



Diseño y desarrollo de una red neuronal de base radial híbrida para la clasificación de los clientes de una PYME: Un caso de estudio

Jonathan Bedolla Guzmán
Departamento de Estudios Multidisciplinarios
División de ingenierías, Campus Irapuato-Salamanca
Universidad de Guanajuato
Av. Universidad S/N, Col. Yacatitas, C.P. 36400, Yuriria, Gto.
j.bedollaguzman@ugto.mx

Resumen

El objetivo del presente trabajo fue desarrollar una red neuronal de función radial híbrida. Se utilizaron los algoritmos K-medias y Backpropagation. Se clasificaron los principales clientes de una PYME en un plano bidimensional. El caso de estudio se centra en una mediana empresa dedicada a la comercialización de bebidas y alimentos en el interior del negocio y con entregas a domicilio, con operaciones en la ciudad de Uriangato al sur de Guanajuato. La metodología consiste en 7 etapas: 1) Obtención de coordenadas geográficas de los clientes por medio de Google Maps, 2) Preprocesamiento de datos, 3) Cálculo de distancias euclidianas, 4) Clasificación con algoritmo K-Medias, 5) Entrenamiento de la red por medio del algoritmo Backpropagation, 6) Validación de la red, 7) Prueba de la red. La arquitectura de la red fue 2-3-1. Se utiliza el 70% de los datos para entrenar a la red, el 20% de los datos para validarla y el 10% para probar su funcionamiento. Como resultado se obtiene un aprendizaje de la red al 100%, una validación del 90% y una prueba de 100%. La red neuronal de función radial facilita una clasificación automatizada por zonas geográficas de los principales clientes y ofrece a la empresa en estudio ventajas en la implementación de estrategias dirigidas a la optimización de rutas que reducen considerablemente los tiempos y los costos de entrega.



Introducción

Las Pymes tienen que estar mejorando constantemente o innovando tanto sus procesos como sus productos para ser más eficientes y efectivas en el mercado en el cual participan. En el estado de Guanajuato se encuentran 108,616 unidades económicas comerciales, de las cuales el 99.8% son MIPyMES y representan el 88.02% del personal ocupado en este sector (INEGI, 2014).

El presente estudio se centra en una mediana empresa con giro de cafetería tipo Gourmet llamada “El Barista”, se encuentra ubicada en el centro de la ciudad de Uriangato, al sur de Guanajuato y se caracteriza por ser la empresa cafetera líder en el mercado local y regional debido a sus altas ventas y a la gran demanda de pedidos tanto en el interior del negocio como en las entregas a domicilio. La empresa tiene una gran cantidad de clientes registrados que constantemente le realizan pedidos de sus productos a domicilio y al mismo tiempo, lo anterior llega a representar una problemática para la empresa afectando las operaciones de repartición y causando un descontrol en las actividades de entrega.

El objetivo central fue diseñar y desarrollar una red neuronal de base radial con aprendizaje híbrido para clasificar en 3 zonas geográficas a los principales 50 clientes del negocio, agrupándolos como clústers de manera precisa.

Se utilizó el algoritmo K-medias como clasificador para formar los grupos correspondientes de clientes y mostrarlos como clústers en un plano bidimensional.

Se desarrolló el algoritmo Backpropagation para automatizar la clasificación de clientes nuevos tomando como base los centros de clasificación obtenidos por el algoritmo k-medias.

Las redes neuronales artificiales se han convertido en soluciones muy buscadas en varias áreas que requieren inteligencia artificial en tareas como el reconocimiento de patrones y la clasificación (Mohamed B., 2015). Las redes neuronales de función de base radial (RBF) ofrecen grandes posibilidades para resolver problemas de procesamiento de señales y clasificación de patrones (George, 2007). La red neuronal con función de base radial se utiliza por sus ventajas de entrenamiento rápido, generalidad y simplicidad sobre la retroalimentación hacia adelante (Hung C., 2002). Comparando la red de función radial RBF con la red perceptrón multicapa MLP. (Hjouji A., 2019) afirma que el enfoque RBF es más intuitivo que el MLP. La formación de redes Radial



Basis Function (RBF) suele ser más rápida que la formación de redes de perceptrón multicapa (MLP) (Kumar R., 2016).

En la presente investigación se implementa una red neuronal de función radial debido a que se caracteriza por tener mayor rapidez en su entrenamiento comparada con otras redes, además de facilitar la clasificación de una gran cantidad de datos. El objetivo del estudio es solucionar los problemas en la logística de salida del negocio en estudio de manera que la red neuronal agrupe de forma automática a los principales clientes y los clasifique en tres zonas geográficas graficándolos como clústers. El desarrollo de la red neuronal de base radial busca impactar en la empresa en estudio mejorado las entregas a domicilio, optimizando los viajes, minimizando las distancias recorridas, reduciendo tiempos, combustibles y en general los costos de entrega de la empresa.

Objetivos

General

El objetivo general de la presente investigación fue diseñar y desarrollar una red neuronal de base radial híbrida para clasificar en tres zonas geográficas a los 50 clientes de la empresa “El Barista”.

Específicos

Los objetivos específicos propuestos para el presente estudios fueron los siguientes:

- a) Optimizar las rutas de reparto para lograr una mejor eficiencia operativa
- b) Minimizar las distancias recorridas para reducir las emisiones de CO₂
- c) Minimizar los tiempos de entrega de los productos para incrementar el nivel de satisfacción de los clientes

Justificación

El desarrollo y aplicación de la presente herramienta impacta en la parte operativa de la empresa en estudio debido a que esta se encuentra en constante cambio y adaptación a las nuevas tecnologías para sus procesos industriales. El incluir herramientas de inteligencia artificial como lo son las redes neuronales en sus operaciones diarias le genera un valor agregado que les garantiza mejores resultados y beneficios duraderos con una sola implementación. Una red neuronal de base radial que clasifique al 100% sus clientes por zonas geográficas le proporcionan una ventaja sobre sus competidores. El funcionamiento y los resultados de la herramienta se muestran claros y concisos, son fáciles de interpretar y son de fácil adaptabilidad. Una de las



ventajas de la herramienta es que una vez desarrollada y entrenada es posible ingresarle datos de clientes nuevos que necesiten ser clasificados y la red neuronal entrenada los clasificará automáticamente sin necesidad de comenzarla desde cero.

La automatización en pequeñas, medianas y grandes empresas es cada vez más indispensable para sus procesos operativos y la incursión de redes neuronales en dichos procesos es la mejor estrategia para conseguir los resultados esperados.

Metodología

a) Ubicación geográfica de los clientes por medio de Google Maps

Se obtuvieron las coordenadas de latitud y longitud de cada uno de los clientes usando la herramienta de Google Maps. Se analizaron 50 clientes de los más significativos para la empresa en estudio. Los datos se utilizaron como entradas para el entrenamiento de la red.

b) Preprocesamiento de los datos

Los datos se sometieron a una codificación en una escala de 2.9 y 0.1 con el objetivo de ayudar a la red a que funcionara de mejor manera y mejorar la visibilidad de los puntos al momento de graficar.

c) Cálculo de distancias euclidianas

Se calcularon las distancias euclidianas entre las entradas y las neuronas en la capa oculta. Se conectan los patrones de entrada con su entorno dentro de la red y se aplica la función radial para obtener los valores de las neuronas en la capa oculta.

d) Clasificación por algoritmo K-Medias

Se estableció la clasificación por zonas de los 50 clientes en estudio y se graficaron los puntos en un plano bidimensional.

e) Entrenamiento de la red por medio del algoritmo Backpropagation

Se resolvió la segunda mitad de la red por medio del algoritmo backpropagation. Se utilizaron 35 clientes (70% de los datos) para entrenar la red.

f) Validación de la red

Se utilizaron 10 clientes (20% de los datos) para validar la red.

g) Prueba de la red

Se utilizaron los 5 clientes restantes (10% de los datos) para probar la red.



Para el desarrollo del presente estudio se diseñó una red neuronal de función radial con el objetivo de clasificar a los principales clientes de una mediana empresa al sur de Guanajuato en tres zonas geográficas. La arquitectura de la red fue 2-3-1. Ver figura 1.

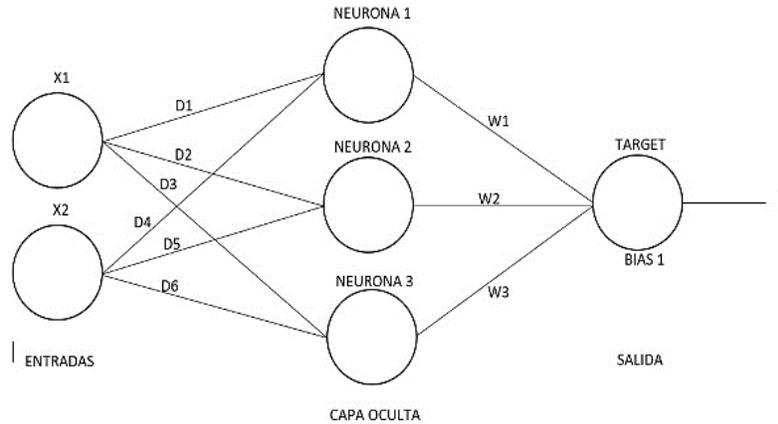


Fig. 1. Arquitectura de la red neuronal 2-3-1

En la presente investigación se habla sobre una Red Neuronal de Base Radial “Híbrida”. Se le llama así porque se resuelve por medio de la combinación de dos algoritmos diferentes; K-medias y Backpropagation.

Para desarrollar la red fue necesario obtener las coordenadas de los clientes por medio de la herramienta de Google Maps, se extrajeron los datos correspondientes a latitud y longitud y se representaron como coordenadas X_1 y X_2 .

Para poder fungir como patrones de entrada de la red los datos deben de cumplir con los requerimientos de codificación. Para esto se hizo un preprocesamiento de los 50 datos y se codificaron en una escala de entre 2.9 y 0.1. Se muestran solo 5 de los 50 datos en estudio. Ver tabla 1.

Tab. 1. Preprocesamiento de datos

| Cientes | X_1 | X_2 | 2.9 | 0.1 | | 2.9 | 0.1 |
|---------|-----------|----------|-----------|-----------|---|------|-----|
| 1 | 20.152849 | -101.171 | pendiente | ordenada | 1 | 2.31 | 2.4 |
| 2 | 20.141199 | -101.178 | 50.36 | -1,012.62 | 2 | 1.73 | 2.2 |
| 3 | 20.142308 | -101.188 | | | 3 | 1.78 | 1.7 |
| 4 | 20.124337 | -101.184 | pendiente | ordenada | 4 | 0.88 | 1.9 |
| 5 | 20.140574 | -101.183 | 45.45159 | 4600.854 | 5 | 1.69 | 1.9 |



Al contar con los datos codificados se procede al entrenamiento de la primera parte de la red que se resuelve por medio del algoritmo K- Medias. Se calcularon las distancias euclidianas entre los nodos y los centroides correspondientes al número de neuronas en la capa oculta. Se utilizaron 35 clientes para entrenar la red, 10 clientes para validarla y 5 para probarla.

A. Red Neuronal de Base Radial con estructura 2-3-1

El algoritmo K-Medias establece el cálculo de un determinado número de iteraciones con el objetivo de encontrar aquellas en las que se obtenga la menor variación entre las distancias euclidianas y los centros correspondientes.

Lo que se busca con el desarrollo del algoritmo K-medias es minimizar las distancias entre los clientes y los centros establecidos. Es aquí donde se lleva a cabo una parte muy importante de la red pues las variables de entrada se conectan con el entorno de la red neuronal permitiendo el cálculo de las neuronas en la capa oculta y aplicando la función radial en las mismas.

Para el entrenamiento de la red, los cálculos del algoritmo K-Medias terminan después de la sexta iteración puesto que en la quinta iteración y en esta última se obtienen los mismos resultados para los centros y las distancias euclidianas. Ver tabla. 2.

Tab. 2. Iteraciones 5 y 6 de la red 2-3-1

| ITERACIÓN 5 | | | | | |
|-------------|------|----------|------|----------|------|
| CENTRO 1 | | CENTRO 2 | | CENTRO 3 | |
| 2.40 | 2.51 | 0.87 | 2.36 | 1.56 | 1.60 |
| 0.11 | 1 | 1.45 | 0 | 1.13 | 0 |
| 0.76 | 0 | 0.88 | 0 | 0.58 | 1 |
| 1.02 | 0 | 1.13 | 0 | 0.24 | 1 |
| 1.65 | 0 | 0.48 | 1 | 0.74 | 0 |
| 0.92 | 0 | 0.94 | 0 | 0.35 | 1 |
| ITERACIÓN 6 | | | | | |
| CENTRO 1 | | CENTRO 2 | | CENTRO 3 | |
| 2.40 | 2.51 | 0.82 | 2.41 | 1.55 | 1.61 |
| 0.11 | 1 | 1.49 | 0 | 1.12 | 0 |
| 0.76 | 0 | 0.94 | 0 | 0.57 | 1 |
| 1.02 | 0 | 1.20 | 0 | 0.25 | 1 |
| 1.65 | 0 | 0.54 | 1 | 0.72 | 0 |
| 0.92 | 0 | 1.00 | 0 | 0.34 | 1 |



Después de obtener las igualdades en las iteraciones, se establece la clasificación por grupos A, B y C, que representarán las zonas geográficas en las que se clasificarán los 35 clientes en estudio establecidas por Suroeste, Noreste y Sur.

La tabla 3 muestra la clasificación de cada cliente en su grupo correspondiente, los valores mostrados representan las coordenadas X_1 y X_2 que se utilizaron para graficar a los clientes en el plano bidimensional. Ver tabla. 3.

Tab. 3. Clasificación por grupos con algoritmo K- medias

| GRUPO A | | GRUPO B | | GRUPO C | | 6 |
|---------|--------|---------|--------|---------|--------|---|
| 2.3118 | 2.4384 | 0 | 0 | 0 | 0 | |
| 0 | 0 | 0 | 0 | 1.7251 | 2.1515 | |
| 0 | 0 | 0 | 0 | 1.7810 | 1.6981 | |
| 0 | 0 | 0.8759 | 1.8785 | 0 | 0 | |
| 0 | 0 | 0 | 0 | 1.6936 | 1.9232 | |

Se graficaron los resultados de los grupos en un plano bidimensional que representa las 3 zonas geográficas de clasificación para los clientes. La gráfica clasifica 3 clientes en la zona Suroeste, 8 en la zona Noreste y 24 en la zona Sur. Ver figura 4.

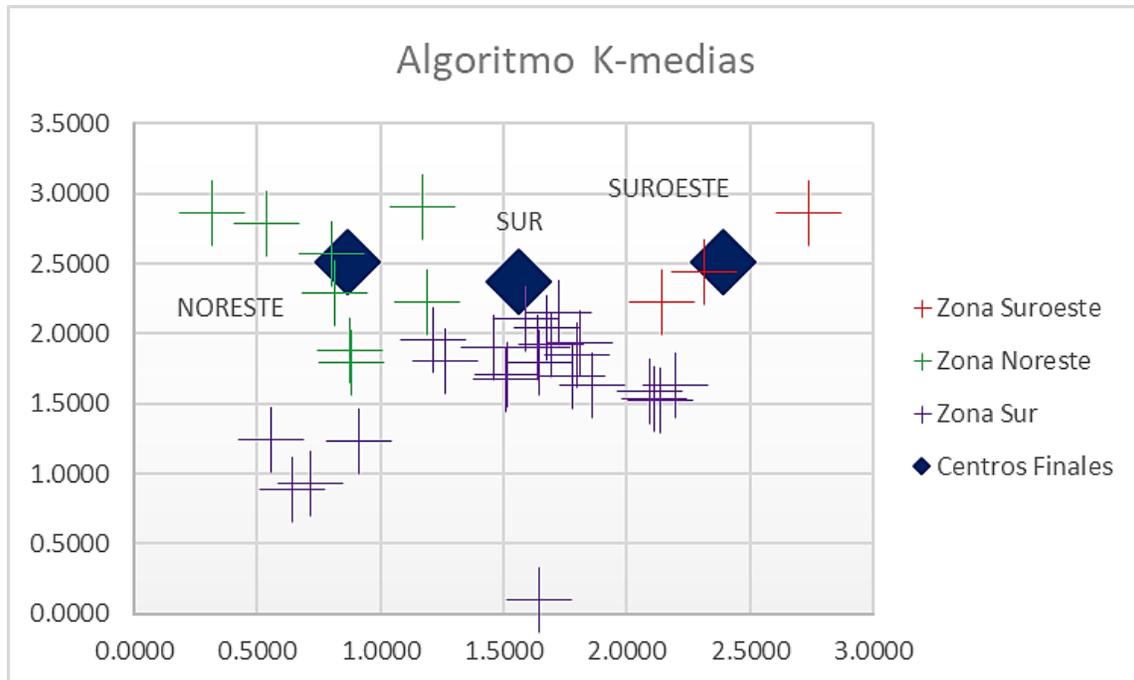


Fig. 4. Clasificación de clientes con algoritmo k-medias



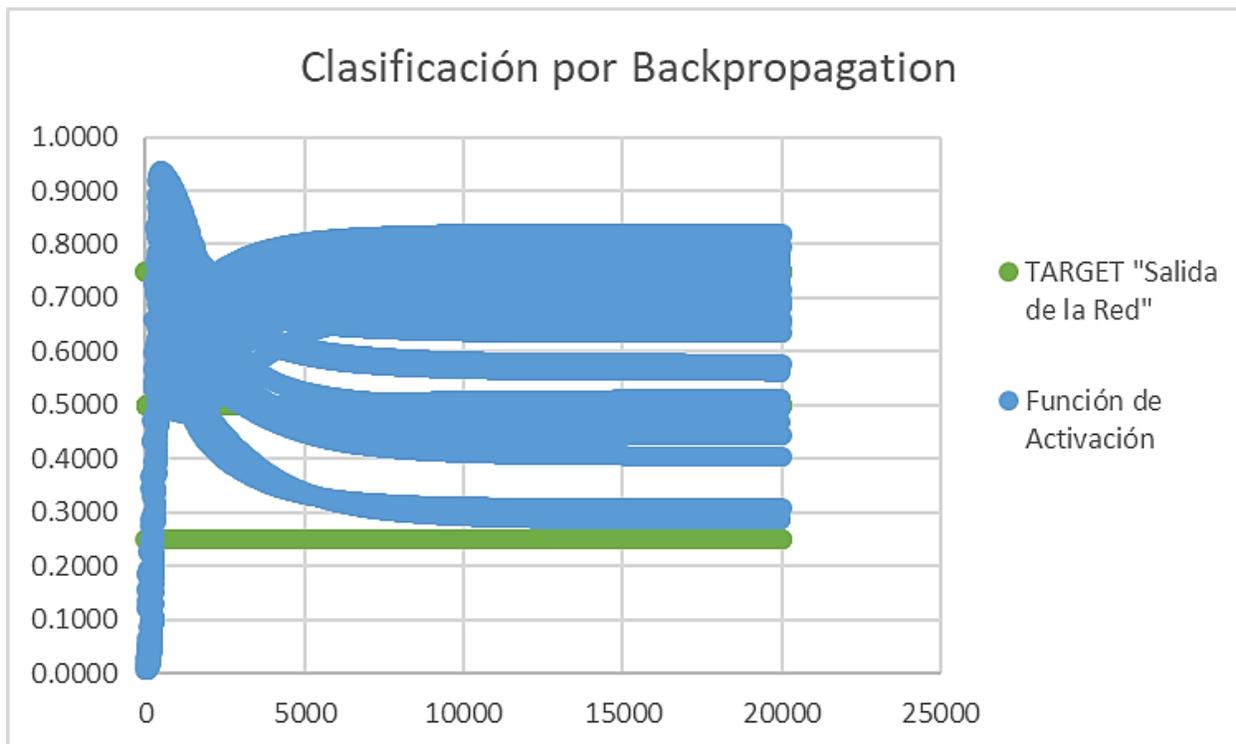
Se establecieron las condiciones de clasificación en la columna del "Target" (T), mismas que se refieren a la clasificación en tres zonas geográficas distintas: 0.25= Suroeste, 0.50= Noreste y 0.75= Sur.

Se resolvió la segunda mitad de la red por medio del algoritmo Backpropagation. Se utilizaron 35 datos como patrones de entrenamiento previamente codificados a una escala de entre 2.9 y 0.1., X_1 y X_2 representan las coordenadas de Latitud y Longitud respectivamente y los datos se calcularon con 20,000 iteraciones con el objetivo de tener una mejor perspectiva del comportamiento de la red. La tabla muestra 5 de los 35 patrones de entrenamiento. Ver tabla 5.

Tab. 5. Entrenamiento de la red mediante backpropagation

| | X1 | X2 | NEU1 | NEU2 | NEU3 | NEU4 | W1 | W2 |
|---|-------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 2.311 | 2.4384 | 0.9879 | 0.1227 | 0.2819 | 3.0000 | 2.0000 | 0.000 |
| 2 | 1.725 | 2.1515 | 0.5607 | 0.4570 | 0.7173 | 2.9996 | 2.0000 | -0.001 |
| 3 | 1.781 | 1.6981 | 0.3545 | 0.2785 | 0.9442 | 2.9995 | 1.9999 | -0.003 |
| 4 | 0.875 | 1.8785 | 0.0666 | 0.7917 | 0.5764 | 2.9993 | 1.9997 | -0.008 |
| 5 | 1.693 | 1.9232 | 0.4328 | 0.4156 | 0.8846 | 2.9992 | 1.9989 | -0.014 |

Finalmente se grafican los datos del "Target" (T), que representa la clasificación deseada y se le agrega otra serie correspondiente a la *función de activación* que corresponde a la salida de la red. La gráfica muestra la clasificación de los clientes en estudio en las tres zonas geográficas propuestas: 0.25= Suroeste, 0.50= Noreste y 0.75= Sur. Ver figura 5.





B. Etapa de validación de la Red

Se utilizaron 10 datos para validar el aprendizaje de la red con 10 clientes diferentes, se calcularon nuevas neuronas con los valores de los centros finales y se graficaron los datos del “*Target*” (clasificación deseada) y los de la “*función de activación*” (salida de la red). Se muestran 5 de los 10 datos en estudio. Ver tabla 7.



La figura. 6 muestra la inclusión de los clientes nuevos representados por un “triangulo” en la clasificación por zonas de la red. La validación clasifica 3 clientes en la zona Noreste, y 7 en la zona Sur. La validación clasifica a los clientes nuevos en un 90%. Ver figura. 6

c. Etapa de Prueba de la Red

Se utilizaron 5 datos correspondientes al 10% de los clientes en estudio para probar la red neuronal. Se muestran 3 de los 5 datos en estudio. Ver tabla. 7.

La figura 7 muestra la clasificación por zonas de todos los clientes en estudio incluyendo los 5 destinados a probar la red que se representan con un “circulo”. Esta última clasifica 2 clientes en la zona Noreste y 3 en la zona Sur. La fase de prueba de la red clasifica a los nuevos clientes nuevos al 100%. Ver figura 7.

Tab. 7. Patrones de entrenamiento de 5 clientes en la etapa de validación y 3 clientes en la etapa de prueba

| VALIDACIÓN | | | | | | | | | | | |
|------------|------|------|------|-------|------|-------|------|------|--------|------|------|
| X1 | X2 | Ne1 | Ne2 | Ne3 | w1 | w2 | w3 | Bias | Agre | FDA | T |
| 2.02 | 1.88 | 0.58 | 0.20 | 0.744 | -1.9 | -1.17 | 1.79 | 0.65 | 0.586 | 0.64 | 0.75 |
| 0.82 | 2.50 | 0.08 | 0.98 | 0.258 | -1.9 | -1.17 | 1.79 | 0.65 | -0.195 | 0.45 | 0.50 |
| 1.28 | 1.59 | 0.12 | 0.46 | 0.924 | -1.9 | -1.17 | 1.79 | 0.65 | 1.522 | 0.82 | 0.75 |
| 1.24 | 1.64 | 0.12 | 0.51 | 0.900 | -1.9 | -1.17 | 1.79 | 0.65 | 1.419 | 0.80 | 0.75 |
| 1.53 | 1.91 | 0.33 | 0.52 | 0.903 | -1.9 | -1.17 | 1.79 | 0.65 | 1.002 | 0.73 | 0.75 |
| PRUEBA | | | | | | | | | | | |
| X | Y | Ne1 | Ne2 | Ne3 | w1 | w2 | w3 | Bias | Agre | FDA | T |
| 0.51 | 2.74 | 0.02 | 0.76 | 0.090 | -1.9 | -1.17 | 1.79 | 0.65 | -0.130 | 0.46 | 0.50 |
| 0.10 | 0.71 | 0.00 | 0.03 | 0.053 | -1.9 | -1.17 | 1.79 | 0.65 | 0.711 | 0.67 | 0.75 |
| 2.90 | 1.89 | 0.53 | 0.01 | 0.153 | -1.9 | -1.17 | 1.79 | 0.65 | -0.133 | 0.46 | 0.50 |

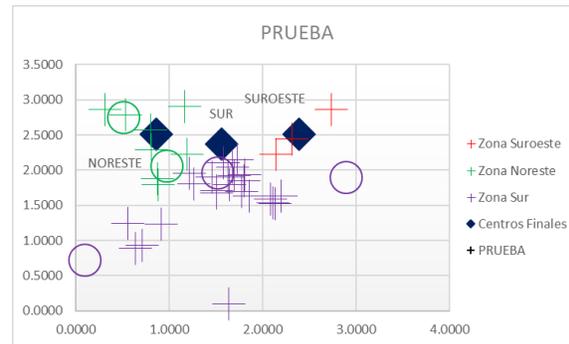
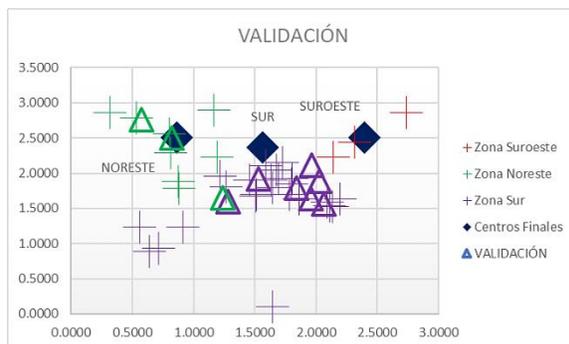


Fig. 6. Clasificación por zonas geográficas fase de validación



Conclusiones

El diseño y desarrollo de la red neuronal aplicada en el presente caso de estudio permite analizar y obtener un panorama específico para la empresa y le facilita la toma de decisiones en cuanto a la ubicación y agrupación geográfica de sus principales clientes. La clasificación de los clientes de la empresa representaba uno de los problemas más significativos para el negocio debido a que sus vehículos repartidores no podían entregar los productos en tiempo y forma, provocando con esto molestias y desconfianza por parte de los clientes. La red neuronal de base radial clasifica 50 clientes en 3, zonas geográficas de manera precisa, minimizando distancias y reduciendo costos de distribución para la empresa cumpliendo así el objetivo general propuesto. Los resultados favorecen a la empresa al momento de tomar decisiones dirigidas a eficientar los servicios de entregas a domicilio con el objetivo de maximizar la satisfacción de los clientes. El estudio proporciona también un enfoque general para que la empresa pueda optimizar la distribución de sus rutas, agrupando a ciertos clientes con base a su ubicación y cercanía geográfica, permitiendo con esto analizar el tema de las emisiones de CO₂ por viaje y tomar decisiones dirigidas a reducir la generación de dichos gases. Los resultados obtenidos al aplicar la presente herramienta permiten el incremento en la satisfacción de los clientes cumpliendo así con el objetivo específico más importante ofreciendo así a la empresa en estudio un beneficio significativo en cuanto a su calidad y mejora en el servicio de entregas a domicilio.

Utilizar inteligencia artificial para controlar los procesos comerciales del negocio lo posiciona por encima de la competencia y le otorga la ventaja competitiva para sobresalir y diferenciarse en su sector empresarial, haciendo que los clientes se fidelicen con sus servicios y los recomienden con mayor seguridad logrando un mejor desempeño del negocio y con ello crecimiento económico tanto interno como externo.

Cabe mencionar y destacar la excelente relación y disposición que presentaron los dueños del negocio que incondicionalmente brindaron la información solicitada, la cual fue indispensable para el desarrollo del presente estudio, se agradece la atención recibida y se reconoce el excelente trabajo que realizan.



Referencias

- George, M. (2007). *Radial Basis Function Neural Networks and Principal Component Analysis for Pattern Classification*. *IEEE Electronic Library*, 200-206.
- Hjouji A., E.-M. J. (2019). *Image retrieval and classification using shifted Legendre invariant moments and Radial Basis Functions Neural Networks*. *Elsevier Science Direct*.
- Hung C., K. Y. (2002). *A comparative study of radial basis function in classification of remotely sensed data neural networks and wavelet neural networks*. *IEEE Electronic Library*.
- INEGI. (2014). *Micro, pequeña, mediana y gran empresa: estratificación de los establecimientos*. *Censos Económicos*.
- Kumar R., S. S. (2016). *Time Series Prediction Using Focused Time Lagged Radial Basis Function Network*. *IEEE Electronic Library*, 4.
- Mohamed B., I. A. (2015). *ECG Image Classification in Real time based on the Haar-like Features and Artificial Neural Networks*. *Elsevier Science Direct*, 32-39.