

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE INGENIERÍA, ARQUITECTURA Y DISEÑO
MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA



TESIS

“Análisis de la calidad del vino proveniente del Valle de Santo
Tomás mediante técnicas de inteligencia artificial”

Que para cubrir parcialmente los requisitos necesarios para obtener el grado de
MAESTRA EN INGENIERÍA

presenta:

Briseida Rojas Villavicencio

Director de tesis

Dr. Diego Alfredo Tlapa Mendoza

Ensenada, Baja California, Junio de 2025

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE INGENIERÍA, ARQUITECTURA Y DISEÑO

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

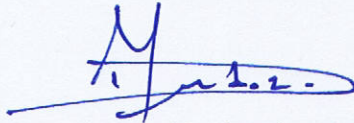
Análisis de la calidad del vino proveniente del Valle de Santo Tomás mediante técnicas de inteligencia artificial

TESIS

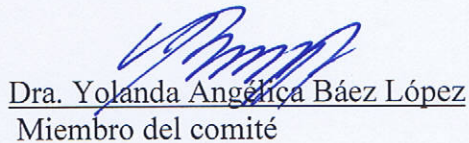
Que para obtener el grado de Maestría en Ingeniería presenta:

Briseida Rojas Villavicencio

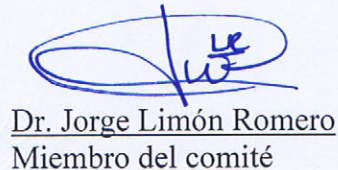
Aprobada por:



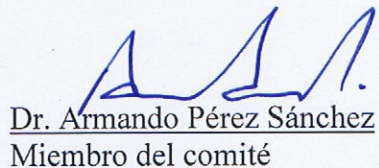
Dr. Diego Alfredo Tlapa Mendoza
Director de tesis



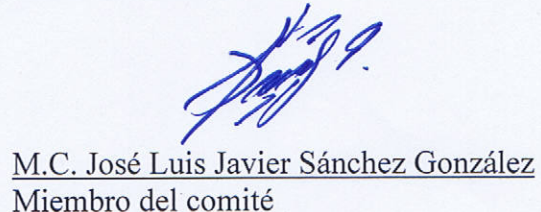
Dra. Yolanda Angélica Báez López
Miembro del comité



Dr. Jorge Limón Romero
Miembro del comité



Dr. Armando Pérez Sánchez
Miembro del comité



M.C. José Luis Javier Sánchez González
Miembro del comité

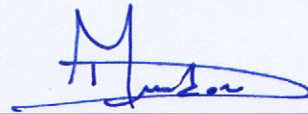
Ensenada Baja California, México. 9 de junio del 2025

RESUMEN

La producción de vino es una actividad económica relevante de Baja California. La calidad del vino toma mayor importancia conforme el mercado crece en competitividad. La calidad del vino se define por metodologías cuantitativas y cualitativas, sin embargo, la subjetividad pone en riesgo la determinación de la calidad de manera eficiente. En el presente proyecto se analizó la calidad del vino del Valle de Santo Tomás, mediante análisis fisicoquímicos, empleando dos técnicas de inteligencia artificial. Se realizó una revisión sistemática siguiendo la metodología PRISMA, misma que arrojó una ausencia de estudios en el país que aborden temas de calidad del vino mediante inteligencia artificial. Se unificaron bases de datos con las variables fisicoquímicas del periodo 2020 a 2024, posteriormente, se diseñaron códigos de Python con técnicas de análisis de datos con relaciones no lineales, en este caso entre las clases de calidad y las variables fisicoquímicas. Se diseñaron dos modelos, el primero fue un árbol de decisión, con una precisión del 93%, es un modelo preciso que logra clasificar la clase 6 con un rendimiento cercano al 100%. El segundo modelo es una Red neuronal, con una precisión del 94% se obtuvo un modelo preciso y estable, en el cual se determinó que las variables fisicoquímicas con mayor influencia en la calidad del vino. El modelo de Redes Neuronales presentó un resultado que se relaciona en mayor medida con el análisis sensorial llevado a cabo por productores de vino de la región. Las tres hipótesis planteadas recibieron respuesta, la primera fue la única que no se aprobó ya que se identificaron 7 de las 8 variables planteadas en la H1 (AT, AV, Gr.AI%, °Brix, AZ, NFA, SO₂L y pH). La hipótesis 2 fue aceptada, las variables °Brix, AT y pH influyen positivamente en la calidad del vino, mientras que AV, AM, SO₂L y SO₂m influyen negativamente. Y finalmente la hipótesis número 3, se desarrolló un modelo de árbol de decisión y uno de redes neuronales, ambos predicen la calidad del vino con una precisión superior al 90%.

Palabras Clave: *Calidad, vino, red neuronal, árbol de decisión, análisis fisicoquímico*

aprobado por:



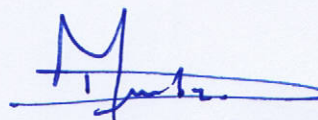
Dr. Diego Alfredo Tlapa Mendoza
Director de Tesis

ABSTRACT

Wine production is a relevant economic activity in Baja California. As the market becomes more competitive, wine quality gains greater importance. Wine quality is defined through both quantitative and qualitative methodologies; however, subjectivity risks determining quality efficiently. This project analyzed the quality of wines from Santo Tomás Valley using physicochemical analyses and implementing two artificial intelligence techniques. A systematic literature review was conducted following the PRISMA methodology, which revealed a lack of studies in the country addressing wine quality through artificial intelligence. Databases containing physicochemical variables from 2020 to 2024 were consolidated, and Python codes were developed to apply data analysis techniques focusing on nonlinear relationships between quality classes and physicochemical variables. Two models were developed: the first one is a decision tree, with an accuracy of 93% and was effective in classifying quality class 6 with nearly 100% of the performance. The second model is a Neural Network, with an accuracy of 94%, providing a precise and stable prediction model that identified the most influential physicochemical variables affecting wine quality. The Neural Network model demonstrated a higher correlation with the sensory analysis performed by local wine producers. All three hypotheses were addressed: the first was the only one not accepted and 7 of the 8 proposed variables in H1 (AT, AV, Alc.Vol, °Brix, AZ, NFA, SO2L, and pH) were identified. Hypothesis 2 was accepted, variables such as °Brix, AT, and pH showed a positive influence on wine quality, while AV, AM, SO2L, and SO2m had a negative impact. Finally, hypothesis 3 was confirmed: both the decision tree and neural network models predict wine quality with accuracy above of the 90%.

Key words: *Quality, wine, neural network, decision tree, physicochemical analysis*

Approved by:



Dr. Diego Alfredo Tlapa Mendoza
Director de Tesis

DEDICATORIA

Dedico este trabajo a:

Brianda, mi incondicional, luz y fortaleza;

William, mi refugio e inspiración;

Kenia, mi motor de motivación.

Gracias por estar siempre conmigo y seguir en el mismo barco.

AGRADECIMIENTOS

Agradezco infinitamente a mi familia, sin su apoyo incondicional nada es posible, son el pilar fundamental de mi vida. Su paciencia y fé en mí me ayudan a alcanzar esta meta. En especial a mi hermana Brianda, quien fue la que me impulsó en todo momento. Gracias Kenia por tus palabras de aliento en el cansancio y tu ánimo en el desvelo.

A la institución UABC que me ayudó a formarme, así como al MyDCI por la oportunidad de ingresar al programa y concluirlo con éxito. A todos los maestros pertenecientes a la FIAD que me dieron espacios de aprendizaje y crecimiento.

A mi asesor el Dr. Diego Alfredo Tlapa Mendoza, sin su acompañamiento, compromiso y guía no hubiera completado esta etapa. Su tiempo y disposición marcaron la diferencia en mi proceso de formación.

A mis amigos y compañeros, quienes estuvieron presentes en todo momento, gracias por su compañía y palabras de aliento. Gracias Wiliam por inspirarme en todo momento y mostrarme cómo debe sentirse ser un profesional. A Amairani por mostrarme que el camino del éxito va de la mano con el profesionalismo.

Agradezco especialmente a la empresa Bodegas de Santo Tomás por ayudarme a concluir este proyecto, por siempre tener la disponibilidad de ayudar y mejorar, hicieron este trayecto más sencillo.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación por el apoyo económico brindado durante todo el periodo de la maestría.

Finalmente, agradezco a todas las personas que, de una u otra manera, formaron parte de este camino, a quienes me hicieron creer en mí y no dejaron que me rindiera.

TABLA DE CONTENIDO

1 INTRODUCCIÓN	1
1.1 Antecedentes	3
1.2 Contexto de la Investigación	5
1.3 Planteamiento del Problema	6
1.4 Preguntas de Investigación	7
1.5 Hipótesis	7
1.6 Objetivos	8
1.6.1. Objetivo general.	8
1.6.2. Objetivos específicos.	8
1.7 Justificación	9
1.8 Delimitación y Limitación	10
2 MARCO TEÓRICO	11
2.1 Definición de vino	11
2.1.1 Calidad en el vino	11
2.1.1.1. Análisis sensorial	12
2.1.1.2. Análisis fisicoquímico	12
2.1.2 Composición del fruto y aportaciones en el vino	13
2.1.2.1. Acidez total (AT)	14
2.1.2.4. Anhídrido sulfuroso libre y molecular (SO ₂ L y SO ₂ m)	15
2.1.2.5. Azúcar residual (AZ)	16
2.1.2.7 Grados Brix (°Brix)	17
2.1.2.8. Nitrógeno fácilmente asimilable (NFA)	17
2.1.2.9. Potencial de hidrógeno (pH)	18
2.2 Posición del vino Mexicano	18
2.2.1 Vitivinicultura en Baja California, México	19
2.2.1.1. Valle de Santo Tomás	19
2.3 Inteligencia artificial	20
2.3.1. Python	20
2.3.2. Modelos de predicción	21
2.3.2.1. Árboles de decisión	21
2.3.2.2. Redes neuronales	22
3 METODOLOGÍA	23
3.1 Definición de las variables independientes y dependientes	23
3.1.1. Realizar revisión sistemática del tema	23
3.1.2. Selección de variables independientes	25
3.2. Medición del problema	28
3.2.1. Generación de datos: Bases de datos	28
3.2.2. Generación de datos: Metodologías de análisis químicos, enzimáticos y	

fisicoquímicos	29
3.2.2.1. Acidez total (AT)	30
3.2.2.2. Grados Brix (°Bx)	31
3.2.2.3. Nitrógeno fácilmente asimilable (NFA)	31
3.2.2.4. Azúcar residual (AZ)	32
3.2.2.5. Acidez volátil (AV)	33
3.2.2.6. Grado alcohólico (Gr.Al%)	34
3.2.2.7. Anhídrido sulfuroso libre	35
3.2.2.8. Anhídrido sulfuroso molecular	35
3.2.2.9. Ácido málico	36
3.2.2.10. Potencial de hidrógeno (pH)	37
3.3. Análisis de datos	38
3.3.1. Correlación de variables.	38
3.3.2. Diseño de árbol de decisión	38
3.3.3. Diseño de redes neuronales	39
3.4 Mejorar la precisión de los modelos obtenidos	40
3.4.1. Ajuste del modelo de árbol de decisión generado	40
3.4.2. Ajuste del modelo de redes neuronales generado	40
4 RESULTADOS Y DISCUSIÓN	41
4.1 Definición de variables independientes	41
4.1.1. Realizar revisión sistemática del tema	41
4.1.2. Selección de variables de interés	43
4.2 Medición del problema	48
4.2.1. Base de datos unificada 2020-2023	48
4.2.2. Análisis de variables 2024	49
4.3 Análisis de datos	50
4.3.1. Correlación de variables fisicoquímicas (Anexo 1, Figura 11)	50
4.3.2. Árbol de decisión	52
4.3.3. Redes neuronales	54
4.4 Mejora en los modelos de predicción	55
4.4.2. Modelo mejorado de Redes neuronales (Anexo 1, Figuras 15-18)	59
4.4.3. Comparación de los modelos de predicción generados	62
5 CONCLUSIONES Y RECOMENDACIONES	65
5.1 Conclusiones	65
5.2 Recomendaciones	66
5.2.1 Trabajo Futuro	67
6 REFERENCIAS	68
7 ANEXOS	81
ANEXO 1. Códigos generados en Python	81

ÍNDICE DE FIGURAS

Figura 1. Diagrama de flujo PRISMA de la revisión sistemática con la información de descarte de trabajos.	42
Figura 2. Correlación entre las variables independientes y calidad mediante escala de colores.	51
Figura 3. Representación gráfica del árbol de decisión obtenido a partir de la base de datos cargada.	52
Figura 4. Resultados del entrenamiento del modelo de redes neuronales, mostrando un modelo sobre ajustado.	55
Figura 5. Árbol de decisión mejorado con validación cruzada y ajuste de hiperparámetros.	56
Figura 6. Matriz de confusión del modelo de árbol de decisión después de las mejoras.	57
Figura 7. Gráfico de la importancia de las variables fisicoquímicas en relación con la calidad a partir del árbol de decisión.	58
Figura 8. Matriz de confusión generada en el modelo de redes neuronales después de las mejoras.	59
Figura 9. Resultados del entrenamiento del modelo mejorado de redes neuronales.	60
Figura 10. Gráfico de la importancia de las variables fisicoquímicas en relación con la calidad a partir del modelo de redes neuronales.	61
Figura 11. Código de Python empleado para la obtención de la correlación de las variables fisicoquímicas.	81
Figura 12. Código de Python empleado para la obtención del árbol de decisión a partir de variables fisicoquímicas.	82
Figura 13. Código de Python empleado para el modelo de predicción de redes neuronales.	83
Figura 14. Modificación en el código de Python del modelo árbol de decisión para aumentar la precisión del modelo.	84
Figura 15. Parte 1 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.	85
Figura 16. Parte 2 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.	86
Figura 17. Parte 3 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.	87
Figura 18. Parte 4 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.	88

ÍNDICE DE TABLAS

Tabla 1. Primera parte del ejercicio de enología para determinar la calidad del vino a partir de análisis fisicoquímicos y sensoriales.....	26
Tabla 2. Segunda parte del ejercicio de enología para determinar la calidad del vino a partir de análisis fisicoquímicos y sensoriales.....	26
Tabla 3. Diseño de tabla empleada en la unificación de datos proporcionados por año.....	28
Tabla 4. Análisis realizados junto con su unidad de medida.....	29
Tabla 5. Extracto de la comparativa de los trabajos realizados sobre la calidad del vino en diferentes países.....	41
Tabla 6. Resultados de la revisión realizada dividida por temas principales.....	43
Tabla 7. Resultados del primer ejercicio de enología para la obtención de variables.....	45
Tabla 8. Resultados del segundo ejercicio de enología para la obtención de variables fisicoquímicas.....	46
Tabla 9. Variables de interés según los objetivos planteados.....	47
Tabla 10. Extracto de la base de datos de análisis químicos, fisicoquímicos y enzimáticos.....	48
Tabla 11. Análisis químicos, fisicoquímicos y enzimáticos a vinos 2024.....	49
Tabla 12. Métricas de evaluación de las técnicas empleadas.....	62
Tabla 13. Comparativa de los métodos empleados para identificar variables.....	63

1 INTRODUCCIÓN

El vino se puede definir como una bebida alcohólica producida por la fermentación del fruto de *Vitis vinífera*. Es la bebida alcohólica con mayor historia y tradición, por lo tanto todo lo referente a este producto genera inquietud en la sociedad (Blouin y Peynaud, 2003).

En México se cuenta con una superficie de viñedos de alrededor de 36,000 hectáreas, con una producción de Baja California de alrededor de 32.3 millones de litros de vino, reportados hasta 2023, es decir, representa el 85% de la producción anual de vino de todo México. El municipio de Baja California donde se encuentra el 90% de la producción del estado es Ensenada, y dentro del municipio hay 7 valles líderes en la producción de vino: Guadalupe, Ojos Negros, Santo Tomás, San Vicente, La Grulla, Uruapan y El Porvenir (Bahena, 2023).

Ser consistente en la calidad del vino que llega al consumidor es indispensable en la actualidad, debido a la competencia en el mercado, se debe una estandarización y automatización de los procesos, para de esta manera cumplir con los estándares de calidad nacionales e internacionales (Oczkowski, 2016). Actualmente la calidad del vino está representada por dos evaluaciones principales. Los análisis sensoriales se realizan con personas expertas en el tema, con las certificaciones necesarias para evaluar la calidad de un vino mediante el empleo de solo sus sentidos. Mientras que el análisis fisicoquímico, es hacer la carta descriptiva de la química del vino y de esta manera, verificar que cumpla con estándares establecidos por organismos nacionales e internacionales. En resumen, se emplea una metodología cualitativa y una cuantitativa (Organización Internacional de la Viña y el Vino [OIV], 2015).

Emplear técnicas de inteligencia artificial actualmente es de gran relevancia en todo proceso de manufactura. Enfocándonos en la industria alimentaria, el analizar relaciones no lineales entre variables es fundamental para muchos de sus procesos. Evaluar la calidad del vino mediante modelos que analicen de esta manera los datos proporciona una evaluación con mayor objetividad (Geetha et al, 2024).

Árboles de decisión y redes neuronales son modelos de aprendizaje automático, supervisados, que requieren de un entrenamiento proporcionado por los datos que se estudiarán, ambos tienen la capacidad de predecir, en este caso, predecir la calidad del vino mediante análisis fisicoquímicos, ya que son las variables cuantitativas que se emplean en el análisis de la calidad en primer lugar. Si bien ambos modelos pueden realizar la predicción, lo hacen de manera diferente y con diferentes consideraciones, por lo que al emplear una misma base de datos, generan información diferente y con diferente precisión (Mahima et al, 2020).

En la revisión de la literatura, ya existen publicaciones que tocan temas sobre la calidad del vino, ya sea mediante análisis fisicoquímicos o análisis sensoriales, y más allá, empleando técnicas de inteligencia artificial para el análisis de los datos. Aunque dentro de estos estudios, hay muchos sobre “vinos” producidos con otro tipo de frutos o estudios que se enfocan a una o dos variedades de *Vitis vinífera*, y sobre todo, enfocándose mayormente en vinos tintos. Una de las maneras más efectivas de evaluar un modelo de predicción es mediante la métrica de precisión, en la revisión se observaron precisiones que van desde el 50% hasta el 99%.

Los trabajos publicados respecto a la predicción de la calidad del vino mediante análisis fisicoquímicos es nula. Por lo que realizar este tipo de estudios proporciona información valiosa a manera de diagnosticar la situación de la región. El objetivo de obtener este modelo va más allá de generar publicaciones científicas, sino de crear una metodología que ayude a los productores de la región, ayudando a pequeñas y medianas empresas que no pueden cubrir los gastos que se generan con una evaluación sensorial de la calidad del vino de manera continua en sus procesos.

Considerando lo anterior el objetivo del presente trabajo es determinar la calidad del vino mediante técnicas de aprendizaje automático, identificando las variables fisicoquímicas más relevantes en vino del Valle de Santo Tomás.

1.1 Antecedentes

Conforme la tecnología sigue su evolución, los cuestionamientos aumentan, un ejemplo es la industria alimentaria, donde el volumen y diversidad de datos que se obtienen mediante los análisis modernos está en un constante crecimiento. Existe un catálogo amplio de las técnicas y metodologías de minería de datos que son aplicados, sólo por repasar algunos, se encuentran los análisis discriminantes de mínimos cuadrados parciales, el modelado independiente suave por analogía de clase, el análisis factorial paralelos, árbol de clasificación y regresión, y los basados en el modelo ANN (red neuronal artificial), entre muchos otros. Las redes neuronales tienen la capacidad de ser tolerantes a fallos en los primeros métodos computacionales, con esta capacidad inherente, si se produce un fallo en un número no muy grande de neuronas, el sistema no sufre una caída repentina, aunque el comportamiento si se ve influenciado (Pletcher y Wills, 2005).

Años más tarde, en 2014, el equipo de Yadav, Malik y Chandel (2014) utilizó este fundamento para desarrollar modelos de predicción de radiación solar empleando el software Waikato Environment for Knowledge Analysis. WEKA es reconocido como un sistema de referencia en minería de datos y aprendizaje automático, ampliamente utilizado por investigadores para implementar nuevos algoritmos sin tener que preocuparse por la infraestructura de apoyo para los datos, manipulación y evaluación de esquemas. Uno de los algoritmos que más se utiliza en el programa es el denominado J48, construye un árbol de decisión a partir del conjunto de datos, el algoritmo considera todas las pruebas posibles a partir de la división de datos y selecciona la que mejor dé la información con los parámetros establecidos (Solanki, 2014).

Una aplicación de la minería de datos es el proyecto desarrollado en Colombia donde el objetivo era establecer por medio de técnicas de inteligencia artificial las variables más influyentes en la calidad sensorial del vino, lográndolo con una efectividad superior al 95% (Galeano-Arias, Aguirre y Castrillón-Gómez, 2021). En 2022 en Nueva Zelanda aplicaron técnicas de inteligencia artificial para estimar la calidad del vino producido con uva Pinot Noir. El estudio que se realizó proporcionó evidencia sobre el uso de minería de datos, se emplearon los modelos RF y AdaBoost

como clasificadores automáticos, de esta manera de las 55 variables estudiadas, se seleccionaron las que tienen un mayor impacto en la calidad del vino. Al final del estudio se inicia el desarrollo de nuevos proyectos que apoyen a enólogos y viticultores, para que puedan predecir la calidad del vino a partir de los compuestos químicos y fisicoquímicos desde la uva hasta la generación del vino (Bhardwaj, Tiwari, Olejar, Parr, y Kulasiri, 2022).

El análisis de la calidad en el vino se puede realizar desde la composición química del vino o de los cultivos, hasta por aspectos económicos o sociales. Hablando de México, en 2012, específicamente en Querétaro se realizó un proyecto donde el propósito era caracterizar fisicoquímicamente y relacionarlo con la calidad, el diseño experimental fue completamente al azar con cuatro repeticiones y arreglo factorial de tratamientos, las variedades de uva que se analizaron fueron: Cabernet Sauvignon, Merlot y Tempranillo. Los resultados se presentaron mediante tablas y gráficas donde se comparó cada dato entre sí, de esta manera se detectó que la variedad Merlot de un cultivo tenía una mayor calidad (Cruz-de Aquino et al, 2012). 10 años más tarde en Chihuahua se desarrolló un proyecto cuyo objetivo era caracterizar los parámetros que determinan la calidad del vino Tempranillo de diferentes cultivos. Un total de 10 variables fueron las que se analizaron de las cuales destacan: color, antocianinas, taninos y grado alcohólico. El análisis que se realizó fue sensorial, es decir, una descripción de las características organolépticas, los resultados preparados fueron gráficos creados para relacionar parámetros como gusto, olfato y color (Fuentes-Verduzco, Lugo-García, Pérez-Leal y Camacho-Inzunza, 2022).

Regionalmente, en 2010 se inició con un proyecto donde se relaciona la calidad del vino, pero con parámetros medioambientales que afectan el proceso. Se eligieron 11 variables, en donde se encuentran análisis de clima, suelo y tipo de nutrientes. Sánchez y Mungaray concluyeron que los factores climáticos, edafológicos, tecnológicos, físicos, educativos y socio institucionales, son propicios para impulsar un crecimiento económico sostenible en el entorno del Valle de Guadalupe, Baja California, la discusión de resultado se fundamenta en una correlación de variables, misma que se pudo lograr con el empleo de un modelo de regresión lineal (Sánchez Zepeda y Mungaray Lagarda, 2010).

1.2 Contexto de la Investigación

El Valle de Santo Tomás es el que tiene mayor antigüedad en la producción de vino respecto a los otros valles vitivinícolas del estado de Baja California, México. En la antigua ruta del vino se encuentran diversas unidades vitivinícolas, desde pequeñas producciones hasta las que superan las 100,000 cajas anuales. Una vinícola que destaca es Bodegas de Santo Tomás, que por más de 136 años se ha posicionado en el mercado con vino de calidad, por lo cuál será la unidad de estudio. Poseen una línea de producción con más de 25 productos diferentes, que se clasifican en su gran mayoría en 3 calidades: Premium, Clásica o varietal y básica.

Bodegas de Santo Tomás proporcionará información de al menos 4 años sobre los análisis fisicoquímicos realizados durante todas las etapas de elaboración del vino, que van desde un seguimiento del fruto hasta el embotellado del vino producido. Estos análisis se realizan en el laboratorio de enología interno que posee la empresa, mismo que cuenta con todas las metodologías necesarias para caracterizar de manera fisicoquímica un vino. Además, el laboratorio realiza más de 30 técnicas analíticas diferentes, desde físicas, químicas, enzimáticas y microbiológicas, éstas últimas necesarias para el correcto seguimiento de la elaboración del vino, y sobre todo, son técnicas que ayudan a evaluar procesos de impacto en la calidad del vino, como es el embotellado del producto, donde la inocuidad es fundamental.

Para realizar un proyecto entre una institución pública y una privada se debe de apegar a los lineamientos de ambas partes. En este caso, para que UABC pueda trabajar con Bodegas de Santo Tomás se firmó un convenio donde se establece que se puede realizar intercambios de capital humano, inteligencia y tecnología. Partiendo de este convenio, se trabajará con las áreas de Enología y Laboratorio de Enología, este último es el que posee las bases de datos con las variables a estudiar. El área de laboratorio tiene diferentes bases de datos, donde marcan una trazabilidad de producto, mismas que se tienen que revisar y realizar un análisis estadístico para encontrar la correlación entre las variables de interés con la calidad.

1.3 Planteamiento del Problema

La calidad del vino actualmente está determinada mediante análisis físicoquímicos y sensoriales, mismos que se evalúan desde el fruto y durante toda la elaboración del vino, en las etapas de crianza cuando el vino está en contacto con madera, y etapas finales del proceso, como son la filtración, estabilización y embotellado del vino. Con base a lo anterior, se emplea un análisis cuantitativo y un análisis cualitativo, el último es llevado a cabo por catadores y expertos en el tema, por medio de degustaciones, donde se analiza al vino desde un punto de vista, es decir, a criterio de la persona que realiza la degustación. Entre más personas externas se incluyen en estos análisis, se sugiere un aumento de gastos en las productoras de vino. Al ser un análisis subjetivo, el control de calidad sensorial genera opiniones divididas dentro de la industria, si bien el análisis sensorial es utilizado en eventos como concursos de vino, ya sean nacionales o internacionales, algunas unidades vitivinícolas tienen personas especializadas en estos análisis, este gasto extra puede generar un desbalance importante en empresas pequeñas y medianas que tengan menos de 5 años produciendo vino.

La dependencia de los factores subjetivos para definir la calidad del vino, conlleva a una afectación en la consistencia de la misma a lo largo del tiempo y en gran medida en la eficiencia operativa de las empresas vinícolas. El mercado competitivo genera mayor presión en posicionar las marcas, y algunos de los factores clave son las fluctuaciones que muchas productoras tienen en la calidad de sus vinos, ya sea por cambios organizacionales o que no tienen establecido un control de calidad eficiente.

A través de la revisión de literatura, se observa que en diferentes regiones vinícolas en el mundo, como Sudamérica han optado por emplear técnicas de inteligencia artificial, con las cuales se han determinado las variables que tienen una mayor influencia en la calidad del vino, los resultados obtenidos sugieren un gran avance en el tema, de igual manera, en Europa, específicamente en España han diversificado el uso de inteligencia artificial a lo largo de todos los procesos de elaboración del vino, pero aún falta determinar si es aplicable en otras áreas

geográficas, por lo que se requiere de una ampliación del espectro muestral, siendo el valle de Santo Tomás en conjunto con el valle de San Vicente una opción viable para aplicar el estudio, ya que tienen condiciones ambientales y de suelo distintivas, aportando diferentes valores nutrimentales que destacan sobre otras regiones vitivinícolas de Baja California y del país en general.

1.4 Preguntas de Investigación

- 1 ¿Cuáles son las variables fisicoquímicas que tienen una mayor influencia en la calidad del vino?
- 2 ¿Cómo influyen las variables fisicoquímicas en la calidad del vino?
- 3 ¿Cómo puede ser desarrollado un modelo de inteligencia artificial con un nivel de precisión comparable a los estándares actuales para predecir la calidad del vino basado en las variables fisicoquímicas seleccionadas?

1.5 Hipótesis

- H1: Las 8 variables con mayor influencia en la calidad del vino son: pH, Acidez total, Azúcares, Acidez volátil, Ácido málico, Nitrógeno fácilmente asimilable, anhídrido sulfuroso libre y molecular.
- H2: Las variables como Grados Brix, acidez total y pH influyen positivamente en la calidad del vino, mientras que acidez volátil, ácido málico, anhídrido sulfuroso libre y molecular influyen negativamente.
- H3: El uso de árboles de decisión o redes neuronales como la primera fase en el desarrollo de un modelo predictivo es el enfoque más adecuado para identificar la relación entre las variables fisicoquímicas y la calidad del vino.

1.6 Objetivos

1.6.1. Objetivo general.

Determinar la calidad del vino mediante técnicas de aprendizaje automático, identificando las variables fisicoquímicas más relevantes en vino del Valle de Santo Tomás.

1.6.2. Objetivos específicos.

- Identificar variables fisicoquímicas relevantes en muestras de vino del Valle de Santo Tomás.
- Desarrollar un modelo predictivo basado en inteligencia artificial para relacionar estas variables con la calidad del vino.
- Probar y validar el modelo con un nivel de precisión comparable con los estándares actuales para predecir la calidad del vino basado en las variables seleccionadas.

1.7 Justificación

La producción de vino en el Valle de Santo Tomás, Baja California, México, ha sido una actividad económica en crecimiento constante, posicionando a la región como un importante jugador en la industria vinícola mexicana. Sin embargo, este sector enfrenta desafíos significativos, especialmente en lo que respecta al control de calidad en el proceso de producción, problemáticas como la cantidad de datos cualitativos o un análisis erróneo de los cuantitativos.

El presente proyecto busca abordar esta problemática mediante la implementación de técnicas de inteligencia artificial para analizar variables fisicoquímicas del vino. El tipo de datos a estudiar tienen relaciones no lineales entre sí y con la calidad del vino, el desarrollo de técnicas como árboles de decisión y redes neuronales hacen que el análisis se pueda efectuar de manera exitosa. Al emplear un enfoque basado en datos cuantitativos, se pretende minimizar la influencia de los parámetros cualitativos en el control de calidad, mejorando así la precisión y la objetividad del proceso de evaluación.

Los resultados de esta investigación no solo tendrán un impacto positivo en las empresas productoras de vino del Valle de Santo Tomás, sino que beneficiarán a todo el sector vitivinícola de México, ya que actualmente no se tienen estudios similares en esta región. Los estudios realizados con mayores similitudes al presente son de otros continentes, principalmente Europa y Asia. Las conclusiones que abordan en la literatura, es que se requiere de una aplicación en otros sectores del mundo de los diferentes modelos de inteligencia artificial desarrollados, ya que las condiciones de elaboración del vino son diferentes en cada región.

La generación de una nueva metodología de análisis de datos mediante inteligencia artificial puede propiciar el desarrollo de proyectos afines y fomentar la innovación en la industria, contribuyendo así al crecimiento sostenible y a la competitividad del sector a nivel nacional e internacional.

1.8 Delimitación y Limitación

En el desarrollo del proyecto solo se toman en cuenta los análisis fisicoquímicos, que es la determinación cuantitativa de la calidad del vino, actualmente, también se emplea un análisis sensorial, si bien, es un análisis subjetivo y variable es empleado en gran medida en la industria vinícola, agregando complejidad a la evaluación de la calidad.

Este proyecto busca generar un modelo predictivo de la calidad de vino a partir de análisis físico químicos, excluyendo factores que afectan en menor medida, como son las condiciones climáticas, el tipo de suelo, prácticas de cultivo, entre otros. La variabilidad de datos puede considerarse como una limitante, ya que al ser diferentes años y diferentes ranchos se puede generar una menor precisión en el modelo resultante. Aunque los valles seleccionados son regiones vitivinícolas reconocidas, son de características específicas, por lo que reproducir el análisis en otro sector puede presentar diferencias significativas en los modelos generados.

El uso de técnicas de inteligencia artificial y sus resultados son en gran medida de la cantidad y calidad de datos de entrada. Los modelos pueden resultar sobre ajustados si no se realizan correctamente, reduciendo de esta manera la precisión obtenida cuando se apliquen a otros datos.

2 MARCO TEÓRICO

2.1 Definición de vino

El vino se describe como una bebida alcohólica resultante de la fermentación de la uva, esta metodología data de hasta 200 años a.C. Como resultados de estas fermentaciones se pueden obtener vinos blancos, rosados o tintos, los cuales difieren de manera sensorial y fisicoquímica, los dos aspectos de calidad que más se evalúan durante un proceso de producción (Blouin y Peynaud, 2003).

2.1.1 Calidad en el vino

Un análisis de calidad sensorial es llevado a cabo mediante degustaciones, donde se estudian aspectos organolépticos del vino, mientras que en un análisis de calidad fisicoquímica intervienen parámetros de química analítica y física, los cuales se interponen de manera directa e indirecta en algunos casos en la calidad del vino, algunos ejemplos más estudiados y reconocidos por la comunidad son pH, grados de azúcar y grados de alcohol, este último parámetro lo podemos encontrar como dato clave en todas las bebidas alcohólicas, de igual manera, hay otros parámetros no tan conocidos pero de gran importancia, como lo es acidez total dada por ácido tartárico, grados Brix, sulfitos añadidos, ácido málico, etc. Si bien, muchos de estos parámetros están relacionados con el proceso de vinificación que se diseñe, también hay influencia por compuestos químicos dados desde la planta, tipo de cosecha, cantidades producidas y el estado fisiológico de la planta pero también del suelo, sin dejar de mencionar los aspectos medioambientales a los cuales un viñedo esté expuesto. Para determinar el efecto en la calidad del vino de una etapa del proceso, actualmente se requiere de un análisis fisicoquímico y de un análisis sensorial. Por ejemplo, para determinar un parámetro de color, se realizan análisis polifenólicos como polímeros pigmentados, antocianinas, taninos, flavanos y/o compuestos fenólicos. Mientras que para el análisis sensorial se tiene parámetros como intensidad y sub calidades de la astringencia, sabor, olor, aroma, mismo que se requiere sea ejecutado por un jurado entrenado, en este tipo de análisis se busca la imparcialidad siendo más efectivo el contratar servicios externos (Valls, Lampreave, Nadal y Arola, 2000).

2.1.1.1. Análisis sensorial

El análisis sensorial es una evaluación de las propiedades organolépticas de un producto perteneciente a la industria alimentaria o en algunas ocasiones en productos de higiene o perfumería. En la industria de las bebidas alcohólicas este tipo de análisis es comúnmente empleado en la detección de fallas en la calidad del producto, como un disciplina científica basados en una evaluación, medición y evaluación de resultados. En el vino, esto puede ser diferente, ya que el análisis sensorial es considerado más un arte que una ciencia, debido a la presencia de atributos que expertos en el tema comentan, es difícil interpretarlos con una respuesta de Si o No (Franco, Naranjo y Moreira, 2021).

Durante el análisis sensorial en vinos, las principales características que son evaluadas son las relacionadas con el aroma, el sabor y el color. La degustación sucede en condiciones de temperatura adecuadas y libre de aromas externos que puedan interferir en el análisis, así como con las copas indicadas para cada tipo de vino. El panel de expertos se compone en su gran mayoría de Sommeliers con las certificaciones correspondientes y enólogos, quienes son los encargados de la elaboración de los vinos (Organización Internacional de la Viña y el Vino [OIV], 2015).

2.1.1.2. Análisis fisicoquímico

Los análisis fisicoquímicos se pueden definir como las técnicas con las que se evalúan las propiedades químicas y físicas del vino. Todas estas en conjunto determinan la estabilidad de los vinos, para que no existan problemas en su elaboración. De este concepto se desprende otro igual de importante, la calidad fisicoquímica, los expertos en el tema comparten que está dada por parámetros que se emplean normalmente en el seguimiento químico del vino y cómo evolucionan en cada etapa de elaboración. Azúcares, alcohol, pH, taninos, acidez total, acidez volátil, nitrógenos, cloruros, por mencionar algunos, son parámetros que pueden ser determinados por diferentes reacciones químicas por metodologías tradicionales o complejas, dependiendo el objetivo de su análisis. Cada analítico está relacionado con parámetros sensoriales, es decir, la presencia de ciertos compuestos químicos producen diferentes colores, aromas y sabores en el vino (Galeano et al, 2021).

2.1.2 Composición del fruto y aportaciones en el vino

La uva se divide en dos tipos principales tejidos, la pulpa y el hollejo, estos contribuyen de manera diferencial a la composición final del mosto y del vino. La pulpa aporta el agua que constituye entre un 80-90 % del volumen del vino y componentes mayoritarios del metabolismo primario como son los azúcares (glucosa y fructosa) y los ácidos orgánicos, fundamentalmente los ácidos málico y tartárico. Durante la fase de maduración, el fruto se convierte en un reservorio de fotoasimilados, un ejemplo es la sacarosa que se encuentre en las hojas y es transformada en hexosas como glucosa y fructosa, localizadas en las vacuolas de las células.(Bejerano y Zapater, 2013).

El hollejo, el cual se define como la capa exterior que recubre la pulpa (piel de la uva), contribuye con un gran número de compuestos del metabolismo secundario que en su conjunto aportan al vino características de gran importancia. Entre ellos los compuestos fenólicos solubles que contribuyen al color y al sabor del vino, y los compuestos aromáticos que contribuyen al sabor y al aroma (Lund y Bohlmann, 2006). Los taninos se sintetizan durante las etapas tempranas del desarrollo de la baya y de la semilla. Entre los compuestos fenólicos no flavonoides se encuentran los estilbenos como el resveratrol, conocido por su elevado poder antioxidante, y diversos compuestos fenólicos volátiles que confieren aromas al vino. Por su parte, el hollejo y también la pulpa contribuyen al aroma del vino que viene determinado por cientos de metabolitos secundarios presentes en la baya en concentraciones variables, pero también por metabolitos y compuestos derivados de los procesos de extracción y tratamiento del mosto, de los procesos de fermentación o de los procesos de crianza en barrica. Los metabolitos aromáticos volátiles o conjugados derivados de la uva, son los que aportan las características varietales del vino. Entre ellos, una de las familias más importantes es la de los terpenos con compuestos como linalool, terpineol o geraniol que confieren aromas frutales y en especial el conocido aroma moscatel (Dennis et al., 2012).

2.1.2.1. Acidez total (AT)

La presencia de ácidos orgánicos en un vino es importante para la estabilidad, color y vida media del producto. La acidez total está dada por la suma de la acidez fija y acidez volátil. Los ácidos más comunes que constituyen la acidez fija son ácido tartárico, málico y cítrico, mientras que el mayor representante de la acidez volátil es el ácido acético. En la uva, fruto precursor del vino, el ácido más característico es el ácido tartárico, por que al realizar la medición correspondiente se emplean gramos de ácido tartárico por cada litro de jugo de uva (mosto) o vino. Para su determinación se emplea el fundamento de titulación ácido-base, la cual se puede realizar de manera tradicional utilizando un indicador de color o potenciométricamente con equipos específicos (Zoecklein, Fugelsang, Gump, & Nury, 2001).

2.1.2.2. Acidez volátil (AV)

La acidez volátil está determinada por la combinación de ácidos orgánicos volátiles, en otras palabras, ácidos que se evaporan con mayor facilidad. El principal compuesto volátil que se encuentra en el vino es el ácido acético, además de que es el que genera mayores problemas a nivel organoléptico, ya que aporta aromas asociados al vinagre, que en concentraciones elevadas genera que los vinos se clasifiquen como de baja calidad y con una vinificación inadecuada. Todos los vinos poseen ácido acético, ya que es un subproducto de la fermentación alcohólica, pero en concentraciones menores a 0.8 g/L no son detectadas por el humano mediante análisis sensorial, este número significa que hay 0.8 gramos de ácido acético cada litro de vino. Según la legislación de cada país, cuando el valor de acidez volátil pasa de 2 g/L, el producto no puede ser vendido como vino, y debe de ser empleado en elaboración de otros productos. Su determinación está dada principalmente por una destilación fraccionada, donde se obtiene un destilado que posteriormente se valora mediante una titulación ácido-base empleando indicador de color (Espíritu y Quiroz, 2019).

2.1.2.3. Ácido málico (AM)

El ácido málico es un compuesto orgánico presente en la uva, durante la maduración del fruto este valor va disminuyendo. La velocidad de su degradación está relacionada con la temperatura, en regiones con altas temperaturas durante la maduración, las uvas tienen menos concentración de ácido málico. Cuando termina la fermentación alcohólica, y todavía no se ha agregado ningún inhibidor microbiano, las bacterias lácticas iniciarán con la fermentación maloláctica, responsable de la degradación de ácido málico en láctico. Esta segunda fermentación baja la acidez total del vino, y según análisis sensoriales realizados vuelve el vino más estable y suave. Sin embargo, estas últimas características no son comunes de encontrar en los vinos blancos, por lo que no es una etapa de su proceso de vinificación. Existen múltiples técnicas de análisis para determinar la concentración de ácido málico, pero las más empleadas son las relacionadas con análisis enzimáticos (Martinez, 2005).

2.1.2.4. Anhídrido sulfuroso libre y molecular (SO₂L y SO₂m)

Es un compuesto químico, combinación de azufre y oxígeno. Empleado en la elaboración del vino como bactericida y antioxidante, además de que posee una relación negativa con el color, ya que en concentraciones elevadas puede tener actividad extractora de color. El anhídrido sulfuroso se encuentra en el vino en dos formas, la primera es como anhídrido sulfuroso libre, el encargo de realizar las actividades antes mencionado normalmente se encuentra en el vino en concentraciones no mayores a 50 partes por millón (ppm), el segundo es el combinado inactivo y como su nombre indica no ejerce ninguna influencia en el vino, la suma de ambos da el anhídrido sulfuroso total que no excede las 200 ppm. La metodología de análisis tradicional es el método de Ripper el cual se basa en una titulación con yodo. Otro concepto es el anhídrido sulfuroso molecular, este es un concepto teórico que explica la eficiencia de la actividad antiséptica, es una relación principalmente entre el anhídrido sulfuroso libre y el pH, pero también se incluyen variables como la temperatura ambiental y el alcohol producido, entre menor sea el valor, menos protección ante contaminaciones se posee (Martinez, 2005).

2.1.2.5. Azúcar residual (AZ)

El azúcar es un sustrato, y en términos generales del vino, es el principal precursor de la fermentación alcohólica. Sin embargo una alta concentración de azúcar puede generar una inhibición en el crecimiento de microorganismos, las levaduras empleadas en la fermentación alcohólica en su mayoría son *Saccharomyces cerevisiae* las cuales tienen una alta tolerancia a la presencia de azúcar y etanol. El azúcar residual se describe como la concentración de azúcares resultante después de la fermentación alcohólica, si se desea tener una concentración alta en azúcares residuales, se puede detener la fermentación prematuramente, con cambios de temperatura o adición de otros compuestos. La metodología más empleada para el análisis de azúcares residuales es Lane-Eynon basada en una reacción de óxido-reducción (Saranraj, Sivasakthivelan y Naveen 2017).

2.1.2.6. Grado alcohólico (Gr.Al%)

El grado alcohólico volumétrico es una medición del contenido de etanol en el vino comúnmente analizado mediante aerometría a una temperatura de 20°C. El alcohol etílico es un producto de la fermentación alcohólica que llevan a cabo las levaduras en presencia de compuestos orgánicos como el azúcar. Este valor es indispensable en el seguimiento de la vinificación, ya que es un parámetro legal que debe de ser colocado en la etiqueta del producto. Es el componente más abundante en el vino después del agua y es de los componentes más característicos e identificados en un análisis sensorial. La mayoría de los vinos contienen entre un 11.5 y 13.5 grados de alcohol, pero según el clima y la ley de producción de cada país este valor puede cambiar (Nogueira y Nascimento, 1999).

2.1.2.7 Grados Brix (°Brix)

La refractometría es una técnica analítica que permite determinar de forma indirecta la concentración de azúcar de un mosto a través de la medida de su índice de refracción, ya que éste aumenta proporcionalmente a medida que lo hace la densidad del mosto (OIV, 2019). En la actualidad, es habitual que los refractómetros proporcionen el índice de refracción, expresado en grados Brix, relacionado con el porcentaje de sólidos solubles contenidos en la solución acuosa, en este caso, el mosto. Dentro de los sólidos solubles se encuentran todos los sólidos disueltos, incluyendo azúcares, sales, proteínas, ácidos, etc. El refractómetro se calibra respecto a la cantidad de azúcar, de modo que los grados Brix de un mosto representan los gramos de azúcar (expresados en gramos de sacarosa) que contienen 100 mL de mosto. El Método hidrométrico es utilizado en mostos donde se tienen grandes cantidades de ácido carbónico. Se basa en la medición de la densidad aparente, dada por la concentración de sólidos disueltos y en suspensión, empleando para el efecto un hidrómetro con escala en grados Brix y calibrado a 293K (20°C) (NMX-F-274-1987, 1987).

2.1.2.8. Nitrógeno fácilmente asimilable (NFA)

Se define el nitrógeno fácilmente asimilable como la combinación de compuestos nitrogenados alfa-aminos presentes en la uva que están relacionados con el éxito de las fermentaciones alcohólicas. El nitrógeno que se encuentra disponible para la fermentación varía según la región y un valor menor a las 150 ppm pueden influir en una fermentación de baja eficiencia. Las metodologías de análisis recomendadas para la obtención de esta variable es por espectrofotometría, pero debido al alto costo de la técnica, se emplean métodos más sencillos, como el de Sorensen donde se emplea formaldehído para bloquear grupos aminos y poder realizar la medición de los compuestos por separado (Cortéz et al, 2009).

2.1.2.9. Potencial de hidrógeno (pH)

El pH es una medida de concentración de iones de hidrógeno (H⁺) cuando se está en presencia de soluciones ácidas y la concentración de iones hidroxilo (OH⁻) en medios alcalinos o básicos. La medición de esta variable es mediante un potenciómetro, donde se mide la presencia de los iones ya mencionados y se relacionan con la escala estandarizada. La medición de pH es en una escala 1 a 14, entre más alto sea el valor más concentración de iones hidroxilo hay en la solución. En los vinos la medición de pH se encuentra entre 3 y 4, es decir es un medio con concentración de acidez mayor. La mayoría de microorganismos (bacterias, levaduras y hongos) crecen en pH cercanos a la neutralidad (valor de 7), pero en algunos casos pueden crecer desde pH de 2.5, por lo tanto, esta variable afecta la estabilidad del vino (Saranraj et al., 2017).

2.2 Posición del vino Mexicano

Enfocándonos en el vino de México, entre los estados productores destacan Baja California, Coahuila, Querétaro, Guanajuato, Aguascalientes y Sonora. El mayor productor es Baja California con un 90% de la producción anual. Al hacer un acercamiento en el estado líder, la región comprendida entre el Valle de Guadalupe a Valle de San Vicente en el Municipio de Ensenada representa alrededor del 85% de Bodegas, en este espacio geográfico se encuentran sectores como San Antonio de Las Minas, Uruapan y Santo Tomás (Andrade y Flores, 2014). En este estado se encuentra la segunda vinícola más antigua del país, Bodegas de Santo Tomás con 135 años de antigüedad tiene una calidad internacional reconocida. La empresa ha sido pionera en adoptar la tecnología de vinificación más vanguardista en México a través de su historia. Se tienen reconocimientos desde nivel regional hasta internacional, posicionando la marca ante la inminente globalización del mercado con 25 etiquetas divididas en 6 líneas diferentes de vino, logrando un crecimiento de dos dígitos en el mercado nacional en los últimos 5 años (Méndez Medina, 2016).

2.2.1 Vitivinicultura en Baja California, México

Cada vez es más frecuente asociar a Baja California con la producción de vinos mexicanos de buena calidad, algunos de ellos con reconocimiento internacional. A partir del año 2000 la fabricación de vinos y el cultivo de la vid han cobrado mayor importancia entre las actividades económicas de esta entidad, cuyos viñedos comprenden 85% del total de vino producido en México. En las últimas décadas se multiplicaron las vinícolas, en su mayoría empresas familiares, aunado a que el gobierno federal ha incentivado la viticultura mediante la inversión destinada a mejorar los procesos posteriores a la cosecha de uva, el establecimiento de huertos y el uso de riego tecnificado (Meléndez, Íñiguez, Puras, Ortiz y Sarabia, 2010). La creciente relevancia de la vitivinicultura en Baja California contrasta con el conocimiento sobre la historia de esta actividad en la entidad. La literatura acerca del tema está dominada por estudios que explican el tránsito de esta zona, de ser un área vitivinícola a convertirse en un corredor turístico, que atraviesa los municipios de Tecate, Tijuana y Ensenada. También se han realizado trabajos centrados en recuperar testimonios de quienes han participado en esta actividad, y algunos estudiosos han descrito el desenvolvimiento de la vitivinicultura en zonas específicas de Baja California (Ruiz Ríos, 2011).

2.2.1.1. Valle de Santo Tomás

De los 7 valles principales de Ensenada, enfocados en la producción vitivinícola, el de mayor antigüedad es el Valle de Santo Tomás, ubicado a 35 kilómetros al sur de la ciudad de Ensenada. Múltiples unidades empresariales han tenido base en este valle, algunas por tiempos menores a 5 años y otras llevan más de un siglo de producción, un ejemplo de ello es Bodegas de Santo Tomás. Fue la primera bodega vinícola en establecerse en el territorio en 1888, y hacia 1941 el valle de Santo Tomás se mantenía entre los principales productores de vid para la fabricación de vino. En correspondencia con la cantidad de uva cosechada, Bodegas de Santo Tomás era la mayor productora de vino, aunque la producción local no era suficiente y tenía que importar una gran cantidad de uva. De acuerdo con la información disponible sobre Bodegas de Santo Tomás, el vino producido en el Territorio Norte que se comercializaba en el país era trasladado en barco desde Ensenada a otros puertos del

Pacífico (Guaymas, Mazatlán y Manzanillo) para distribuirse en el Distrito Federal y otras entidades cercanas a la capital (Méndez, 2016; Ruiz, 2011).

2.3 Inteligencia artificial

Al hablar de inteligencia artificial (IA) entramos en definiciones muy variadas y que se adaptan al área de estudio, se puede considerar como una disciplina de la informática que se dedica a realizar prototipos, ya sea de máquinas y sistemas que puedan desempeñar actividades realizadas por la inteligencia humana. Al 2023 la IA se encuentra más presente día a día, pronosticando que en unos años se convierta en la mayor oportunidad de negocio (Tatú et al., 2023). Si bien, es una nueva herramienta que vino a revolucionar la gestión de los macrodatos (Big Data), esto puede dar lugar a conflictos de privacidad, responsabilidad o intereses. Es por ello que cada gobierno debe abordar el tema estableciendo los marcos legales necesarios para favorecer su desarrollo. Las técnicas de IA tienen diversas aplicaciones, una de ellas ha sido en las problemáticas de la industria alimentaria, en donde se destacan metodologías con máquinas de soporte vectorial y árboles, tanto de clasificación como de regresión, todos estos aplicados en la determinación de la calidad de un producto. Es importante mencionar que en los últimos 5 años se han realizado acercamientos a la industria del vino, específicamente en Sudamérica donde se han iniciado con minería de datos para lograr mejores extracciones colorantes, selección específica de características y detección de problemas en el proceso de vinificación (Galeano et al., 2021; Schober et al., 2022).

2.3.1. Python

Python es un lenguaje de programación ampliamente utilizado para el desarrollo de software. Es una herramienta de acceso libre que basa la construcción de programas, que se considera tiene un lenguaje multi-paradigmas y de alto nivel. Python cuenta con una gran fortaleza, su librería estándar, con la que se cubren gran cantidad de temas, desde estructura de datos hasta servicios de sistemas operativos (Challenger, 2014). Spyder es un entorno de desarrollo integrado multiplataforma dirigido a la implementación de programas, es compatible con los principales sistemas operativos. Este entorno es ideal para la escritura y depuración

de código Python, y para la generación de prototipos de análisis exploratorio y aprendizaje automático (Cubero Ledezma, 2020).

2.3.2. Modelos de predicción

El aprovechamiento de las capacidades de la inteligencia artificial ayuda a generar algoritmos avanzados de predicción a partir de grandes y complejos volúmenes de datos. Los modelos de predicción representan un enfoque que asiste en la comprensión y optimización de los procesos de producción. Los modelos ajustan iterativamente sus parámetros, de esta manera se disminuyen significativamente los errores en la predicción aumentando así su precisión. Para su evaluación se emplean métricas adaptadas a tareas de regresión o clasificación, y existen técnicas de ajuste que optimizan el modelo, generando aumentos en la precisión y rendimiento. En la industria del vino, se emplean modelos de predicción en proyectos de análisis de mercado y ventas, pero, también pueden emplearse para analizar la calidad del vino, y de esta manera los productores puedan mantener una consistencia en sus productos, mejorando la experiencia del consumidor y marcar una diferencia en el mercado competitivo (Geetha et al, 2024; Aiello y Tosi 2024).

2.3.2.1. Árboles de decisión

Dentro de los algoritmos de aprendizaje automático están los árboles de decisión que son empleados en diferentes temas, son parte fundamental del concepto de aprendizaje automático, y por la estructura de representación tienen similitudes a un árbol, los nodos principales denotan las condiciones o características a cumplir, mientras que las ramas son las reglas de decisión. Los árboles de decisión se forman de manera iterativa y con segmentaciones. En diferentes estudios se observa el uso de árboles de decisión o Random Forests, estos últimos se caracterizan por combinar varios árboles de decisión, de esta manera se reduce el sobreajuste que se genera al momento de aumentar la precisión (Jain, 2023).

2.3.2.2. Redes neuronales

Dentro de los modelos de aprendizaje automático, las redes neuronales destacan, diseñadas para tomar decisiones de manera comparable al humano. Los nodos son unidades de procesamiento que transmiten información entre sí. Son de análisis profundo por lo que se recomienda para procesos con relaciones complejas entre los datos de estudio, ya que es un modelo que es capaz de aprender y modelar relaciones no lineales entre las variables, de entrada y de salida. Debido a esto, han sido empleadas en diferentes actividades económicas, como en la agricultura para la predicción de rendimientos en cultivos. Su aplicación puede ser en cualquier proceso de manufactura, por ejemplo se puede emplear para predecir de calidad, rendimientos y detección de fallas (Mahima et al, 2020).

Existen múltiples ventajas al utilizar este tipo de modelos, ya que es adaptativo, crea una propia organización de los datos a estudiar, así como la tolerancia a fallos parciales, mismos que pueden ser detectados al llevar una operación en tiempo real. Al poder tener diferentes tipos de conexiones se pueden encontrar variantes de las redes neuronales como las recurrentes que pueden ejecutar conexiones de retroalimentación y las no recurrentes que trabajan en un solo sentido. También se pueden encontrar redes neuronales multicapas, donde se trabaja con capas ocultas (Carvajal, 2023).

3 METODOLOGÍA

Para el logro de los objetivos, se diseñó y desarrolló una metodología de trabajo con las siguientes fases y actividades, mismas que son ampliadas en el presente capítulo.

Fase 1. Determinar las variables independientes y la variable dependiente del análisis.

1.1. Realizar revisión sistemática del tema

1.2. Selección de 10 variables influyentes en la calidad del vino por enólogos de la región.

1.3. Determinación de los años de estudio.

Fase 2. Generación y análisis de las variables fisicoquímicas

2.1. Obtención de información a través de bases de datos anuales.

2.2. Desarrollo de metodologías de análisis químicos para la obtención de los datos del año 2024.

Fase 3. Análisis de los datos obtenidos mediante técnicas de inteligencia artificial.

3.1. Generación de árboles de decisión.

3.2. Desarrollo de modelo predictivo mediante redes neuronales.

Fase 4. Mejorar la precisión de los modelos obtenidos.

4.1. Aumentar la precisión en ambos modelos de predicción.

4.2. Aplicar los cambios en un nuevo análisis de la base de datos unificada.

3.1 Definición de las variables independientes y dependientes

3.1.1. Realizar revisión sistemática del tema

Una revisión sistemática es una técnica de búsqueda de información secundaria con metodología rigurosa, de esta manera garantizamos la confiabilidad de los resultados, y que puedan ser aplicados de una manera objetiva y clara. La revisión fue realizada conforme a la metodología PRISMA (Preferred reporting items for systematic reviews and meta-analyses) para documentar de manera adecuada el proceso, se emplearon bases de datos confiables y la extracción de la información se hizo en una tabla con los datos más relevantes, tanto como fuente de información como para clasificar correctamente cada trabajo colectado (Gonzalez y Rodriguez, 2018). La búsqueda se extendió a cualquier región del mundo y se tomaron en cuenta vinos que fueran realizados con una base del fruto de *Vitis vinífera* y los producidos con frutos de otras plantas. A partir de esta búsqueda se establecieron los siguientes criterios de inclusión.

- Estudios publicados en español o inglés con una antigüedad menor a 30 años.
- Metodologías de análisis fisicoquímico y/o sensorial de mínimo 3 parámetros.
- Métricas de evaluación presentes en los resultados y discusiones.

Se emplearon bases de datos académicas reconocidas como Scopus, ScienceDirect y Google Scholar, así mismo, se empleó la biblioteca Digital de UABC, y editoriales científicas con gran reconocimiento como Elsevier, IEEE y Wiley. Los resultados de las búsquedas se filtraron con los criterios descritos anteriormente, en una primera etapa se eligieron artículos por el título y el resumen, posteriormente se eliminaron los duplicados, seleccionando artículos con acceso disponible, ya sea por acceso libre gratuito o por parte de la institución afiliada (Universidad autónoma de Baja California). En una siguiente etapa se descartaron las revisiones y los trabajos sin metodología clara.

La información que se extrajo de cada artículo se colocó en una base de datos, donde las columnas que se llenaron fueron:

- Año de publicación
- Autores
- País donde se realizó la metodología
- Si el vino tenía base en *Vitis vinífera* especificar si era vino blanco, rosado, tinto o todos incluidos.

- Un resumen de la metodología aplicada.
- Parámetros de evaluación de la metodología.
- Análisis fisicoquímicos (pH, acidez total o fija, acidez volátil, azúcares residuales, grados Brix o densidad, grado alcohólico, anhídridos sulfurosos y antocianinas o compuestos fenólicos).
- Análisis sensoriales (Aroma, sabor y color).
- Cita en formato Apa séptima edición.

Al finalizar la base de datos se agruparon los hallazgos por tipos de parámetros evaluados y su relación con la clasificación o percepción de la calidad del vino de cada autor.

3.1.2. Selección de variables independientes

Para iniciar con la generación del modelo predictivo de la calidad del vino mediante variables fisicoquímicas, se definieron cuáles son las 10 variables que se consideran con mayor influencia en la calidad del vino, como se muestran en la tabla 2, esto fue proporcionado por el área de enología de la empresa Bodegas de Santo Tomás. Se reunieron 10 enólogos del municipio de Ensenada, México, quienes de manera anónima contestaron un cuestionario que constaba de dos etapas. La primera etapa era una cata a ciegas, como se muestra en la tabla 1, se presentaron 20 vinos que correspondían a 3 líneas de vinos de la empresa ya mencionada, 5 vinos de cada calidad (Premium, clásica, básica y baja), se les entregó una hoja donde tenían que elegir la calidad para cada vino. En una segunda etapa, ponderaron del 1 al 15 la importancia de cada variable fisicoquímica según su criterio, siendo el 1 la más importante y el número 15 la de menor relevancia, como se muestra en la tabla 2. La sesión de la degustación fue programada para 4 horas, el tiempo promedio que se emplea por día, en los concursos internacionales de vino (Fernández-Cuesta, 2014).

Tabla 1. Primera parte del ejercicio de enología para determinar la calidad del vino a partir de análisis fisicoquímicos y sensoriales.

1. Deguste los vinos que se enumeran a continuación, considerando solo valores organolépticos (aroma, sabor y color) ¿Cómo clasificaría el vino? Si considera que el vino es por debajo de la clasificación "Calidad básica" deje sin rellenar las casillas

Muestra	Calidad premium	Calidad clásica	Calidad básica
Vino 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 13	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 14	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 15	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 16	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 17	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 18	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 19	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vino 20	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Tabla 2. Segunda parte del ejercicio de enología para determinar la calidad del vino a partir de análisis fisicoquímicos y sensoriales.

2. Enumere los siguientes análisis en orden de relevancia a su criterio, siendo el 1 el más importante y el 15 el que considere con menor importancia

Análisis fisicoquímico	Número
pH	
Acidez Volatil	
Acidez total	
Anhídrido total	
Anhídrido molecular	
Anhídrido Libre	
Azúcares residuales	
°Brix	
Nitrógeno fácilmente A.	
Antocianinas	
Polifenoles	
Cloruros	
Grado alcohólico	
Intensidad colorimétrica	
Ácido Máfico	

La información requerida para cumplir con los objetivos generados, se proporcionó a lo largo de la colecta de uva 2024, que comprendió de agosto a diciembre del mismo año y fueron de diferentes variedades de *Vitis vinífera*. Las variables fisicoquímicas que se estudiaron fueron: acidez total, acidez volátil, ácido málico, anhídrido sulfuroso libre, anhídrido sulfuroso molecular, azúcar residual, grados Brix, grado alcohólico, nitrógeno fácilmente asimilable y pH. La variable que se consideró como dependiente es “Calidad”, tomando en cuenta que son 4 categorías: Calidad premium, Calidad varietal, Calidad básica y Baja calidad, esta última se tomó al identificar vinos que no entraban en las primeras 3 categorías. Las variables independientes fueron analizadas en conjunto con los registros de trazabilidad proporcionados desde el año 2020, por el área de Laboratorio de Bodegas de Santo Tomás.

De igual manera, con el área de enología se definieron las variedades de uva que son más representativas, tomando en cuenta dos criterios. El primero, es a la superficie plantada, es decir se seleccionaron las variedades de *Vitis vinífera* que tienen una mayor producción en los dos valles. El siguiente criterio se enfoca en los análisis fisicoquímicos que se proporcionaron, ya que son las variedades que son más distintivas en la empresa que proporcionó la información.

Para realizar un proyecto entre una institución pública y una privada, este se debe de apegar a los lineamientos de ambas partes. En este caso, para que UABC pudiera trabajar con Bodegas de Santo Tomás se firmó un convenio de colaboración donde se estableció que se podían realizar intercambios de capital humano, inteligencia y tecnología. El área de laboratorio de Bodegas de Santo Tomás posee diferentes bases de datos, donde marcan una trazabilidad de producto, desde la recepción de la uva hasta el embotellado del vino, estas mismas bases se analizaron y adecuaron para realizar de manera adecuada el análisis planteado en los siguientes puntos.

3.2. Medición del problema

Las variables fueron obtenidas de dos maneras. La primera, de bases de datos proporcionadas por la empresa de análisis que van desde el año 2020 a 2023 y la segunda, por metodologías de análisis químico, enzimático y fisicoquímico, con metodologías que se apegan a las regulaciones nacionales e internacionales, para los datos del año 2024, mismas que fueron empleadas en los años antes mencionados. Las variables fisicoquímicas seleccionadas para el estudio fueron dadas por el ejercicio de enología que se menciona a continuación.

3.2.1. Generación de datos: Bases de datos

El área de laboratorio de Bodegas de Santo Tomás tiene 4 archivos de excel (programa perteneciente a la paquetería de Office), de cada año de análisis, como parte de la trazabilidad que llevan por cada variedad del fruto de *Vitis vinífera* recibido. Se solicitaron copias de los 16 archivos de interés, es decir, del año 2020 a 2023, para adecuarlas a nuestro estudio. Se diseñó una tabla dónde colocar toda la información de manera unificada, en la tabla 3 se muestra el diseño general de las columnas donde se colocaron los datos.

Tabla 3. Diseño de tabla empleada en la unificación de datos proporcionados por año.

Analítica del vino															
Variedad	Rancho	Año	°Brix	pH	AT	SO2L	SO2m	NFA	AZ	Gr.Al%	AV	AM	Calidad Premium	Calidad Varietal	Calidad Básica

Los datos fueron cargados por año, es decir, iniciando con los análisis del 2020, asimismo, se colocaron primero las variedades de uva blanca y finalmente las variedades tintas.

Las abreviaciones de los análisis y unidades empleadas se pueden verificar en la tabla 4, a partir de esta sección las variables fisicoquímicas se colocarán en el texto con la abreviatura asignada.

Tabla 4. Análisis realizados junto con su unidad de medida

Análisis	Abreviación	Medición	Tolerancia (+/-)
Ácido Málico	AM	Gramos por litro (g/L)	0.02
Acidez total	AT	Gramos por litro (g/L)	1.00
Acidez volátil	AV	Gramos por litro (g/L)	0.02
Azúcares residuales	AZ	Gramos por litro (g/L)	0.30
Grado alcohólico	Gr.Al%	%	0.30
Grados Brix	°Bx	Grados (°)	0.20
Nitrógeno Fácilmente Asimilable	NFA	Lectura de pH	0.05
Unidades de pH	pH	u. pH	0.08
Anhídrido sulfuroso libre	SO ₂ L	Miligramos por litro (mg/L)	1.50
Anhídrido sulfuroso molecular	SO ₂ m	Miligramos por litro (mg/L)	No aplica

Las variables de calidad, en la base de datos se cargaron con “SI” y “NO”, siendo así, un vino de “Calidad Premium” tendría “SI” en las 3 categorías, mientras que los vinos que obtuvieron 3 “No” fueron etiquetados como vinos de “Baja Calidad”.

3.2.2. Generación de datos: Metodologías de análisis químicos, enzimáticos y fisicoquímicos

Una vez que la uva pasó por la primera etapa de vinificación, prensado en el caso de vinos blancos y despalillado en vinos tintos, lo producido pasó de llamarse uva a mosto, cuando se encontraron en un depósito de acero inoxidable previo a la fermentación alcohólica se analizaron los °Brix, AT y NFA. Cuando la fermentación alcohólica finalizó, los análisis que prosiguieron fueron pH, AZ, AV, SO₂L, SO₂m y Gr.Al%, mientras que cuando finalizó la fermentación maloláctica los analíticos fueron pH, SO₂L, SO₂m y AM. Siendo pH, la única variable que puede ser analizada

en todas las etapas, pero el dato final se considera después de la fermentación maloláctica.

3.2.2.1. Acidez total (AT)

Los materiales que se emplearon fueron:

- Imán magnético
- Probeta de 50 mL
- Matraz Kitasato de 500 mL
- Vasos de reacción de 100 mL
- Titrador automático TitraLab 1000 Series.
- Buffer 4.01
- Buffer 7.01
- Buffer 10.01
- NaOH 0.25 N

Para realizar la prueba, fue necesario desgasificar la muestra. Para ello, se utilizó una bomba de vacío de marca Rocket y un matraz Kitasato de 500 mL, se vació la muestra y se agitó vigorosamente de manera manual el mosto. Esta operación se repitió tres veces. Luego, se llevó la muestra desgasificada a un vaso de reacción (30 mL de muestra) junto con un imán magnético.

El equipo (Titralab) se encendió y, en su panel de control, se seleccionó la opción de “calibración”. Se siguieron los pasos correspondientes para calibrar el electrodo con las soluciones de pH 4.01, 7.01 y 10.01, validando la calibración (>97%). A continuación, se colocó la muestra y el electrodo correspondiente (PHC725) en su interior. En el panel de control, se seleccionó la opción de “pH/AT”, luego “muestra” e iniciar. Se esperó a que el equipo proporcionara el resultado y se realizó el mismo procedimiento por triplicado, lavando el electrodo con agua destilada cada vez que se utilizó con una nueva muestra. Finalmente, se registró el resultado.

3.2.2.2. Grados Brix (°Bx)

Los materiales que se emplearon fueron:

- Probeta graduada de 250 mL
- Termómetro calibrado de 0 a 40 °C
- Hidrómetros de -0.5 a 5 °Brix y de 0 a 35°Brix
- Jarra de plástico de 1 litro
- Refractómetro digital

La medición de grados brix se tomó con refractómetro digital y por duplicado. Primero, se verificó el equipo con una gota de agua destilada, donde la medición siempre debe ser cero, de no ser así, se debe de presionar el botón “Zero”, después de mezclar el mosto (jugo de uva), se tomó aproximadamente 1 mL de muestra con una pipeta pasteur de plástico y se colocó en el lente del refractómetro previamente lavado con agua destilada. Después de esto, se realizó la medición y se registraron los resultados.

3.2.2.3. Nitrógeno fácilmente asimilable (NFA)

Los materiales que se emplearon fueron:

- Pipeta serológica de 2 mL
- Pipeta serológica de 10 mL
- Vaso de precipitado de 250 mL
- Imán magnético
- Potenciómetro Orion Star
- Agitador magnético
- NaOH 0.1 N
- NaOH 0.5 N
- Formaldehído (37%) (Ajustado a pH 8)
- Peróxido de hidrógeno 3%
- Agua destilada

Se añadieron 4 mL de la muestra de mosto a un vaso de precipitado con 100 ml de agua destilada. Luego, se agregó 1 mL de H₂O₂ al 3% y tituló hasta un pH de 8.0 con solución de NaOH 0.5 N.

Posteriormente, se adicionaron 2 mL de formaldehído previamente neutralizado a pH 8.0. Una vez realizado, fue necesario esperar 1 minuto en agitación y titular a pH 8.0 con solución NaOH 0.1 N.

Se anotó el gasto de NaOH 0.1 N en mL. Después, se realizó el cálculo con la siguiente fórmula:

$$NFA=(V1) \times (1400) \times (V2)$$

Donde:

- V1: Volumen en mL de NaOH 0.1 N gastados en la titulación
- V2: Volumen en mL de la muestra agregada

3.2.2.4. Azúcar residual (AZ)

Los materiales que se emplearon fueron:

- Soporte universal
- Bureta de 50 ml
- Matraz Erlenmeyer de 250 mL
- Vaso de precipitado 1 Litro
- Pipeta volumétrica de 25 mL
- Pipeta volumétrica de 10 mL
- Pinzas para matraz
- Papel filtro Whatman No.1
- Embudo de plástico
- Placa de calentamiento
- Dextrosa al 1%
- Fehling A
- Fehling B
- Azul de metileno
- Carbón activado
- Agua destilada

Los vinos tintos y rosados se tuvieron que decolorar antes de analizar la concentración de azúcares residuales. A 100 mL de vino se agregaron 5 gramos de

carbón activado. Se agitó y dejó reposar 15 minutos. Pasado este tiempo se filtró utilizando un embudo de plástico y papel Whatman No.1.

En un vaso de precipitado de 1 litro, se agregaron 500 mL de Fehling A y 500 mL Fehling B, simultáneamente, la mezcla se almacenó en un frasco de plástico rotulado con el nombre Soxhlet.

En un matraz Erlenmeyer de 125 mL se agregaron 25 mL de Soxhlet y 10 mL de la muestra. Se colocó el matraz sobre la placa caliente y al primer hervor, se agregaron 3 gotas de azul de metileno. Se tituló con dextrosa al 1%, agitando constantemente el matraz con ayuda de las pinzas y manteniendo el matraz siempre en contacto con la placa. El vire de azul a cobrizo indicó el final de la titulación. Como testigo, se utilizó un matraz sin muestra, es decir, se titularon solo los 25 mL de Soxhlet.

Para obtener la concentración de azúcares residuales en g/L se hizo de la siguiente manera:

$$[\text{Azúcares reductores}] = V1 - V2$$

Donde:

- V1= Volumen en mL de dextrosa gastados en la titulación del testigo
- V2= Volumen en mL de dextrosa gastados en la titulación de la muestra

3.2.2.5. Acidez volátil (AV)

Los materiales que se emplearon son:

- Pipeta pasteur de 1 mL
- Celdas de análisis de 1 mL
- Analizador automático Y15
- Kit enzimático Ácido acético líquido A + B
- Agua destilada

Primero se encendió el analizador automático Y15 de la marca Biosystems, se verificó su calibración y estado general, para esto se inició con el “Warm Up”

siguiendo las indicaciones del monitor. Luego, se seleccionó el método para la determinación de ácido acético (acidez volátil) en el panel de control y se ajustaron los parámetros según las especificaciones del análisis. Se colocaron las muestras en las celdas de análisis y los reactivos en los compartimentos correspondientes del analizador, asegurando una correcta disposición. Se inició el análisis desde el panel de control, y el equipo realizó automáticamente las mediciones. Una vez completado el análisis, se obtuvieron los resultados en el panel de control. Finalmente, se realizó la limpieza de los compartimentos del analizador y el mantenimiento de cierre que solicita el equipo para su correcto apagado.

3.2.2.6. Grado alcohólico (Gr.Al%)

Los materiales que se emplearon fueron:

- Matraz bola con fondo plano y cuello
- Sistema de destilación (serpentín y manguera)
- Perlas de ebullición
- Matraz kohlrusch 200 ml
- Tapón de hule no. 7
- Probeta de vidrio 250 ml sin graduación
- Densímetro de grado alcohólico del rango necesario
- Termómetro digital
- Tablas de equivalencia de grado alcohólico
- Placa de calentamiento
- Refrigerador
- Agua destilada

Se colocó la muestra de vino dentro del matraz kohlrusch y se aforó, siendo ésta la marca que indica el volumen calibrado a 20°C. Posteriormente se agregó la muestra dentro del matraz de destilación colocando previamente perlas de ebullición, y se posicionó sobre la placa de calentamiento encendida a 500°C, conectando el matraz al sistema de destilación. Una vez alcanzado el nivel de aforo con destilado, se retiró el matraz kohlrusch, desconectando el matraz de destilación y cerrando la llave de agua. El matraz con destilado se aforó a 20°C y posteriormente se llevó a una temperatura de 15°C, para lograr esto se colocó en el refrigerador a 4°C por periodos

no mayores a 5 minutos, tomando la temperatura en cada intervalo. Después se vació todo el contenido del matraz kohlrusch a una probeta de vidrio de 250 mL. Por último se introdujo el hidrómetro de grado alcohólico, midiendo la temperatura y determinando el grado alcohólico utilizando las tablas de conversión del fabricante.

3.2.2.7. Anhídrido sulfuroso libre

Los materiales que se emplearon fueron:

- Imán magnético
- Probeta de 50 ml
- Vasos de reacción de 100 mL
- Titrador automático TitraLab 1000 Series.
- Solución de titrante de yodo 0.025 N
- Solución estándar H₂SO₄ 10 N
- Yoduro de potasio
- Agua destilada

Primero, se enciende el equipo Titralab 1000 series y, desde el panel de control, se selecciona la opción "SO₂ libre". Luego, se coloca una muestra de 20 ml en el agitador magnético, se añade el electrodo MTC695, y en el panel de control se elige la opción "muestra" y luego "iniciar". Se espera a que el equipo se prepare y, a continuación, se agrega una pizca de yoduro de potasio sólido. Se continúa con la prueba hasta obtener el resultado, se realiza un duplicado lavando el electrodo con agua destilada entre cada muestra nueva. Se registra el promedio de dos análisis.

3.2.2.8. Anhídrido sulfuroso molecular

Se calcula el SO₂ molecular con los valores de pH y SO₂ libre utilizando la siguiente fórmula:

$$\text{SO}_2 \text{ Molecular (mg/L)} = \frac{100}{10^{(pH-pK_1)} + 1} \times \frac{SO_2 L}{100}$$

Donde pK_1 es:

$$pK_1 = 1.9499 + (T - 20) \times 0.0322 + (A - 10) \times 0.01971$$

Variables a considerar:

- pH: Valor de pH del vino.
- SO_2L : Concentración de dióxido de azufre libre (mg/L).
- T: Temperatura de la muestra en °C.
- A: Grado alcohólico (% v/v).

3.2.2.9. Ácido málico

Los materiales que se emplearon fueron:

- Pipeta pasteur de 1 mL
- Celdas de análisis de 1 mL
- Analizador automático Y15
- Kit enzimático Ácido málico A + B
- Agua destilada

Primero se encendió el analizador automático Y15 de la marca Biosystems, se verificó su calibración y estado general, para esto se inició con el “Warm Up” siguiendo las indicaciones del monitor. Luego, se seleccionó el método para la determinación de ácido málico en el panel de control y se ajustaron los parámetros según las especificaciones del análisis. Se colocaron las muestras en las celdas de análisis y los reactivos en los compartimentos correspondientes del analizador, asegurando una correcta disposición. Se inició el análisis desde el panel de control, y el equipo realizó automáticamente las mediciones. Una vez completado el análisis, se obtuvieron los resultados en el panel de control. Finalmente, se realizó la limpieza de los compartimentos del analizador y el mantenimiento de cierre que solicita el equipo para su correcto apagado.

3.2.2.10. Potencial de hidrógeno (pH)

Los materiales que se emplearon fueron:

- Vasos de reacción de 100 mL
- Probeta de plástico
- Imán magnético
- Titrador automático TitraLab 1000 Series
- Buffer 4.01
- Buffer 7.01
- Buffer 10.01
- Agua destilada

Para realizar la prueba, fue necesario desgasificar la muestra. Para ello, se utilizó una bomba de vacío de marca Rocket y un matraz Kitasato de 500 mL, se vació la muestra y se agitó vigorosamente de manera manual el mosto. Esta operación se repitió tres veces, después, se llevó la muestra desgasificada a un vaso de reacción (30 mL de muestra) junto con un imán magnético.

El equipo (titralab) se encendió, en su panel de control seleccionamos la opción de “calibración” y se siguieron los pasos correspondientes a la calibración del electrodo con las soluciones de pH: 4.01, 7.01 y 10.01. se valida la calibración (>97%); enseguida se colocó la muestra y dentro de ella el electrodo correspondiente (PHC725), seleccionamos en el panel de control la opción de “mantenimiento”, en seguida “medición directa”, esperamos de 30 a 60 segundos para que se estabilizara y obtener un resultado preciso, registramos el resultado obtenido.

3.3. Análisis de datos

Al obtener una base de datos con 16 columnas y más de 350 filas, se optó por emplear técnicas de inteligencia artificial para el análisis de la información.

3.3.1. Correlación de variables.

Para encontrar la correlación entre las variables se empleó Spyder, entorno desarrollado para Python con orientación en la ciencia de datos y el aprendizaje automático. El código inició empleando una biblioteca de Python llamada “Pandas” para cargar la base de datos unificada que se encontraba en formato Excel. Se convirtieron las columnas de las categorías de calidad de “SI y No” a “1 y 0”.

Se validaron todas las columnas de interés, en este caso las 10 columnas de las variables independientes, 3 columnas con la información general del vino (Variedad, añada y rancho de procedencia) y las 3 columnas de la variable dependiente. Posteriormente, se visualizó la correlación mediante un mapa de calor que representó gráficamente la intensidad de correlación entre las variables mediante una escala de colores, identificando la variable con una mayor correlación con la calidad.

3.3.2. Diseño de árbol de decisión

Para la generación de árboles de decisión se dividieron los datos en conjuntos de entrenamiento con un 80% y con un 20% de prueba. Después, se seleccionó un clasificador inicial, y se estableció el criterio de entropía, esto para manejar los datos desbalanceados. La estructura del árbol de decisión se visualizó con “Plot_tree” dónde se interpretó el modelo. Además, se realizó un ajuste de la variable objetivo que en este caso es la calidad, definiendo clases:

- Clase 0: Baja calidad
- Clase 1: Calidad Básica
- Clase 3: Calidad Varietal
- Clase 6: Calidad Premium

El modelo se evaluó con dos pruebas. La primera es la “Precisión”, se calcula mediante una división entre las predicciones verdaderas entre el total de predicciones. La segunda prueba es “Recall”, ésta también es resultado de una división, en este caso de las predicciones verdaderas entre la suma de estas mismas predicciones más los falsos negativos. Por último está la métrica “F1-score” este es un cálculo con promedio de las métricas Precisión y Recall, es una media armónica entre la precisión y la capacidad del modelo para identificar los casos positivos dentro de cada clase.

3.3.3. Diseño de redes neuronales

El código inició empleando una biblioteca de Python llamada “Pandas” para cargar la base de datos unificada que se encontraba en formato Excel. Se convirtieron las columnas de las categorías de calidad de “SI y No” a “1 y 0”.

Se validaron todas las columnas de interés, en este caso las 10 columnas de las variables independientes y las 3 columnas de la variable dependiente. Se establecieron las clasificaciones de clases para calidad:

- Clase 0: Baja calidad
- Clase 1: Calidad Básica
- Clase 3: Calidad Varietal
- Clase 6: Calidad Premium

El criterio empleado para la definición de estas clases se basó en la combinación de las 3 columnas designadas para calidad.

Se dividieron los datos en conjuntos de entrenamiento con un 80% y con un 20% de prueba usando Random Forest como clasificador inicial, balance sugerido en el análisis realizado por Alhassan y Asamoah en 2023. Posteriormente, se entrenó el modelo con los datos de entrenamiento y las predicciones con el conjunto de prueba

La evaluación del modelo fue mediante la precisión, midiendo el porcentaje de predicciones correctas y una matriz de confusión.

3.4 Mejorar la precisión de los modelos obtenidos

3.4.1. Ajuste del modelo de árbol de decisión generado

Al obtener una precisión en el modelo menor al 90% se optó por mejorar el modelo. Se realizó una optimización de hiperparámetros mediante “GridSearchCV”, seleccionando la mejor configuración del árbol de decisión.

Se realizó una validación cruzada de 5 pliegues, esto para evitar que el modelo se sobreajuste, limitando la profundidad del árbol y el número mínimo de muestras por nodo. Se hizo un análisis de importancia de variables, de esta manera se permitió el identificar cuáles variables tienen una mayor influencia en la calidad del vino.

3.4.2. Ajuste del modelo de redes neuronales generado

Al obtener un valor de precisión debajo del 90% se mejoró el modelo, se simplificó usando una sola etiqueta multiclase eliminando los problemas de clasificación. Posteriormente se consideró el desbalance de clases que se presenta en el primer código generado, para esto se empleó la técnica de SMOTE eliminando clases con menos de 3 clases, además de emplear el peso de clases durante el entrenamiento (`class_weight`). El modelo generado en la metodología 3.3.3 es una red neuronal simple de 3 capas, mientras que el modelo mejorado es una red más profunda, pero, para evitar un sobreajuste se emplean funciones como Dropout. Se incorporaron técnicas avanzadas que optimizaron el proceso de entrenamiento, sobre todo en las últimas rondas (Early Stopping y Reduce ROn Plateau), un ejemplo del empleo de estas técnicas es en el trabajo de Makhmudov y colaboradores en 2025.

En el modelo mejorado no solo se muestran métricas de evaluación, sino se añade una matriz de confusión donde se interpreta de una manera más eficiente el rendimiento por clase, y la importancia de las variables estudiadas en la predicción de la calidad. Estas mejoras permitieron evaluar de forma más robusta el problema de clasificación e interpretabilidad que se tuvieron en el primer código de redes neuronales.

4 RESULTADOS Y DISCUSIÓN

A continuación en el siguiente bloque se muestran los resultados obtenidos a partir de la metodología que se aplicó:

4.1 Definición de variables independientes

4.1.1. Realizar revisión sistemática del tema

Al realizar una búsqueda de trabajos científicos enfocados en la calidad del vino, ya sea con análisis fisicoquímicos, sensoriales o modelos de inteligencia artificial, se obtuvieron más de 250 resultados, después de aplicar los criterios de depuración del apartado 3.1.1. quedaron un total de 165 resultados, de los cuales fueron revisados un total de 125. Después de seguir con la metodología descrita se redujeron a 78 trabajos, en la tabla 5 se muestra un extracto de los resultados.

Tabla 5. Extracto de la comparativa de los trabajos realizados sobre la calidad del vino en diferentes países.

Año	Autores	País	Título	Artículos revisados							
				Clasificación	¿Vitis vinifera?	¿Tinto o Blanco?	Resumen	Metodología	Evaluación	Análisis	Cita
2024	Aiello, G., & Tosi, D.	Italia	An Artificial Intelligence-based tool to predict "unhealthy" wine and olive oil	Todos	SI	Tinto	El artículo presenta una herramienta basada en inteligencia artificial diseñada para predecir la calidad sanitaria de vinos y aceites de oliva, clasificándolos como "saludables" o "no saludables". Utilizando técnicas avanzadas de machine learning.	Árbol de decisión, SVM y ANN	Predicción de calidad del vino > 0.65 (más del 65%)	SI	Aiello, G., & Tosi, D. (2024). An artificial intelligence-based tool to predict "unhealthy" wine and olive oil. <i>Journal of Agriculture and Food Research</i> , 16, 101179. https://doi.org/10.1016/j.jaf.2024.101179
2024	Geetha, C., Anunsundar, B., Vasumathi, G.	India	Decoding AI: Experimental analysis of artificial intelligence based wine quality prediction using usno	Todos	SI	Tinto	El artículo presenta un estudio experimental sobre la aplicación de una estrategia de clasificación profunda convolucional para predecir la calidad del vino a partir de sus características fisicoquímicas. Los autores utilizaron un conjunto de datos	CNN	Predicción de Accuracy 0.97	SI	Geetha, C., Anunsundar, B., Vasumathi, G., Thamichachakan, K., & Venkata Sudhakar, C. (2024). Decoding AI: Experimental analysis of artificial intelligence based wine quality prediction using usno convoluted deeo classification strategy. In <i>Proceedings of the 2024 Ninth</i>
2022	Bhardwaj, P., Tiwan, P., Ojear, J. K., Parr, W., & Kulaasi, D.	Nueva Zelanda y China	A machine learning application in wine quality prediction	Fisicoquímico/Inteligencia Artificial	SI	Ambos	Este estudio presenta una aplicación de técnicas de aprendizaje automático para predecir la calidad del vino a partir de sus características fisicoquímicas.	Random Forest, SVM, ANN	MSE, Error Classifier Precision Recall F1 ROC AUC		Bhardwaj, P., Tiwan, P., Ojear, J. K., Parr, W., & Kulaasi, D. (2022). A machine learning application in wine quality prediction. <i>Machine Learning with Applications</i> , 5, 100261. https://doi.org/10.1016/j.mla.2022.100261
2023	Mani, S., Krishnakumary, H. A., Swarnathan, S., & Theerthajagi, P.	India	An investigation of wine quality testing using machine learning techniques	Fisicoquímico/Inteligencia Artificial	NO	NA	El artículo "An investigation of wine quality testing using machine learning techniques" explora cómo los métodos de aprendizaje automático pueden	Random Forest, Naive Bayes y SVM	Logistic Regression, 0.921		Mani, S., Krishnakumary, H. A., Swarnathan, S., & Theerthajagi, P. (2023). An investigation of wine quality testing using machine learning techniques. <i>International Journal of Artificial Intelligence (IJ-AI)</i> , 12(2), 747-754. https://doi.org/10.11591/ijai.v12i2.pp747-754
2019	Khalafyan, A. A., Temerdashev, Z. A., Yakuba, Y. F., &	Rusia	Computer analysis of the sensory qualities of red wines as a method to	Sensorial/Inteligencia Artificial	SI	Tinto	El artículo explora la aplicación del análisis computacional para optimizar las mezclas de vinos tintos basándose en sus cualidades sensoriales.	Análisis sensorial-Estadística ANN, ANN, Aprendizaje supervisado	Modelo cuadrático	SI	Khalafyan, A. A., Temerdashev, Z. A., Yakuba, Y. F., & Guquchkina, T. I. (2019). Computer analysis of the sensory qualities of red wines as a method to optimize their blend formulation. <i>Helvion</i> , 5(5), e01602. https://doi.org/10.1016/j.helion.2019.e01602
2019	Zeng, Y., Liu, Y., Wu, L., Ding, H., Zhang, Y., Guo, H., Guo, Z., Wang, S., & Lan, Y.	China	Evaluation and Analysis Model of Wine Quality based on Mathematical	Todos	NO	NA	El artículo presenta un modelo de evaluación y análisis de la calidad del vino utilizando herramientas matemáticas. El objetivo principal fue	ANN, Aprendizaje supervisado	Coefficiente de determinación (R ²)	SI	Zeng, Y., Liu, Y., Wu, L., Ding, H., Zhang, Y., Guo, H., Guo, Z., Wang, S., & Lan, Y. (2019). Evaluation and analysis model of wine quality based on mathematical model. <i>Studies in Engineering and Technology</i> , 6(1), (artículo n.º 3628). https://doi.org/10.11143set.v6i1.3628

De los 165 trabajos, solo 5 de ellos fueron descartados por estar duplicados, pero de los artículos seleccionados a revisión el 25% fueron descartados ya que no se tenía acceso al texto completo. Partiendo de los 78 trabajos seleccionados, un 33.5% abordan el análisis de la calidad del vino sólo partiendo de análisis fisicoquímicos, y 40% sólo de análisis sensoriales. Un total de 20 artículos de investigación plantean un análisis de la calidad del vino ya sea mediante análisis fisicoquímicos y/o sensoriales pero mediante técnicas de inteligencia artificial, donde predominan

técnicas como Árboles de decisión, redes neuronales y en menor medida máquina de vectores de soporte.

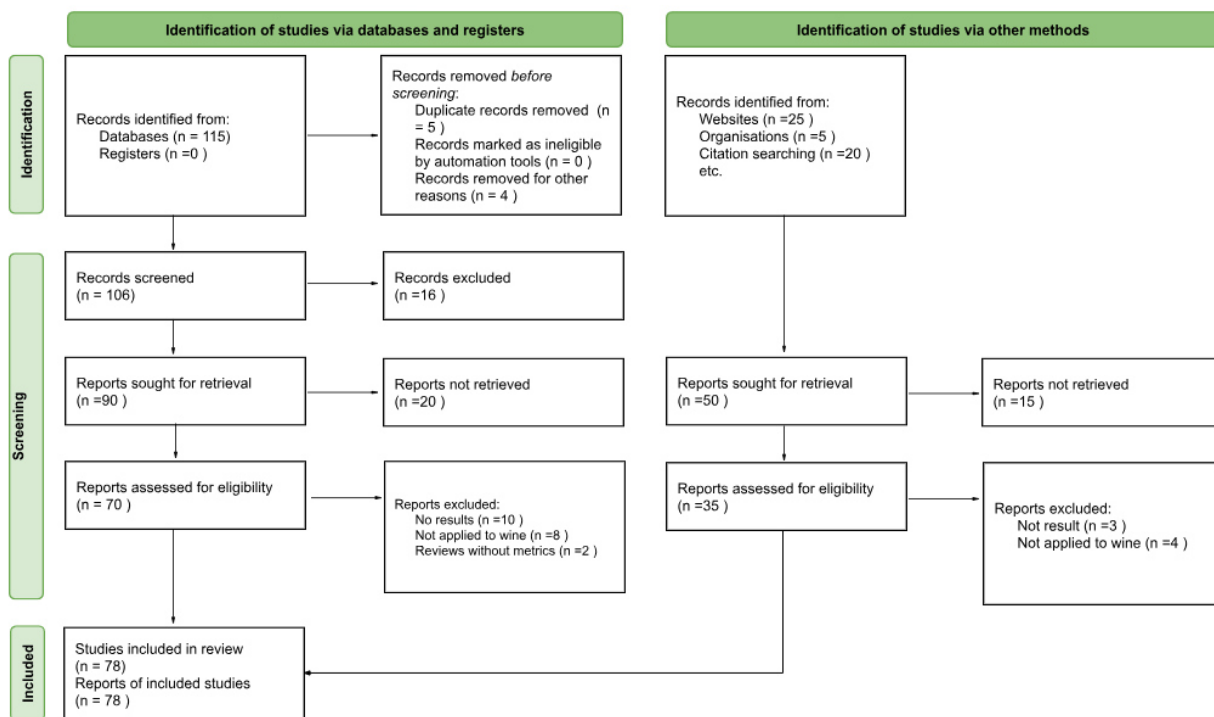


Figura 1. Diagrama de flujo PRISMA de la revisión sistemática con la información de descarte de trabajos.

En la figura 1 se muestra el diagrama de flujo PRISMA que resultó de la revisión sistemática, donde se observan los aspectos principales por los que se excluyeron artículos de investigación, como el que no se abordara el tema del vino, que no tuvieran la metodología completa o que no incluyeran métricas de medición y/o evaluación. Los hallazgos encontrados en la revisión se clasificaron en 6 temas principales, la cantidad de artículos revisados de cada tema se explica detalladamente en la tabla 6.

Tabla 6. Resultados de la revisión realizada dividida por temas principales.

Revisión sistemática: Calidad del vino mediante variables fisicoquímicas y sensoriales		
Temas abordados	Número de artículos	Porcentaje
Análisis sensoriales	25	32.05%
Análisis fisicoquímicos	14	17.95%
Análisis sensoriales y fisicoquímicos	19	24.36%
Análisis sensoriales e inteligencia artificial	3	3.85%
Análisis fisicoquímicos e inteligencia artificial	13	16.67%
Análisis sensoriales, fisicoquímicos e inteligencia artificial	4	5.13%
TOTAL	78	100.00%

De igual manera, se obtuvieron más resultados al respecto, se describen a continuación:

- 56 de los 78 estudios se centran en diferentes variedades de *Vitis vinífera*, el fruto tinto se estudia en 48 de los artículos, mientras que de fruto blanco se estudian en 27 artículos, pero de manera exclusiva solo en 8 artículos.
- El 27% de los artículos revisados emplean vinos con base en frutos como moras, granadas, litchi, arroz, arándanos, etc. Principalmente de China y de Sudamérica.
- Acidez total, pH, Gr.Al% y AZ son los principales análisis de los estudios. Mientras que en los análisis sensoriales el 58% de los artículos que emplean análisis sensoriales analizan aroma, sabor y color.
- Las conclusiones respecto a la calidad del vino se centran en las características organolépticas (análisis sensorial) que presentan.

4.1.2. Selección de variables de interés

La degustación a ciegas que llevaron a cabo los enólogos duró un tiempo de tres horas y veinte minutos, donde discutieron las razones por las que eligieron la calidad de cada vino. A diferencia de otras metodologías como la Pivot descrita en el trabajo de Pearson y colaboradores en 2020, no se empleó una muestra de referencia y se ocultó cualquier información referente al vino que pudiera servir para definir la

calidad, de esta manera se aseguró que el ensayo se desarrollara por completo como un análisis sensorial. Todas las botellas eran de color verde ambar, 10 vinos blancos y 10 vinos tintos, fueron analizados de manera continua del 1 al 20, tomando en cuenta solo parámetros sensoriales como aroma, sabor y color, estas tres características son base de estudio en múltiples trabajos, por ejemplo, en 2014 Gómez, Lasanta y Cobo, presentaron una propuesta educativa sobre el análisis sensorial y como llevarlo a cabo sin necesidad de acudir a expertos, profesionalizando el sector, como en este estudio, al recibir a 10 enólogos de la región se unificaron las metodologías individuales y se convirtió en un estudio comparativo. Se tuvo un porcentaje promedio de coincidencia del 91% tomando como referencia que cada calidad se relaciona con una línea de producción diferente, en la tabla 7 se observan los resultados de manera detallada, destacando que en sólo 7 vinos existió una concordancia entre los sujetos del 100%.

El porcentaje de coincidencia más bajo es del 70%, era un vino blanco con calidad clásica, específicamente un Chenin Blanc 2023, dentro de las principales razones, si analizamos la tabla 7 los últimos 4 vinos son los que tienen un porcentaje más bajo, lo que puede relacionarse con una saturación o sobreexposición debido a la cantidad de muestras a degustar. Si bien, expertos recomiendan no exceder de 7 vinos a degustar de manera continua, para que el ejercicio sea con objetividad y no tengan problemas como la fatiga sensorial, el estudio del presente proyecto se basó en lo establecido para análisis sensoriales realizados en concursos internacionales, donde las jornadas de análisis son de 4 a 5 horas, y el número de vinos analizados es superior a 20 (Fernández-Cuesta, 2014). En este estudio, después de los resultados mencionados anteriormente, si existe una sobre exposición de estímulos sensoriales que superan la capacidad de procesamiento de muestras (Lesschaeve y Noble, 2022).

Tabla 7. Resultados del primer ejercicio de enología para la obtención de variables

Ejercicio 1 de enología: Degustación a ciegas		
<i>Muestra</i>	<i>Calidad BST*</i>	<i>Porcentaje de coincidencia</i>
Vino 1	Baja calidad	100%
Vino 2	Calidad básica	90%
Vino 3	Calidad Premium	100%
Vino 4	Baja calidad	90%
Vino 5	Calidad clásica	90%
Vino 6	Calidad Premium	90%
Vino 7	Calidad Premium	100%
Vino 8	Calidad básica	90%
Vino 9	Baja calidad	90%
Vino 10	Calidad clásica	90%
Vino 11	Calidad básica	100%
Vino 12	Calidad clásica	90%
Vino 13	Baja calidad	100%
Vino 14	Calidad Premium	100%
Vino 15	Calidad clásica	100%
Vino 16	Baja calidad	90%
Vino 17	Calidad básica	80%
Vino 18	Calidad básica	80%
Vino 19	Calidad clásica	70%
Vino 20	Calidad Premium	80%

*Calidad BST: Calidad comparada con las 3 principales líneas del caso de estudio

Mediante el análisis que realizaron los 10 enólogos de la región, se seleccionaron las 10 variables que consideraron, son las que tienen una mayor influencia en la calidad del vino. Siendo la AV, AZ y Gr.Al%, las más influyentes respectivamente en la calidad del vino. Mientras que variables como Antocianinas, polifenoles, cloruros e intensidad colorimétrica son los que menos influencia tienen.

Tabla 8. Resultados del segundo ejercicio de enología para la obtención de variables fisicoquímicas.

Ejercicio 2 de enología: Variables fisicoquímicas	
<i>Análisis fisicoquímico</i>	<i>Número</i>
pH	9
Acidez Volatil	1
Acidez total	8
Anhídrido total	11
Anhídrido molecular	6
Anhídrido Libre	5
Azúcares residuales	2
°Brix	4
Nitrógeno fácilmente A.	7
Antocianinas	12
Polifenoles	13
Cloruros	14
Grado alcohólico	3
Intensidad colorimétrica	15
Ácido Málico	10

Número: Valor proporcionado por el ejercicio de enología, siendo 1 el más influyente en la calidad y 15 el menos.

A partir de las 10 variables seleccionadas, que se observan en la tabla 8, se desarrolló la planeación de los análisis, esto se hizo tomando en cuenta el histórico de análisis proporcionado por el área de laboratorio del caso de estudio. Se verificaron que las metodologías para obtener las variables fisicoquímicas coincidiera con lo establecido en la NOM-142-SSA1/SCFI-2014 (Secretaría de Salud y Secretaría de Economía, 2014) y que fueran las mismas metodologías de los análisis realizados desde 2020, para que existiera una trazabilidad legítima de los resultados y pudieran ser analizados con las mismas condiciones de años anteriores, las metodologías para la obtención de cada variable fisicoquímica están presentadas en la tabla 9.

Tabla 9. Variables de interés según los objetivos planteados

Análisis	Método de análisis	Descripción
Acidez total	Titulación volumétrica	Cuantificación de ácido tartárico en uva, mosto y vino
Acidez volátil	Reacción enzimática	Cuantificación de ácido acético por litro en vino
Anhidrido sulfuroso	Reacción REDOX	Determinación de la concentración de SO ₂ , presente en vinos como conservante.
Ácido málico	Reacción enzimática	Determinación de la concentración de ácido málico post fermentación maloláctica en vinos
Azúcares reductores	Reacción enzimática	Cuantificación de azúcares residuales presentes en vinos espumosos, semidulces y secos
°Brix	Refractometría, densitometría	Análisis de azúcares presentes en la uva y mosto*
Grado alcohólico	Destilación	Determinación del porcentaje de alcohol presente en un vino después de la fermentación alcohólica
Nitrógeno fácilmente asimilable	Titulación volumétrica	Cuantificación de nitrógeno presente en mosto pre fermentación alcohólica
pH	Potenciometría	Determinación del grado de acidez y/o alcalinidad de la uva, mosto y vino.

4.2 Medición del problema

4.2.1. Base de datos unificada 2020-2023

Tabla 10. Extracto de la base de datos de análisis químicos, fisicoquímicos y enzimáticos

Variedad	Rancho	Año	°Brix	pH	AT	SO2L	SO2m	NFA	Azúcar	Alcohol	AV	AM	Calidad Premium	Calidad Varietal	Calidad Básica
Chenin Blanc	Los Dolores	2020	20.80	3.44	6.60	4.75	0.11	200.00	0.75	11.75	0.20	1.45	SI	SI	SI
Chenin Blanc	Los Dolores	2020	20.90	3.40	6.55	4.73	0.12	200.00	0.75	11.75	0.20	1.39	SI	SI	SI
Chenin Blanc	Los Dolores	2020	20.80	3.42	6.59	4.68	0.11	200.00	0.74	11.75	0.19	1.46	SI	SI	SI
Chenin Blanc	Los Dolores	2020	22.00	3.46	5.14	5.31	0.12	215.00	0.60	12.15	0.22	1.30	NO	SI	SI
Chenin Blanc	Los Dolores	2020	22.00	3.45	5.23	5.30	0.12	211.00	0.60	12.10	0.22	1.32	NO	SI	SI
Chenin Blanc	Los Dolores	2020	22.10	3.40	5.15	5.33	0.14	215.00	0.62	12.15	0.22	1.34	NO	SI	SI
Chenin Blanc	Los Dolores	2021	20.30	3.50	6.31	6.00	0.12	215.00	1.10	12.30	0.17	1.50	NO	SI	SI
Chenin Blanc	Los Dolores	2021	20.40	3.54	6.40	6.05	0.11	215.00	1.12	12.25	0.19	1.56	NO	SI	SI
Chenin Blanc	Los Dolores	2021	20.40	3.52	6.39	6.12	0.12	220.00	1.12	12.30	0.19	1.69	NO	SI	SI
Chenin Blanc	Los Dolores	2021	22.30	3.33	6.06	7.60	0.23	230.00	1.00	13.60	0.12	1.24	SI	SI	SI
Chenin Blanc	Los Dolores	2021	22.50	3.30	6.05	7.55	0.24	230.00	1.05	13.60	0.10	1.32	SI	SI	SI
Chenin Blanc	Los Dolores	2021	22.40	3.32	6.00	7.53	0.23	230.00	1.05	13.60	0.11	1.24	SI	SI	SI
Chenin Blanc	Los Dolores	2022	21.50	3.35	4.42	5.36	0.15	105.00	1.40	12.45	0.19	2.10	SI	SI	SI
Chenin Blanc	Los Dolores	2022	21.50	3.30	4.39	5.20	0.17	105.00	1.40	12.50	0.19	2.03	SI	SI	SI
Chenin Blanc	Los Dolores	2022	21.50	3.32	4.46	5.87	0.18	105.00	1.35	12.45	0.19	2.08	SI	SI	SI
Chenin Blanc	Los Dolores	2022	19.50	3.50	4.97	4.75	0.10	231.00	2.30	12.50	0.34	1.02	NO	SI	SI

En la tabla 10 se muestran 16 filas de las 366 que se obtuvieron en la unificación de bases de datos. Se trabajó en Excel y la actividad duró cuatro meses, donde se revisaron los archivos que se proporcionaron en orden de añada, se identificó la información duplicada y se modificó el orden de presentación de los análisis, ya que era diferente en cada archivo. Se enriqueció la información recolectada con el rancho de procedencia, dicha información se obtuvo de la bitácora de trabajo de enología, tanto la bitácora como el 50% de la información analítica del año 2020 se encontraban de manera física en cuadernos de trabajo, por lo que su digitalización llevó tres de los cuatro meses de la actividad.

En la base de datos unificada se encuentra toda la información que estaba distribuída en 16 archivos diferentes, esto conlleva una disminución de tiempo en la búsqueda de datos, ya que al aplicar los filtros correctos en las columnas de la base de datos se puede encontrar la información requerida en menos de un minuto. Este cambio es importante para el área de Enología y de Laboratorio de la empresa de estudio, ya que vuelve más eficiente la trazabilidad de los lotes, necesaria para certificaciones de la industria alimentaria y en específico, del vino (Rivera y Peña, 2022).

4.2.2. Análisis de variables 2024

Tabla 11. Análisis químicos, fisicoquímicos y enzimáticos a vinos 2024.

Variedad	Rancho	Año	°Brix	pH	AT	SO2L	SO2m	NFA	Azúcar	Alcohol	AV	AM
Chenin Blanc	Los Dolores	2024	23.40	3.34	5.92	2.45	0.07	231.00	1.65	13.00	0.22	1.38
Chenin Blanc	Los Dolores	2024	22.80	3.55	5.12	21.19	0.39	193.00	1.35	12.80	0.33	1.89
Chardonnay	Los Dolores	2024	24.10	3.32	6.24	15.32	0.47	231.00	3.42	14.10	0.50	2.14
Colombard	Los Dolores	2024	20.10	3.22	7.28	8.45	0.33	178.00	3.20	13.85	0.32	2.17
Viognier	Los Dolores	2024	22.20	3.28	7.28	9.28	0.31	188.00	2.30	13.75	0.09	2.01
Sauvignon Blanc	Los Llanitos	2024	22.80	3.08	6.14	9.90	0.53	224.00	2.75	13.85	0.22	1.29
Grenache	Santa Isabel	2024	22.00	3.20	7.91	3.54	0.14	161.00	1.95	12.85	0.28	1.92
Barbera	Los Dolores	2024	25.40	3.22	7.13	24.32	0.95	262.00	1.95	12.45	0.54	0.23
Barbera	La Changa	2024	24.60	3.20	7.23	2.56	0.10	241.00	1.67	13.30	0.45	0.28
Cabernet Sauvignon	Los Dolores	2024	18.50	3.52	6.50	4.90	0.10	248.00	2.75	12.05	0.45	0.10
Cabernet Sauvignon	La Changa	2024	21.50	3.55	6.53	4.50	0.08	178.00	2.95	12.90	0.23	0.03
Cabernet Sauvignon	Los Dolores	2024	22.50	3.49	5.65	9.23	0.19	164.00	3.45	13.10	0.63	0.05
Cabernet Sauvignon	Los Llanitos	2024	21.50	3.50	6.93	4.56	0.09	210.00	3.25	12.05	0.60	0.02
Cabernet Sauvignon	Los Llanitos	2024	25.50	3.20	7.79	23.12	0.94	185.00	1.95	13.20	0.34	0.03
Merlot	Los Dolores	2024	23.60	3.14	6.50	9.19	0.43	224.00	2.85	12.45	0.37	0.08
Merlot	Los Dolores	2024	23.60	3.35	5.42	11.34	0.33	234.00	3.05	14.10	0.38	0.02
Merlot	Los Dolores	2024	23.70	3.37	6.24	16.92	0.47	222.00	1.20	13.70	0.41	0.09
Mision	Los Dolores	2024	24.50	3.39	6.23	19.54	0.51	178.00	1.05	14.25	0.54	0.05
Mision	Los Dolores	2024	24.10	3.28	6.34	20.58	0.70	246.00	2.36	13.75	0.50	0.10
Syrah	Los Dolores	2024	25.50	3.50	7.24	22.31	0.46	178.00	2.95	13.80	0.23	0.05
Syrah	Los Dolores	2024	24.80	3.39	6.42	11.39	0.30	205.00	2.65	13.80	0.27	0.09
Tempranillo	Los Dolores	2024	24.50	3.71	5.74	9.87	0.12	231.00	3.05	13.95	0.30	0.01
Tempranillo	Los Dolores	2024	23.40	3.72	4.75	5.82	0.07	248.00	1.95	13.80	0.35	0.04
Tempranillo	Los Dolores	2024	24.00	3.80	4.40	6.20	0.06	188.00	1.00	14.00	0.40	0.03

La etapa de análisis de laboratorio tuvo un periodo de 6 meses, este tiempo fue determinado por el proceso de vinificación, los resultados se encuentran en la tabla 11. Si bien, los análisis de °Brix y NFA se obtuvieron en los dos primeros meses de este tiempo, tenemos datos como AM que se obtiene hasta finalizar la fermentación maloláctica, el último lote en concluir esta fermentación fue en enero de 2025.

Los tiempos de fermentación alcohólica tienen efectos sobre la calidad del vino, como el aumento de compuestos volátiles, en específico la presencia de ácido acético, ácido orgánico que en la industria del vino es relacionado al olor y sabor a vinagre, este subproducto de la fermentación en cantidades superiores a 1.2 g/L es considerado un vino que fue sometido a un proceso de vinificación inadecuado. El tiempo promedio de la fermentación de los vinos 2024 de la empresa de estudio fue de 20 días, un tiempo adecuado según las cinéticas de fermentación de cada lote analizado (Hao et al, 2024).

4.3 Análisis de datos

4.3.1. Correlación de variables fisicoquímicas (*Anexo 1, Figura 11*)

Los análisis de correlación mediante inteligencia artificial ayudan a entender relaciones que debido a la misma complejidad de los datos que se manejan no es posible observarlas en análisis estadísticos. Aplicando esto al vino, las técnicas estadísticas son empleadas sobre todo en análisis sensoriales, y la relación entre la presencia de compuestos aromáticos y cómo influyen, ya sea positiva o negativamente en la percepción del vino (Zeng et al, 2019).

La correlación de las variables fisicoquímicas y la calidad es una herramienta aplicable a diferentes procesos de la industria alimentaria, por ejemplo en la producción de aceite de oliva, en ambos procesos la calidad sanitaria es indispensable, ya que se deben de cumplir los lineamientos legales establecidos en cada país, además de asegurar la calidad del producto (Aiello y Tosi, 2024). El ingreso de la inteligencia artificial a la industria alimentaria genera que se apoyen procesos análisis complejos, como es el observado en este proyecto, donde la correlación se representó en un mapa de color, con relaciones positivas o negativas, valores que van de -1 a 1 (Figura 2).

Se encontraron relaciones esperadas, como la existente entre los °Brix y el Gr.Al%, la primera es una técnica para determinar la cantidad de azúcar existente, la cual es precursor de la fermentación alcohólica, sobre todo en temperaturas mayores a los 24°C y por consiguiente de la producción de etanol, por lo tanto el valor de 0.72 entre °Brix y Gr.Al% observado en la figura 2 es totalmente esperado, se traduce a que entre mayor concentración de °Brix mayor presencia de Gr.Al% (Gómez et al, 2010). Siguiendo con los °Brix, la relación de 0.14 con la calidad nos dice que no se tiene una relación directa entre estas dos variables. Lo anterior mencionado tiene lógica, ya los azúcares residuales están relacionados con la eficiencia de la fermentación alcohólica, ya que representan la concentración de gramos de azúcar por cada litro de vino que queda disponible después de la terminada la fermentación, al estar relacionados de esta manera, es congruente que tampoco tenga una relación directa con la calidad (Izquierdo et al, 2024).

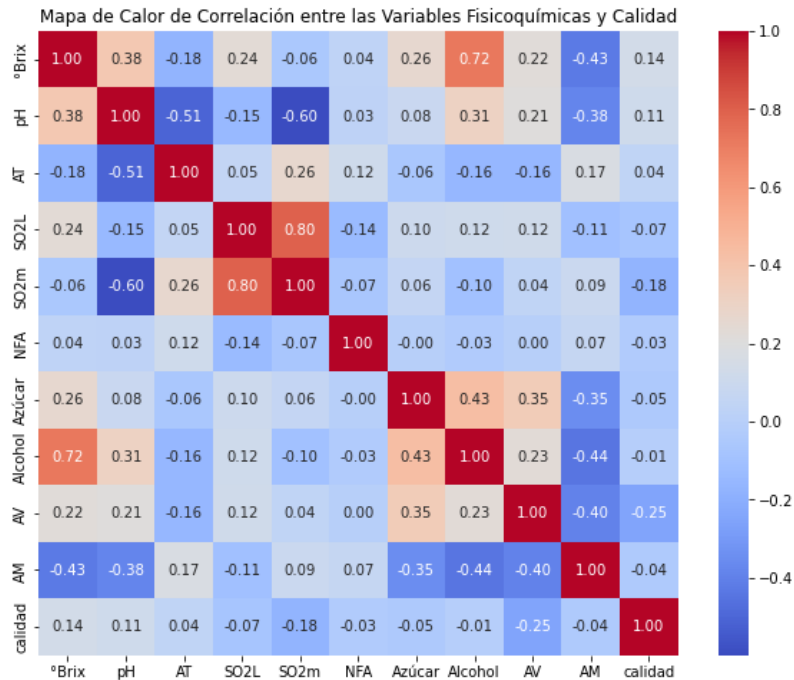


Figura 2. Correlación entre las variables independientes y calidad mediante escala de colores.

Las relaciones que existen entre el valor de pH y SO_{2m} (valor de -0.60), cuando el potencial de hidrógeno es alto el SO_{2m} es bajo, esto está fundamentado en que obtención del último es mediante una fórmula como se observa en el apartado 3.2.2.8, en dicha fórmula se toman en cuenta variables como SO_{2L} mismo caso se observa con su relación con el SO_{2L}. Un variable a considerar es el NFA y su relación negativa con los analíticos de sulfatos, el nitrógeno fácilmente asimilable es parte fundamental de la nutrición de la fermentación, es decir, si las levaduras empleadas para la fermentación alcohólica y cuya función principal es la conversión de azúcar en alcohol no tienen el nutriente suficiente, no consumieran su materia prima, teniendo una fermentación ineficiente y por tanto, afectando la calidad del vino. Ahora bien, los sulfatos son empleados en la industria alimentaria como antimicrobianos, es decir, que inhiben el crecimiento de microorganismos como bacterias o levaduras, entonces, entre mayor NFA existe una mejor cinética de fermentación con una población fuerte de levaduras y para que se inhiba el crecimiento se tendrían que tener grandes concentraciones de sulfatos, justificando la relación que el modelo proporciona (Cortéz et al, 2009).

Una correlación negativa interesante de destacar es la de valor -0.60 del pH con el Nitrógeno fácilmente asimilable, entre mayor sea el potencial del hidrógeno menor es la concentración en partes por millón de NFA, aunque es una relación fuerte, individualmente no afectan a la calidad del vino. Una correlación negativa con la calidad fue el valor de AV, es decir, entre mayor sea la concentración de ácido volátiles, disminuye la calidad de los vinos, esto es ampliamente estudiado, y está relacionado a lo mencionado anteriormente (Espíritu y Quiroz, 2019). Otra correlación a destacar, es la presente entre los °Brix y la calidad, el valor numérico es de 0.14, aunque significa una relación débil es de los valores más altos que se correlacionan con la calidad.

4.3.2. Árbol de decisión

El árbol de decisión generado (Anexo 1, Figura 12) se empleó para clasificar las variables fisicoquímicas en función de predecir la calidad del vino, clasificando este último en las 4 clases de vino ya mencionadas en la metodología (Apartado 3.3.2).

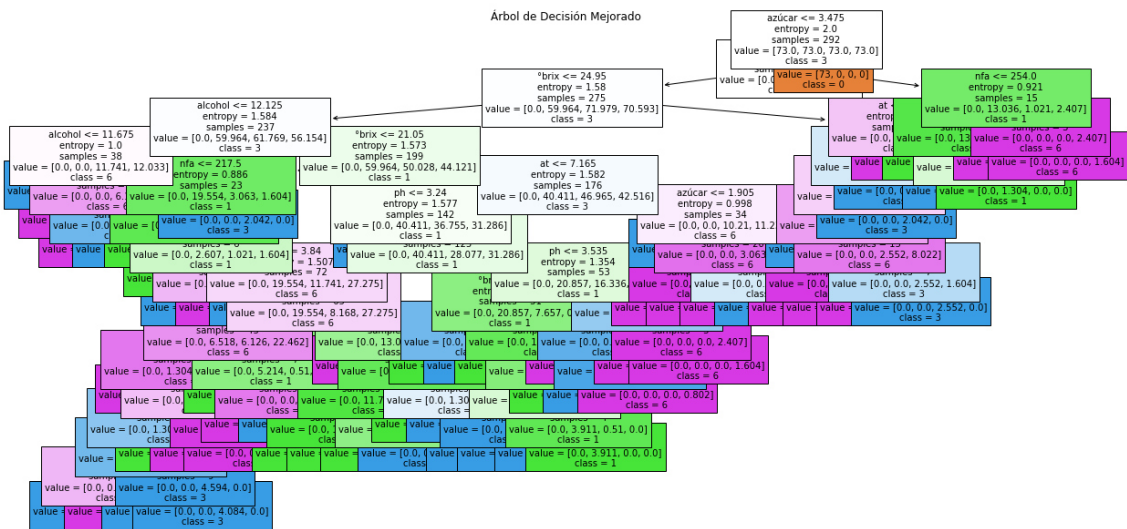


Figura 3. Representación gráfica del árbol de decisión obtenido a partir de la base de datos cargada.

Los nodos, representados en la figura 3 por rectángulos de color blanco, son las primeras condiciones de clasificación, específicamente en este caso la condición inicial está dada por un valor de azúcar menor o igual a 3.475. Cuando el valor del vino cumple la condición, después se consideran variables como Gr.Al%, clasificando los vinos en la clase 6, es decir, calidad premium. Caso contrario se presenta al presentar un azúcar residual mayor a la condición se clasifica como clase 6, vinos de baja calidad. Variables que también son influyentes son el Gr.Al%, el cual aparece en los primeros niveles de clasificación, seguidos por NFA, pH, °Brix y AT. Los árboles de decisión clasifican casi todos los valores, pero algunos con errores, son usados ampliamente en este tipo de estudios según la revisión realizada, pero se tiene una problemática, tiene una limitación de análisis, no clasifican valores que estén en extremos, es decir, valores que no coinciden en su totalidad con las clasificaciones que tienen más datos durante el entrenamiento (Mahima et al, 2020).

La precisión general del modelo para predecir una correcta clasificación de calidad en el vino es del 89%, es decir que el modelo acierta en el 89% de los casos que se presentan en la base de datos generada (3.2.1), el valor de F1-Score fue de 0.88, es considerado aceptable, pero con rango de mejora, como en el trabajo abordado por Jain y colaboradores en 2023, quienes obtuvieron un valor de 0.61 en la métrica de evaluación F1-Score, concluyendo que aunque su modelo lograba predecir la calidad, se requerían de técnicas para mejorar la precisión y de un mayor entrenamiento, que se esperaba lograr al aumentar el conjunto de datos analizados.

Si bien, el modelo presenta un alto rendimiento de manera general, hay clases que individualmente tienen mejores resultados, como la 3 y la 6, calidad clásica y calidad premium respectivamente. Mientras que en clases 0 y 1 las métricas tienen una confiabilidad menor. El valor de la precisión está representado por un porcentaje, que sea inferior a 90% no significa que el modelo no funciona, sino que requiere de modificaciones y en su gran mayoría, de mayor entrenamiento, pero que sea 100% sigue sin ser una señal de éxito, en este caso se puede presentar un sobreajuste de los datos, como se muestra en el trabajo de Bhardwaj y Colaboradores en 2022, donde emplearon diferentes técnicas de inteligencia artificial para determinar la calidad del vino mediante análisis fisicoquímicos, algunas de estas técnicas que al ser evaluadas obtuvieron valores de 100% en Precisión, Recall y F1.

Mientras que en el modelo obtenido de árboles de decisión tuvieron valores por debajo de 0.80, un valor menor al obtenido en el árbol que se muestra en la figura 3.

4.3.3. Redes neuronales

En la generación del modelo se tuvo una precisión de 85%, (Anexo 1, Figura 13) un valor relativamente menor al compararlo con estudios como el publicado por Arshad y colaboradores en 2024, modelo que generó una precisión del 95%, la diferencia más considerable entre los estudios es la región y año de los datos analizados, ya que el estudio realizado en Pakistán fue empleando variables fisicoquímicas recolectadas en Portugal entre 2004 y 2007. Aun así es una precisión que nos dice que el modelo que presentan clasifica correctamente 1,500 muestras de las 1,600 analizadas.

Se evaluó el desempeño con dos gráficos, uno evaluando la precisión durante el entrenamiento y el segundo sobre la pérdida también durante el entrenamiento. Se seleccionaron hacer 50 rondas, es decir, que el modelo emplea 50 veces los mismos datos pero en diferente orden. Partiendo de esto, como se muestra en la figura 4, durante las primeras 10 rondas se mantiene una precisión del 100%, pero en las últimas 5 rondas hay una disminución del 20%, que puede ser debido a factores como un sobreajuste de la información, sugiriendo que aunque el modelo aprende correctamente no se está monitoreado de manera correcta, por lo que se aconsejan ajustes en el código y revisión de la base de datos empleada. En el gráfico de pérdida durante el entrenamiento maneja un valor constante de cero, confirmando que el modelo tiene errores. Esta clase de problemas en el entrenamiento se puede comparar con los se identificaron en el estudio de Agrawal realizado en Corea del Sur en 2018, donde también se empleó la misma base de datos que Arshad y colaboradores, pero la diferencia es que la precisión de clasificación fue baja, teniendo un máximo del 53% en las métricas de evaluación del modelo.

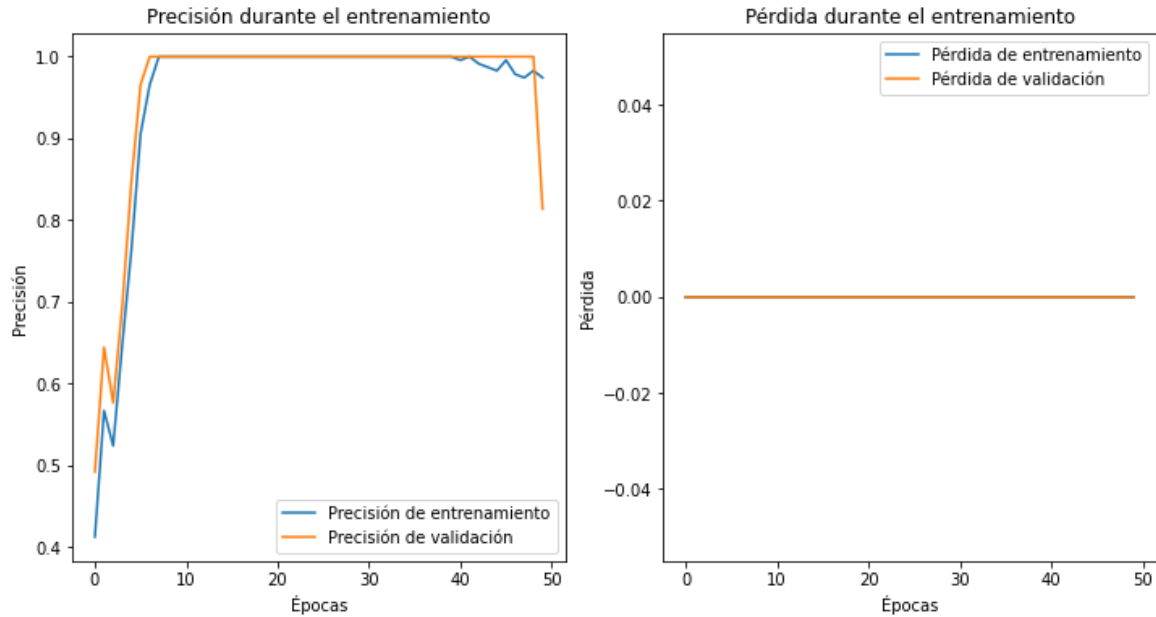


Figura 4. Resultados del entrenamiento del modelo de redes neuronales, mostrando un modelo sobre ajustado.

4.4 Mejora en los modelos de predicción

4.4.1. Modelo mejorado de Árbol de decisión

Al realizar las mejoras en el modelo de árbol de decisión a partir de las modificaciones del código de Python mostrados en el Anexo 1 (Figura 14), se obtuvo un árbol simplificado donde cambió la variable de condición. En la figura 5 se observa que la acidez volátil con un valor igual o menor que 0.285 es la condición de clasificación, en un segundo nivel tenemos variables con influencia relevante como el anhídrido sulfuroso libre y los grados Brix. Otras variables que destacar son NFA, pH, AT, Azúcar y AV de nuevo, pero la última con un valor mayor.

Cuando se cumple la condición principal y a lo largo de esa bifurcación tenemos vinos con calidad Premium y Clásico, mientras que del otro lado del árbol están vinos de clase Clásica y Básica. En comparación con la figura 3, donde se muestra el modelo generado antes de las mejoras, el valor de la entropía era mayor a 1.0, con picos de 2.0, en el árbol de decisión mejorado se tienen valores menores a 1.0 y con un pico de 1.5, concluyendo que existe una mejora respecto a la impureza. Trabajos

publicados sugieren que este tipo de modelos pueden lograr automatizar el proceso de evaluación de calidad, aun con precisiones que no superen el 95% o con valores de entropía no tan cercanos a cero, la recomendación es usar más datos y aumentar el entrenamiento (Mani et al, 2023).

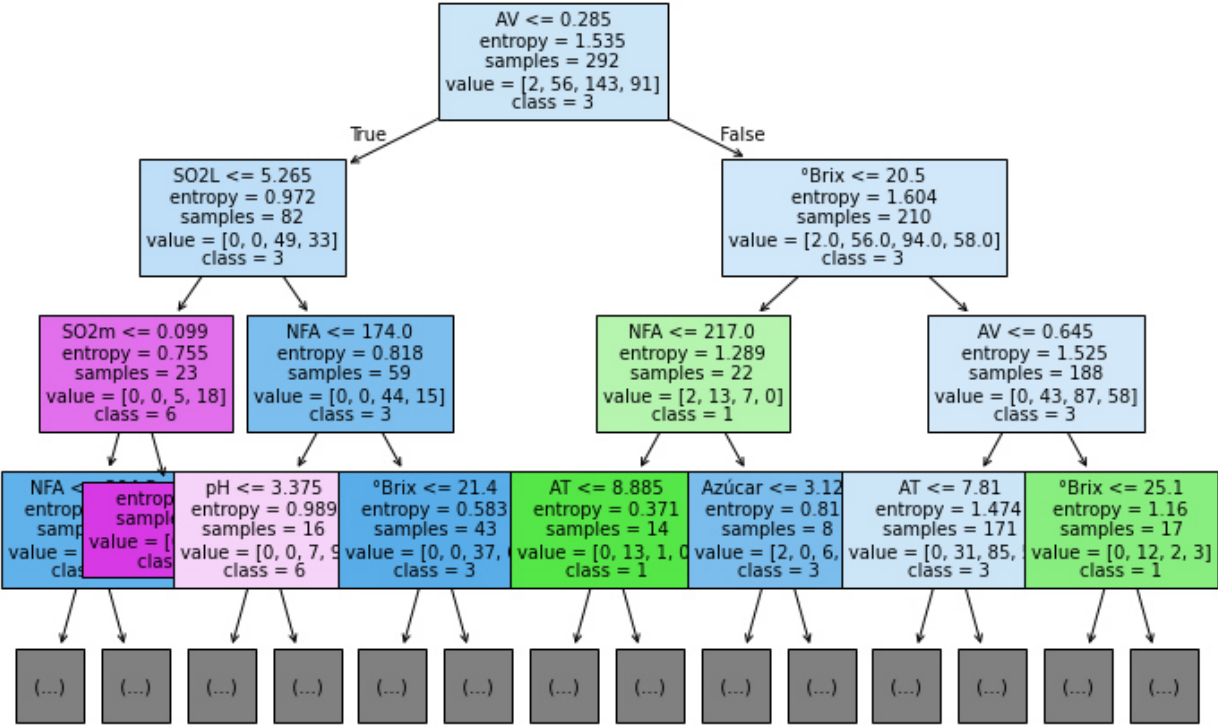


Figura 5. Árbol de decisión mejorado con validación cruzada y ajuste de hiperparámetros.

Según el reporte de clasificación del Árbol de decisión generado posterior a las mejoras, tuvo una precisión general del 93%. No solo se considera el aumento de la precisión del modelo, sino que las variables que son consideradas para clasificar la calidad del vino coincide con las seleccionadas en el ejercicio de enología realizado (4.1.2). Las clases 3 y 6 siguen teniendo una precisión mayor, es decir que las clasificaciones coinciden con los datos proporcionados, mientras que la clase 1 aunque tiene un valor de precisión más bajo, tiene un porcentaje de precisión mayor que en trabajos como el encabezado por Er y Atasoy en 2026 donde manejan una precisión alrededor del 70% para la predicción de la calidad en el vino, sin embargo

su modelo logró un 99% de precisión para la clasificación del tipo de vino, si era blanco o tinto.

Otra de las herramientas que se emplearon para evaluar el rendimiento del modelo de clasificación fue la matriz de confusión que se presenta en la figura 6. En la clase 3 se tuvieron solo 2 errores en la clasificación y 41 aciertos, lo mismo se repite en las clases 3 y 1 que tuvieron 2 errores, pero con una menor cantidad de aciertos. Una característica que comparten las clases 3 y 6 en este análisis, es que los errores de clasificación fueron detectados como una clase anterior a la real, es decir, los dos errores de la clase 3 fueron clasificados como clase 1. Sin embargo, esto no sucedió con la clase 1, ya que el error fue detectado como clase 3 (Ünlü, 2020).

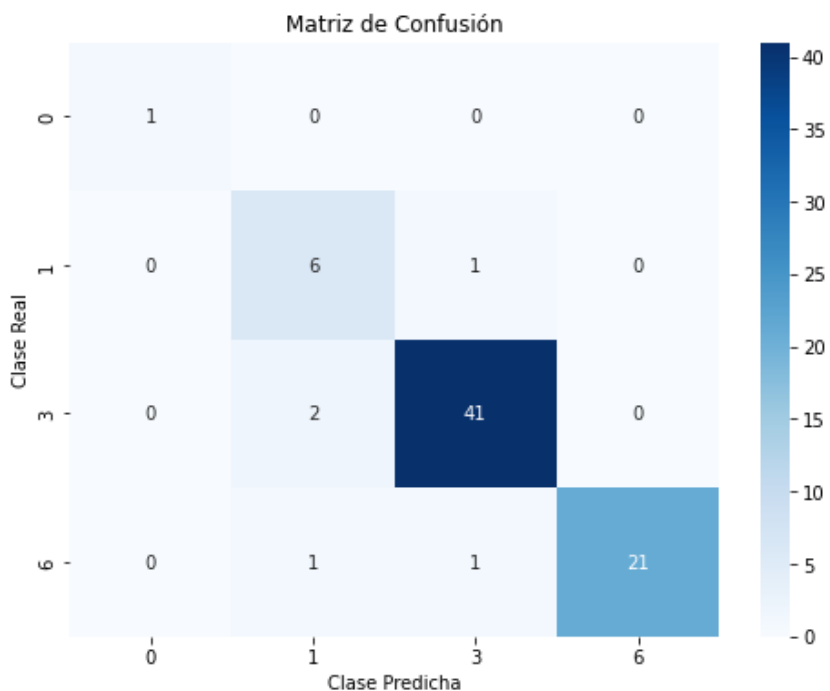


Figura 6. Matriz de confusión del modelo de árbol de decisión después de las mejoras.

Uno de los cambios más significativos en el código del modelo de árbol de decisión, fue que se incluyó un gráfico donde se representara la importancia de las variables fisicoquímicas en la calidad del vino. Como se muestra en la figura 7, las más influyentes son °Brix, AV, NFA, y AT, en orden decreciente. Al compararlos con

los resultados obtenidos en el ejercicio de enología tienen el lugar 4, 1, 7 y 8 respectivamente, que si bien, está la variable considerada como la más importante en el apartado de resultados 4.2.1, también están variables que están cercanas al puesto 10.

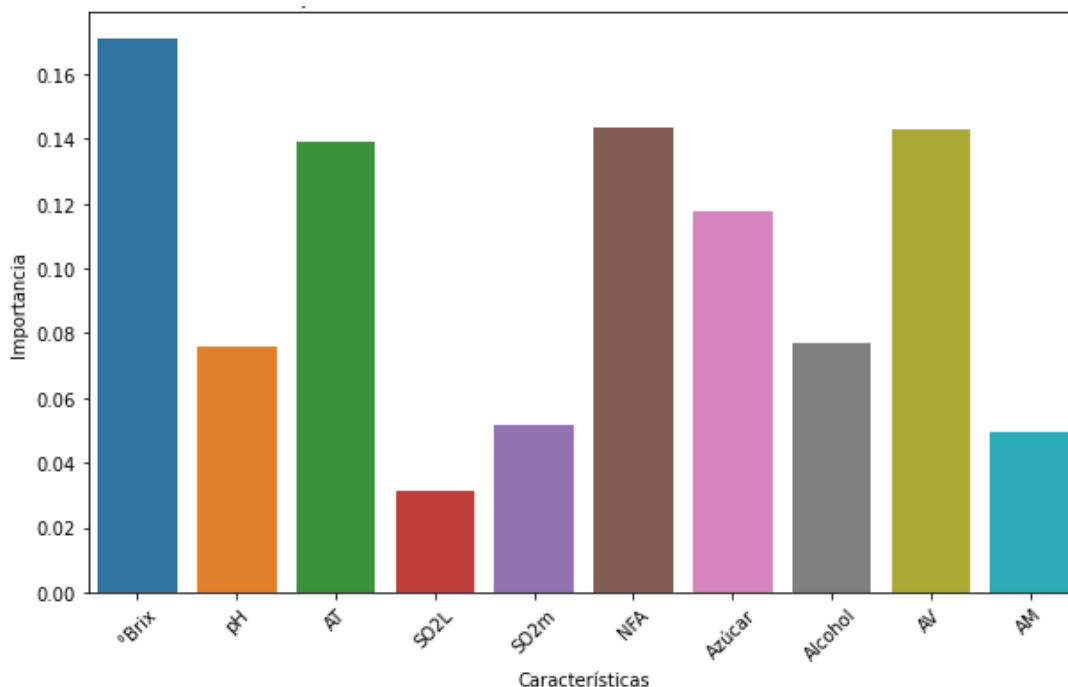


Figura 7. Gráfico de la importancia de las variables fisicoquímicas en relación con la calidad a partir del árbol de decisión.

Estos resultados se pueden comparar con los obtenidos en el trabajo de Galeano, Aguirre y Castrillón de 2021, en 300 de las muestras evaluadas se tuvieron un total de 14 clasificaciones incorrectas, de igual manera se manejaron 4 clases y la que tuvo más errores fue la primera con 7 errores en total, con una precisión general de 95%. El modelo propuesto en el presente proyecto con una precisión del 94% es comparable con los estudios mencionados, ya que emplean técnicas similares, la diferencia radica en la cantidad de variables estudiadas, el estudio es realizado en Colombia, la diferencia de región es una de las razones por la que los criterios de selección de variables cambian.

4.4.2. Modelo mejorado de Redes neuronales (*Anexo 1, Figuras 15-18*)

En la evaluación del modelo mejorado se obtuvo una precisión del modelo de 94%, un aumento del 9%, además de obtener un F1-score y Recall con el mismo valor. Las clases 0, 3 y 6 tuvieron una precisión mayor al 95%, no fue el mismo caso de la clase 1, teniendo un porcentaje menor, pero considerando que la base de datos con la que se realizó el estudio, la clase 1 tiene un 35% menos de datos que las otras clases. El entrenamiento fue correcto, se obtuvieron buenas métricas, agregar más datos de las clases con menor precisión proporciona mayor confiabilidad en los resultados.

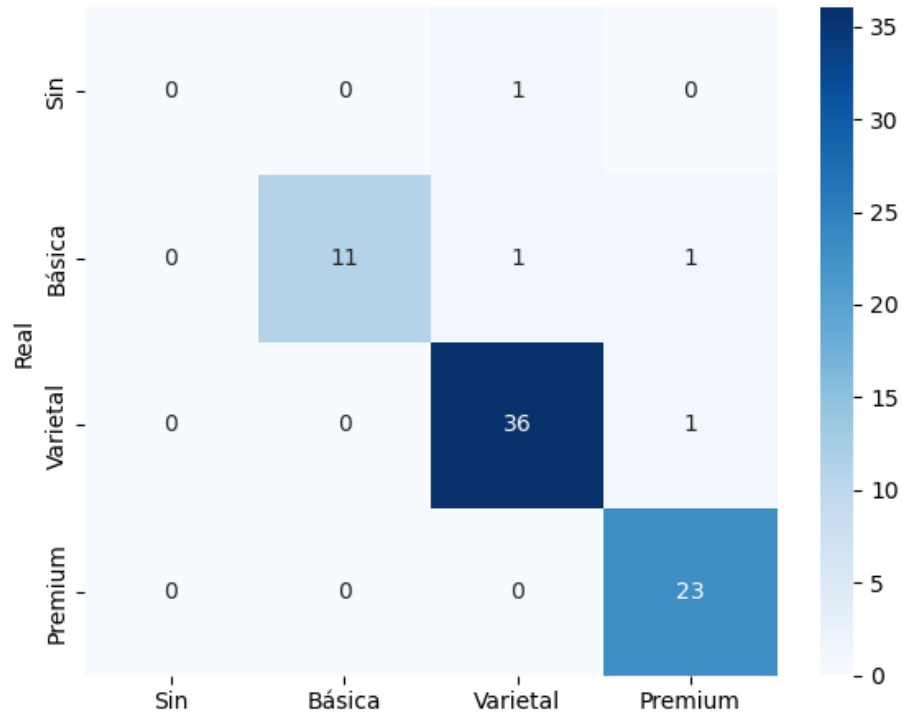


Figura 8. Matriz de confusión generada en el modelo de redes neuronales después de las mejoras.

Al igual que el modelo generado de árbol de decisión se empleó una matriz de confusión (Figura 8) como evaluación del modelo después de las mejoras implementadas. Solo existieron 4 errores en general, donde se predijo la clase 1, dos eran clase 3 y dos clase 6. Esto nos arroja información que no se obtuvo antes,

cuando el modelo comete errores suele ser a las clases donde se tiene menor información, en este caso es la clase 0 y 1.

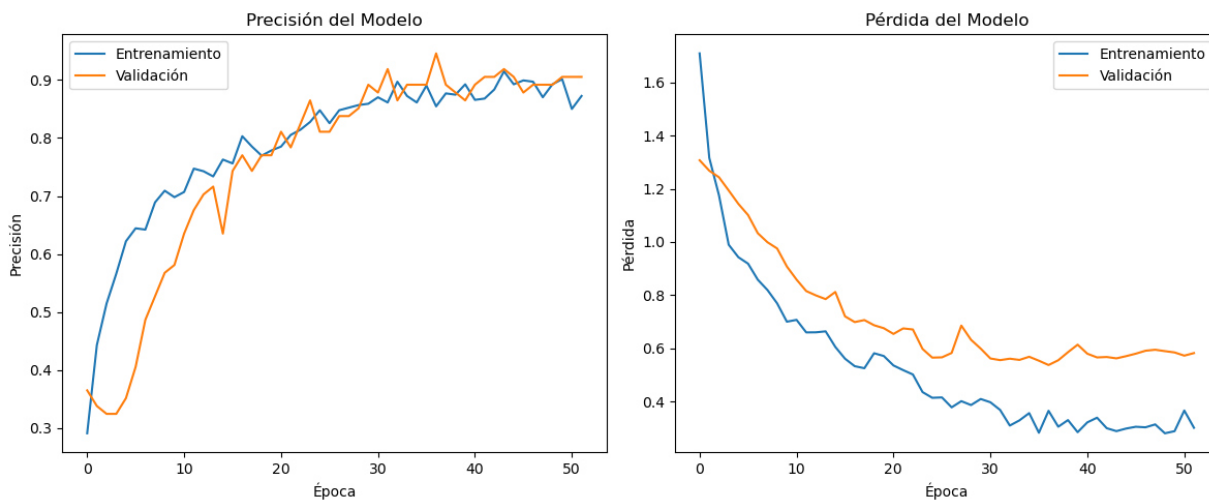


Figura 9. Resultados del entrenamiento del modelo mejorado de redes neuronales.

En la figura 9, en el gráfico de la izquierda se muestra el rendimiento del modelo de redes neuronales generado durante 50 rondas de entrenamiento, alcanzando valores superiores al 90% de precisión. A partir de esto, se entiende que el modelo ha aprendido de manera correcta y que puede generalizar adecuadamente. Por otra parte, en el gráfico de pérdida del modelo se tiene información relevante, como que se tiene una disminución progresiva de la pérdida durante las primeras 20 rondas, se puede observar que la pérdida del entrenamiento es menor que la de validación, pero es muy poca la diferencia además que es algo esperado en este tipo de modelos. Los resultados son favorables, ya que se refleja un modelo preciso y estable, sin embargo, se puede mejorar con un aumento en los datos analizados (Makhmudov et al., 2025).

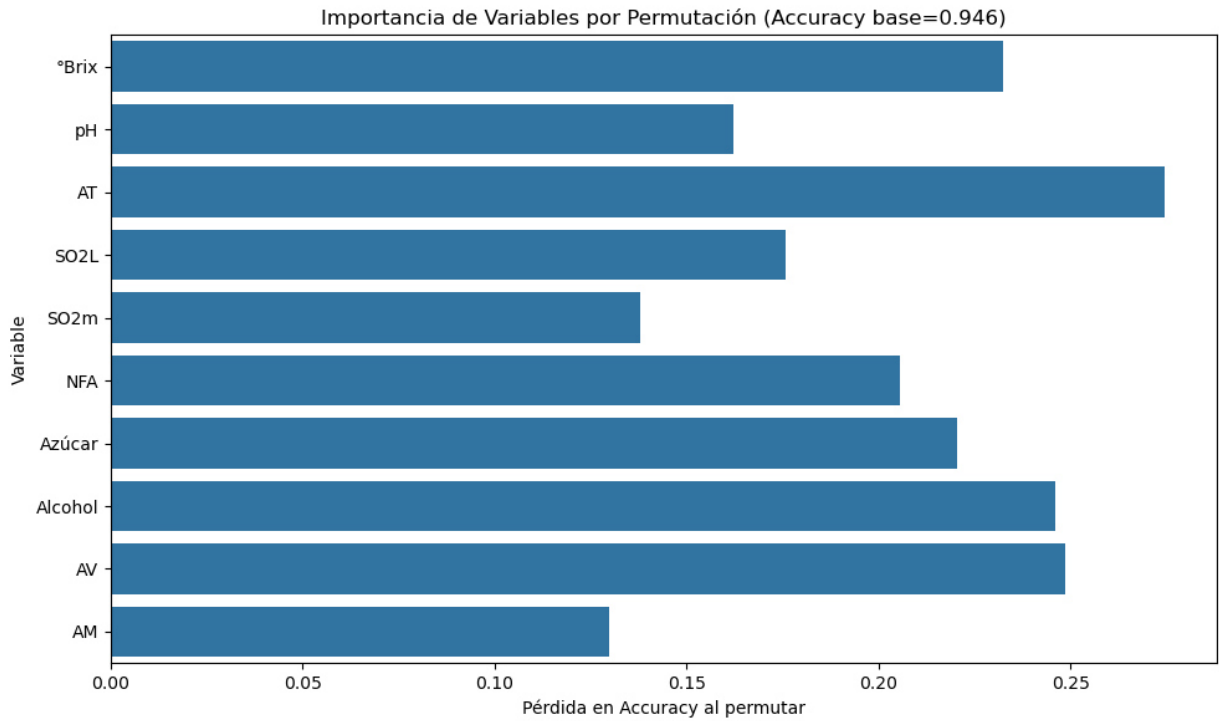


Figura 10. Gráfico de la importancia de las variables fisicoquímicas en relación con la calidad a partir del modelo de redes neuronales.

Al realizar el análisis de las variables fisicoquímicas que tienen mayor influencia en la calidad del vino, como se muestra en la figura 10 las variables que se obtuvieron fueron diferentes a las mencionadas anteriormente en el punto 4.3.1. Las variables identificadas son AT, AV, Gr.Al%, °Brix, NFA y Azúcar, los resultados son comparables con los obtenidos en trabajos similares como el realizado en 2023 por Jain et al., donde determinaron que las variables con una mayor influencia son Gr.Al%, SO₂L, AV y °Brix.

4.4.3. Comparación de los modelos de predicción generados

Según la tabla 12, los modelos generados cumplieron con el objetivo de predecir la calidad del vino según la información proporcionada por la empresa de estudio. Al realizar una comparación entre el árbol de decisión y el modelo por redes neuronales se pueden encontrar diferencias significativas, si bien la precisión (Accuracy) es muy similar al solo diferir por 1%, existieron más errores de clasificación en el árbol de decisión en diferentes clases y sin tendencia clara. Con el modelo de redes neuronales se encontró una posible tendencia de error a futuro que debe de ser corregida, al seguir alimentando la base de datos empleada. Múltiples estudios valoran dos o más técnicas para la predicción de la calidad del vino mediante análisis fisicoquímicos, muchos de ellos ya fueron mencionados en el texto, pero, al momento de elegir la mejor técnica aun cuando la precisión es mejorable, se elige el uso de redes neuronales, ya que capturan relaciones más complejas, aprenden patrones no lineales de alta complejidad, en cambio, el árbol de decisión es más limitado, específicamente en el tema del vino, existen relaciones no lineales entre las variables fisicoquímicas, por lo tanto el árbol de decisión no las capta bien (Gupta, 2018).

Tabla 12. Métricas de evaluación de las técnicas empleadas

Métricas de evaluación del modelo		
Métrica	Árbol de decisión	Redes Neuronales
Accuracy	0.93	0.94
F1-Score	0.91	0.94
Recall	0.93	0.91

Uno de los objetivos del proyecto era identificar las variables fisicoquímicas con mayor influencia en la calidad del vino, tomando en cuenta los resultados del ejercicio de enología ya mencionado y el cual es considerado por la Organización Internacional de la Viña y del Vino, un ejercicio válido para la determinación de la calidad del vino, el modelo que coincide en identificar las 8 variables es el de redes neuronales, si bien, algunas posiciones son diferentes, tiene una mayor coincidencia que la presentada en el modelo de árbol de decisión. Como se muestra en la tabla 13,

las variables que se consideran con un mayor influencia en la calidad del vino son AT, AV, Gr.Al%, °Brix, AZ, NFA, SO₂L y pH. (Organización Internacional de la Viña y el Vino [OIV], 2015).

Tabla 13. Comparativa de los métodos empleados para identificar variables

Comparativa de los métodos para determinar las variables con mayor influencia en la calidad del vino			
<i>Análisis fisicoquímico</i>	<i>Degustación</i>	<i>Árbol de decisión</i>	<i>Redes neuronales</i>
AV	1	2	2
Azúcares	2	5	5
Alcohol	3	6	3
°Brix	4	1	4
SO ₂ L	5	10	7
SO ₂ m	6	8	9
NFA	7	3	6
AT	8	4	1
pH	9	7	8
AM	10	9	10

*El número 1 es la variable fisicoquímica con mayor relevancia en la calidad del vino, siendo el 10 el que menos influencia tiene.

En la tabla 13 aparece una comparativa entre las tres metodologías descritas en este trabajo, en la columna de degustación se hace referencia al análisis sensorial, es decir, datos obtenidos de manera cualitativa, que no es el caso con las últimas dos columnas, que son obtenidos a través de las técnicas de inteligencia artificial y con datos cuantitativos. El número 1, es decir, la variable más influyente en la calidad del vino, es diferente en los 3 casos, en la degustación se tiene la AV, que como se ha explicado a lo largo del texto la presencia de altas concentraciones pone en riesgo la calidad del vino, sin embargo, tanto en el árbol de decisión como en redes neuronales es una variable que está en segundo lugar, con solo un lugar de diferencia se comprueba que es una variable determinante en la calidad del vino (Espíritu y Quiroz, 2019). La AT es la variable que el modelo de redes presentó como la más influyente, analizando los estudios sensoriales publicados, como el de Yang y

colaboradores en 2025, donde comunican que expertos pudieron identificar vinos sólo con atributos como la acidez tartárica (AT), misma que es constantemente evaluada para determinar la calidad del vino, como en el trabajo realizado por Zhao y colaboradores en 2023, que al estudiar una nueva técnica de análisis sensorial ultrasónico, la AT fue de los principales parámetros medidos antes y después de aplicar la técnica, empleando esta variable como una métrica de evaluación, se puede concluir que la variable es altamente estudiada por ser influyente en la calidad del vino, lo que coincide con lo obtenido en el modelo que se generó.

Una variable que es considerada por las tres metodologías, de las menos influyentes es el AM, es un resultado razonable ya que en estudios extensos realizados para caracterizar al vino de manera fisicoquímica, no se menciona una evaluación de la concentración de ácido málico, si bien, la fermentación maloláctica sucede de manera espontánea en gran porcentaje, hay vinos que nunca la tienen, como la mayoría de los vinos blancos, por lo que su influencia en la calidad del vino es menor que otras variables (Mani et al., 2023). La variable AM, junto con °Brix y Gr.Al% son las concordancias entre el análisis sensorial y el modelo de redes neuronales, pero también se encuentran variables que difieren por uno o dos puestos, es decir, que en una metodología están en cuarto lugar y en la otra en quinto lugar, en este caso se tienen variables como NFA y pH, comprobando la importancia del primero, función en el vino que ya fue descrita anteriormente, lo que no coincide con la literatura es el pH, aun estando en octavo lugar es la variable que se encuentran en al 80% de los estudios que abordan análisis fisicoquímicos, pero, es una técnica que se tiene al alcance y de bajo costo, por lo que se justifica el que sea tan empleada (Bayoï et al., 2024).

Tanto el análisis sensorial y el árbol de decisión tienen una mayor concordancia con el modelo de redes neuronales, pero no entre sí, ya que no coinciden en ninguna de las variables y existen grandes diferencias, como en la variable de SO₂L, en el análisis sensorial está en la quinta posición mientras que en el árbol de decisión está en el lugar 10. Estas diferencias sugieren que se analiza de manera diferente la influencia de las variables fisicoquímicas en la calidad del vino, y que un análisis cuantitativo coincide con lo establecido por lo publicado en el mundo (Gupta, 2018).

5 CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

En función de los resultados obtenidos, se presentan las siguientes conclusiones:

- La revisión sistemática realizada arrojó que de los 78 artículos seleccionados, en el 70%, el objeto de estudio es vino proveniente de *Vitis vinífera*, de los cuáles sólo el 30% abordaban en su metodología algún análisis de datos con técnicas de inteligencia artificial.
- La calidad del vino actualmente está determinada por análisis fisicoquímicos y sensoriales, siendo estos últimos el foco principal en más del 60% de los estudios que emplean *Vitis vinífera*, el doble de los artículos que se enfocan en análisis fisicoquímicos como las principales variables que tienen influencia en la calidad del vino.
- El análisis sensorial es actualmente la metodología más replicada para determinar la calidad del vino, como evaluación en Consejos Vitivinícolas así como en concursos de vinos, donde se realizan degustaciones extensas. Según el ensayo realizado, a partir del vino número 16 se muestra una sobre exposición de vino, lo que genera una menor eficiencia al momento de clasificar, aumentando considerablemente los errores al momento de elegir la calidad del vino.
- Las 8 variables que tienen una mayor relevancia en la predicción de la calidad del vino son AT, AV, Gr.Al%, °Brix, AZ, NFA, SO₂L y pH. Se responde la primer pregunta de investigación pero se rechaza la H1, ya que las variables que se establecieron en la hipótesis fueron propuestas al realizar una revisión de literatura de manera parcial, en donde se incluyó el AM y SO₂m dentro de las 8 variables más influyentes y se dejó fuera los °Brix.
- Con el análisis de correlación que se realizó se determinaron las relaciones positivas y negativas que existen entre las 10 variables elegidas para el estudio y la calidad del vino. En correlaciones positivas están variables como °Brix, pH y AT, es decir que cuando existe un concentración mayor de estos variables contribuyen en el aumento de la calidad, en contraparte, las variables que

cuando se tienen en altas concentraciones disminuyen la calidad están SO₂L, SO₂m, AV y AM aceptando la H₂ en su totalidad y dando respuesta la segunda pregunta de investigación.

- Tanto el modelo generado de árbol de decisión como el de redes neuronales predicen la calidad del vino con una precisión superior al 90%, comparable con los trabajos científicos disponibles a lo largo del mundo. Sin embargo, se considera que Redes Neuronales es el enfoque con mejores resultados, y por consiguiente el más adecuado para el caso de estudio. Se acepta la hipótesis número tres que da respuesta a la tercera pregunta de investigación.

5.2 Recomendaciones

Con la finalidad de que este trabajo pueda ser replicado, se presentan las siguientes recomendaciones:

- Se recomienda realizar una reestructuración de las clases de calidad al momento de replicar el estudio, dado que las definidas en este proyecto son establecidas por el caso de estudio, esto permitirá mantener coherencia y relevancia de los resultados que se obtengan.
- Para generar un modelo preciso y estable, ya sea por árbol de decisión o redes neuronales, se sugiere tener grandes conjuntos de datos, ya que hacen más eficiente el entrenamiento de los modelos.
- Se sugiere replicar este estudio en otras regiones vitivinícolas del estado de Baja California, con el fin de enriquecer la información que se tiene disponible en el tema de la calidad del vino.
- Con el fin de garantizar la validez y reproducibilidad del estudio, se recomienda emplear las metodologías de análisis estandarizadas y reconocidas por organismos oficiales, como la NOM-142-SSA1/SCFI-2014 en México y lineamientos internacionales como los establecidos por la OIV.
- El valle de Santo Tomás, Ensenada, fue el foco de estudio, pero existen otros valles en el municipio de Ensenada que se pueden considerar, se sugiere ampliar la investigación a estos valles, tomando en cuenta que poseen diferencias de suelo y clima, características importantes a considerar al momento de incluirlos en el estudio.

5.2.1 Trabajo Futuro

Como toda investigación, el presente proyecto aborda áreas con gran oportunidad de crecimiento, pero al emplear una actividad económica de gran relevancia en el estado, que también tiene un impacto cultural y social, las oportunidades son aún mayores. Emplear inteligencia artificial para determinar la calidad del vino es una gran oportunidad para empresas de todo tamaño, pero sobre todo para empresas pequeñas que no tienen el recurso necesario para gastar en análisis sensoriales que pueden estar fuera de su presupuesto anual. Pero, para poder llegar al punto de establecer este modelo de predicción de calidad como una nueva metodología se requiere de un mayor análisis, sobre todo, de seguir alimentando el modelo con nuevos datos y nuevas variables.

Solo en el municipio de Ensenada hay al menos 7 valles reconocidos por su producción vitivinícola, cada uno con condiciones edafoclimáticas diferentes. Para aplicar un modelo que pueda abarcar todo el territorio de producción se deben de tomar en cuenta los datos regionales de los vinos producidos en estos valles, así como los análisis fisicoquímicos que de cada uno.

Como trabajos futuros a realizar se tiene:

- Un modelamiento integral para la evaluación de la calidad del vino mediante inteligencia artificial empleando sistemas supervisados y no supervisados.
- Análisis de la calidad del vino producido en el municipio de Ensenada, México mediante análisis fisicoquímicos empleando técnicas de inteligencia artificial.
- Aplicación de un modelamiento integral para la evaluación de la calidad del vino mediante inteligencia artificial en otros estados productores de vino de México.
- Publicaciones científicas relacionadas a revisiones sistemáticas sobre la determinación de la calidad del vino mediante análisis fisicoquímicos y sensoriales.

6 REFERENCIAS

Adamashvili, N., Zhizhilashvili, N., & Tricase, C. (2024). The integration of the Internet of Things, Artificial Intelligence, and Blockchain Technology for advancing the wine supply chain. *Computers*, 13(3), 72. <https://doi.org/10.3390/computers13030072>

Agrawal, G., & Kang, D.-K. (2018). Wine Quality Classification with Multilayer Perceptron. *International Journal of Internet, Broadcasting and Communication*, 10(2), 25–30. <https://doi.org/10.7236/IJIBC.2018.10.2.5>

Aiello, G., & Tosi, D. (2024). An artificial intelligence-based tool to predict “unhealthy” wine and olive oil. *Journal of Agriculture and Food Research*, 16, 101179. <https://doi.org/10.1016/j.jafr.2024.101179>

Alhassan, R., & Asamoah, D. (2023). Utilizing random forest algorithm for early detection of academic underperformance in open learning environments. *PLOS ONE*, 18(12), e0281234. <https://doi.org/10.1371/journal.pone.0281234>

Alves, J. A., Lima, L. C. O., Nunes, C. A., Diaz, D. R., & Freitas, R. (2011). Chemical, physical-chemical, and sensory characteristics of lychee (*Litchi chinensis* Sonn) wines. *Journal of Food Science*, 76(5). <https://doi.org/10.1111/j.1750-3841.2011.02188.x>

Amorocho-Cruz, C.-M., Soto-Mora, J.-E., & Charry-Roa, S. (2022). Caracterización fisicoquímica, microbiológica y sensorial de vino de curuba (*Passiflora mollissima* var. Bailey). *Revista Biotecnología en el Sector Agropecuario y Agroindustrial*, 20(2), 45–59. <https://doi.org/10.18684/rbsaa.v20.n2.2022.1562>

Andrade, S. G., & Flores, N. F. (2014). Matriz de insumo producto vitivinícola de Baja California México. *Contaduría y Administración*, 30(81), 57. [https://doi.org/10.1016/S0186-1042\(14\)71166-9](https://doi.org/10.1016/S0186-1042(14)71166-9)

Andreu-Sevilla, A. J., Mena, P., Martí, N., García Viguera, C., & Carbonell-Barrachina, Á. A. (2013). Volatile composition and descriptive sensory analysis of pomegranate juice and wine. *Food Research International*, 54, 246–254. <https://doi.org/10.1016/j.foodres.2013.07.007>

Ao, H., Tang, C., Lu, Y., Zhang, Y., He, L., Qiu, S., Yan, Y., & Li, C. (2025). Characterization of physicochemical properties, sensory characteristics, and volatile compounds with a special focus on the terpene profile of commercial Chinese kiwifruit wines. *Journal of Food Composition and Analysis*, 140, 107187. <https://doi.org/10.1016/j.jfca.2025.107187>

Arshad, H., Naveed, H., Nasim, F., Ahmad, J., Ahmed, M., & Liaqat, M. S. (2024). Wine Quality Prediction Using Machine Learning. *JICET*, 4(2).

Ashton, R. H. (2013). Is There Consensus Among Wine Quality Ratings of Prominent Critics? An Empirical Analysis of Red Bordeaux, 2004–2010. *Journal of Wine Economics*, 8(2), 225–234. <https://doi.org/10.1017/jwe.2013.18>

Atwal, G., Bryson, D., & Williams, A. (2021). An exploratory study of the adoption of artificial intelligence in Burgundy's wine industry. *Strategic Change*, 30(3), 299–306. <https://doi.org/10.1002/jsc.2413>

Bayoï, J. R., & Etoa, F.-X. (2024). Formulation, physicochemical, sensory, bioactive properties, and microbiological safety assessment of aged traditional “bekleu” (*Balanites aegyptiaca*) wine. *Journal of Agriculture and Food Research*, 15, 101003. <https://doi.org/10.1016/j.jafr.2024.101003>

Bahena, L. (2023). Panorama de la producción de vino en México. *The Food Tech*. Recuperado de <https://thefoodtech.com/tendencias-de-consumo/16-estados-de-mexico-estan-pr-oduciendo-vino-de-alta-calidad/>

Bejerano, P. C., & Zapater, J. M. (2013). Estructura y composición de la uva y su contribución al vino. *Revista de Enología Científica y Profesional*, 139(8). Recuperado de https://www.acenologia.com/estructura_composicion_vino_cienc1013/

Beltrán Balarezo, C., Carchi Tandazo, T. A., Yáñez Romero, M. E., Lam, A. M., & Flores Acosta, A. R. (2023). Influencia de las propiedades fisicoquímicas del vino de mora, determinadas mediante análisis sensorial. *Ciencia Latina Revista Científica Multidisciplinar*, 7(4). https://doi.org/10.37811/cl_rcm.v7i4.7125

Beltrán Balarezo, C., Carchi Tandazo, T. A., Yáñez Romero, M. E., Lam, A. M., & Flores Acosta, A. R. (2023). Influencia de las propiedades fisicoquímicas del vino de mora, determinadas mediante análisis sensorial. *Ciencia Latina Revista Científica Multidisciplinar*, 7(4). https://doi.org/10.37811/cl_rcm.v7i4.7125

Bhardwaj, P., Tiwari, P., Olejar Jr, K., Parr, W., & Kulasiri, D. (2022). A machine learning application in wine quality prediction. *Machine Learning with Applications*, 8, 100261. <https://doi.org/10.1016/j.mlwa.2022.100261>

Blouin, J., & Peynaud, É. (2003). *Enología práctica: Conocimiento y elaboración del vino* (4.ª ed.). Ediciones Mundi-Prensa. ISBN 978-84-8476-160-0.

Blouin, J., Papet, N., & Stonestreet, E. (2000). Study of the polyphenolic structure of red wines by physical-chemical and sensory analyses. *Journal International des Sciences de la Vigne et du Vin*, 34(1), 33–40.

Cadot, Y., Caillé, S., Samson, A., Barbeau, G., & Cheynier, V. (2010). Sensory dimension of wine typicality related to a terroir by Quantitative Descriptive Analysis, Just About Right analysis and typicality assessment. *Analytica Chimica Acta*, 660(1-2), 53–62. <https://doi.org/10.1016/j.aca.2009.10.006>

Caissie, A. F., Riquier, L., de Revel, G., & Tempere, S. (2023). Sensory dominances depend on the wine quality dimension. *Food Quality and Preference*, 112, 104998. <https://doi.org/10.1016/j.foodqual.2023.104998>

Carvajal Chávez, C. A. (2023). Uso de técnicas como la regresión y redes neuronales para anticipar el rendimiento del maíz. *RECIMUNDO*, 8(4), 126–135. [https://doi.org/10.26820/recimundo/8.\(4\).diciembre.2024.126-135](https://doi.org/10.26820/recimundo/8.(4).diciembre.2024.126-135)

Castro-López, L. R., Velasco-Aulcy, L., Chávez-Márquez, A., de Lira-García, C., Mata-Miranda, C., & Castillo, G. (2024). Physicochemical characteristics of red wines (Cabernet Sauvignon and Tempranillo) from Ensenada, Baja California. *Agrociencia*, 58(3). <https://doi.org/10.47163/agrociencia.v58i3.3103>

Césari, M., Césari, R., Gámbaro, A., & Arnaud, E. (2013). Análisis de datos provenientes de pruebas sensoriales del vino, utilizando la lógica borrosa. Séptimo Encuentro de Investigadores y Docentes de Ingeniería (EnIDI 2013), Los Reyunos, San Rafael, Mendoza, Argentina.

Cevallos-Mendoza, M. J., Chompoy-Salazar, S. M., & Barre-Zambrano, R. (2022). Propiedades fisicoquímicas y sensoriales del vino del pseudofruto de marañón (*Anacardium occidentale*). *Revista Científica INGENIAR: Ingeniería, Tecnología e Investigación*, 5(9). <https://doi.org/10.46296/ig.v5i9.0043>

Challenger-Pérez, I., Díaz-Ricardo, Y., & Becerra-García, R. A. (2014). El lenguaje de programación Python. *Ciencias Holguín*, 20(2), 1-13.

Cholassery, S., Krishna, V., Sethuraj, S. V., Rehina, S. S., Ranganathan, V., Dileep, L. C., Sabu, K. R., Rajesh, B. R., & Chandran, R. P. (2019). Analysis of physicochemical and sensory parameters of wine produced from Carica papaya. *Journal of Applied Biology & Biotechnology*, 7(5), 74–78. <https://doi.org/10.7324/JABB.2019.70512>

Christaki, T., & Tzia, C. (2002). Quality and safety assurance in winemaking. *Food Control*, 14(4), 289–297. [https://doi.org/10.1016/S0956-7135\(02\)00030-3](https://doi.org/10.1016/S0956-7135(02)00030-3)

Chvanova, M. S., & Bakalets, I. A. (2022). Neural network technology for quality control of the milk packaging process. En *Proceedings of the 2022 6th Scientific School Dynamics of Complex Networks and their Applications (DCNA)* (pp. 73–74). IEEE. <https://doi.org/10.1109/DCNA56428.2022.9923283>

Claypo, N., Jaiyen, S., & Hanskunatai, A. (2023). Inspection system for glass bottle defect classification based on deep neural network. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(7), 339–346. <https://doi.org/10.14569/IJACSA.2023.0140739>

Cliff, M. A., King, M. C., & Schlosser, J. (2007). Anthocyanin, phenolic composition, colour measurement and sensory analysis of BC commercial red wines. *Food Research International*, 40(1), 92–100. <https://doi.org/10.1016/j.foodres.2006.08.002>

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. <https://doi.org/10.1016/j.dss.2009.05.016>

Cruz-de Aquino, M. A. de la, Martínez-Peniche, R. A., Becerril-Román, A. E., & Chávaro-Ortiz, M. del S. (2012). Caracterización física y química de vinos tintos producidos en Querétaro. *Revista Fitotecnia Mexicana*, 35(Especial 5), 61–67. <https://doi.org/10.35196/rfm.2012.especial.5.61>

Cubero Ledezma, H. (2020). Python: Concepto, instalación y sus aplicaciones. Recuperado de <https://repositorio.uned.ac.cr/handle/120809/2151>

Dahal, K. R., Dahal, J. N., Banjade, H., & Gaire, S. (2021). Prediction of Wine Quality Using Machine Learning Algorithms. *Open Journal of Statistics*, 11(2), 278–289. <https://doi.org/10.4236/ojs.2021.112015>

Dennis, E. G., Keyzers, R. A., Kalua, C. M., Maffei, S. M., Nicholson, E. L., & Boss, P. K. (2012). Grape contribution to wine aroma: Production of hexyl acetate, octyl acetate, and benzyl acetate during yeast fermentation is dependent upon precursors in the must. *Journal of Agricultural and Food Chemistry*, 60(10), 2638–2646. <https://doi.org/10.1021/jf2045849>

Ding, H., Xie, Z., Yu, W., Cui, X., & Wilson, D. I. (2025). Artificial intelligence enhances food testing process: A comprehensive review. *Food Bioscience*, 68, 106404. <https://doi.org/10.1016/j.fbio.2025.106404>

Dzedze, N., Van Breda, V., Hart, R. S., & Van Wyk, J. (2019). Wine chemical, sensory, aroma compound and protein analysis of wines produced from chemical and biological fungicide treated Chenin blanc grapes. *Food Control*, 105, 265–276. <https://doi.org/10.1016/j.foodcont.2019.06.007>

Ebeler, S. E. (1999). Linking flavor chemistry to sensory analysis of wine. In T. Teranishi, et al. (Eds.), *Flavor Chemistry: 30 Years of Progress* (pp. 410–421). Kluwer Academic/Plenum Publishers.

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*, 4(Special Issue), 23–26.

Espíritu Quiroz, L., & Díaz Ortega, J. L. (2019). Análisis de calidad de vinos artesanales procedentes de Cascas. *Cientifi-k*, 7(2), 11–17. <https://doi.org/10.18050/cientifi-k.v7i2.815>

Esti, M., González Airola, R. L., Moneta, E., Paperaio, M., & Sinesio, F. (2010). Qualitative data analysis for an exploratory sensory study of Grechetto wine. *Analytica Chimica Acta*, 660(1-2), 63–67. <https://doi.org/10.1016/j.aca.2009.09.031>

Fandiño, M., Vilanova, M., Caldeira, I., Silvestre, J. M., Rey, B. J., Mirás-Avalos, J. M., & Cancela, J. J. (2020). Chemical composition and sensory properties of Albariño wine: Fertigation effects. *Food Research International*, 137, 109533. <https://doi.org/10.1016/j.foodres.2020.109533>

Farhangi, O., Sheidaee, E., & Kisalaei, A. (2024). Machine vision for detecting defects in liquid bottles: An industrial application for food and packaging sector. *Computers, Communications, and Distributed Systems*, 5(2), 1-15. <https://doi.org/10.37256/ccds.5220244756>

Fernández-Cuesta, J. (2014, marzo 30). Los secretos de los concursos de vinos: cómo se cata, cómo se vota. ABC. <https://www.abc.es/viajar/restaurantes/20140330/abci-concurso-vinos-201403261257.html>

Feng, Z., Martínez-Lapuente, L., Ayestarán, B., & Guadalupe, Z. (2023). Volatile and sensory characterization of Tempranillo wines aged in Quercus alba oak barrels of different geographical origins in USA. *LWT - Food Science and Technology*, 173, 114328. <https://doi.org/10.1016/j.lwt.2022.114328>

Franco Pérez, A. F., Naranjo Armijo, F. G., & Moreira Rosales, L. V. (2021). Análisis educativo sobre la valoración sensorial en catación de vinos. *Revista Conrado*, 17(78), 178–182.

Fuentes-Verduzco, C., Lugo-García, G. A., Pérez-Leal, R., & Camacho-Inzunza, F. A. (2022). Calidad de vinos de la variedad tempranillo, cultivada en tres viñedos en Chihuahua, México. *Estudios Sociales. Revista de Alimentación Contemporánea y Desarrollo Regional*, 32(59), 1–13. <https://doi.org/10.24836/es.v32i59.1195>.

Galeano-Arias, L. F., Aguirre, S. G., & Castrillón-Gómez, O. D. (2021). Análisis de calidad del vino por medio de técnicas de inteligencia artificial. *Información Tecnológica*, 32(1), 17–26. <https://doi.org/10.4067/S0718-07642021000100017>

García Zapateiro, L. A., Florez Mendoza, C. I., & Marrugo Ligardo, Y. (2016). Elaboración y caracterización fisicoquímica de un vino joven de fruta de borjón (*Bactris patinoi* Cuatrec). *Universidad Nacional de Entre Ríos*, 27(52), 507-519.

Geetha, C., Arunsundar, B., Vasumathi, G., Thamizhazhakan, K., & Venkata Sudhakar, C. (2024). Decoding AI: Experimental analysis of artificial intelligence based wine quality prediction logic using convoluted deep classification strategy. In *Proceedings of the 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. xxx-xxx). IEEE. <https://doi.org/10.1109/ICONSTEM60960.2024.10568783>

Gómez Benítez, J., Lasanta Melero, C., & Cobo Heredia, M. A. (2014). Aprendizaje a distancia del análisis sensorial de vinos. 37th World Congress of Vine and Wine and 12th General Assembly of the OIV, 07005. <https://doi.org/10.1051/oivconf/201407005>

Gómez García-Carpintero, E., Sánchez-Palomo, E., & González Viñas, M. A. (2010). Influence of co-winemaking technique in sensory characteristics of new Spanish red wines. *Food Quality and Preference*, 21(6), 705–710. <https://doi.org/10.1016/j.foodqual.2010.05.013>

Gomis-Bellmunt, A., Claret, A., Guerrero, L., & Pérez-Elortondo, F. J. (2024). Sensory evaluation of Protected Designation of origin Wines: Development of olfactive descriptive profile and references. *Food Research International*, 176, 113828. <https://doi.org/10.1016/j.foodres.2023.113828>

Gonzaga, L. S., Capone, D. L., Bastian, S. E. P., Danner, L., & Jeffery, D. W. (2020). Sensory typicity of regional Australian Cabernet Sauvignon wines according to expert evaluations and descriptive analysis. *Food Research International*, 138, 109760. <https://doi.org/10.1016/j.foodres.2020.109760>

González, M. A., & Rodríguez, J. (2018). Metodología de una revisión sistemática. *Actas Urológicas Españolas*, 42(8), 499–506. <https://doi.org/10.1016/j.acuro.2018.01.010>

Granados, C., Torrenegra, M. E., Acevedo, D., & Romero, P. (2013). Evaluación fisicoquímica y microbiológica del aperitivo vínico de lulo (*Solanum quitoense* L.). *Información Tecnológica*, 24(6), 35-40. <https://doi.org/10.4067/S0718-07642013000600006>

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.2017.12.042>

Hao, S., Wang, Z., Ji, S., Yuan, L., Xu, H., & Yue, X. (2025). Impact of ripeness stages on the chemical and sensory profiles of jujube wine: An analysis of physicochemical parameters, antioxidant activity, and volatile compounds. *Food Chemistry*, 477, 143394. <https://doi.org/10.1016/j.foodchem.2025.143394>

Hao, Y., Li, J., Zhao, Z., Xu, W., Wang, L., Lin, X., Hu, X., & Li, C. (2024). Flavor characteristics of Shanlan rice wines fermented for different time based on HS-SPME-GC-MS-O, HS-GC-IMS, and electronic sensory analyses. *Food Chemistry*, 432, 137150. <https://doi.org/10.1016/j.foodchem.2023.137150>

Hosnedlova, B., Sochor, J., Baron, M., Bjørklund, G., & Kizek, R. (2019). Application of nanotechnology based-biosensors in analysis of wine compounds and control of wine quality and safety: A critical review. *Critical Reviews in Food Science and Nutrition*. <https://doi.org/10.1080/10408398.2019.1682965>

Iorizzo, M., Macciola, V., Testa, B., Lombardi, S. J., & De Leonardis, A. (2014). Physicochemical and sensory characteristics of red wines from the rediscovered autochthonous *Tintilia* grapevine grown in the Molise region (Italy). *European Food Research and Technology*, 238(6), 1037–1048. <https://doi.org/10.1007/s00217-014-2186-z>

Izquierdo-Bueno, I., Moraga, J., Cantoral, J. M., Carbú, M., Garrido, C., & González-Rodríguez, V. E. (2024). Smart viticulture: Applying artificial intelligence for improved winemaking and risk management. *Applied Sciences*, 14(22), 10277. <https://doi.org/10.3390/app142210277>

Jain, K., Kaushik, K., Gupta, S. K., Mahajan, S., & Kadry, S. (2023). Machine learning based predictive modelling for the enhancement of wine quality. *Scientific Reports*, 13, 17042. <https://doi.org/10.1038/s41598-023-44111-9>

Khalafyan, A. A., Temerdashev, Z. A., Akin'shina, V. A., & Yakuba, Y. F. (2021). Study of consistency of expert evaluations of wine sensory characteristics by positional analysis. *Heliyon*, 7(e06162). <https://doi.org/10.1016/j.heliyon.2021.e06162>

Khalafyan, A. A., Temerdashev, Z. A., Yakuba, Y. F., & Guguchkina, T. I. (2019). Computer analysis of the sensory qualities of red wines as a method to optimize their blend formulation. *Heliyon*, 5(9), e01602. <https://doi.org/10.1016/j.heliyon.2019.e01602>

Kumar, Y., Cassano, A., Conidi, C., Gottardi, D., Ricci, A., Parpinello, G. P., & Versari, A. (2025). Evaluation of physicochemical characteristics, color and volatile profile of low alcohol beverage based on concentrated white wine produced by NF and RO membranes. *Separation and Purification Technology*, 367, 132847. <https://doi.org/10.1016/j.seppur.2025.132847>

Kurtanjek, Ž. (2023). Wine quality analysis by the structural causal model (SCM). *Croatian Journal of Food Science and Technology*, 15(2), 173-184. <https://doi.org/10.17508/CJFST.2023.15.2.05>

Lesschaeve, I., & Noble, A. C. (2022). Sensory analysis of wine. En A. G. Reynolds (Ed.), *Managing wine quality* (pp. 243–277). Woodhead Publishing. <https://doi.org/10.1016/B978-0-12-824378-2.00013-6>

Li, C., Pu, X., Ye, P., Fu, Q., Zhou, Y., Zhou, X., Shi, X., & Wang, B. (2025). Flavor characterization of six white wines from Xinjiang: Physicochemical property, antioxidant activity, and volatile compounds analysis. *Journal of Agriculture and Food Research*, 20, 101725. <https://doi.org/10.1016/j.jafr.2025.101725>

Lund, S. T., & Bohlmann, J. (2006). The molecular basis for wine grape quality—a volatile subject. *Science*, 311(5762), 804–805. <https://doi.org/10.1126/science.1121734>

Mahima, Gupta, U., Patidar, Y., Agarwal, A., & Singh, K. P. (2020). Wine Quality Analysis Using Machine Learning Algorithms. In D. K. Sharma et al. (Eds.), *Micro-Electronics and Telecommunication Engineering* (pp. 11–18). Lecture Notes in Networks and Systems, 106. https://doi.org/10.1007/978-981-15-2329-8_2

Makhmudov, F., Ravshanov, N., Akhmedov, D., Pekos, O., Turimov, D., & Cho, Y.-I. (2025). A Multitask Deep Learning Model for Predicting Myocardial Infarction Complications. *Bioengineering*, 12(5), 520. <https://doi.org/10.3390/bioengineering12050520>

Mani, S., Krishnankutty, R. A., Swaminathan, S., & Theerthagiri, P. (2023). An investigation of wine quality testing using machine learning techniques. *International Journal of Artificial Intelligence (IJ-AI)*, 12(2), 747–754. <https://doi.org/10.11591/ijai.v12.i2.pp747-754>

Manso-Martínez, C., Sáenz-Navajas, M. P., Hernández, M. M., & Menéndez, C. M. (2020). Sensory profiling and quality assessment of wines derived from Graciano × Tempranillo selections. *LWT - Food Science and Technology*, 127, 109394. <https://doi.org/10.1016/j.lwt.2020.109394>

Manzano, M. H. (2015). Determinación del origen geográfico del vino variedad Malbec en Mendoza y su correlación con provincias vitivinícolas del país a través del análisis físico-químico y estadístico (Tesis doctoral). Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales. <https://digital.cic.gba.gob.ar/handle/11746/5307>

Martínez de Aparicio, M. (2005). *El vino de la A a la Z*. Santiago, Chile: PlanetaVino.

Martínez-Martínez, V., Nevares, I., & del Alamo-Sanza, M. (2020). Artificial intelligence methods for constructing wine barrels with a controlled oxygen transmission rate. *Molecules*, 25(14), 3312. <https://doi.org/10.3390/molecules25143312>

Meléndez, E., Íñiguez, M., Puras, P., Ortiz, M. C., & Sarabia, L. A. (2010). Las uvas también se catan. *Cuaderno de Campo*, (45), 34–39.

Méndez Medina, D. L. (2016). Entre intenciones y limitantes: la industria vitivinícola en Baja California (1935-1943). *Signos Históricos*, 18(36), 148–179. Universidad Autónoma Metropolitana, Unidad Iztapalapa. Recuperado de <https://signoshistoricos.izt.uam.mx/index.php/historicos/article/view/488>

Nanou, E., Metafa, M., Bastian, S. E. P., Tempère, S., & Kotseridis, Y. (2025). Aroma profiling of Greek red wines: A comparative sensory and chemical analysis of Agiorgitiko and Xinomavro varieties. *Journal of Food Composition and Analysis*, 144, 107694. <https://doi.org/10.1016/j.jfca.2025.107694>

Nguyen, T. H., & Durner, D. (2023). Sensory evaluation of wine aroma: Should color-driven descriptors be used? *Food Quality and Preference*, 107, 104844. <https://doi.org/10.1016/j.foodqual.2023.104844>

NMX-F-274-1987. (1987). Determinación del grado Brix en muestras de meladura; masas cocidas; mieles "a" y "b" de refinería y miel final. Por método hidrométrico. Dirección General de Normas, Secretaría de Economía, México.

Noble, A. C. (1988). Analysis of wine sensory properties. En H.-F. Linskens & J. F. Jackson (Eds.), *Wine Analysis* (pp. 315–328). Springer-Verlag. https://doi.org/10.1007/978-3-642-72634-4_17

Nogueira, J. M. F., & Nascimento, A. M. D. (1999). Analytical characterization of Madeira wine. *Journal of Agricultural and Food Chemistry*, 47(2), 566–575. <https://doi.org/10.1021/jf9804490>

Oczkowski, E. (2016). The effect of weather on wine quality and prices: An Australian spatial analysis. *Journal of Wine Economics*, 11(1), 48–65. <https://doi.org/10.1017/jwe.2015.14>

Organisation Internationale de la Vigne et du Vin (OIV). (2023). State of the world vine and wine sector in 2023. <https://www.oiv.int/public/medias/9211/oiv-report-state-of-the-world-vine-and-wine-sector-2023.pdf>

Organización Internacional de la Viña y el Vino (OIV). (2015). Documento de revisión del análisis sensorial del vino. Dijon, Francia: OIV. Recuperado de <https://www.oiv.int/public/medias/3307/review-on-sensory-analysis-of-wine.pdf>

Organización Internacional de la Viña y el Vino (OIV). (2023). FOCUS OIV 2023: Evolution of World Wine Production and Consumption by Colour. Recuperado de https://www.oiv.int/sites/default/files/documents/OIV-FOCUS-2023_Evolution_of_the_world_wine_production_and_consumption_by_colour_4.pdf

Ortega-Heras, M., Curiel-Fernández, M., Pérez-Magariño, S., & González-Sanjosé, M. L. (2024). Generation of wine sensory descriptors by one-step method based on free profile and their use to establish the typicity of PDO wines. *LWT - Food Science and Technology*, 203, 116410. <https://doi.org/10.1016/j.lwt.2024.116410>

Pájaro-Escobar, H. A., Benedetti, J., & García-Zapateiro, L. A. (2018). Caracterización fisicoquímica y microbiológica de un vino de frutas a base de tamarindo (*Tamarindus indica* L.) y carambola (*Averrhoa carambola* L.). *Información Tecnológica*, 29(5), 123–130. <https://doi.org/10.4067/S0718-07642018000500123>

Pearson, W., Schmidtke, L., Francis, I. L., & Blackman, J. W. (2020). An investigation of the Pivot© Profile sensory analysis method using wine experts: Comparison with descriptive analysis and results from two expert panels. *Food Quality and Preference*, 83, 103858. <https://doi.org/10.1016/j.foodqual.2019.103858>

Petropoulos, S., Karavas, C. S., Balafoutis, A. T., Paraskevopoulos, I., Kallithraka, S., & Kotseridis, Y. (2017). Fuzzy logic tool for wine quality classification. *Computers and Electronics in Agriculture*, 142, 552–562. <https://doi.org/10.1016/j.compag.2017.11.015>

Pletcher, D., & Wills, R. (2005). A novel flow battery—A lead acid battery based on an electrolyte with soluble lead(II). Part III: The influence of conditions on battery performance. *Journal of Power Sources*, 149, 96–102. <https://doi.org/10.1016/j.jpowsour.2005.01.048>

Ríos-Reina, R., Azcarate, S. M., Camiña, J. M., & Callejón, R. M. (2020). Sensory and spectroscopic characterization of Argentinean wine and balsamic vinegars: A comparative study with European vinegars. *Food Chemistry*, 323, 126791. <https://doi.org/10.1016/j.foodchem.2020.126791>

Rivera, E., & Peña, L. (2022). Trazabilidad en la cadena de suministro de alimentos: un modelo teórico para la mejora de la seguridad alimentaria. *Revista Gente Clave*, 6(1), 55–70. <https://revistas.ulatina.edu.pa/index.php/genteclave/article/view/129>

Rochfort, S., Ezernieks, V., Bastian, S. E. P., & Downey, M. O. (2010). Sensory attributes of wine influenced by variety and berry shading discriminated by NMR

metabolomics. *Food Chemistry*, 121(4), 1296–1304.
<https://doi.org/10.1016/j.foodchem.2010.01.067>

Rousseau, J., & Delteil, D. (2000). Présentation d'une méthode d'analyse sensorielle des baies de raisin: Principe, méthode, interprétation. *Revue Française d'Oenologie*, (183), 10–13.

Ruiz Ríos, R. E. (2011). Tierra y vid: Bodegas de Santo Tomás y los inicios de la vitivinicultura en el Valle de Guadalupe, Baja California. En L. C. León Velazco (Ed.), *Estudios históricos sobre Baja California* (pp. 139–186). Consejo Nacional para la Cultura y las Artes.

Sánchez Zepeda, L., & Mungaray Lagarda, A. (2010). Vino de calidad: Base de desarrollo endógeno en el Valle de Guadalupe, Baja California. *Frontera Norte*, 22(44), 109–132. <https://doi.org/10.17428/rfn.v22i44.852>

Saranraj, P., Sivasakthivelan, P., & Naveen, M. (2017). Fermentation of fruit wine and its quality analysis: A review. *Australian Journal of Science and Technology*, 1(2), [sin número de páginas]. ISSN 2208-6404.

Schober, D., Gilmore, A., Chen, L., Zincker, J., & Gonzalez, A. (2022). Determination of Cabernet Sauvignon wine quality parameters in Chile by Absorbance-Transmission and fluorescence Excitation Emission Matrix (A-TEEM) spectroscopy. *Food Chemistry*, 392, 133101.
<https://doi.org/10.1016/j.foodchem.2022.133101>

Secretaría de Salud & Secretaría de Economía. (2014). Bebidas alcohólicas. Especificaciones sanitarias. Etiquetado sanitario y comercial (NOM-142-SSA1/SCFI-2014). *Diario Oficial de la Federación*.
https://www.dof.gob.mx/nota_detalle.php?codigo=5372591&fecha=23/03/2015

Solanki, A. V. (2014). Data mining techniques using WEKA classification for sickle cell disease. *International Journal of Computer Science and Information Technologies*, 5(4), 5857–5860.
<http://www.ijcsit.com/docs/Volume%205/vol5issue04/ijcsit20140504222.pdf>

Tang, A., & Peng, B. (2024). Uncovering the flavor differences between black rice wine and glutinous rice wine by GC-MS, GC-IMS, HPLC, and electronic sensory analysis. *Food Bioscience*, 60, 104235. <https://doi.org/10.1016/j.fbio.2024.104235>

Tapsoba, F., Savadogo, A., Somda, K. M., Zongo, C., Barro, N., & Traoré, S. A. (2011). Biodiversité microbienne et paramètres physicochimiques de quelques vins de ronier (*Borassus akeassii*) produits traditionnellement au Burkina Faso. *Revue de Microbiologie Industrielle, Sanitaire et Environnementale*, 5(2), 1-22.

- Tatu, S. I., Bansod, S., Singh, S. K., & Pal, P. (2023). Fault detection in bottle caps and label alignment using convolutional neural network. En Proceedings of the 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) (p. 10118182). IEEE. <https://doi.org/10.1109/ICAECT57570.2023.10118182>
- Tesfaye, W., Morales, M. L., Callejón, R. M., Cerezo, A. B., González, A. G., García-Parrilla, M. C., & Troncoso, A. M. (2009). Descriptive sensory analysis of wine vinegar: Tasting procedure and reliability of new attributes. *Journal of Sensory Studies*, <https://doi.org/10.1111/j.1745-459X.2009.00253.x>
- Tian, H., Xiong, J., Sun, J., Du, F., Xu, G., Yu, H., Chen, C., & Lou, X. (2024). Dynamic transformation in flavor during hawthorn wine fermentation: Sensory properties and profiles of nonvolatile and volatile aroma compounds coupled with multivariate analysis. *Food Chemistry*, 456, 139982. <https://doi.org/10.1016/j.foodchem.2024.139982>
- Tong, W., Zhai, H., Qi, M., Hua, Y., Shi, T., Shang, H., Shi, Y., Duan, C., & Lan, Y. (2024). Characterization of chemical and sensory properties of Cabernet Sauvignon and Marselan wines made by flash d'etente technique. *Food Research International*, 184, 114229. <https://doi.org/10.1016/j.foodres.2024.114229>
- Tong, W., Zhai, H., Qi, M., Hua, Y., Shi, T., Shang, H., Shi, Y., Duan, C., & Lan, Y. (2024). Characterization of chemical and sensory properties of Cabernet Sauvignon and Marselan wines made by flash détente technique. *Food Research International*, 184, 114229. <https://doi.org/10.1016/j.foodres.2024.114229>
- Ünlü, R. (2020). The role of machine learning in productivity: A case study of wine quality prediction. *European Journal of Science and Technology*, (20), 280–286. <https://doi.org/10.31590/ejosat.773736>
- Ureña, M., Chanut, J., Bottreau, V., Bellat, J.-P., Gougeon, R. D., Lagorce, A., & Karbowski, T. (2024). Wine bottle overcapping wax: An aesthetic or functional element? *Food Packaging and Shelf Life*, 46, 101367. <https://doi.org/10.1016/j.fpsl.2024.101367>
- Valls, J., Lampreave, M., Nadal, M., & Arola, L. (2000). Importancia de los compuestos fenólicos en la calidad de los vinos tintos de crianza. *Alimentación, Equipos y Tecnología*, 19(2), 119–124.
- Vannier, A., Brun, O. X., & Feinberg, M. H. (1999). Application of sensory analysis to champagne wine characterisation and discrimination. *Food Quality and Preference*, 10(2), 101–107. [https://doi.org/10.1016/S0950-3293\(98\)00047-0](https://doi.org/10.1016/S0950-3293(98)00047-0)

Vannier, A., Brun, O. X., & Feinberg, M. H. (1999). Application of sensory analysis to champagne wine characterisation and discrimination. *Food Quality and Preference*, 10(2), 101–107. [https://doi.org/10.1016/S0950-3293\(98\)00047-0](https://doi.org/10.1016/S0950-3293(98)00047-0)

Venturi, F., Sanmartin, C., Taglieri, I., Xiaoguo, Y., Quartacci, M. F., Sgherri, C., Andrich, G., & Zinnai, A. (2017). A kinetic approach to describe the time evolution of red wine as a function of packaging conditions adopted: Influence of closure and storage position. *Food Packaging and Shelf Life*, 13, 44–48. <https://doi.org/10.1016/j.fpsl.2017.07.001>

Wang, A., Zhu, Y., Zou, L., Zhu, H., Cao, R., & Zhao, G. (2022). Combination of machine learning and intelligent sensors in real-time quality control of alcoholic beverages. *Food Science and Technology*, 42, e54622. <https://doi.org/10.1590/fst.54622>

Wang, Y., Zhou, B., Zhang, H., & Ge, J. (2011). A vision-based intelligent inspector for wine production. *Machine Vision and Applications*, 23(6), 1033–1045. <https://doi.org/10.1007/s00138-011-0341-0>

Wu, Q., Li, H., Tian, C., Wen, L., & Li, X. (2024). AEKD: Unsupervised auto-encoder knowledge distillation for industrial anomaly detection. *Journal of Manufacturing Systems*, 73, 159–169. <https://doi.org/10.1016/j.jmsy.2024.02.001>

Yadav, A. K., Malik, H., & Chandel, S. S. (2014). Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. *Renewable and Sustainable Energy Reviews*, 31, 509–519. <https://doi.org/10.1016/j.rser.2013.12.008>

Yang, Y., Ye, Z., Araujo, L. D., Rutan, T., Deed, R. C., & Kilmartin, P. A. (2025). A dataset on the sensory analyses of 116 New Zealand Pinot Noir wines from Central Otago, Marlborough, and Martinborough. *Data in Brief*, 60, 111638. <https://doi.org/10.1016/j.dib.2025.111638>

Zeng, Y., Liu, Y., Wu, L., Dong, H., Zhang, Y., Guo, H., Guo, Z., Wang, S., & Lan, Y. (2019). Evaluation and analysis model of wine quality based on mathematical model. *Studies in Engineering and Technology*, 6(1), [artículo n.º 3626]. <https://doi.org/10.11114/set.v6i1.3626>

Zhao, Y., Zeng, Y., Li, X., Yuan, K., Li, Y., Tian, L., Sun, J., & Bai, W. (2023). Modeling and application of sensory evaluation of blueberry wine based on principal component analysis. *Current Research in Food Science*, 6, 100403. <https://doi.org/10.1016/j.crfs.2022.11.022>

Zoecklein, B. W., Fugelsang, K. C., Gump, B. H., & Nury, F. S. (2001). *Análisis y producción de vino*. Zaragoza: Editorial Acribia.

7 ANEXOS

ANEXO 1. Códigos generados en Python

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

file_path = r"C:\Users\PCENSV16\Desktop\Briseida\BASE DE DATOS CALIDAD.xlsx"
data = pd.read_excel(file_path)

calidad_cols = ['Calidad Premium', 'Calidad Varietal', 'Calidad Básica']
if not all(col in data.columns for col in calidad_cols):
    raise KeyError(f"Las columnas necesarias no están en los datos: {calidad_cols}")

data[calidad_cols] = data[calidad_cols].replace({'SI': 1, 'NO': 0})

data['calidad'] = (
    data['Calidad Premium'] * 3 +
    data['Calidad Varietal'] * 2 +
    data['Calidad Básica']
)

fisicoquimicas = ['°Brix', 'pH', 'AT', 'SO2L', 'SO2m', 'NFA', 'Azúcar', 'Alcohol', 'AV', 'AM']
correlation_matrix = data[fisicoquimicas + ['calidad']].corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)
plt.title("Mapa de Calor de Correlación entre Las Variables Fisicoquímicas y Calidad")
plt.show()

correlation_with_calidad = correlation_matrix['calidad'].drop('calidad')
max_correlation_feature = correlation_with_calidad.idxmax()
print(f"La variable con mayor correlación con 'calidad' es: {max_correlation_feature}")

correlation_table = correlation_with_calidad.reset_index()
correlation_table.columns = ['Variable', 'Correlación con Calidad']
correlation_table = correlation_table.sort_values(by='Correlación con Calidad', ascending=False)

print("\nTabla de Correlación de Las Variables Fisicoquímicas con La Calidad:")
print(correlation_table)
```

Figura 11. Código de Python empleado para la obtención de la correlación de las variables fisicoquímicas.

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.metrics import classification_report, accuracy_score

file_path = r"C:\Users\PCENSV16\Desktop\Brisetida\BASE DE DATOS CALIDAD.xlsx"
data = pd.read_excel(file_path)

calidad_cols = ['calidad_premium', 'calidad_varietal', 'calidad_básica']
data[calidad_cols] = data[calidad_cols].replace({'SI': 1, 'NO': 0})

data['calidad'] = (
    data['calidad_premium'] * 3 +
    data['calidad_varietal'] * 2 +
    data['calidad_básica']
)

fisicoquimicas = ['brlx', 'ph', 'at', 'so2L', 'so2m', 'nfa', 'azúcar', 'alcohol']
correlation_matrix = data[fisicoquimicas].corr()

X = data[fisicoquimicas]
y = data['calidad']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

dt_model = DecisionTreeClassifier(
    random_state=42,
    criterion='entropy', # Usar 'entropy' para mejorar las divisiones
    class_weight='balanced' # Balancear las clases en caso de desbalance
)

param_grid = {
    'max_depth': [4, 6, 8, 10, None], # Limitar la profundidad
    'min_samples_split': [2, 5, 10], # Mínimo de muestras para dividir un nodo
    'min_samples_leaf': [1, 2, 4], # Mínimo de muestras por hoja
}

grid_search = GridSearchCV(estimator=dt_model, param_grid=param_grid, cv=5, scoring='accuracy', n_jobs=-1)
grid_search.fit(X_train, y_train)

best_dt_model = grid_search.best_estimator_

y_pred_dt = best_dt_model.predict(X_test)
print("\nReporte de clasificación (Árbol de Decisión):")
print(classification_report(y_test, y_pred_dt))

accuracy_dt = accuracy_score(y_test, y_pred_dt)
print(f"Precisión del modelo Árbol de Decisión: {accuracy_dt:.4f}")

plt.figure(figsize=(20, 10))
plot_tree(best_dt_model, feature_names=fisicoquimicas, class_names=[str(c) for c in sorted(data['calidad'].unique())], filled=True, fontsize=10)
plt.title("Árbol de Decisión Mejorado")
plt.show()

```

Figura 12. Código de Python empleado para la obtención del árbol de decisión a partir de variables fisicoquímicas.

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report

file_path = r'C:\Users\PCENS\16\Desktop\Briseida\BASE DE DATOS CALIDAD.xlsx' # Ruta al archivo
data = pd.read_excel(file_path)
data['Calidad Premium'] = data['Calidad Premium'].apply(lambda x: 1 if x == 'Si' else 0)
data['Calidad Varietal'] = data['Calidad Varietal'].apply(lambda x: 1 if x == 'Si' else 0)
data['Calidad Básica'] = data['Calidad Básica'].apply(lambda x: 1 if x == 'Si' else 0)
for column in categorical_columns:
    if column in data.columns:
        data[column] = label_encoder.fit_transform(data[column])

X = data[numeric_columns] # Características numéricas
y = data[['Calidad Premium', 'Calidad Varietal', 'Calidad Básica']] # Etiquetas de calidad

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

nn_model = Sequential([
    Dense(64, activation='relu', input_dim=X_train_scaled.shape[1]), # Capa de entrada
    Dense(32, activation='relu'), # Capa oculta
    Dense(16, activation='relu'), # Capa oculta
    Dense(3, activation='softmax') # Capa de salida (softmax para clasificación multiclase)
])

nn_model.compile(optimizer=Adam(), loss='categorical_crossentropy', metrics=['accuracy'])
y_train_encoded = pd.get_dummies(y_train)
y_test_encoded = pd.get_dummies(y_test)

history = nn_model.fit(X_train_scaled, y_train_encoded, epochs=50, batch_size=32, validation_split=0.2, verbose=1)
y_pred_nn = nn_model.predict(X_test_scaled)
y_pred_classes = y_pred_nn.argmax(axis=1)
y_test_classes = y_test_encoded.values.argmax(axis=1)

print("Reporte de clasificación (Red Neuronal Multiclase):")
print(classification_report(y_test_classes, y_pred_classes))

plt.figure(figsize=(12, 6))

plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'], label='Precisión de entrenamiento')
plt.plot(history.history['val_accuracy'], label='Precisión de validación')
plt.title('Precisión durante el entrenamiento')
plt.xlabel('Épocas')
plt.ylabel('Precisión')
plt.legend()

plt.subplot(1, 2, 2)
plt.plot(history.history['Loss'], label='Pérdida de entrenamiento')
plt.plot(history.history['val_Loss'], label='Pérdida de validación')
plt.title('Pérdida durante el entrenamiento')
plt.xlabel('Épocas')
plt.ylabel('Pérdida')
plt.legend()
plt.show()

```

Figura 13. Código de Python empleado para el modelo de predicción de redes neuronales.

```

dt_model = DecisionTreeClassifier(random_state=42, criterion='entropy')

param_grid = {
    'max_depth': [3, 5, 10, 20, None],
    'min_samples_split': [2, 5, 10, 20],
    'min_samples_leaf': [1, 2, 5, 10],
    'class_weight': ['balanced', None]
}

grid_search = GridSearchCV(estimator=dt_model, param_grid=param_grid, cv=5, scoring='accuracy', n_jobs=-1)
grid_search.fit(X_train, y_train)

best_dt_model = grid_search.best_estimator_

y_pred_dt = best_dt_model.predict(X_test)

print("\nReporte de Clasificación (Árbol de Decisión):")
print(classification_report(y_test, y_pred_dt))

accuracy_dt = accuracy_score(y_test, y_pred_dt)
f1_dt = f1_score(y_test, y_pred_dt, average='weighted')
print(f"Precisión del modelo Árbol de Decisión: {accuracy_dt:.4f}")
print(f"F1-score del modelo Árbol de Decisión: {f1_dt:.4f}")

conf_matrix = confusion_matrix(y_test, y_pred_dt)
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues', xticklabels=[str(i) for i in sorted(data['calidad'])])
plt.title('Matriz de Confusión')
plt.xlabel('Clase Predicha')
plt.ylabel('Clase Real')
plt.show()

feature_importances = best_dt_model.feature_importances_
features = fisicoquimicas

plt.figure(figsize=(10, 6))
sns.barplot(x=features, y=feature_importances)
plt.title('Importancia de Las Características en el Árbol de Decisión')
plt.ylabel('Importancia')
plt.xlabel('Características')
plt.xticks(rotation=45)
plt.show()

plt.figure(figsize=(12, 8))
plot_tree(best_dt_model,
          feature_names=fisicoquimicas,
          class_names=[str(c) for c in sorted(data['calidad'].unique())],
          filled=True,
          fontsize=10,
          max_depth=3)
plt.title("Árbol de Decisión Mejorado con Validación Cruzada y Ajuste de Hiperparámetros")
plt.show()

```

Figura 14. Modificación en el código de Python del modelo árbol de decisión para aumentar la precisión del modelo.

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from collections import Counter

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
from sklearn.utils.class_weight import compute_class_weight

from imblearn.over_sampling import SMOTE
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Input, Dropout, BatchNormalization, LeakyReLU
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau
from tensorflow.keras.optimizers import Adam

file_path = r"C:\Users\PCENSV16\Desktop\Briseida\BASE DE DATOS CALIDAD.xlsx"
data = pd.read_excel(file_path)

calidad_cols = ['Calidad Premium', 'Calidad Varietal', 'Calidad Básica']
data[calidad_cols] = data[calidad_cols].replace({'SI': 1, 'NO': 0})

def map_calidad(row):
    if row['Calidad Premium'] == 1:
        return 6
    elif row['Calidad Varietal'] == 1:
        return 3
    elif row['Calidad Básica'] == 1:
        return 1
    else:
        return 0 # Sin calidad

y_raw = data.apply(map_calidad, axis=1)
class_map = {0: 0, 1: 1, 3: 2, 6: 3}
y = y_raw.map(class_map)

y.value_counts().sort_index().plot(kind='bar', title='Distribución de Clases')
plt.xlabel('Clase (0:Sin, 1:Básica, 2:Varietal, 3:Premium)')
plt.ylabel('Frecuencia')
plt.tight_layout()
plt.show()

X = data.drop(columns=calidad_cols + ['Variedad', 'Rancho', 'Año'])

```

Figura 15. Parte 1 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.

```

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42, stratify=y
)

counts = Counter(y_train)
clases_validas = [cl for cl, c in counts.items() if c >= 3]

if len(clases_validas) < len(set(y_train)):

mask = y_train.isin(clases_validas)
X_train = X_train[mask]
y_train = y_train[mask]

smote = SMOTE(random_state=42, k_neighbors=3)
X_train_smote, y_train_smote = smote.fit_resample(X_train, y_train)
print("Distribución tras SMOTE:", Counter(y_train_smote))

class_weights = compute_class_weight(class_weight='balanced', classes=np.unique(y_train_smote))
class_weights_dict = dict(enumerate(class_weights))
print("Class weights:", class_weights_dict)

model = Sequential([
    Input(shape=(X_train_smote.shape[1],)),
    Dense(512),
    LeakyReLU(alpha=0.1),
    BatchNormalization(),
    Dropout(0.3),

    Dense(256),
    LeakyReLU(alpha=0.1),
    BatchNormalization(),
    Dropout(0.25),

    Dense(128),
    LeakyReLU(alpha=0.1),
    BatchNormalization(),
    Dropout(0.2),

    Dense(64),
    LeakyReLU(alpha=0.1),
    BatchNormalization(),

```

Figura 16. Parte 2 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.

```

        Dense(4, activation='softmax')
    ])

    optimizer = Adam(learning_rate=0.0003)
    model.compile(optimizer=optimizer, loss='sparse_categorical_crossentropy', metrics=['accuracy'])

    early_stop = EarlyStopping(monitor='val_loss', patience=15, restore_best_weights=True, verbose=1)
    reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=5, verbose=1)

    history = model.fit(
        X_train_smote, y_train_smote,
        epochs=200,
        batch_size=16,
        validation_data=(X_test, y_test),
        callbacks=[early_stop, reduce_lr],
        class_weight=class_weights_dict,
        verbose=1
    )

    plt.figure(figsize=(12, 5))

    plt.subplot(1, 2, 1)
    plt.plot(history.history['accuracy'], label='Entrenamiento')
    plt.plot(history.history['val_accuracy'], label='Validación')
    plt.title('Precisión del Modelo')
    plt.xlabel('Época')
    plt.ylabel('Precisión')
    plt.legend()

    plt.subplot(1, 2, 2)
    plt.plot(history.history['loss'], label='Entrenamiento')
    plt.plot(history.history['val_loss'], label='Validación')
    plt.title('Pérdida del Modelo')
    plt.xlabel('Época')
    plt.ylabel('Pérdida')
    plt.legend()

    plt.tight_layout()
    plt.show()

    y_pred_probs = model.predict(X_test)
    y_pred = np.argmax(y_pred_probs, axis=1)

    print("\nReporte de Clasificación:")

```

Figura 17. Parte 3 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.

```

print(classification_report(y_test, y_pred, target_names=[
    'Sin calidad', 'Básica', 'Varietal', 'Premium'
]))

cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 5))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=['Sin', 'Básica', 'Varietal', 'Premium'],
            yticklabels=['Sin', 'Básica', 'Varietal', 'Premium'])
plt.xlabel('Predicho')
plt.ylabel('Real')
plt.title('Matriz de Confusión')
plt.tight_layout()
plt.show()

feature_names = X.columns
X_test_df = pd.DataFrame(scaler.inverse_transform(X_test), columns=feature_names)

def permutation_importance_keras(model, X, y, metric, n_repeats=10, random_state=42):
    np.random.seed(random_state)
    baseline_preds = np.argmax(model.predict(scaler.transform(X)), axis=1)
    baseline_score = metric(y, baseline_preds)
    importances = []

    for col in X.columns:
        scores = []
        for _ in range(n_repeats):
            X_permuted = X.copy()
            X_permuted[col] = np.random.permutation(X_permuted[col])
            preds = np.argmax(model.predict(scaler.transform(X_permuted)), axis=1)
            score = metric(y, preds)
            scores.append(baseline_score - score) # cuánto baja el score al permutar
        importances.append(np.mean(scores))

    return np.array(importances), baseline_score

importances, base_score = permutation_importance_keras(model, X_test_df, y_test, accuracy_score)

plt.figure(figsize=(10,6))
sns.barplot(x=importances, y=feature_names)
plt.title(f'Importancia de Variables por Permutación (Accuracy base={base_score:.3f})')
plt.xlabel('Pérdida en Accuracy al permutar')
plt.ylabel('Variable')
plt.tight_layout()

```

Figura 18. Parte 4 de la modificación en el código de Python del modelo redes neuronales para aumentar la precisión del modelo.