

Wine quality classification strategy using the SOM neural network based on chemical analysis

Estrategia de clasificación de calidad de vino usando la red neuronal SOM basado en análisis químico

Axel Adrián Culqui Sánchez, Bachiller¹, Miguel Andres Seclen Chavez, Bachiller¹, Erick Irwin Pajuelo Ojeda, Bachiller¹ and Huarote Zegarra Raúl Eduardo, Maestro, Maestro²

¹Universidad Privada del Norte, Perú, n00224983@upn.pe, n00231418@upn.pe, n00207976@upn.pe

²Universidad San Ignacio de Loyola, Perú, rhuarotez@gmail.com

Abstract– *The present investigation covers the need to classify the quality of a wine based on its chemical analysis. To achieve this, the strategy consists of preparing the specific data extracted from a data set and entering it into a SOM (Self-Organizing Map) neural network so that it can be classified. To corroborate this, there are 6 qualities of wine (3, 4, 5, 6, 7, 8). For the implementation of this strategy, a public database is used, which has 4898 number of instances and 12 number of attributes. In the present investigation, we are working with two types of red and white wines, which will be classified according to an analysis of their chemical attributes, thus achieving with this strategy to demonstrate an effectiveness of 91.6%, sensibility 83.3% and 100% of specificity in qualifying the wines.*

Keywords-- Wine, Strategy, SOM Neural Network, wine quality.

Resumen- La presente investigación cubre la necesidad de clasificar la calidad de un vino basado en el análisis químico de este. Para lograr esto, la estrategia consiste en preparar la data específica extraída de un set de datos e ingresarlos a una red neuronal SOM (Mapa autoorganizativo) para que pueda ser clasificada. Para corroborar esto se tiene 6 calidades de vino (3, 4, 5, 6, 7, 8), siendo 3 el rango de más baja calidad y 8 el de mayor calidad. Para la implementación de esta estrategia se hace uso de una base de datos pública, la cual cuenta con 1599 número de instancias y 12 número de atributos. En la presente investigación se está trabajando con dos tipos de vinos tinto y blanco, los cuales serán clasificados según un análisis a sus atributos químicos, logrando así con esta estrategia demostrar una efectividad del 91.6%, sensibilidad de 83.3% y 100% de especificidad en calificar los vinos.

Palabras Clave – Vino, Estrategia, Red Neuronal SOM, calidad de vino.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

I. INTRODUCCIÓN

La evaluación de la calidad de los vinos es una tarea compleja y crucial para la industria vitivinícola. Diversos estudios han demostrado que el análisis químico puede proporcionar información precisa sobre la calidad del vino [1, 2]. Sin embargo, la gran cantidad de variables involucradas en los análisis químicos hace que la interpretación manual de los datos sea difícil y propensa a errores. La clasificación de vinos mediante estrategias de redes neuronales puede proporcionar múltiples beneficios a la industria vitivinícola, tales como la identificación de patrones, automatización en el proceso de clasificación, la mejora del proceso de calificación y el apoyo al marketing, lo cual ayudaría a las empresas a desarrollar un producto que satisfagan al cliente.

En este sentido, el uso de técnicas de aprendizaje automático, como las redes neuronales, puede resultar de gran utilidad para clasificar los vinos según su calidad. En particular, la red neuronal de mapas autoorganizados (SOM) ha demostrado ser efectiva para este fin [3].

Por lo tanto, en este proyecto se propone la evaluación de la calidad de vinos tintos mediante una red neuronal SOM entrenada con datos de análisis químico.

En resumen, este proyecto tiene como objetivo utilizar técnicas de aprendizaje automático para clasificar los vinos según su calidad. En particular, se utilizará la red neuronal SOM para identificar patrones y relaciones entre las variables del análisis químico. Los resultados obtenidos muestran la eficacia de la técnica en la clasificación de los vinos y su potencial para mejorar la evaluación de la calidad de estos.

II. ESTRATEGIA

En la inteligencia artificial se conoce distintas estrategias con las cuales podríamos cubrir nuestra necesidad, en la presente investigación se considerará las redes neuronales artificiales, como es usado en [4] primero se transformó los datos identificados las variables dependientes e independientes,

debido a que se tiene el campo de calidad. Otra estrategia elaborada para este sector es de [5] la cual consiste en transformar el data set en orden para incluir distintas muestras de vino por fila. En el mismo sentido según [6] resalta que las características físicas químicas de las uvas se predicen, mediante técnicas de aprendizaje automático y análisis de textura, logrando de esta forma muy buenos resultados. Para entender esto, en [7] explica que “Actualmente existen diferentes técnicas para el desarrollo de análisis multivariado basadas en inteligencia artificial y otras técnicas: árboles de clasificación, reglas de inferencia, algoritmos evolutivos, redes artificiales, máquinas de aprendizaje automático, algoritmos de Clúster, algoritmos bayesianos, técnicas matemáticas y estadísticas entre muchas otras técnicas.”. Entre ellas resalta al aprendizaje automático como el más relevante, debido a que permite construir modelos a partir de los datos deduciendo patrones configurados ya por ellos y de esta manera pudiendo determinar la calidad del vino. También en [8] se aplica la regresión lineal para determinar la dependencia de la calidad del vino en otras variables independientes.

Como se visualiza en la Fig. 1 se muestra un gráfico de la red neuronal artificial.

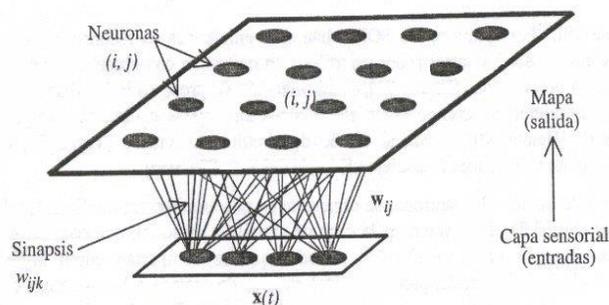


Fig. 1 Red Neuronal Artificial SOM

<https://archive.ics.uci.edu/ml/datasets/Wine+Quality> es un conjunto de datos de atributos químicos de dos tipos de vino (rojo y blanco) este consta de 12 atributos los cuales son: Fixed Acidity, Volatile Acidity, Citric Acid, Residual Sugar, Chlorides, Free Sulfur Dioxide, Total Sulfur Dioxide, Density, pH, Sulphates, Alcohol and Quality.

Por lo tanto, son dos tipos de vino que se considerarán para esta investigación. Los datos comprenden una calidad del 0 al 10 la cual va a ser usada para entrenar al sistema bajo los criterios establecidos.

II. INTELIGENCIA ARTIFICIAL

Según [9] “la inteligencia artificial es la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana”, también [10] considera que “La IA tiene como objeto que los ordenadores hagan lo misma clases de cosas que puede hacer la mente humana.”, así mismo [11] “No

existe una definición universalmente aceptada de la IA. Generalmente, la IA se considera una disciplina de la informática que tiene por objetivo elaborar máquinas y sistemas que puedan desempeñar tareas que requieren una inteligencia humana. El aprendizaje automático y el aprendizaje profundo son dos esferas de la IA. En los últimos años, con el desarrollo de las nuevas técnicas y equipos informáticos basados en redes neuronales, la IA se ha venido entendiendo como un sinónimo de “aprendizaje automático profundo supervisado”. Otra opinión al respecto define [12] “La inteligencia artificial (IA) puede definirse como el medio por el cual las computadoras, los robots y otros dispositivos

realizan tareas que normalmente requieren de la inteligencia humana. Por ejemplo, la resolución de cierto tipo de problemas, la capacidad de discriminar entre distintos objetos o el responder a órdenes verbales.”.

Por lo tanto, todas las opiniones comparten el mismo punto el cual es que la inteligencia artificial consiste en emular lo más posible el funcionamiento de la inteligencia humana, usando hardware y software para llevarlo a cabo.

A. Red neuronal artificial

De acuerdo con [13] “Estas neuronas artificiales se agrupan en capas o niveles y poseen un alto grado de conectividad entre ellas, conectividad que es ponderada por los pesos.”, del mismo modo [14, 15] aseguran que “se han convertido en modelos efectivos de predicción, han demostrado mejor efectividad frente a modelos estadísticos de regresión sin necesidad de cumplir condiciones de linealidad, normalidad, tamaño muestral, entre otros.”.

Por lo tanto, el funcionamiento de una red neuronal artificial se basa en el modelo biológico del humano, capaz de aprender a través del aprendizaje, previamente de un entrenamiento iterativo.

B. Red neuronal SOM

La IA entrega técnicas que facilitan a la comprensión y entendimiento de los patrones que componen la realidad. Por ello, [16] destaca más a las redes neuronales artificiales y concretamente los Mapas Auto-organizados (SOM), debido a su capacidad de cartografiar la realidad, representando sus objetivos distribuidos estructurados bidimensionalmente, a partir únicamente de sus propiedades. Asimismo, mencionan a los Mapas Auto-organizados como útiles para generar prototipos de conocimiento al descubrir patrones y tendencias en los datos, utilizándose para ello por ejemplo en las investigaciones sobre procesos de auto organización, entre otros.

Por lo tanto [17] especifica que, la red neuronal SOM sirve para realizar una transformación de un espacio multi dimensional en una serie de neuronas, de tal forma que las distancias relativas entre los puntos del espacio de entrada se conservan. Las neuronas suelen estar dispuestas en un plano.

IV. PROCESO DE APRENDIZAJE

En el proceso de aprendizaje de la red neuronal SOM al aplicar la estrategia anteriormente explicada, se carga el data set en formato csv (comma separated value), de la cual se tiene los siguientes atributos referenciados en la Tabla I.

TABLA I
Nombre, conteo y tipo de datos de la dataset

Columna	Conteos No nulos	Dtype
Fixed Acidity	1599	float64
Volatile Acidity	1599	float64
Citric Acid	1599	float64
Residual Sugar	1599	float64
Chlorides	1599	float64
Free sulfur dioxide	1599	float64
Total sulfur dioxide	1599	float64
Density	1599	float64
Ph	1599	float64
Sulphates	1599	float64
Alcohol	1599	float64
Quality	1599	int64

En la Fig 2 podemos ver en un gráfico de barras los puntajes de calidad agrupados y el conteo de estos.

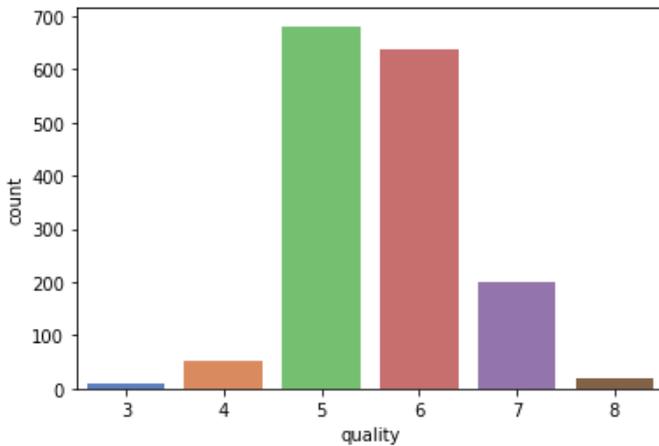


Fig. 2 Histograma de Calidad

La información de la data set nos indica que el atributo de calidad puede ir de 3 a 8 con respecto al nivel de calidad, pero en este grafico se evidencia que nuestra data va en un rango del 3 al 8 siendo los más comunes el 5 y 6. En la Fig 3 podemos ver la relación que tiene calidad con los atributos químicos para así saber cómo afecta cada uno de estos atributos a la calidad, ejemplo de estos son Fig 3a donde fixed acidity afecta de manera irregular la calidad, lo que pasa de manera contraria en la Fig 3b donde volatile acidity al disminuir se ve un aumento significativo en la calidad.

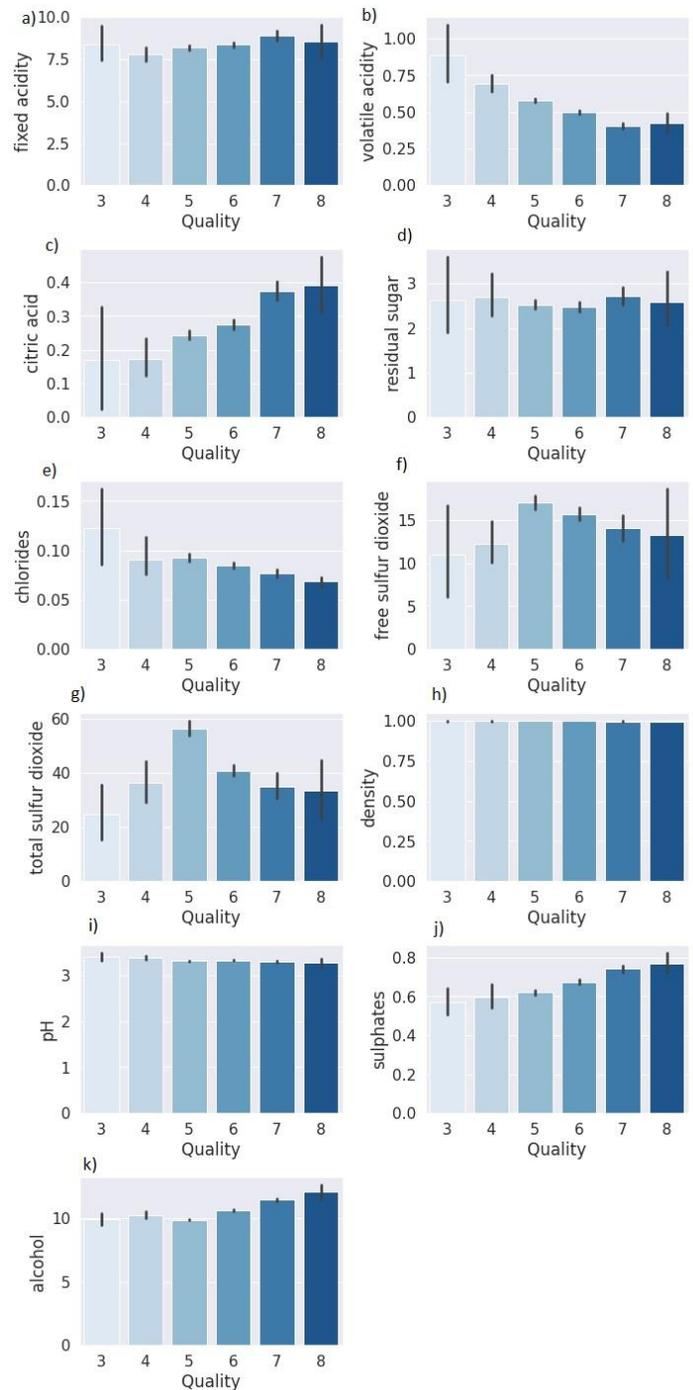


Fig. 3 gráficos de Barras

La Fig 4 muestra una representación 2D de los datos de entrada en un espacio definido por la red neuronal SOM entrenada. Cada punto en el gráfico representa una instancia de datos, y el color del punto indica si se trata de un vino “bueno” (naranja) o “malo” (azul), según se defina en el estudio. La posición del punto en el gráfico indica qué neurona de la red neuronal SOM está más cerca de la instancia de datos

correspondiente. Por lo tanto, el gráfico proporciona información sobre cómo se agrupan las diferentes instancias de datos en función de su calidad, y cómo se relacionan en el espacio definido por la red neuronal SOM.

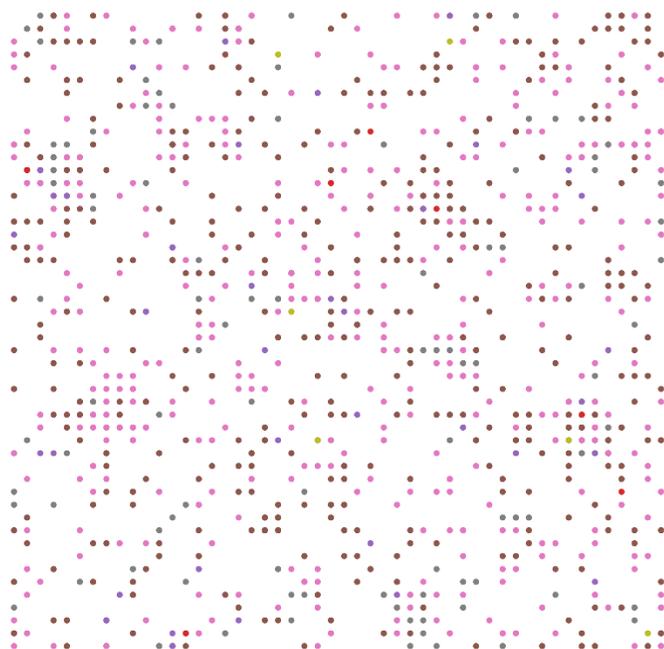


Fig. 4 Grafico de Dispersión de dos dimensiones 50x50

En la figura 5 podemos visualizar el mismo grafico solo que esta vez aplicado a las predicciones.

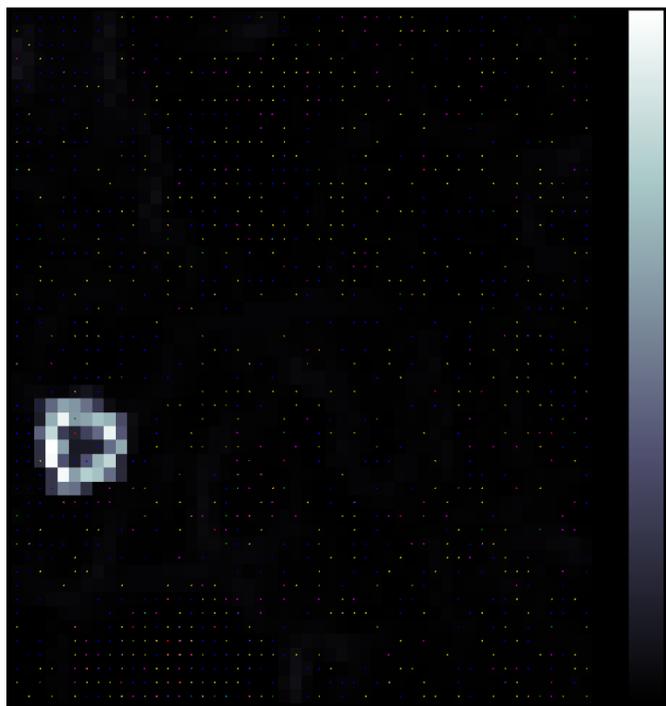


Fig. 5 Grafico de Dispersión de dos dimensiones 50x50

Con la red neuronal SOM identificamos fraudes los cuales son los vinos de acuerdo al tipo de calidad, la calificación por medio de la red neuronal SOM como se explica en la tabla II.

TABLA II
Etiquetas de calificación del algoritmo

Calidad	Significado
3	Calidad de tipo 3
4	Calidad de tipo 4
5	Calidad de tipo 5
6	Calidad de tipo 6
7	Calidad de tipo 7
7	Calidad de tipo 8

Se ha ajustado el nivel de iteraciones a 100000 ya que esto es importante para mejorar la precisión del modelo, el Alpha máximo a 0.5 y la dimensión de la matriz a 50x50, la distancia de activación es la euclídeana, la topología es exagonal y la función de afectación a la vecindad es gaussiana.

Finalmente, poniendo a prueba la red neuronal SOM para calcular su nivel de precisión siendo este de 91.6% con la matriz de confusión en la figura 6 donde se evidencia la sensibilidad y especificidad para cada calidad de en el rango correctamente.

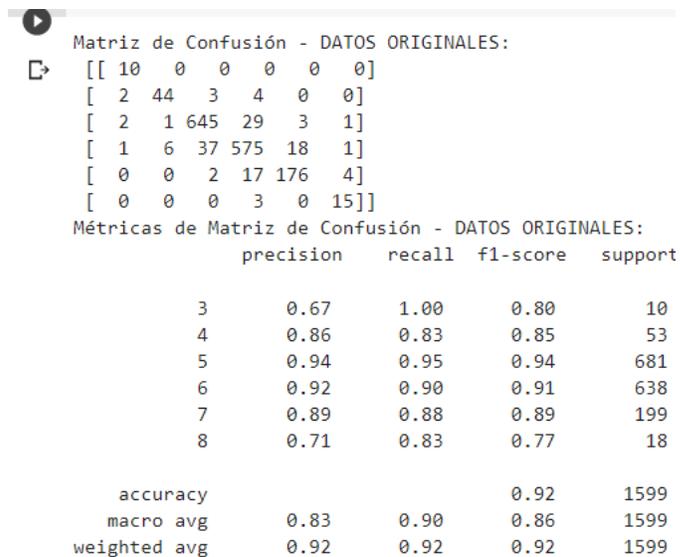


Fig. 6 Matriz de Confusión

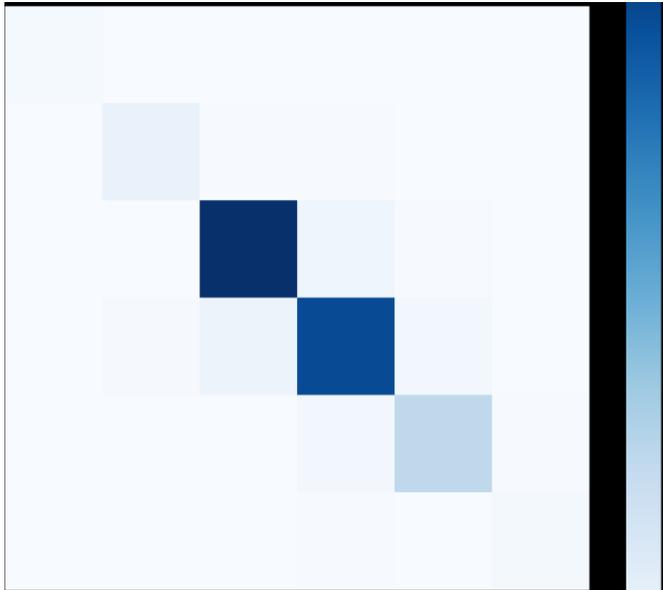


Fig. 7 Matriz de Confusión

En la figura 7 se muestra la gráfica de la matriz de confusión, confirmando los resultados obtenidos.

Con el 25% de la data dirigida a predicción se obtuvo una precisión de 90.15% y en la figura 8 se ve su matriz de confusión, confirmando los resultados obtenidos con una porción de prueba.

Matriz de confusión en el conjunto de prueba

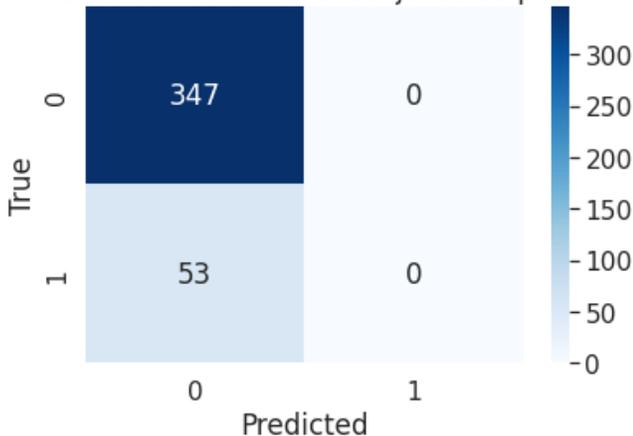


Fig. 8 Matriz de Confusión predicción

V. CONCLUSIONES

Al seguir la estrategia explicada para esta investigación usando la red neuronal SOM, se obtuvo una efectividad de 91.6% para el conjunto de entrenamiento y un 83.3% para el conjunto de pruebas y 100% en sensibilidad y especificidad en la clasificación respectivamente, con los vinos según sus atributos químicos, para obtener este resultado se usó una data

set de 1599 registros obtenida de la base de datos publica UCI para el entrenamiento del modelo. Al aplicar la red neuronal SOM para la clasificación de vinos aporta a la optimización de procesos industriales futuros para la creación de vinos, ya que se puede reducir la mano de obra requerida para clasificar estos. Se usó la librería minisom en el entorno Google Colab, donde tenemos los suficientes recursos para poder desarrollar nuestro modelo tal como se visualiza en la figura 9 el resumen del código fuente.

```

vino = pd.read_csv('winequality-red.csv', sep=',')
vino.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1599 entries, 0 to 1598
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   fixed acidity       1599 non-null   object
 1   volatile acidity   1599 non-null   object
 2   citric acid        1599 non-null   object
 3   residual sugar     1599 non-null   object
 4   chlorides         1599 non-null   object
 5   free sulfur dioxide 1599 non-null   object
 6   total sulfur dioxide 1599 non-null   object
 7   density            1599 non-null   object
 8   pH                1599 non-null   object
 9   sulphates         1599 non-null   object
10   alcohol           1599 non-null   object
11   quality            1599 non-null   object
dtypes: object(12)
memory usage: 150.0+ KB

Vino = []
for i in range(len(vino)):
    Vino.append([f.replace(',','.') for f in vino.values[i]])
Vino = np.array(Vino,dtype=float)
X = Vino[:,0:11]
Y = np.array(Vino[:,11],dtype=int)

max_iter = 100000
alpha_max = 0.5
som = MiniSom(x=50, y=50, input_len=11, sigma=1.0, learning_rate=alpha_max, activation_distance='euclidean',
topology='hexagonal', neighborhood_function='gaussian')
som.random_weights_init(X)
som.train_random(data=X, num_iteration=max_iter)

[98] Ysimulado = []
N = M.astype(np.str)
for x in X:
    w = som.winner(x)
    Ysimulado.append(N[w].astype(np.float64))
print(Ysimulado)

[100] score = 0
for i in range(len(Y)):
    for j in range(len(Y)):
        if Y[i]==Ysimulado[j]:
            score = score + 1
score = 100*score/len(Y)
print(score)

91.61976235146967

```

Fig. 9 Entorno Colab de la aplicación de la red SOM en la selección de vino por calidad

En la figura 9 se evidencia la creación de la red SOM en este caso se eligió para la clasificación, recalando que los valores de los parámetros han sido realizado mediante las pruebas, encontrando un valor adecuado.

VI. DISCUSIÓN

Podemos visualizar que la precisión del modelo se vio aumentada en un aproximado de 3% en comparación con los resultados obtenidos por [4] que utiliza la red neuronal densa, por lo cual se comprueba que la red neuronal SOM tiene una mayor precisión, sin embargo consideramos que el modelo de clasificación de vinos puede mejorar extendiendo con más datos para el aprendizaje, ya que la data set utilizada no era suficiente para darle una calificación de calidad en un rango del 3 a 8, por lo que si se implementa el mismo modelo con una data set más populada los resultados obtenidos serán óptimos.

REFERENCES

- [1] Fernández, J.M., Gómez, J.A., y García, E. (2013). Using near infrared spectroscopy (NIRS) to predict wine authenticity and quality. *Talanta*, 116, 755-762. <https://doi.org/10.1016/j.talanta.2013.07.068>
- [2] Kılıç, K., y Altun, E. (2017). Analysis of wine quality using artificial neural network based on quality control chart. *Journal of Food Processing and Preservation*, 41(2), e12863. <https://doi.org/10.1111/jfpp.12863>
- [3] Escudero, A., y Torga, R. (2012). Wine quality assessment by self-organizing maps. *Expert Systems with Applications*, 39(3), 3014-3020. <https://doi.org/10.1016/j.eswa.2011.08.095>
- [4] Rivera, D., Huamani, C., y Charca, Y. (2022). Sistema automático para la calificación de vino mediante Redes Neuronales. *Innovación y Software*, 3(1), 30-46. <https://www.redalyc.org/journal/6738/673870840003/html/>
- [5] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., y Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4), 547-553. <https://doi.org/10.1016/j.dss.2009.05.016>
- [6] Brillante, L., Gaiotti, F., Lovat, L., Vincenzi, S., Giacosa, S., Torchio, F., Rio, S., Rolle, L., y Tomasi, D. (2015). Investigating the Use of Gradient Boosting Machine, Random Forest and their Ensemble to Predict Skin Flavonoid Content from Berry Physical-Mechanical Characteristics in Wine Grapes. *Computers and Electronics in Agriculture*, 117, 186-193. <http://dx.doi.org/10.1016/j.compag.2015.07.017>
- [7] Galeano, L. F., Aguirre, S. G., y Castrillón, O. D. (2021). Análisis de calidad del vino por medio de técnicas de inteligencia artificial. *Información Tecnológica*, 32(1), 17-26. <http://dx.doi.org/10.4067/S0718-07642021000100017>
- [8] Gupta, Y. (2018). Selection of important features and predicting wine quality using machine learning techniques. *Procedia Computer Science*, 125, 305-312. <https://doi.org/10.1016/j.procs.2017.12.041>
- [9] Rouhiainen, L. (2018). *Inteligencia artificial* (4ta ed.). Planeta, S.A. <https://www.planetadelibros.com/libro-inteligencia-artificial/280581>
- [10] Boden, M. A. (2017). *Inteligencia artificial* (1ra ed.). Turner Publicaciones S.L. <https://cutt.ly/q37rmu6>
- [11] Estupiñan, J., Yelandi, M., Peñafiel, A., y El Assafiri, Y. (2021). Inteligencia artificial y propiedad intelectual. *Aportes Educación, Sociedad y Empresa en América Latina*, 13(S3). <https://rus.ucf.edu.cu/index.php/rus/article/view/2490>
- [12] Solano, A. Y., y Emanuel, A. (s.f). Inteligencia artificial un peligro latente. <https://cutt.ly/137i0KG>
- [13] Salas, R. (2004). Redes neuronales artificiales. *Universidad de Valparaíso-Departamento de Computación*, 1(1), 1-7. <https://cutt.ly/P37yhNH>
- [14] Aranda, G. A., Calmet, J., y Martín, F. J. (2014). *Artificial Intelligence and Symbolic Computation*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-13770-4>
- [15] Samani, T., Porayska, K., y Luckin, R. (2017). *Bridging the gap between high and low performing pupils through performance learning online analysis and curricula*. Springer International Publishing. https://doi.org/10.1007/978-3-319-61425-0_82
- [16] Abarca, F. J., y Osuna, F. (2013). Cartografías semánticas mediante redes neuronales: los mapas auto-organizados (SOM) como representación de patrones y campos. *EGA Revista de expresión gráfica arquitectónica*, 18(22), 154-163. <https://doi.org/10.4995/ega.2013.1692>
- [17] Saludes, S., Vargas, A., y Miguel, L. J. (2015). DETECCIÓN DE FALLOS EN GRUPOS HIDROELÉCTRICOS MEDIANTE LA RED NEURONAL SOM. <https://cutt.ly/437irdO>