

# **Análisis y Reporte Sobre el Desempeño del Modelo**

Por : Sebastian Mora  
A01746423  
10 de Septiembre del 2023  
Inteligencia Artificial Avanzada  
ITESM

## **1. Justificación del Dataset**

El conjunto de datos de calidad del vino representa una amalgama de características químicas y físicas que describen el perfil de diferentes vinos. Estas características, aunque individuales en su naturaleza, colectivamente determinan la percepción de la calidad de un vino, lo que convierte a este conjunto de datos en un desafío multidimensional y complejo. Dado que cada característica puede influir de manera no lineal en la determinación final de la calidad, se requiere un modelo capaz de capturar estas sutiles interacciones entre características, y es aquí donde una red neuronal brilla con fuerza. Las redes neuronales, por su diseño intrínseco, son excepcionales al modelar interacciones complejas entre múltiples variables, y su capacidad para aprender y generalizar patrones a partir de datos de entrenamiento las

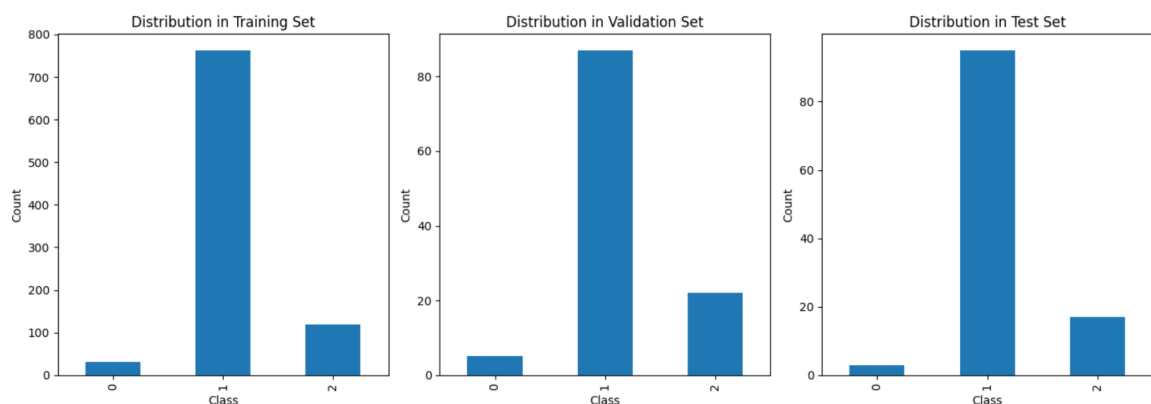
hace especialmente aptas para este tipo de tareas. Además, al transformar la métrica de calidad del vino en una tarea de clasificación tridimensional (malo, regular, bueno), se introduce una claridad adicional en el objetivo, permitiendo que el modelo se centre en la categorización en lugar de una regresión numérica exacta. En resumen, dada la complejidad inherente del conjunto de datos de calidad del vino y la formidable habilidad de las redes neuronales para modelar relaciones no lineales, la elección de este modelo para este dataset es no sólo apropiada, sino también estratégica para lograr una clasificación precisa.

## 2. Separación y evaluación del modelo con un conjunto de prueba y un conjunto de validación

El dataset completo consta de 1144 renglones, cada renglón consta de diversas características químicas y físicas que describen el perfil de diferentes vinos. El dataset será dividido en 3 subconjuntos; *test* (80% del total de renglones), *test* (10% del total de renglones) y *valid* (10% del total de renglones). Dejándonos con 3 sectores de la siguientes dimensiones :

```
Train set: X: (914, 11), y: (914,)
Validation set: X: (114, 11), y: (114,)
Test set: X: (115, 11), y: (115,)
```

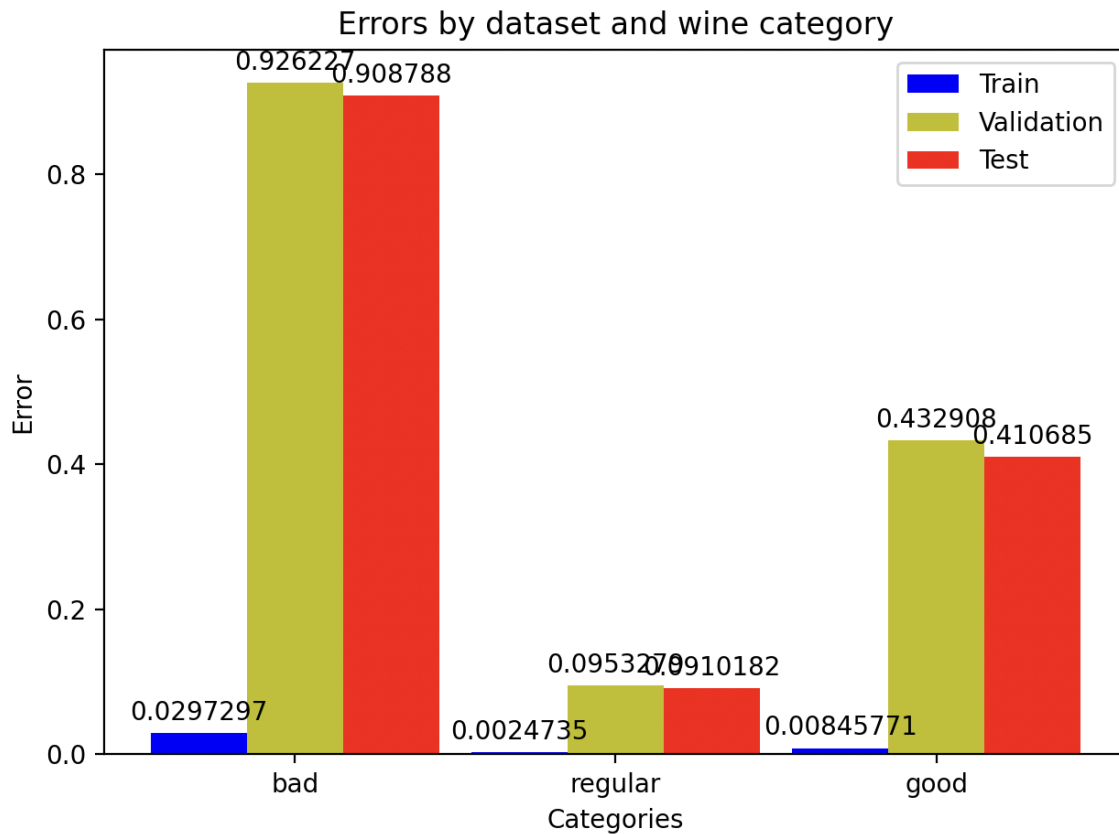
Por motivos de simplicidad y poder hacer un mejor análisis del funcionamiento del modelo, se ha convertido de un problema de regresión a un problema de clasificación, por lo que la *y* del dataset ha sido clasificada (de acuerdo a la calificación de cada vino) en uno de tres grupos; malo, regular y bueno. A continuación podemos ver una gráfica de la distribución de las clases en los 3 subconjuntos descritos anteriormente :



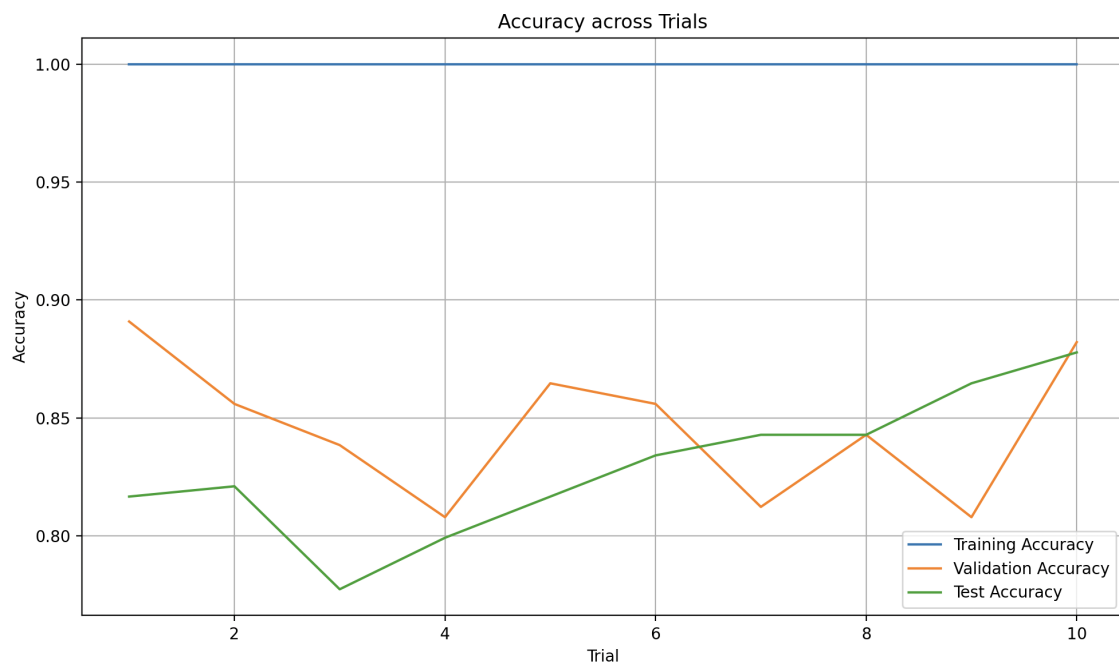
Gráficas de distribución de las clases en cada subconjunto; 0 malo, 1 regular, 2 bueno.

### 3. Diagnóstico de bias y fitting

Para poder hacer el diagnóstico, se entrenaron y evaluaron 10 modelos de red neuronal, recolectando los datos sobre su valor de f1 para cada uno de los subconjuntos de los dataset descritos anteriormente. A continuación, se muestran los gráficos para su posterior análisis y conclusión.



*Valor de f1 para cada categoría, promediando todos los modelos.*



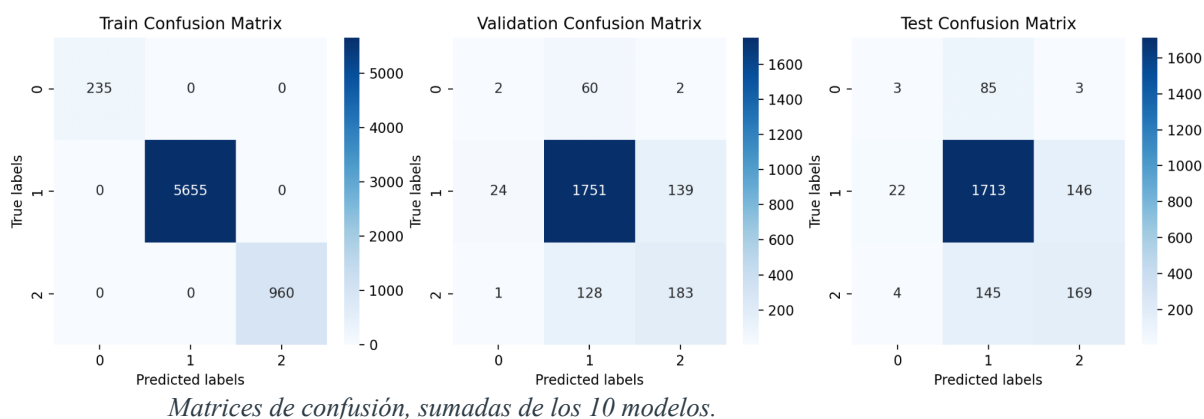
*Precisión de cada modelo, en cada subconjunto del dataset.*

El modelo de clasificación presenta ciertos patrones que vale la pena destacar. En primer lugar, el desempeño al clasificar vinos de calidad "bad" es destacable, alcanzando un F1 score de 0.92. Esto sugiere una eficaz identificación de esta categoría. Sin embargo, el rendimiento para la clasificación de vinos "regular" es considerablemente bajo, con un F1 score cercano a 0.09, indicando dificultades en su correcta clasificación. En cuanto a los vinos de calidad "good", el modelo muestra un rendimiento moderado con un F1 score de 0.4.

Un aspecto crítico del modelo es su capacidad para ajustarse perfectamente a los datos de entrenamiento, logrando una precisión del 100%. Aunque esto podría parecer positivo a primera vista, es un claro indicador de sobreajuste, ya que el modelo está memorizando los datos en lugar de aprender patrones generalizables. Esta hipótesis se refuerza con la disminución observada en la precisión en los conjuntos de validación y prueba, que fluctúa entre 0.8 y 0.9.

En términos generales, el modelo tiene un bajo bias, dado su alto rendimiento en el conjunto de entrenamiento, pero una alta varianza, reflejada en la discrepancia entre los resultados de entrenamiento y validación/prueba. Para abordar estos desafíos, sería recomendable considerar técnicas de regularización, posiblemente obtener o generar más datos, simplificar la arquitectura del modelo y revisar el balance de clases y las características proporcionadas.

#### 4. Diagnóstico y explicación el grado de varianza



El análisis de las matrices de confusión consolidadas para los subconjuntos de entrenamiento, validación y prueba revela información crucial sobre la varianza y el rendimiento general del modelo. Es notable que el cuadrante central de la matriz de confusión del conjunto de entrenamiento presenta un tono significativamente más oscuro, lo que indica que el modelo tiene una alta precisión al predecir la clase central, es decir, vinos de calidad "regular". Esto es coherente con las altas tasas de precisión observadas en el conjunto de entrenamiento.

No obstante, al examinar las matrices de validación y prueba, la intensidad del cuadrante central disminuye, sugiriendo una disminución en la precisión para la clase "regular". Aunque el modelo parece estar sobre ajustando a la clase "regular" durante el entrenamiento, este sobreajuste no se transfiere completamente a los conjuntos de validación y prueba.

El cuadrante inferior derecho, que representa la predicción de vinos de calidad "buena", muestra valores más altos que el cuadrante superior izquierdo, indicativo de vinos de calidad "mala". Esto sugiere que el modelo tiene un rendimiento relativamente mejor al predecir vinos de buena calidad en comparación con los de mala calidad. Sin embargo, aún es notable que la capacidad del modelo para predecir vinos de calidad "buena" y "mala" es inferior a su capacidad para identificar vinos "regulares".

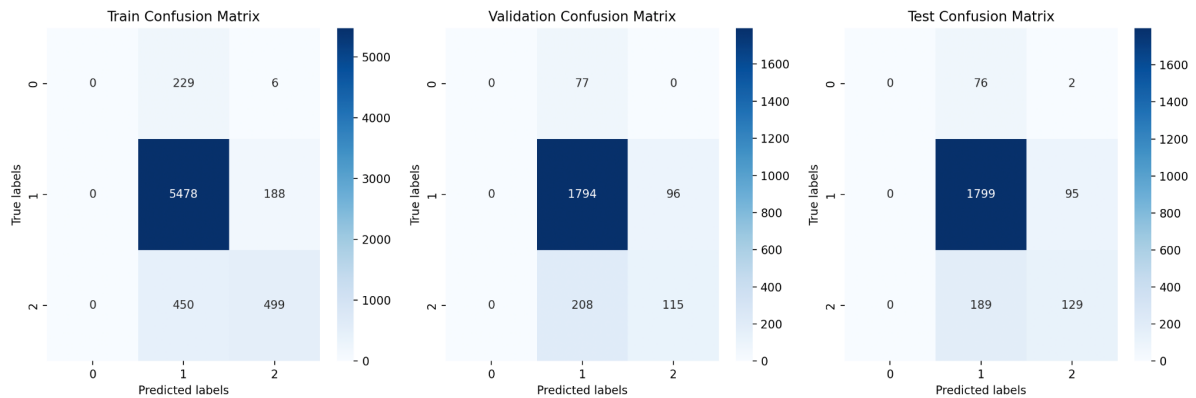
Si bien el modelo exhibe una notable precisión en el conjunto de entrenamiento, especialmente para vinos de calidad "regular", este rendimiento no se mantiene completamente en los conjuntos de validación y prueba, lo que indica una cierta varianza en el modelo. Además, la capacidad del modelo para predecir las clases "buena" y "mala" es inferior a su rendimiento con la clase "regular", lo que refleja posibles áreas de mejora en la clasificación multiclase.

## **5. Regularización para la mejora del modelo**

En el código, se han introducido diversas técnicas de regularización y ajuste con el propósito de optimizar el rendimiento del modelo y abordar posibles problemas, como el sobreajuste. En primer lugar, se ha incorporado la regularización L2, que añade un término de penalización al modelo basado en los pesos de las neuronas. Esta penalización desincentiva los valores extremadamente grandes de los pesos, lo que a su vez evita que la red neuronal se vuelva excesivamente compleja y sea propensa al sobreajuste.

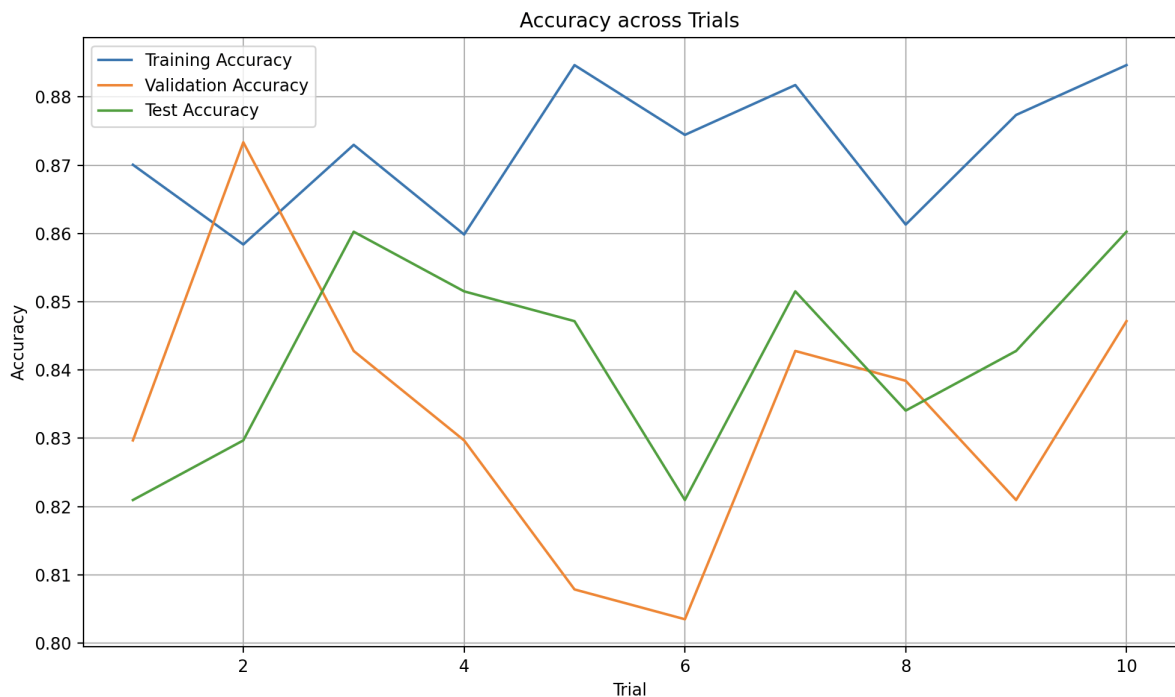
Adicionalmente, se ha incluido la técnica de "Dropout", que durante el entrenamiento desactiva aleatoriamente un porcentaje de neuronas en las capas específicas. Esta desactivación impide que el modelo dependa demasiado de cualquier neurona individual, promoviendo una distribución más uniforme del aprendizaje y, por ende, mejorando la generalización en datos no vistos.

Por último, se ha modificado el optimizador Adam, ajustando la tasa de aprendizaje. El ajuste de la tasa de aprendizaje es fundamental para la convergencia del entrenamiento. Una tasa de aprendizaje demasiado alta podría hacer que el modelo oscile y no converja, mientras que una tasa demasiado baja podría hacer que el aprendizaje sea extremadamente lento. Al modificar esta tasa, buscamos un equilibrio que permita al modelo aprender de manera efectiva y eficiente.



*Matriz de confusión, después de las modificaciones mencionadas.*

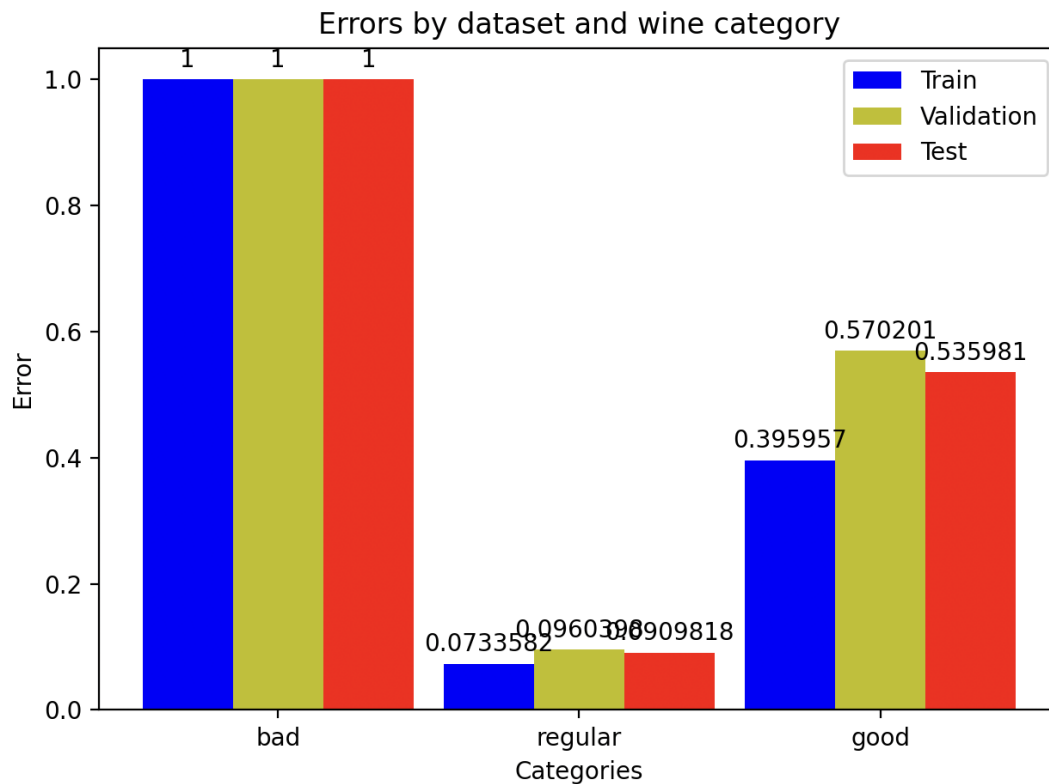
A partir de las modificaciones realizadas en el modelo, es posible discernir una serie de cambios positivos en su desempeño. En un análisis detallado de la matriz de confusión, aunque la clasificación aún se centra primordialmente en los vinos de categoría regular, lo que es realmente notable es la disminución apreciable en las confusiones de esta categoría con las de vinos malos y buenos. Esto se evidencia en los cuadrantes laterales de la matriz, que muestran valores más bajos en comparación con pruebas anteriores, reflejando una mejora en la capacidad discriminativa del modelo.



*Precisión de cada modelo, después de las mejores, en cada subconjunto del dataset*

En cuanto a la gráfica de precisión, se observa una mitigación del sobreajuste que se manifestaba en iteraciones previas. Esto es evidente al ver que la precisión durante el entrenamiento ya no se mantiene constante, sino que presenta ciertas fluctuaciones, lo cual es un indicador de que el modelo está generalizando mejor y no simplemente

memorizando los datos de entrenamiento. Pese a este ajuste, las métricas de validación y prueba mantienen una consistencia con las evaluaciones anteriores, lo que sugiere que las regularizaciones y ajustes no comprometen la capacidad del modelo para hacer predicciones acertadas en datos no vistos previamente.



*Valor de f1 para cada categoría, promediando todos los modelos.*

Finalmente, al observar la gráfica de error por categoría y subconjunto, es destacable el avance significativo en la métrica F1 para la categoría de vinos malos. Esta mejora es evidencia de que el modelo ha fortalecido su capacidad para identificar correctamente esta categoría, equilibrando la precisión y la exhaustividad. Además, también se nota una mejora sutil, pero importante, en la categoría regular. Estas mejoras en conjunto reflejan la eficacia de las técnicas de regularización y ajuste de parámetros aplicadas, consolidando un modelo más robusto y preciso.