

White Wine Quality

Predicción utilizando redes neuronales feedforward

Maite Sánchez-Fornaris¹[0000–0003–3709–9668] and Ernesto
Avila-Domenech²[0000–0002–4797–289X]

¹ Universidad de Camagüey “Ignacio Agramonte Loynaz”, Camagüey, Cuba.
{maite.sanchez}@reduc.edu.cu

² Universidad de Granma, Carretera Central vía Holguín Km $\frac{1}{2}$, Granma, Cuba
{eadomenech}@gmail.com

Abstract. Wine quality certification is of vital importance for wine companies as it guarantees a higher sales and price of the product, as well as preventing it from being adulterated in the distribution chain. Machine learning techniques offer numerous advantages to improve wine quality certification, among the most used models are linear regressions, neighborhood support machines (SVM) and artificial neural networks, which so far have obtained the best results. In this paper, two neural network architectures are proposed for predicting the quality of white wine using the dataset published by the Portuguese company Vinho Verde, which offers 4898 wine samples with 11 characteristics for predicting quality, according to a scale of 7 categories. The expected results were quite good according to the training error calculated through cross entropy, with a high precision.

Keywords: Artificial Neural Network · White Wine Quality.

1 Introducción

El mercado actual tanto a nivel local como global está saturado de productos de diferentes marcas en cantidad abrumadoras. Para cualquier empresa constituye todo un reto posicionar sus productos a la vanguardia de estos mercados, siendo requisito indispensable garantizar la calidad del bien que se está comercializando. Entonces, ¿cómo lo hacen? Pues a nivel internacional existen ciertos organismos que certifican y avalan la calidad de los productos, procesos o servicios. Este es un proceso engorroso por la cantidad de compañías que los solicitan y la poca disponibilidad de personal para realizar la tarea, además de que los datos son subjetivos hasta cierto punto para algunos productos (Ye, Li, & Jia, 2020).

La industria del vino, representada fuertemente por Francia, España y Portugal, se destaca en este contexto por la complejidad de su proceso para certificar la calidad del vino y la necesidad que tienen de él. La certificación de la calidad del vino aumenta exponencialmente el valor del producto, protegiendo y garantizando su calidad, impide la adulteración del vino y permite un mejor control en la línea de producción (Er & Atasoy, 2016). Las compañías vinícolas se aseguran

de certificar la calidad del vino que producen según dos tipos de pruebas: primero una físico-química y la segunda sensorial (Radosavljević, Ilić, & Pitulić, n.d.), la primera se realiza mediante exámenes de laboratorio, sin embargo la segunda requiere de la participación de personal experto calificado y sus resultados son subjetivos y aún no se han comprendido del todo (Ye et al., 2020).

En este contexto, muchas de estas empresas se han apoyado en las técnicas de machine learning para mejorar los resultados de la certificación de calidad del vino y acelerar el proceso en función de obtener un mejor producto en el menor tiempo posible con la calidad necesaria (Ramazan, n.d.). Este es el caso de (Vlassides, Ferrier, & Block, 2001) que utilizaron una red neuronal para clasificar vino en California, dado el nivel de maduración de la uva y los análisis químicos con una muestra de 36 ejemplos y obtuvieron un 6% de error. En (Yu et al., 2008) clasificaron 147 botellas de vino de arroz en 3 categorías utilizando mediciones espectrales. Las técnicas más utilizadas han sido las de regresión lineal, las máquinas de soporte vectorial (SVM) y últimamente las redes neuronales. La regresión lineal tiene la ventaja de ser sencilla de implementar y de interpretar los resultados, sin embargo no es la más exacta. Las máquinas de soporte vectorial dan buenos resultados, pero son fácilmente superadas por las redes neuronales artificiales que hasta el momento son las que mejores resultados han obtenido en este tipo de problemas de clasificación (Er & Atasoy, 2016).

2 Descripción del problema

El problema actual fue planteado por primera vez por (Cortez, Cerdeira, Almeida, Matos, & Reis, 2009) y está dado por la necesidad de utilizar un sistema de apoyo a la toma de decisiones para mejorar el proceso de producción del vino por la compañía portuguesa Vinho Verde. Portugal es uno de los países más importantes en la industria del vino, controlando el 3.17% en ese mercado. La compañía Vinho Verde en el momento de la recogida de los datos experimentaba un crecimiento en sus ventas de un 36%, viéndose en la posición de requerir tecnología que mejore el proceso de producción, aumente la calidad de sus productos y reduzca los costos de producción.

La empresa contaba con datos históricos de la calificación de sus vinos blancos y rojos según los análisis físico-químicos efectuados en el laboratorio, así como la valoración de los expertos en el análisis sensorial, que esta última no se publicó por razones de respeto a la privacidad. En esta investigación se utiliza el conjunto de 4898 muestras del vino blanco Vinho Verde para determinar su calidad, que se describe mediante 7 categorías.

Las variables de entrada identificadas en el problema son:

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar
5. chlorides
6. free sulfur dioxide

7. total sulfur dioxide
8. density
9. pH
10. sulphates
11. alcohol3

En el dataset se evidencia poca correlación entre las variables de entrada, excepto entre los atributos alcohol, sulphates y pH, aunque tampoco es muy alta. La cantidad de alcohol muestra una fuerte relación positiva con la calidad del vino. En cambio citric acid, free sulfur dioxide, sulphates tiene muy poca o ninguna relación con la calidad.

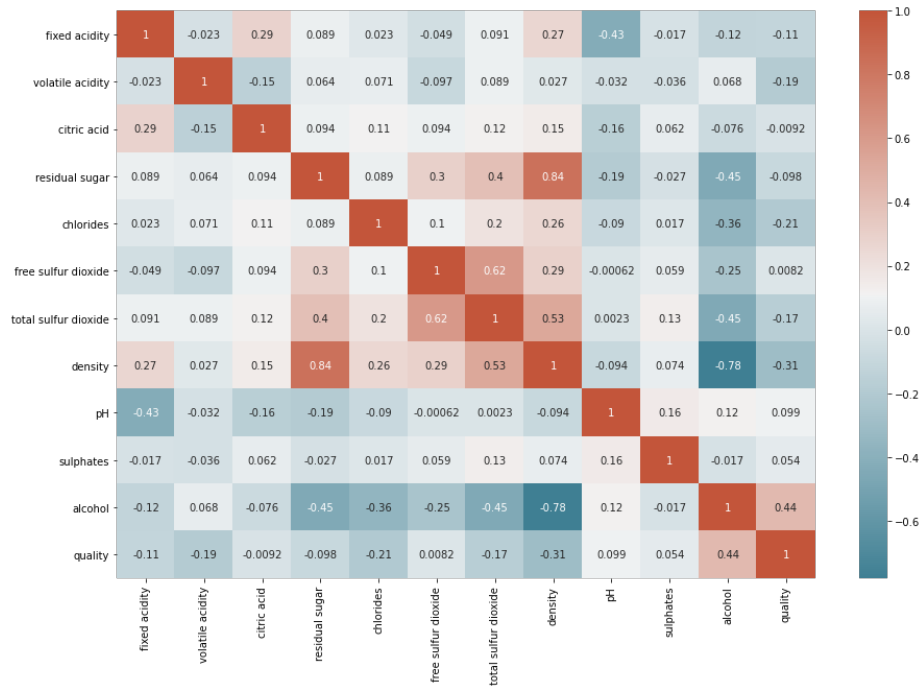


Fig. 1. Correlación entre los atributos del dataset del vino blanco.

La variable de salida identificada en el problema es:

quality: Describe la calidad del vino con valores comprendidos en el conjunto {3, 4, 5, 6, 7, 8, 9}.

El dataset no contiene valores nulos. Las 11 variables de entrada son de dominio continuo y la variable de salida es de tipo ordinal, con una relación de orden ascendente y el valor más representativo en la categoría 6, como se muestra en la siguiente figura:

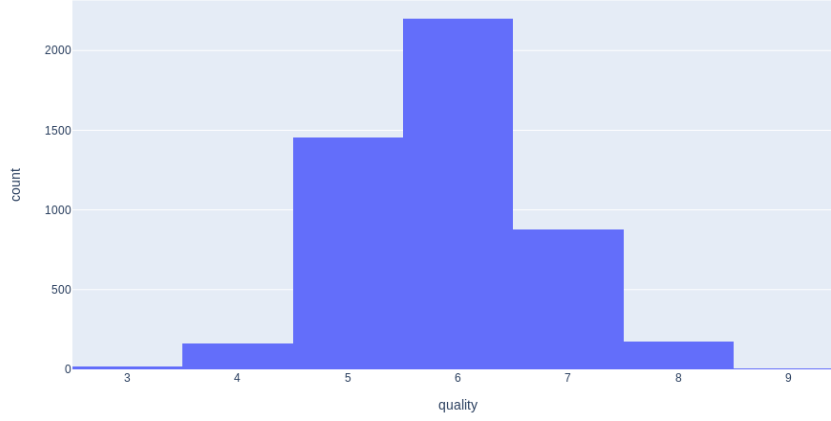


Fig. 2. Distribución de la clasificación de la calidad del vino blanco en el dataset.

3 Descripción de la solución.

3.1 Preprocesamiento de los datos

Se aplicó una normalización a los datos debidos a la amplitud y diversidad de rangos de valores en los datos de entrada. Estos se hicieron mediante el método `fit_transform(x)` de la biblioteca `sklearn`, este método combina `fit()` y `transform()` que se utilizan para centrar o caracterizar la escala de un dato dado. Básicamente, ayuda a normalizar los datos dentro de un rango particular. Para ello, utilizan el método Z-score (Raschka, 2015).

$$z = \frac{x - \mu}{\sigma}$$

Y básicamente encadenan ambos métodos de la siguiente forma:

1. `Fit()`: el método calcula los parámetros μ y σ y los guarda como objetos internos.
2. `Transform()`: el método que utiliza los parámetros μ y σ calculados previamente, luego aplica la transformación a un conjunto de datos en particular.
3. `Fit_transform()`: une el método `fit()` y `transform()` para la transformación del conjunto de datos.

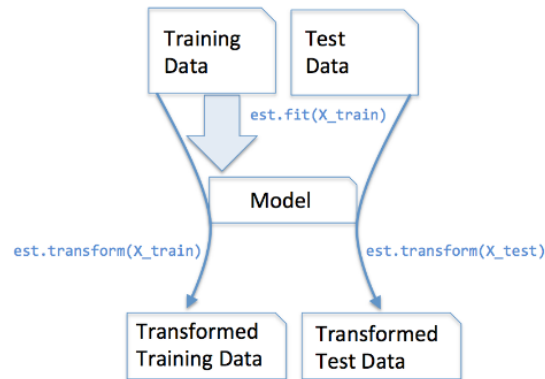


Fig. 3. Funcionamiento del método `fit_transform(x)` de sklearn.

3.2 Arquitectura de la Red Neuronal Artificial

Se proponen dos arquitecturas de redes neuronales, la primera de ellas comprende 4 capas completamente conectadas:

- Capa 1:** 100 neuronas
- Capa 2:** 320 neuronas
- Capa 3:** 100 neuronas
- Capa 4:** 7 neuronas

A cada capa se aplicó una función de activación ReLU, para la función de costo se utilizó Cross Entropy Loss y como optimizador se empleó descenso del gradiente estocástico (SGD).

Como función de activación se escogió ReLU (Rectified Linear Unit) que transforma los valores introducidos anulando los valores negativos y dejando los positivos tal y como entran. Esta función añade no linealidad a la neurona para mejorar la salida de la red neuronal (Jurado & Fellman, 2020).

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

Fig. 4. Función ReLU.

Características de la función ReLU:

- Activación Sparse solo se activa si son positivos.
- No está acotada.
- Se pueden morir demasiadas neuronas.
- Se comporta bien con imágenes.
- Buen desempeño en redes convolucionales.

La función de pérdida cross entropy permite comparar el valor estimado por la red con el valor real de nuestro conjunto de entrenamiento cuando se ha calculado la salida (Gonzalez & Miikkulainen, 2020):

$$C_{CE}(W, B, S^r, E^r) = - \sum_j [E_j^r \ln a_j^L + (1 - E_j^r) \ln (1 - a_j^L)]$$

Fig. 5. Función de costo cross entropy.

La función cross entropy tiene la propiedad de cancelar uno de los términos dependiendo del valor de la variable de salida y real. El primero cuando el valor real es 0 y el segundo cuando es 1. Existen otras funciones de pérdida que se asocian principalmente a redes neuronales profundas.

Una vez que se conoce el error de la estimación con respecto de la salida real se procede a calcular el ajuste necesario para los pesos de la red, es decir mejorar los parámetros W y b de las neuronas de la red para que produzcan mejores estimaciones.

El método cross entropy tiene una estructura sencilla que aplicarse a cualquier problema. Funciona para problemas simples, como la regresión y también para problemas más complejos, como el de clasificación. Es posible notar que para la regresión no produce mejoras significativas pero esto se debe a que son problemas sencillos que ya están optimizados al máximo con los métodos tradicionales. Sin embargo, en clasificación se obtienen resultados mejores que con las soluciones clásicas. Además, cuanto más complejo sea el problema (más clases, más variables y diferentes tipos de variables) mayor es la mejora del error con este método.

Como defectos de este método, sólo cabe destacar uno, que es el tiempo de ejecución, que en parte se puede reducir algo optimizando la red.

En la segunda arquitectura se usaron 3 capas de neuronas completamente conectadas:

Capa 1: 100 neuronas

Capa 2: 50 neuronas

Capa 3: 7 neuronas

La primera capa se activó con una función ReLU y la segunda con una softmax, como función de pérdida se usó también cross entropy y como optimizador el método de Adam.

La función softmax es una generalización de la función logística y convierte variables independientes de rango casi infinito en probabilidades simples entre $[0,1]$.

$$\phi(x) = \frac{1}{1 + e^{-x}} \quad \forall x \in \mathbb{R}$$

$$Im: \phi(x) \in [0, 1]$$

Fig. 6. Función de activación softmax.

Soporta sistemas de clasificación multinomial, por lo que se convierte en el recurso principal utilizado en las capas de salida de un clasificador. Esta función de activación devuelve la distribución de probabilidad de cada una de las clases soportadas en el modelo. La función softmax calcula la distribución de probabilidades del evento sobre n eventos diferentes. En términos generales, esta función calculará las probabilidades de cada clase objetivo sobre todas las clases objetivo posibles. Más tarde, las probabilidades calculadas serán útiles para determinar la clase objetivo para las entradas dadas.

La principal ventaja de usar softmax es el rango de probabilidades de salida. El rango será de 0 a 1, y la suma de todas las probabilidades será igual a uno. Este tipo de función de activación es muy utilizado en el modelo de regresión logística de clasificación múltiple y en diferentes niveles de capa de cara a la construcción de redes neuronales (Ambrogioni, Güçlü, van Gerven, & Maris, 2017).

El optimizador Adam trata de solventar el problema con la fijación de el ratio de aprendizaje del SGD, para ello adapta el ratio de aprendizaje en función de cómo estén distribuidos los parámetros. Si los parámetros están muy dispersos (sparse) el ratio de aprendizaje aumentará. Adam es una actualización del Optimizador RMSProp (Root Mean Square Propagation) que se basa en un ratio de aprendizaje adaptativa.

El algoritmo ADAM según (Kingma & Ba, 2014) calcula la dirección de descenso usando momentum y utiliza una estrategia similar para calcular adaptar el tamaño de paso. Es decir, utiliza momentum para actualizar el paso, lo que evita cambios bruscos en el mismo. Esto lo hace muy estable para su uso en una estrategia tipo Gradiente Estocástico (SGD) donde las muestras pueden provocar cambios grandes en la magnitud del gradiente, además calcula un paso global en vez de usar un paso para cada variable. También adecuado en estrategias de

entrenamiento tipo estocásticas o por lotes, como en el caso de Redes Neuronales Profundas (Deep Learning). Una mejora importante es la corrección del sesgo (bias) en la estimación de los momentos. Generalmente las razones de aprendizaje (momentum) son cercanas a 1.

4 Resultados

En la primera red entrenada con SGD y ratio de aprendizaje de 0.003 con 20 iteraciones se obtuvo un error de entrenamiento de 0.6933326721191406.

Para la segunda variante entrenada con Adam y un ratio de aprendizaje de 0.01 en 100 iteraciones se obtuvo un error de entrenamiento de 0.9620 y una precisión de 56%.

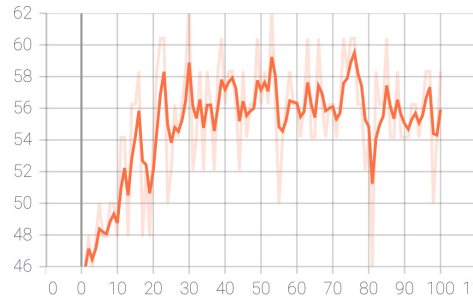


Fig. 7. Comportamiento de la precisión durante la predicción en el conjunto de prueba.

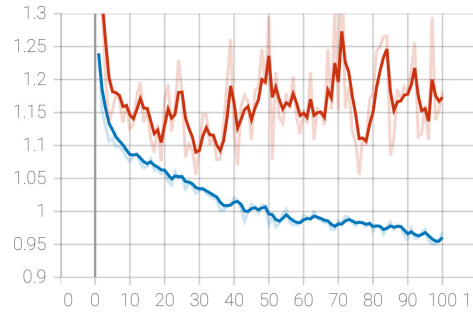


Fig. 8. Comportamiento de los errores de entrenamiento y validación.

5 Conclusiones

Una comprensión cabal de las propiedades físico-químicas del vino blanco es fundamental para tener éxito en la predicción de su calidad utilizando redes neuronales. Se comprobó que el grado de alcohol es la característica que más se tiene en cuenta para valorar la calidad del vino, en cambio el ácido cítrico, el dióxido de azufre libre y los sulfatos tienen poca o ninguna influencia en la calidad. Como algunas de estas variables físico-químicas pueden controlarse desde el proceso de producción, la información obtenida puede utilizarse para mejorar la calidad del vino desde la fábrica. Se propusieron dos modelos de redes neuronales y se implementaron utilizando el framework PyTorch, visualizándose los gráficos con Tensorboard. Con ambos modelos se obtuvieron resultados aceptables.

References

- Ambrogioni, L., Güçlü, U., van Gerven, M. A., & Maris, E. (2017). The kernel mixture network: A nonparametric method for conditional density estimation of continuous random variables. *arXiv preprint arXiv:1705.07111*.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4), 547–553.
- Er, Y., & Atasoy, A. (2016). The classification of white wine and red wine according to their physicochemical qualities. *International Journal of Intelligent Systems and Applications in Engineering*, 23–26.
- Gonzalez, S., & Miikkulainen, R. (2020). Improved training speed, accuracy, and data utilization through loss function optimization. In *2020 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1–8).
- Jurado, M., & Fellman, H. (2020). Modelos de redes neuronales artificiales, como sustento evaluativo al crecimiento pedagógico virtual en educación superior. *Educación Superior*, 7(2), 25–36.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Radosavljević, D., Ilić, S., & Pitulić, S. (n.d.). A data mining approach to wine quality prediction.
- Ramazan, Ü. (n.d.). The role of artificial intelligence in productivity: A case study of wine quality prediction. *Avrupa Bilim ve Teknoloji Dergisi*(20), 280–286.
- Raschka, S. (2015). *Python machine learning*. Packt publishing ltd.
- Vlassides, S., Ferrier, J. G., & Block, D. E. (2001). Using historical data for bioprocess optimization: modeling wine characteristics using artificial neural networks and archived process information. *Biotechnology and Bioengineering*, 73(1), 55–68.
- Ye, C., Li, K., & Jia, G.-z. (2020). A new red wine prediction framework using machine learning. In *Journal of physics: Conference series* (Vol. 1684, p. 012067).

- Yu, H., Lin, H., Xu, H., Ying, Y., Li, B., & Pan, X. (2008). Prediction of enological parameters and discrimination of rice wine age using least-squares support vector machines and near infrared spectroscopy. *Journal of agricultural and food chemistry*, 56(2), 307–313.