hablamos del Aprendizaje automático, podemos recurrir a una subrama que se encarga utilizar redes neuronales artificiales para modelar y analizar datos complejos. Este enfoque está inspirado en cómo funciona el cerebro humano para procesar información.

El propósito de este trabajo será utilizar la Encuesta Nacional de Ocupación y Empleo (ENOE) al 2 trimestre del 2024, proporcionada por el Instituto Nacional de Estadística y Geografía (INEGI), y mediante la técnica Autoaprendizaje competitivo del Aprendizaje profundo, construir un modelo clustering que nos permita determinar, a partir de ciertas características, si una persona es asignada con un salario bajo, alto o promedio.

Análisis descriptivo

Cabe mencionar que hay variables que fueron precodificadas en la ENOE de tal forma que podemos asegurar nuestra muestra el enfoque correcto. Si estamos hablando de salario, estamos hablando de una persona que recibe ingresos por ejercer una actividad laboral.

En los datos precodificados nos aseguramos de trabajar con la menor cantidad de datos nulos posibles, nos enfocamos en que los encuestados formen parte de la población económicamente activa, y que tengan la edad mínima legal para trabajar (15 años).

Variable	Nombre	Descripción contenido
ent	Estado residencia	Cualitativa 1 (Aguascalientes) 2 (Baja California) 3 (Baja California Sur) 4 (Campeche) 5 (Coahuila de Zaragoza) 6 (Colima) 7 (Chiapas) 8 (Chihuahua) 9 (Ciudad de México) 10 (Durango) 11 (Guanajuato) 12 (Guerrero) 13 (Hidalgo) 14 (Jalisco) 15 (México) 16 (Michoacán de Ocampo) 17 (Morelos) 18 (Nayarit) 19 (Nuevo León) 20 (Oaxaca) 21 (Puebla) 22 (Querétaro) 23 (Quintana Roo) 24 (San Luis Potosí) 25 (Sinaloa) 27 (Tabasco) 28 (Tamaulipas) 29 (Tlaxcala) 30 (Veracruz de Ignacio de la Llave) 31 (Yucatán) 32 (Zacatecas)

sex	Sexo	Cualitativa 1 (Hombre) 2 (Mujer)
eda	Edad	Cuantitativa 15 – 92 (Número de años cumplidos del integrante del hogar)
cs_p13_1	Nivel escolar hasta el que estudió	Cualitativa 0 (Ninguna) 2 (Primaria) 3 (Secundaria) 4 (Preparatoria) 5 (Normal) 6 (Carrera técnica) 7 (Profesional) 8 (Maestría) 9 (Doctorado)
cs_p13_2	Años aprobados en el nivel escolar	Cuantitativa 0 – 6 (Años aprobados según el nivel escolar)
n_hij	Total hijos o hijas que nacieron vivos	Cualitativa 0 (Ninguno) 1 – 12 (Número de hijos o hijas) 99 No especificado
e_con	Estado conyugal	Cualitativa 1 (Vive con su pareja en unión libre) 2 (Separado(a)) 3 (Divorciado(a)) 4 (Viudo(a)) 5 (Casado(a)) 6 (Soltero(a))
hrsocup	Horas trabajadas en la semana	Cuantitativa 0 - 119 (Número de horas trabajadas durante la semana)

ingocup	Ingreso mensual	<p>Cuantitativa</p> <p>0 – 219000 (Ingreso mensual capturado)</p>
scian	<p>Clasificación de actividades económicas con base en el sistema de clasificación industrial de América del norte</p>	<p>Cualitativa</p> <p>1 (Agricultura, ganadería, aprovechamiento forestal, pesca y caza)</p> <p>2 (Minería)</p> <p>3 (Generación y distribución de electricidad, suministro de agua y gas)</p> <p>4 (Construcción)</p> <p>5 (Industrias manufactureras)</p> <p>6 (Comercio al por mayor)</p> <p>7 (Comercio al por menor)</p> <p>8 (Transportes, correos y almacenamiento)</p> <p>9 (Información en medios masivos)</p> <p>10 (Servicios financieros y de seguros)</p> <p>11 (Servicios inmobiliarios de alquiler de bienes)</p> <p>12 (Servicios profesionales, científicos y técnicos)</p> <p>13 (Corporativos)</p> <p>14 (Servicios de apoyo a los negocios y manejo de desechos)</p> <p>15 (Servicios educativos)</p> <p>16 (Servicios de salud y asistencia social)</p> <p>17 (Servicios de esparcimiento culturales y deportivos)</p> <p>18 (Servicios de hospedaje y preparación de alimentos y bebidas)</p> <p>19 (Otros servicios, excepto actividades gubernamentales)</p> <p>20 (Actividades gubernamentales y de organismos internacionales)</p>
t_tra	Total de trabajos	<p>Cualitativa</p> <p>1 (Un solo trabajo)</p> <p>2 (Dos trabajos)</p>

Para los análisis siguientes, se utilizaron 2 lenguajes de programación: Python y Spark. Buscaremos determinar si una persona tiene un salario bajo (2), alto (1) o medio (3) tomando en cuenta las variables sociodemográficas anteriores.

Los análisis consisten en 2 comparativas:

1. Asignar los niveles salariales tomando solo en cuenta las divisiones marcadas por el INEGI (Spark) (asignación tradicional).
2. Asignar los niveles salariales tomando en cuenta todas las variables mediante el método Autoaprendizaje competitivo.

Spark

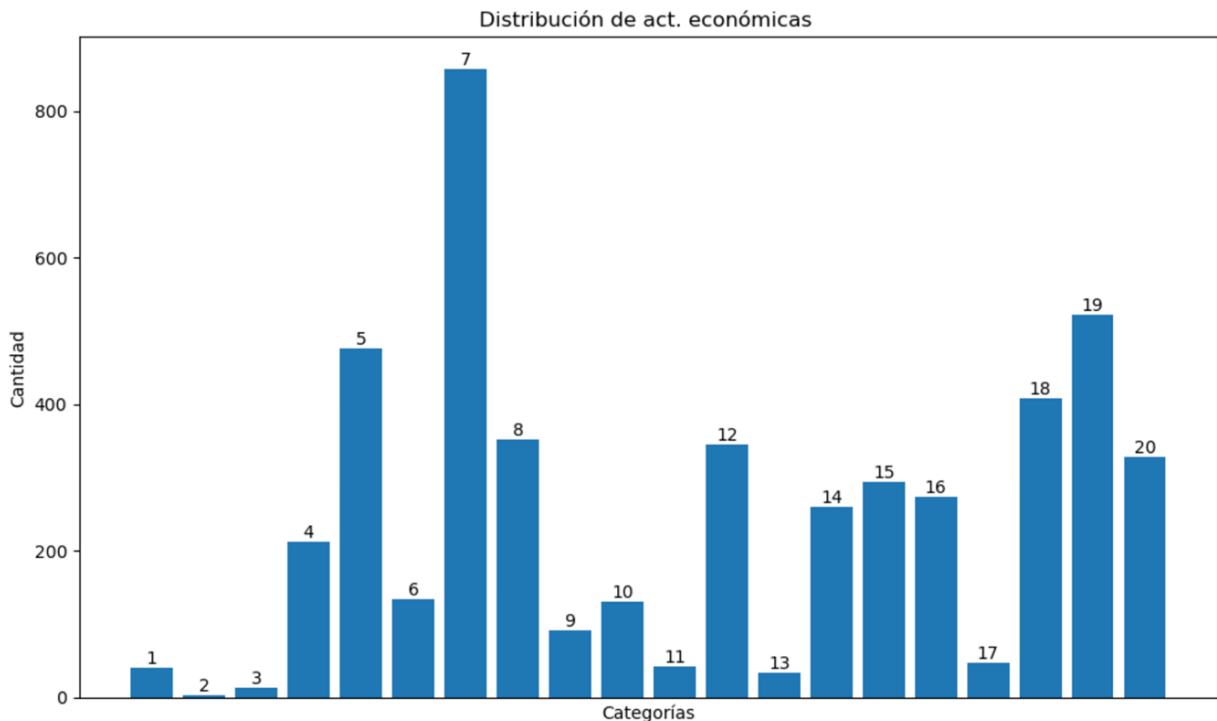
Podemos comenzar observando una imagen general del conjunto. Pero recordemos que el análisis solo será para Nuevo León. Por lo que filtramos para únicamente tomar la entidad "9". Para propósitos más sencillos, eliminaremos la columna "ent" de la visualización pero no del filtrado, e imprimiremos el total de encuestados.

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|eda|n_hij|cs_p13_2|hrsocup|ingocup|sex|t_tra|cs_p13_1|e_con|scian|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 33|    2|    1|    30|  20000|  2|    1|    7|    5|   12|
| 52|    2|    3|    0|    0|    2|    1|    6|    5|    4|
| 24|    0|    6|   66|   5160|  2|    1|    2|    6|   18|
| 24|    0|    5|   30|  10000|  2|    1|    7|    6|   12|
| 50|    2|    3|   25|    0|    2|    1|    6|    5|    7|
| 52|    2|    3|    0|   7000|  2|    1|    4|    5|   15|
| 63|    2|    4|   24|    0|    2|    1|    7|    5|   12|
| 27|   99|    4|   30|    0|    1|    1|    7|    6|    4|
| 17|   99|    6|   36|    0|    1|    1|    2|    6|   18|
| 38|    0|    3|   16|    0|    2|    1|    8|    6|   20|
| 33|   99|    4|   24|    0|    1|    1|    7|    6|    9|
| 25|   99|    3|   27|  24000|  1|    1|    7|    6|   14|
| 31|    0|    3|   24|    0|    2|    1|    7|    6|   14|
| 24|   99|    3|   54|   9000|  1|    1|    4|    5|    6|
| 57|   99|    3|   40|   3870|  1|    1|    3|    1|   19|
| 29|    0|    4|   54|  12000|  2|    1|    7|    1|   16|
| 34|    1|    4|    0|  25000|  2|    1|    7|    1|   10|
| 38|   99|    2|   30|   8600|  1|    2|    4|    1|    5|
| 35|   99|    3|   54|    0|    1|    1|    4|    6|    7|
| 38|    2|    3|    0|   6450|  2|    1|    6|    1|   19|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
```

4870

Tenemos un total de 4870 individuos neoloneses que contestaron la ENOE.

Situación laboral en Nuevo León



```
[Row(scian=7, count=858),  
Row(scian=19, count=522),  
Row(scian=5, count=477),  
Row(scian=18, count=408),  
Row(scian=8, count=352),  
Row(scian=12, count=345),  
Row(scian=20, count=328),  
Row(scian=15, count=295),  
Row(scian=16, count=273),  
Row(scian=14, count=261),  
Row(scian=4, count=213),  
Row(scian=6, count=134),  
Row(scian=10, count=131),  
Row(scian=9, count=91),  
Row(scian=17, count=47),  
Row(scian=11, count=42),  
Row(scian=1, count=41),  
Row(scian=13, count=34),  
Row(scian=3, count=14),  
Row(scian=2, count=4)]
```

De acuerdo con la gráfica y la agrupación. La actividad económica marcada en “7” (Comercio al por menor), es la que tiene registrada más trabajadores en Nuevo León con un total de 858.

Recordemos que el comercio al por menor son todos aquellos establecimientos como: HEB, Soriana, OXXO, 7-eleven, etc.

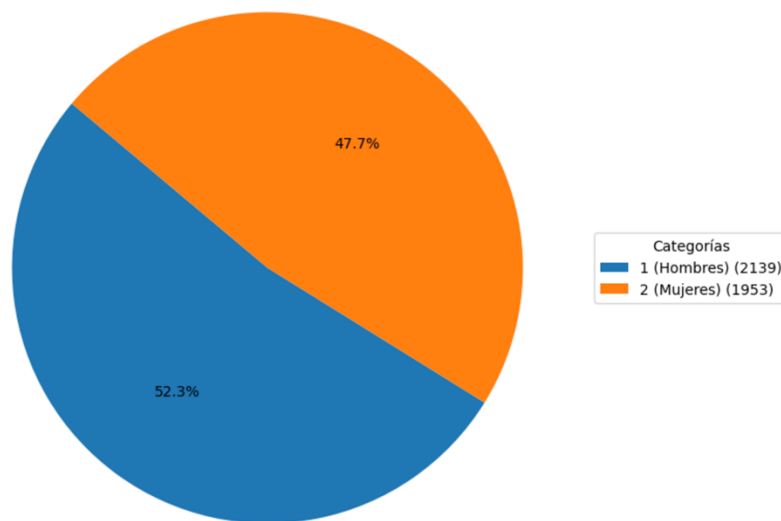
Vamos a dividir a los individuos que trabajan en “comercio al por menor”, de acuerdo con divisiones establecidas por el INEGI, de la siguiente manera.

- Salario baja \leq \$11,343 MXN mensuales
- Salario media \geq \$11,343 MXN mensuales y \leq \$77,975 MXN mensuales
- Salario alta \geq \$77,975 MXN mensuales

Y hagamos una gráfica pastel de cómo están distribuidos sus salarios de acuerdo con los sexos.

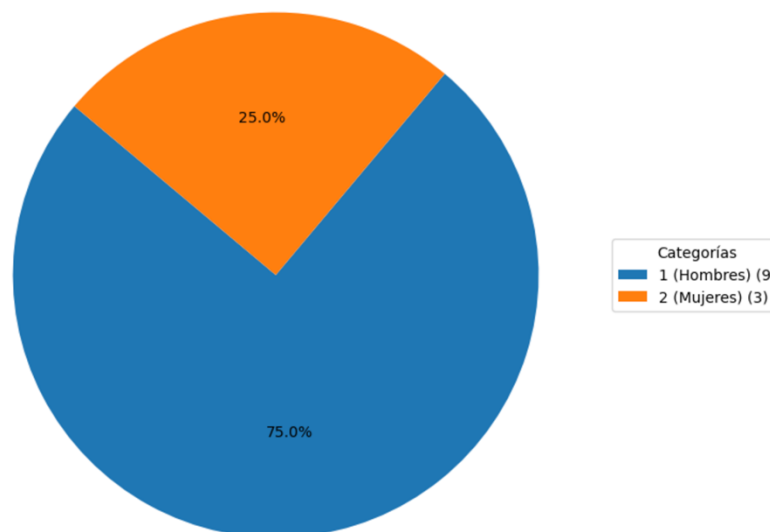
Clase baja

Distribución de valores en la columna 'sex'



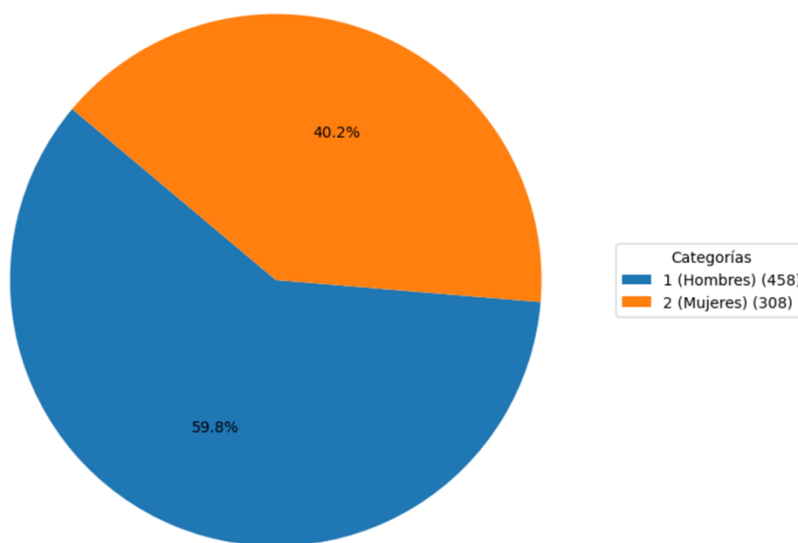
Distribución de valores en la columna 'sex'

Clase alta



Clase media

Distribución de valores en la columna 'sex'



Dentro de la actividad económica más ejercida, vemos como en los 3 niveles el hombre es el que sobresale en ocupación. Esto implica que el hombre es más contratado que la mujer, inclusive dentro de las subactividades que tengan un salario menos remunerado.

Asignación tradicional

Recordemos que en la asignación tradicional se emplearán las divisiones salariales de acuerdo con lo establecido por el INEGI:

- Salario bajo (2).
- Salario alto (1).
- Salario medio (3).

Aquí la impresión de las asignaciones:

```
+-----+-----+-----+
|Clase_1|Clase_2|Clase_3|
+-----+-----+-----+
|      12 |    4092 |    766 |
+-----+-----+-----+
```

Podemos ver que el salario bajo es el más dominante en todos los individuos neoloneses que tienen su salario registrado.

Ahora observemos la ocupación que tienen los individuos con salario bajo en sus respectivas actividades económicas.

Pertenencia	eda	n_hij	cs_p13_2	hrsocup	ingocup	sex	t_tra	cs_p13_1	e_con	scian
2	52	2	3	0	0	2	1	6	5	4
2	24	0	6	66	5160	2	1	2	6	18
2	24	0	5	30	10000	2	1	7	6	12
2	50	2	3	25	0	2	1	6	5	7
2	52	2	3	0	7000	2	1	4	5	15
2	63	2	4	24	0	2	1	7	5	12
2	27	99	4	30	0	1	1	7	6	4
2	17	99	6	36	0	1	1	2	6	18
2	38	0	3	16	0	2	1	8	6	20
2	33	99	4	24	0	1	1	7	6	9
2	31	0	3	24	0	2	1	7	6	14
2	24	99	3	54	9000	1	1	4	5	6
2	57	99	3	40	3870	1	1	3	1	19
2	38	99	2	30	8600	1	2	4	1	5
2	35	99	3	54	0	1	1	4	6	7
2	38	2	3	0	6450	2	1	6	1	19
2	35	0	4	70	0	2	1	7	6	7
2	50	2	4	42	0	2	1	7	5	6
2	51	99	3	60	9460	1	1	3	5	8
2	17	0	2	18	8000	2	1	4	6	7

only showing top 20 rows

scian	count
7	785
19	483
5	413
18	364
8	275
12	269
20	247
15	224
14	220
16	208
4	176
6	118
10	100
9	67
1	38
17	37
11	35
13	21
3	10
2	2

De acuerdo con la gráfica y la agrupación. La actividad económica marcada en “7” (Comercio al por menor), es la que tiene registrada más trabajadores en Nuevo León con un salario bajo; con un total de 785.

Son datos alarmantes, porque vemos que el trabajo con más ocupación es el trabajo con la mayor cantidad de trabajadores que tienen asignado un salario bajo.

Estos son indicios de que Nuevo León tiene altos números pobreza.

Python

Asignación Autoaprendizaje competitivo

El autoaprendizaje competitivo es un modelo utilizado en aprendizaje profundo que consiste en realizar divisiones de grupos (**Neuronas**) de acuerdo con las características de la variable que queremos analizar, pero tomando en cuenta también las demás variables de la base datos.

Como queremos estudiar los 3 niveles de salario existente, en el primer grupo asignaremos los valores máximos de todas las variables sociodemográficas; en el segundo grupo, asignaremos los valores mínimos de todas las variables sociodemográficas; y en el tercer grupo, asignaremos los valores promedio de todas las variables sociodemográficas.

El modelo consiste en que vamos a sacar la distancia euclidiana de cada uno de los datos registrados en la base datos, contra cada uno de los valores guardados en los 3 grupos anteriores. De tal modo que, dependiendo cuál de las 3 distancias es la menor, es la asignación salarial que se le dará al individuo. Si la distancia euclidiana entre el primer valor que hay en la base datos y el primer valor del grupo valores máximos, es el menor, entonces a ese individuo se le asigna un salario alto; Si la distancia euclidiana entre el primer valor que hay en la base datos y el primer valor del grupo valores mínimos, es el menor, entonces a ese individuo se le asigna un salario bajo; Si la distancia euclidiana entre el primer valor que hay en la base datos y el primer valor del grupo valores medios, es el menor, entonces a ese individuo se le asigna un salario medio. Y así para cada uno de los demás valores registrados en los grupos y en la base datos.

Fórmula matemática del modelo

$$\text{Min}_{c_k} E = \sum_{p=1}^N \sum_{k=1}^K \mu_{kp} |x_p - c_k|^2$$

Como manejamos variables de distinto tipo, debemos de escalarlas. Para ello, a las variables cuantitativas las estandarizamos, y a las cualitativas les aplicamos el método one-hot encoding. El método one-hot encoding también es muy útil en variables cualitativas para que el modelo no tenga inconvenientes a la hora de leer los datos ingresados.

	eda	n_hij	cs_p13_2	hrsocup	ingocup	sex	t_tra	cs_p13_1_0	cs_p13_1_2	cs_p13_1_3	...	scian_11	scian_12	scian_13	scian_14	scian_15	scian_16	scian_17
1	-0.651373	-1.061371	-1.941007	-0.567502	1.286999	2	1	0	0	0	...	0	1	0	0	0	0	0
2	0.682275	-1.061371	-0.271803	-2.174480	-0.504793	2	1	0	0	0	...	0	0	0	0	0	0	0
3	-1.283101	-1.102469	2.232004	1.360872	-0.042510	2	1	0	1	0	...	0	0	0	0	0	0	0
4	-1.283101	-1.102469	1.397402	-0.567502	0.391103	2	1	0	0	0	...	0	1	0	0	0	0	0
5	0.541891	-1.061371	-0.271803	-0.835332	-0.504793	2	1	0	0	0	...	0	0	0	0	0	0	0
...
4866	-0.861949	-1.081920	-0.271803	0.396685	2.098681	2	1	0	0	0	...	0	0	0	0	0	0	0
4867	-0.230221	-1.081920	0.562800	-2.174480	-0.504793	2	1	0	0	0	...	0	0	0	0	1	0	0
4868	-0.721565	-1.081920	0.562800	0.396685	-0.504793	2	1	0	0	0	...	0	0	0	0	0	0	0
4869	0.261123	-1.081920	0.562800	-0.299672	-0.504793	2	1	0	0	0	...	0	0	0	0	1	0	0
4870	-0.019645	0.931891	-0.271803	1.039476	-0.504793	1	1	0	0	1	...	0	0	0	0	0	0	0

Recordemos que el método one-hot encoding lo que hace es que los elementos de las variables cualitativas se dividen en columnas, y por cada que se presente alguno de sus elementos se le asigna a la respectiva columna un 1, por cada que no se presente se le asigna un 0.

Originalmente teníamos solo 10 variables, ahora tenemos 42 columnas.

Tras aplicar el modelo autoaprendizaje competitivo, tenemos los siguientes resultados para las asignaciones de los individuos.

	Pertenencia	eda	n_hij	cs_p13_2	hrsocup	ingocup	sex	t_tra	cs_p13_1	e_con	scian
1	3.0	33	2	1	30	20000	2	1	7	5	12
2	3.0	52	2	3	0	0	2	1	6	5	4
3	3.0	24	0	6	66	5160	2	1	2	6	18
4	3.0	24	0	5	30	10000	2	1	7	6	12
5	3.0	50	2	3	25	0	2	1	6	5	7
...
4866	2.0	30	1	3	48	29060	2	1	4	2	18
4867	3.0	39	1	4	0	0	2	1	7	1	15
4868	2.0	32	1	4	48	0	2	1	7	1	7
4869	1.0	46	1	4	35	0	2	1	7	5	15
4870	3.0	42	99	3	60	0	1	1	3	1	8

	Neurona 1	Neurona 2	Neurona 3
1	1461.0	1862.0	1547.0

Vemos que el modelo autoaprendizaje competitivo hizo ahora más balanceada la asignación de los grupos salarios a comparación del método tradicional. Sin embargo, vemos que aun así el grupo que más domina en Nuevo León es el de la gente con salario bajo.

Ahora observemos la ocupación que tienen los individuos con salario bajo en sus respectivas actividades económicas.

		Pertenencia	eda	n_hij	cs_p13_2	hrsocup	ingocup	sex	t_tra	cs_p13_1	e_con	scian	
		102	2.0	58	99	3	40	0	1	1	4	6	14
		105	2.0	28	3	2	40	0	2	1	4	1	7
		109	2.0	53	4	3	27	2580	2	1	3	4	19
		115	2.0	39	99	5	40	0	1	1	7	1	9
		117	2.0	50	2	2	40	9000	2	1	6	4	20
	
		4858	2.0	47	99	3	50	17000	1	1	4	5	20
		4864	2.0	26	0	4	40	0	2	1	7	6	12
		4865	2.0	24	2	3	24	5160	2	1	3	6	7
scian		4866	2.0	30	1	3	48	29060	2	1	4	2	18
7	328	4868	2.0	32	1	4	48	0	2	1	7	1	7
5	206												
19	201												
18	143												
8	133												
12	124												
16	114												
20	111												
15	106												
14	93												
4	81												
6	61												
10	52												
9	33												
1	20												
13	19												
17	14												
11	12												
3	9												
2	2												
..													

De acuerdo con la gráfica y la agrupación. La actividad económica marcada en “7” (Comercio al por menor), es la que tiene registrada más trabajadores en Nuevo León con un salario bajo; con un total de 328.

Siguen siendo datos alarmantes, porque vemos que el trabajo con más ocupación es el trabajo con la mayor cantidad de trabajadores que tienen asignado un salario bajo. Hay altos índices de pobreza en Nuevo León.

Podemos ver como el método autoaprendizaje competitivo es más exacto en cuanto a resultados, comparado con el método tradicional. Ya que el primero sí tomó en cuenta a todas las variables para el análisis, mientras que el segundo solo tomó en cuenta al salario mensual.

De acuerdo con la gráfica y la agrupación. La actividad económica marcada en “7” (Comercio al por menor), es la que tiene registrada más trabajadores en Nuevo León con un salario bajo; con un total de 328.

Siguen siendo datos alarmantes, porque vemos que el trabajo con más ocupación es el trabajo con la mayor cantidad de trabajadores que tienen asignado un salario bajo. Hay altos índices de pobreza en Nuevo León.

Podemos ver como el método autoaprendizaje competitivo es más exacto en cuanto a resultados, comparado con el método tradicional. Ya que el primero sí tomó en cuenta a todas las variables para el análisis, mientras que el segundo solo tomó en cuenta al salario mensual.