



**DIVISIÓN DE ESTUDIOS POSGRADO E INVESTIGACIÓN**

Opción 1.- TESIS  
**TRABAJO PROFESIONAL**

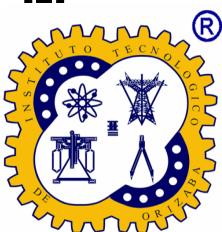
**“PREDICCIÓN DE LA CALIDAD DEL CAFÉ Y SUS PROPIEDADES  
SENSORIALES UTILIZANDO ANÁLISIS ESPECTROSCÓPICOS Y  
FÍSICOS DEL CAFÉ A TRAVÉS DE METODOLOGÍAS DE  
INTELIGENCIA ARTIFICIAL”**

**QUE PARA OBTENER EL GRADO DE:  
DOCTOR EN CIENCIAS  
DE LA INGENIERÍA**

**PRESENTA:**  
BLANCA ESTELA GONZÁLEZ SÁNCHEZ

**DIRECTOR DE TESIS:**  
OSCAR OSVALDO SANDOVAL GONZÁLEZ

**CODIRECTOR DE TESIS:**  
JOSÉ DE JESÚS AGUSTÍN FLORES CUAUTLE





Orizaba, Veracruz, 11/03/2024  
Dependencia: División de Estudios de  
Posgrado e Investigación  
Asunto: Autorización de Impresión  
OPCIÓN: I

C. BLANCA ESTELA GONZÁLEZ SÁNCHEZ  
Candidato a Grado de Doctor en:  
**CIENCIAS DE LA INGENIERÍA**  
**P R E S E N T E.**

De acuerdo con el Reglamento de Titulación vigente de los Centros e Institutos Tecnológicos Federales del Tecnológico Nacional de México, de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

**"PREDICCIÓN DE LA CALIDAD DEL CAFÉ Y SUS PROPIEDADES SENSORIALES UTILIZANDO ANÁLISIS ESPECTROSCÓPICOS Y FÍSICOS DEL CAFÉ A TRAVÉS DE METODOLOGÍAS DE INTELIGENCIA ARTIFICIAL"**

Comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

ATENTAMENTE  
Excelencia en Educación Tecnológica®  
CIENCIA - TÉCNICA - CULTURA®

DRA. OFELIA LANDETA ESCAMILLA  
JEFA DE LA DIVISIÓN DE ESTUDIOS  
DE POSGRADO E INVESTIGACIÓN





Orizaba, Veracruz, Marzo 06, de 2024.  
Asunto: Revisión de trabajo escrito

C. OFELIA LANDETA ESCAMILLA  
JEFA DE LA DIVISIÓN DE ESTUDIOS  
DE POSGRADO E INVESTIGACIÓN  
P R E S E N T E.-

Los que suscriben, miembros del jurado, han realizado la revisión de la Tesis del (la) C.

**BLANCA ESTELA GONZÁLEZ SÁNCHEZ**

La cual lleva el título de:

**"PREDICCIÓN DE LA CALIDAD DEL CAFÉ Y SUS PROPIEDADES SENSORIALES UTILIZANDO  
ANÁLISIS ESPECTROSCÓPICOS Y FÍSICOS DEL CAFÉ A TRAVÉS DE METODOLOGÍAS DE  
INTELIGENCIA ARTIFICIAL"**

Y concluyen que se acepta.

ATENTAMENTE  
Excelencia en Educación Tecnológica®  
CIENCIA - TÉCNICA - CULTURA®

PRESIDENTE: DR. OSCAR OSVALDO SANDOVAL GONZALEZ

FIRMA

SECRETARIO: DR. JOSÉ DE JESÚS AGUSTÍN FLORES  
CUAUTLE

FIRMA

VOCAL: DR. ALFONSO AGUILAR LASERRE

FIRMA

VOCAL: DR. EDUARDO ROLDAN REYES

FIRMA

VOCAL: DR. GUILLERMO CORTES ROBLES

FIRMA

VOCAL  
SUPLENTE: DR. OTNIEL PORTILLO RODRÍGUEZ.

FIRMA

TA-09 -18



## **Declaración de originalidad y cesión de derechos**

Orizaba, Veracruz, a. 11 de marzo del 2024.

La que suscribe

C. BLANCA ESTELA GONZÁLEZ SÁNCHEZ

Declaro que esta tesis, que tiene una extensión de 108 cuartillas, ha sido escrita por mí y constituye el registro escrito del trabajo de la tesis titulada

**"PREDICCIÓN DE LA CALIDAD DEL CAFÉ Y SUS PROPIEDADES  
SENSORIALES UTILIZANDO ANÁLISIS ESPECTROSCÓPICOS Y  
FÍSICOS DEL CAFÉ A TRAVÉS DE METODOLOGÍAS DE  
INTELIGENCIA ARTIFICIAL"**

del programa: Doctorado en Ciencias de la Ingeniería bajo la asesoría y dirección del Dr. Oscar Osvaldo Sandoval González y coasesor el Dr. José de Jesús Agustín Flores Cuautle y no ha sido sometida en ninguna otra institución previamente.

Todos los datos y las referencias a materiales ya publicados están debidamente identificados con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan como tal y, en los casos que así lo requieran, cuento con las debidas autorizaciones de quienes poseen los derechos patrimoniales. Por lo tanto, me hago responsable de cualquier litigio o reclamación relacionada con derechos de propiedad intelectual, exonerando de toda responsabilidad al Tecnológico Nacional de México campus Orizaba.

También declaro que, al presentar esta tesis, cedo los derechos del trabajo al Tecnológico Nacional de México campus Orizaba para su difusión, con fines académicos y de investigación, bajo las regulaciones propias de la institución y que si existe algún acuerdo de confidencialidad de la información lo haré saber en forma escrita para que se omitan las secciones correspondientes.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y del director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección: depi\_orizaba@tecnm.mx. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente de este.



Blanca Estela González Sánchez

# CONTENIDO

AGRADECIMIENTOS .....	7
RESUMEN .....	8
ABSTRACT .....	9
INTRODUCCIÓN .....	10
PLANTEAMIENTO DEL PROBLEMA .....	11
JUSTIFICACIÓN .....	12
HIPÓTESIS .....	14
OBJETIVOS .....	15
CAPÍTULO 1: ESTADO DEL ARTE .....	16
1.1   Estado del arte .....	16
CAPÍTULO 2: Estudio de las características físicas y los defectos del café verde que influyen en las notas sensoriales mediante modelos de Machine Learning .....	30
2.1   Introducción .....	30
2.2   Marco Teórico .....	32
2.2.1   Variedades de café .....	32
2.2.2   Tipos de Café .....	33
2.2.2.1   Arábica .....	33
2.2.2.2   Robusta .....	34
2.2.2.3   Libérica .....	34
2.2.2.4   Excelsa .....	34
2.2.3   Café verde .....	34
2.2.4   Composición Química .....	34
2.2.5   Propiedades Físicas .....	35

2.2.6 Catación .....	36
2.2.6.1 Aroma .....	37
2.2.6.2 Sabor.....	37
2.2.6.3 Acidez .....	37
2.3 Materiales y métodos .....	37
2.3.1 Muestras de café .....	38
2.3.2 Metodología de la catación del café .....	39
2.3.3 Metodología de aprendizaje automático .....	41
2.4 Resultados .....	43
2.4.1 Predicción de las notas sensoriales.....	44
2.4.2 Resultados del análisis físico.....	44
2.4.3 Resultados de la cata de café .....	46
2.4.4 Resultados de la importancia de las características del análisis sensorial.....	48
2.4.5 Resultados de la predicción de las notas sensoriales.....	53
2.5 Conclusiones .....	60
CAPÍTULO 3: Estudios de espectroscopía NIR en granos de café verde para la determinación de las notas sensoriales utilizando aprendizaje profundo .....	61
3.1 Introducción .....	61
3.2 Marco Teórico .....	61
3.2.1 Espectrofotometría.....	61
3.2.2 Instrumentos de medición .....	61
3.3 Metodología .....	63
3.3.1 Metodología Espectrofotometría NIR .....	63
3.3.2 Obtención de las muestras .....	63
3.3.2.1 Preparación de la muestra .....	63

3.3.2.2	<b>Valoración de la muestra.</b>	64
3.3.2.3	<b>Sistema inteligente</b>	64
3.4	<b>Experimentación</b>	64
3.4.1	<b>Obtención de las señales</b>	64
3.4.2	<b>Cuantificación de las muestras</b>	65
3.4.3	<b>Determinación de Características Importantes</b>	72
3.4.4	<b>Modelos de Inteligencia Artificial para la predicción de Propiedades Sensoriales y calidad del café utilizando datos de espectroscopía NIR</b>	82
3.5	<b>Conclusiones</b>	88
	<b>CONCLUSIONES</b>	89
	<b>RECOMENDACIONES</b>	90
	<b>PRODUCTOS ACADÉMICOS</b>	96
	<b>ANEXOS</b>	97

## LISTA DE FIGURAS

Figura 2.1. Muestra de café verde Coatepec-Veracruz, Tipo Pergamino, Variedad:Typica - Costa Rica	35
Figura 2.2. Diagrama esquemático de la metodología aplicada.	38
Figura 2.3. Variedades de café verde. (a) Bourbon, (b) Typica, (c) Colombia, (d) Caturra, (e) Costa Rica, (f) mundo novo, (g) Garnica, (h) Catuai, y (i) Marseillaise	39
Figura 2.4. (a) Análisis de defectos en granos de café verdes, (b) preparación de la muestra de granos de café tostados para la cata, y (c) preparación de la taza de café y proceso de cata.	40
Figura 2.5. Diagrama esquemático de la metodología de Machine Learning implementada y utilizada para el análisis de variables importantes y la estimación de propiedades sensoriales.	41
Figura 2.6. Diagrama esquemático que representa la metodología de aprendizaje automático mediante validación cruzada para la determinación de variables importantes.	43
Figura 2.7. Diagrama representativo de la metodología de aprendizaje automático para la generación de modelos de predicción de las notas sensoriales del café.	43
Figura 2.8. Resultados que indican el número de granos de café verde según su tipo de defecto.	45
Figura 2.9. Resultados que indican el número de granos de café verde según su tamaño.	46
Figura 2.10. Resultados de las notas sensoriales de fragancia y aroma de la cata de café.	47
Figura 2.11. Resultados de las notas sensoriales del sabor y el regusto de la cata de café.	48
Figura 2.12. Número de concordancias de las variables de importancia obtenidas por los métodos utilizados café en las notas sensoriales e acuerdo a su tipo y variedad.	49
Figura 2.13. Importancia de las propiedades físicas en las notas sensoriales.	51
Figura 2.14. Resultados de precisión de los modelos ML utilizados en la predicción de notas sensoriales de aroma.	53
Figura 2.15. Resultados de precisión de los modelos ML utilizados en la predicción de notas sensoriales de fragancias.	54
Figura 2.16. Resultados de precisión del algoritmo ML utilizado en la predicción de notas sensoriales de sabor.	54
Figura 2.17. Resultados de precisión del algoritmo ML utilizado en la predicción de las notas sensoriales de regusto.	55
Figura 3.1. Metodología para la determinación de las notas sensoriales del café utilizando datos de espectroscopía NIR a través de Redes Neuronales Artificiales.	62
Figura 3.2. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo pergamino.	67
Figura 3.3. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo CR95.	67
Figura 3.4. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Bourbon.	68
Figura 3.5. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Geisha.	68
Figura 3.6. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Marago.	69
Figura 3.7. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Marse-	

Ilesa.	69
Figura 3.8. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Mundo Nôvo.	70
Figura 3.9. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Oro Azteca.	70
Figura 3.10. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Paka-mara.	71
Figura 3.11. Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Typica.	71
Figura 3.12. Frecuencia de los datos en cada longitud de onda.	72
Figura 3.13. Metodología de Variables de Importancia relacionadas a los niveles de absorbancia de diferentes longitudes de onda respecto a las notas sensoriales del café	73
Figura 3.14. Variables de Importancia en las longitudes de onda de 500nm a 1200nm respeto a las notas sensoriales de aroma, fragancia, sabor y resabio. Donde FI_A es aroma, FI_FR es fragancia, FI_FL es sabor y FI_AT es resabio.	75
Figura 3.15. Variables de Importancia en las longitudes de onda de 1200nm a 1695nm respeto a las notas sensoriales de aroma, fragancia, sabor y resabio. Donde FI_A es aroma, FI_FR es fragancia, FI_FL es sabor y FI_AT es resabio.	76
Figura 3.16. Variables de Importancia en las longitudes de onda de 1600nm a 3600nm respeto a las notas sensoriales de aroma, fragancia, sabor y resabio. Donde FI_A es aroma, FI_FR es fragancia, FI_FL es sabor y FI_AT es resabio.	77
Figura 3.17. Resultados ordenados por importancia de las variables de todas las notas sensoriales del Aroma respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro.	78
Figura 3.18. Resultados ordenados por importancia de las variables de todas las notas sensoriales del Fragancia respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro.	79
Figura 3.19. Resultados ordenados por importancia de las variables de todas las notas sensoriales del Sabor respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro	80
Figura 3.20. Resultados ordenados por importancia de las variables de todas las notas sensoriales del Resabio respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro	81
Figura 3.21. Precisión obtenida por las redes neuronales. Valores estimados vs valores deseados. a) Aroma Floral, b) Aroma chocolate, c) Aroma nueces, d) Aroma floral.	82

## LISTA DE TABLAS

Tabla 2.1. Variedades de Café arábica de portes bajo y alto evaluadas en Zentla, Veracruz. *Brote terminal con hojas verdes; †Brote terminal con hojas cafés; §Con tres ramas plágiotropicas por nudo, en lugar de dos.....	33
Tabla 2.2. Componentes químicos de café Arábica y Robusta .....	36
Tabla 2.3. Variables físicas y defectos del grano que afectan a cada nota sensorial del café en su aroma, fragancia sabor y resabio. ....	50
Tabla 2.4. Variables físicas y defectos del grano que afectan a cada nota sensorial del café en su aroma, fragancia sabor y resabio. ....	52
Tabla 2.5. Métricas de evaluación utilizadas en la determinación de las notas sensoriales de aroma y fragancia notas sensoriales .....	55
Tabla 2.6. Métricas de evaluación utilizadas en la determinación de las notas sensoriales de aroma y fragancia notas sensoriales .....	57
Tabla 2.7. Los mejores algoritmos ML hiperparámetros para la estimación de notas sensoriales de Aroma .....	59
Tabla 2.8. Los mejores algoritmos ML hiperparámetros para la estimación de notas sensoriales de Aroma .....	59
Tabla 3.1. Rango de longitudes de onda con su correspondiente estructura orgánica utilizados en la es- pectroscopía NIR .....	65
Tabla 3.2. variables de importancia del aroma, fragancia, sabor y resabio que impactan en las notas sensoriales (caramelo, chocolate, terroso, fermentado, floral) .....	81
Tabla 3.3. variables de importancia del aroma, fragancia, sabor y resabio que impactan en las notas sensoriales (herbal, nueces, pirolítico, resinoso, especies) .....	82
Tabla 3.4. Resultados de la precisión para la estimación del Aroma del café utilizando redes neuronales con diferentes configuraciones de hiperparámetros .....	84
Tabla 3.5. Resultados de la precisión para la estimación de la Fragancia del café utilizando redes neuro- nales con diferentes configuraciones de los hiperparámetros .....	85
Tabla 3.6. Resultados de la precisión para la estimación del sabor del café utilizando redes neuronales con diferentes configuraciones de hiperparámetros .....	86
Tabla 3.7. Resultados de la precisión para la estimación del resabio del café utilizando redes neuronales con diferentes configuraciones de hiperparámetros .....	87

## AGRADECIMIENTOS

“Los logros de una organización son el resultado de los esfuerzos combinados de cada individuo.”

**Vince Lombardi**

Agradezco a:

**COVEICYDET**; por el apoyo otorgado a través del proyecto 15 2243/2021, lo cual permitió obtener resultados relacionados al tema de tesis.

**El Tecnológico Nacional de México**; por otorgar las facilidades para la continuación en la formación Doctoral.

**CAFECOL A.C**; por ser parte fundamental en la experimentación del café y compartir sus conocimientos a través de estancias de investigación.

Mi Asesor el **Dr. Oscar Osvaldo Sandoval González**; por ser parte esencial en el desarrollo del tema de tesis. Agradezco su experiencia, dedicación y tiempo, los cuales contribuyeron para la obtención de resultados y adquirir el conocimiento durante el tiempo de mi formación.

Mi Codirector **Dr. José de Jesús Agustín Flores Cuautle**; por sus observaciones y recomendaciones para la mejora del tema y del escrito.

El **comité revisor**; por contribuir con la retroalimentación durante el desarrollo de la tesis.

**Mi familia**; por ser el soporte en los momentos mas difíciles en la trayectoria de esta etapa.

**Mis amigos**; por estar presentes en periodos complicados y siempre tener las palabras correctas para animarme.

## RESUMEN

El presente trabajo doctoral se enfoca en la determinación de las propiedades sensoriales del café verde mediante el análisis de sus variables físicas, deformaciones de los granos y datos de espectroscopía NIR.

La primera contribución destacada se centra en el estudio de las características físicas y los defectos del café verde que influyen en sus notas sensoriales, utilizando modelos de Machine Learning. Los resultados muestran la posibilidad de comprender cómo las variables físicas y las deformaciones de los granos impactan en dichas notas. Se llevó a cabo un análisis detallado para estimar un perfil sensorial preciso del café verde, considerando su variedad, tipo y defectos en grano.

La segunda contribución se basa en los estudios de espectroscopía NIR en granos de café verde para determinar las notas sensoriales mediante aprendizaje profundo. Esto permitió identificar la relación entre ciertas combinaciones de estructuras orgánicas del café verde y sus notas sensoriales en aroma, fragancia, sabor y resabio.

En ambos casos, se emplearon modelos de inteligencia artificial, incluyendo Machine Learning y Deep Learning, para determinar las variables clave y desarrollar sistemas inteligentes capaces de predecir las notas sensoriales del café utilizando tanto información física como datos de espectroscopía NIR.

**Palabras Clave:** Café verde, Espectroscopia, Perfil Sensorial, Inteligencia Artificial.

## ABSTRACT

The present doctoral work focuses on the determination of the sensory properties of green coffee through the analysis of its physical variables, bean deformations and NIR spectroscopy data.

The first outstanding contribution focuses on the study of the physical characteristics and defects of green coffee that influence its sensory notes, using Machine Learning models. The results show the possibility of understanding how physical variables and bean deformations impact these notes. A detailed analysis was carried out to estimate an accurate sensory profile of green coffee, considering its variety, type and bean defects.

The second contribution is based on NIR spectroscopy studies on green coffee beans to determine sensory notes through deep learning. This allowed the identification of the relationship between certain combinations of organic structures of green coffee and its sensory notes in aroma, fragrance, flavor and aftertaste.

In both cases, artificial intelligence models, including Machine Learning and Deep Learning, were used to determine the key variables and develop intelligent systems capable of predicting the sensory notes of coffee using both physical information and NIR spectroscopy data.

**Keywords:** Green coffee, Spectroscopy, Sensory Profiling, Artificial Intelligence.

## INTRODUCCIÓN

En México el cultivo de café (caficultura) es de relevancia debido a la integración de cadenas productivas importantes, el producto llamado oro verde vio un incremento en la producción de acuerdo a datos de SAGARPA entre 1999 y 2000 CEDRSSA, 2018, pero no fue hasta el año 2017 cuando acorde a las estadísticas del servicio de información agroalimentaria y pesquera *Agrícola Nacional CAFÉ*, 2017 que el café cereza tuvo un incremento en su producción. La producción generada a partir de ese año fue de 839 mil toneladas. En el país uno de los principales productores de café es el estado de Chiapas que genera una producción del 39 por ciento del volumen nacional, seguido por Veracruz 30 por ciento y Oaxaca con un 13 por ciento. La producción y cultivo de café se ha extendido a 12 estados de la república mexicana, con lo cual las grandes empresas transnacionales están motivando la producción de café en algunas de estas zonas, aumentando su volumen de producción, pero también reduciendo la calidad de este. A pesar de la gran importancia que representa el café, existe una alta pobreza en los municipios cafetaleros. Las estadísticas indican que, en municipios cafetaleros siete de cada diez habitantes viven en condiciones de alta y muy alta marginación, por lo cual se vuelve imposible que en estas zonas se utilice tecnología que pueda monitorear la calidad del café que se está produciendo. Hoy en día organismos especializados como el Centro Nacional de Investigación, Innovación y Desarrollo Tecnológico del Café (CENACAFE) han concluido que en la caficultura el incremento de la calidad del café es mucho más redituable que el incremento en su producción. Por lo cual se han generado varias estrategias para impulsar a los cafeticultores a producir café de alta calidad. El café de alta calidad es altamente codiciado y se vende a precios muy superiores a los que las empresas transnacionales pagan actualmente a los cafeticultores por la producción de café de baja calidad. A pesar de la crisis del sector cafetalero en el cual ha estado inmerso en el mercado internacional, las perspectivas del desarrollo del sector cafetalero se siguen impulsando como eje de desarrollo comunitario y regional. Debido a las bajas oportunidades de los cafeticultores para determinar la calidad de su producto, es necesario desarrollar ventajas competitivas que radican en la calidad del café. Esta producción tendrá una mayor oportunidad en el mercado, con el uso de herramientas que determinen procesos que potencialicen su calidad. Para poder elevar la calidad del café, se requiere de una disciplina, capacitación y conocimiento, al igual que la implementación de tecnología que ayude a los cafeticultores a lograr esta meta. Abordando el tema de la calidad del café, esta calidad es analizada a través de pruebas de laboratorio que permiten determinar las propiedades fisicoquímicas y sensoriales del mismo. La investigación aborda la implementación de diversas metodologías de inteligencia artificial para determinar las notas sensoriales del café utilizando datos de espectrometría y análisis físicos del café.

“Las necesidades de investigación y desarrollo del sector cafetalero orgánico son diversas; entre las más importantes están el incremento de los rendimientos, la reducción en los costos de producción, y el mejoramiento de la calidad del grano y la bebida, factores asociados con el proceso de producción, beneficiado y comercialización”CEDRSSA, 2018.

## PLANTEAMIENTO DEL PROBLEMA

Uno de los mayores desafíos en la industria del café es garantizar la consistencia en el perfil sensorial y la calidad del producto. Factores como las condiciones climáticas, el suelo, las prácticas agrícolas y el procesamiento pueden influir en las características sensoriales del café, lo que dificulta mantener un estándar uniforme. Además, la subjetividad en la catación y la evaluación sensorial puede llevar a discrepancias en la percepción de la calidad del café entre los catadores.

Otro problema común es la falta de transparencia y trazabilidad en la cadena de suministro del café. Los agricultores suelen carecer de acceso a tecnologías que les permitan monitorear y rastrear el proceso de producción, lo que dificulta la identificación de posibles problemas que puedan afectar la calidad del café. Esto también puede contribuir a prácticas comerciales injustas, donde los agricultores reciben compensaciones injustas por su trabajo.

La tecnología puede desempeñar un papel fundamental en la solución de las problemáticas del café en cuanto a su perfil sensorial y calidad. El desarrollo de herramientas de análisis de datos y modelos de inteligencia artificial puede ayudar a los productores a comprender mejor la relación entre las propiedades físicas y las características sensoriales del café. Estos modelos pueden identificar patrones y tendencias que no serían evidentes para el ojo humano, lo que permite a los agricultores optimizar sus prácticas de cultivo para producir cafés de alta calidad de manera más consistente.

En cuanto a la evaluación sensorial, la tecnología puede ayudar a reducir la subjetividad y mejorar la precisión en la catación del café. El uso de dispositivos espectroscopía puede proporcionar mediciones objetivas del perfil sensorial del café, lo que ayuda a estandarizar los criterios de calidad y facilita la comparación entre diferentes lotes de café.

Estas son las razones por las cuales esta propuesta doctoral se enfoca al estudio profundo de las relaciones que existen entre las propiedades físicas y las señales de espectroscopía respecto al perfil sensorial del café.

## JUSTIFICACIÓN

La industria del café emplea a millones de personas en todo el mundo, desde agricultores y recolectores hasta trabajadores en plantas de procesamiento, distribución y venta al por menor. Estos empleos no solo proporcionan medios de vida para las comunidades locales, sino que también ayudan a impulsar la economía de las regiones cafetaleras.

El cultivo del café suele ser una actividad central en muchas áreas rurales, especialmente en países en desarrollo. La producción de café a menudo impulsa el desarrollo de infraestructura y servicios básicos en estas regiones, como carreteras, escuelas y centros de salud.

Los movimientos de comercio justo y producción sostenible en la industria del café buscan garantizar condiciones laborales justas y precios equitativos para los agricultores. Estas iniciativas promueven el bienestar social al apoyar a comunidades agrícolas marginadas y fomentar prácticas agrícolas responsables que protegen el medio ambiente y los recursos naturales.

El café no es solo una bebida; es una parte integral de muchas culturas y tradiciones en todo el mundo. Las cafeterías y las reuniones alrededor de una taza de café son lugares comunes para la socialización y el intercambio de ideas en muchas sociedades. Esta conexión cultural contribuye a la cohesión social y al sentido de comunidad.

La tecnología desempeña un papel fundamental en el desarrollo del café por varias razones:

**Mejora de la productividad agrícola:** La tecnología agrícola, como sistemas de riego automatizados, drones para monitoreo de cultivos, sensores de humedad del suelo y herramientas de análisis de datos, ayuda a los agricultores a optimizar el cultivo del café. Estas tecnologías permiten una gestión más eficiente de los recursos, aumentando así la productividad y la calidad del café.

**Control de calidad:** La tecnología permite implementar sistemas de monitoreo y control de calidad en todas las etapas de producción, desde el cultivo hasta el procesamiento y el envasado. Esto garantiza que el café cumpla con los estándares de calidad requeridos por los consumidores y los mercados internacionales.

**Innovación en procesamiento y tostado:** La tecnología también ha revolucionado el procesamiento y el tostado del café, permitiendo la creación de métodos más eficientes y precisos. Por ejemplo, los sistemas de tostado controlados por computadora pueden garantizar una temperatura y un tiempo de tostado óptimos, lo que resulta en un café de mejor calidad y sabor.

El aumento de la oferta de café a nivel mundial hace necesario dar un valor agregado a este grano, por lo cual, incrementar la calidad del café otorga a los productores una ventaja competitiva. La tecnología continúa desempeñando un papel crucial en la mejora y la innovación de la industria del café, y es por esta razón que la investigación y experimentación con diferentes técnicas de espectroscopía NIR, podrían tener la capacidad de

ayudar a la creación de nuevas metodologías para la determinación de la calidad del café con la ayuda de algoritmos de inteligencia artificial, que logren realizar una estimación cuantitativa y cualitativa de las propiedades sensoriales. Todo esto con el objetivo principal de reducir los tiempos de análisis, eliminar la subjetividad de los catadores y reducir los costos al brindar información del perfil sensorial del café a los cafeticultores.

## HIPÓTESIS

Mediante el uso de modelos de Machine Learning y Deep Learning, es posible determinar las propiedades sensoriales del café verde a partir del estudio de sus variables físicas y de deformación de los granos, así como de los datos de espectroscopía NIR. Se plantea que estas propiedades físicas y estructurales, así como las características detectadas por la espectroscopía NIR, tienen una influencia significativa en las notas sensoriales del café verde, como aroma, fragancia, sabor y resabio. Además, se sugiere que los modelos de inteligencia artificial pueden ser eficaces para estimar perfiles sensoriales detallados del café verde, considerando su variedad, tipo y defectos en grano.

# OBJETIVOS

## OBJETIVO GENERAL

Desarrollar un sistema de estimación de las notas sensoriales del café en grano verde mediante la implementación de modelos predictivos avanzados basados en aprendizaje automático, aprendizaje profundo y análisis de datos, empleando información de las propiedades físicas, defectos en los granos y señales de espectroscopía NIR del café verde

## OBJETIVOS ESPECÍFICOS

1. Investigación exhaustiva de las propiedades físicas y los defectos presentes en los granos de café verde, y su correlación con las notas sensoriales clave como fragancia, aroma, sabor y resabio, empleando técnicas avanzadas de Machine Learning.
2. Desarrollo de un modelo predictivo para estimar las propiedades sensoriales del café verde, integrando información variada como la variedad del café, su tipo, así como las propiedades físicas y los defectos de los granos, mediante enfoques innovadores de análisis de datos y Machine Learning.
3. Realización de un análisis detallado para comprender el impacto individual de las variables de espectroscopia y las propiedades físicas del café verde en relación con cada una de las propiedades sensoriales y la calidad general del café, utilizando métodos estadísticos y herramientas de Machine Learning avanzadas.
4. Implementación de un sistema basado en Redes Neuronales Artificiales para la estimación precisa de las propiedades sensoriales del café verde a partir de datos de espectroscopía NIR, con el fin de proporcionar resultados confiables y consistentes para mejorar la calidad y la consistencia del producto final.

# CAPÍTULO 1: ESTADO DEL ARTE

En este capítulo se denota un análisis del conocimiento acumulado del tema, basado en investigación científica, metodológica y de resultados, los cuales han sido plasmados a través de referencias citadas en el trabajo.

## 1.1 Estado del arte

La calidad del café es uno de los factores más importantes para introducirlo al mercado, por ello la importancia de evaluar el café, el cual es basado en el análisis sensorial. Una de las funciones primarias en el hombre es la conocida como evaluación sensorial, “conocida como un análisis normalizado de los alimentos que se realiza con los sentidos”, se usa para describir, medir, analizar e interpretar aquellas características (organolépticas) del café percibida por los sentidos (vista, olfato, gusto, tacto y oído) Sensorial y Cafe, 2015. Dentro del proceso de la degustación del café se considera la calidad, por eso la importancia de las funciones de la degustación. Degustar o catar es la acción de emitir un juicio en bases a: Clasificar, ordenar, describir, analizar e integrar. La importancia de un buen análisis a través de consideraciones científicas, permite elegir entre las variedades del café y su calidad. Las características varían debido al grano de café, su origen, la humedad, la temperatura etc. Esto determina la calidad final del café, la cual se confirma a través de los evaluadores sensoriales (catadores) del café. La composición química del café es cambiante de acuerdo a sus diversidades de cultivo y tostado. Otro factor importante son los aromas, los cuales están compuestos de varias sustancias volátiles, dentro de los cuales se encuentran aproximadamente 850 compuestos, una sustancia puede presentar varios olores dependiendo de sus características químicas Quintero, 2011.

La constitución química del café es de agua y materia seca (minerales, carbohidratos, lípidos, proteínas, alcaloides y compuestos volátiles), las cuales se ven afectadas durante la tostación del café. Otros factores que influyen en la alteración química y en la calidad del café son: la madurez, la especie, el secado, el almacenamiento y la preparación de la bebida Franca et al., 2005. Una de las cosas que más afectan la calidad del café es la presencia de granos de café defectuosos esto es muy relevante para la calidad del café. El resultado presentado a través del estudio está encaminado a una evaluación de los atributos físicos y químicos de los cafés de diferentes calidades, tanto en verde y tostado. Las pruebas se realizaron con diferentes muestras de café previamente clasificadas en cuanto a sus atributos, densidad y el volumen de la muestra (frijol). La calidad y propiedades del café varían de acuerdo a la presencia de granos defectuosos. La acidez aumenta y los niveles de pH disminuyó a medida que disminuye la calidad de taza, probablemente debido al efecto de granos defectuosos (frijoles agria). Otro factor para evaluar la calidad del café se evalúa comúnmente de acuerdo con criterios tales como el tamaño del grano, el color y la forma. El método de procesamiento a través de clasificación permitió determinar en función del número de defectos presentes en las muestras. El estudio evaluó los atributos físicos y químicos de los cafés de diferentes calidades. Los resultados obtenidos muestran que los granos defectuosos juegan un papel importante en la calidad del café, antes y después de la torrefacción. Ferreira et al., 2016 El objetivo de realizar un estudio

considerando distribuciones probabilísticas con los consumidores de café y teniendo en cuenta el análisis sensorial de cuatro cafés especiales producidos con diferentes procesos y en diferentes alturas de regiones montañosa, dio como resultado comprobar la calidad o aceptabilidad de un producto como el café. Vale la pena mencionar que la heterogeneidad entre las características del café por su lugar de origen y sus componentes químicos son factores importantes que afectan la calidad sensorial del mismo. De acuerdo a los resultados obtenidos se concluyó que los consumidores requieren de capacitación para poder realizar una discriminación entre cafés de especialidad y por ello las técnicas de obtención de las características organolépticas del café son cada vez más confiables para los resultados sensoriales. Otra manera de realizar la evaluación sensorial fue a través de 5 catadores que determinaron la acidez, aroma cuerpo y sabor del café con un panel de cinco catadores, los cuales determinaron por medio de una escala de 1 al 15 las características del café. Los resultados fueron analizados por el software Statistica 6.1 análisis de datos estadísticos de la empresa StatSoft. Cirillo et al., 2019.

Las interpretaciones sensoriales en los experimentos son difíciles de implementar, ya que los factores fisicoquímicos que intervienen en cada proceso son diferentes e individuales. Otro factor importante son sus características físicas las cuales son esenciales para determinar su calidad. Se considera que hay propiedades físicas como el volumen, la densidad, la masa y porosidad en el café verde que son diferencias altamente significativas «Discrimination of defective (Full Black, Full Sour and Immature) and nondefective coffee beans by their physical properties», 2014. Este hecho se hace más relevante en los experimentos que involucran mezclas de cafés con diferentes calidades, variedades, orígenes y formas de procesamiento y preparación. Resultados presentados para la discriminación de café, demuestran el impacto significativo en su calidad. Brasil es un país que comercializa cafés arábica y robusta; la discriminación de café defectuoso y no defectuoso impacta significativamente en su calidad. Las características físicas del café difieren según su tipo; la técnica de separación de color según su tueste, el cual sólo fue eficaz para el café robusta, determinando una mejor calidad Wondimkun et al., 2020. Un producto de mala calidad demerita la venta, por eso la evaluación de los productos antes de salir al mercado requiere una técnica de selección Chandrasekar y Viswanathan, 1999, se presenta una metodología que a través de los atributos físicos y químicos determinados permitió diferenciar los tipos estudiados de frijoles defectuosos y no defectuosos. Los tipos de café Arábica y Robusta presentan propiedades físicas tales como: tamaño, densidad, masa y resistencia física. Se demostró una variante en la técnica aplicada en la resistencia al aplastamiento de los granos entre un café y otro. Otra característica a destacar fue la humedad que interfirió en los resultados para ambas variantes de café Sinaga y Julianti, 2021. Métodos implementados en regiones debido a sus condiciones geográficas y climáticas únicas que son necesarios para controlar la calidad del café. Un estudio demostró que la calidad física de los granos verdes de café Gayo utilizando el método indonesio de procesado semilavado en comparación con el método de procesado, método de elaboración totalmente lavado (método tradicional) ofrece resultados preliminares para evaluar el método de elaboración, su grado de tueste y la calidad del café ensayado Mendonça et al., 2009. Según los resultados, el proceso de tostado tiene un impacto directo en las propiedades físicas de Sidikalang robusta. Para las propiedades físicas del café robusta Sidikalang y los diferentes colores del café cereza, se determinó que los resultados cambian con respecto a la condición inicial de los granos, incluyen-

do cambios en el contenido de agua, la masa, el valor de porosidad y la densidad aparente Yusibani et al., 2022. Entre los componentes relevantes para determinar la calidad en los alimentos se encuentran las características sensoriales porque estas características, como el sabor, los defectos y la textura determinan cómo perciben los consumidores los productos Titova et al., 2015. En cuanto al café, una de las bebidas más consumidas en el mundo, con un consumo global de 166 millones de sacos de 60 kg en 2020 ICO, 2021, su calidad ha sido evaluada por expertos catadores que determinan su aspecto, textura, sabor, aroma, etc. Feria-Morales, 2002. El café arábica es uno de los máspreciados en la región de Veracruz, la producción de este café se lleva a cabo debido a las características de suelo para su crecimiento «Revista fitotecnia mexicana publicada por la Sociedad Mexicana de Fitogenética.», 2016. La calidad de sus frutos se da en la pre y pos cosecha, de ahí la calidad y variedad en tamaños. Debido a las características del suelo del estado de Veracruz su aporte es uno de los principales en México llegando al 21.25 por ciento de la producción total.

La espectrofotometría es una herramienta muy sencilla y accesible a todo el mundo, por eso, el uso como análisis cuantitativo y cualitativo en diferentes sectores alimenticios. Varios son los componentes fisicoquímicos que componen el grano de café verde como la cafeína, el cual es un alcaloide del grupo de las xantinas, sólido cristalino, blanco y de sabor amargo Bhawani et al., 2015, usado también como ingrediente en varios productos del sector alimenticio y farmacéutico. Uno de los aspectos importantes estudiados fue el mantener un nivel óptimo en los diferentes productos que así lo requieren. En la aportación científica se incorporan diversos métodos espectrofotométricos utilizados en el análisis de la cafeína en diversas muestras ambientales. La determinación de la cafeína se ha dado por diferentes métodos quimiométricos, aumentando el alcance a través de la espectrofotometría, la cual es una herramienta muy sencilla y accesible a todo el mundo. En la actualidad el uso de los sistemas espectrales se han vuelto herramientas fiables para la valoración objetiva de la calidad del café Barbin et al., 2014, la contribución metodológica deja ver de manera clara como a través del infrarrojo medio se puede obtener la composición química y las propiedades relacionadas de café, así como su calidad y algunas de las características importantes de café tales como: humedad, lípidos y contenido de cafeína, Los beneficios del café a través de su autenticación basado en los parámetros de calidad, trae bienes económicos en la industria del café, aumentando en los consumidores la confianza del producto. Weldegebre et al., 2017 A través de la experimentación mediante FT-IR-ATR y espectrofotometría de fluorescencia se logró determinar la cafeína en una solución acuosa de grano de café verde. La cafeína también se determinó directamente en solución en dimetilformamida usando espectroscopía NIR con técnica de calibración univariante. Se determinó que el porcentaje de cafeína en el café verde es variante de acuerdo con el método usado (FT-IR-ATR, NIR o espectroscopía de fluorescencia). Una manera rápida y sencilla para determinar el contenido de cafeína es usando espectrofotometría NIR en muestras de granos de café verdes con cantidad reducida de disolvente orgánico. La sensibilidad de las mediciones espectrométricas se basa en intensidades de banda, incluso los espectros obtenidos para la medición NIR de cafeína fue muy intensa con respecto a otras. da Silva Araújo et al., 2020 se evaluó el rendimiento de los modelos de calibración construida a partir de datos espectrales obtenidos por espectroscopía de infrarrojo cercano (FT-NIR) para determinar los valores de pH y la acidez en los granos de

café de una manera práctica y no destructiva. Los resultados obtenidos mostraron que los mejores métodos de preprocesamiento eran la primera derivada de la variable de pH y la variable normal estándar para la acidez. Los modelos obtenidos fueron los adecuados de acuerdo a las relaciones de la predicción de la desviación. El porcentaje promedio de error se encontró entre los valores predichos y experimentales, los cuales fueron inferiores a 7 porciento. Se determinó que NIR se aplicó con éxito para predecir las propiedades relacionadas con la calidad del café. Como resultado final se determinó que la aplicabilidad de la espectroscopía NIR es confiable para la determinación de las propiedades fisicoquímicas, pH y acidez, y como tal la calidad de los granos de café verde, de una manera rápida y de forma no destructiva. Dando así que NIR es una herramienta útil para la industria del café como un medio para evaluar rápidamente la calidad de la materia prima. Las propiedades fisicoquímicas del café como los ácidos clorogénicos, cafeína y propiedades antioxidantes de los extractos de café verde son importantes para la formación de sabor «Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation», 2016, la determinación de las propiedades se determinó por análisis HPLC-DAD (cromatografía equipada con la unidad de HPLC analítico modelo 1100 - Agilent Technologies, CA, EE.UU.) en doce muestras de café de Robusta y Arábica tipos de diferente origen geográfico incluyendo cafés al vapor y descafeinado. La determinación de las propiedades fisicoquímicas del café por otros métodos, nos deja ver la diferencia con otras herramientas y procesos para una mejora en la investigación. La importancia del café Arábica y Robusta por su sabor uno más fino que el otro, han adquirido calidad económica en todo el mundo Briandet et al., 1996. Los resultados presentados son basados en la espectroscopía infrarroja por transformada de Fourier, la cual es una herramienta que proporciona una alternativa rápida para mejorar los métodos químicos para la autenticación y cuantificación de los productos de café. La espectroscopía IR es usada como un método para discriminar y cuantificar el café arábica y robusta. Los resultados de este método con respecto a su composición química son determinados por la absorción en la “huella digital” región espectral (diferentes frecuencias). Los granos de café verdes contienen una amplia gama de diferentes compuestos químicos, que reaccionan e interactúan en todas las etapas de procesamiento de café para producir un producto final con una diversidad aún mayor y la complejidad de la estructura. Los compuestos más importantes de café liofilizado son carbohidratos, minerales, cafeína, ácido clorogénico, proteínas, y lípidos. Algunas diferencias físicas se pueden apreciar en los espectros de ambas especies, todo ello relacionado con sus propiedades fisicoquímicas. Los resultados muestran que la espectroscopía FTIR se puede utilizar para identificar y cuantificar las propiedades fisicoquímicas del café arábica y robusta a través de su muestreo de espectros. Los cuales dieron excelentes resultados en el análisis discriminante, con 100 por ciento de los espectros en los dos conjuntos de prueba de formación e independientes clasificadas correctamente. Una aportación científica a través de modelos quimométricos, daría a los productores de café un análisis sensorial descriptivo usando la espectroscopía de infrarrojo cercano Ribeiro et al., 2011, la cual indica que es posible estimar la calidad del café utilizando modelos de regresión PLS obtenidos mediante el uso de espectros NIR de análisis de café. Los parámetros predichos fueron puntuaciones sensoriales de sabor, acidez, amargura, cuerpo, limpieza y calidad general. La selección de variables mediante el uso del algoritmo OPS fue un paso primario para determinar las mejores regiones espec-

trales que describen cada atributo sensorial estudiado. Alessandrini et al., 2008. La composición química del café es sin duda una de las más complejas, por tal razón se muestra el estudio en relación de algunas variables de torrefacción de café (pérdida de peso, la densidad y la humedad), los métodos usados habitualmente a nivel industrial son de mayor costo e implican más elaboración, es por ello que se ha considerado la técnica con NIR (espectroscopia de infrarrojo cercano) para demostrar el grado de tostado del café, los modelos de calibración y de validación independiente basadas en modelos de regresión parcial de mínimos cuadrados (PLS), ya que es una técnica rápida, simple y barata. Se realizó un análisis con NIRS para predecir y discriminar diferentes grados de tostado de café mediante el uso de las características tradicionales del proceso de tostado, tales como pérdida de peso, la densidad y la humedad. Para este objetivo, se llevó a cabo la medición de variables de tueste tradicional en un subconjunto representativo de muestras de café verde y tostado de manera diferente, así como modelos de regresión de mínimos cuadrados parciales separadas, los cuales se construyeron en base al espectro NIR de café. Los modelos de calibración de alta calidad resultantes desarrollados sobre la base de los datos espectrales NIR podrían ser utilizados para fines de control de calidad del café. Es posible mediante esta técnica obtener resultados confiables para determinar las propiedades del café en función a la torrefacción. Otro método para la obtención de las propiedades fisicoquímicas del café es a través de las mediciones dieléctricas, por el cual es posible determinar la humedad en los granos de café Orduña et al., 2016. Se usaron dos métodos, el de los sensores capacitivos y de detección de carga mecánica, para estimación de la misma. Los resultados mostraron que es posible determinar la humedad en el café verde en un rango de 40 a 10 porciento, el beneficio es la mejora en las condiciones de almacenamiento del grano de café. Esta aportación tiene la característica de incrementar la calidad en el café, lo que significa mantener las propiedades físicas y sensoriales del mismo. Una de las características de la capacitancia es permitir dar a conocer cuando los granos de café han alcanzado el nivel recomendado de humedad para la conservación, dando al productor la oportunidad de mejorar la calidad requerida en el grano. La necesidad de estudios de campo ha permitido el desarrollo de los dispositivos hiperespectrales para determinar de manera rápida y fácil la información espectral, en algunos casos, han tomado gran importancia de acuerdo a los rápidos resultados del análisis de cultivos, una ventaja es la rápida captura de la información y la no invasión de la misma Aitkenhead et al., 2017. La importancia de análisis de datos en campo ha dado lugar a hacer uso de equipos portátiles, los cuales nos proporcionen información en los rangos de longitud de onda visible e infrarrojo cercano (VIS+NIR). La determinación de características de algunos cultivos como: detección de enfermedades, nutrientes en plantas, concentraciones de clorofila, calidad del café etc. Es un método de análisis rápido y preciso para la determinación cuantitativa de los componentes principales de productos alimenticios. El procesamiento de las señales espectrales nos proporciona la interpretación de los diferentes datos utilizados, así como la visualización a través de los diferentes programas o software los cuales nos permiten la integración, manipulación y edición de los mismos. La importancia de la espectroscopía IR (infrarroja) en las prácticas agrícolas en países en desarrollo fue subrayado por Shepherd y Walsh, 2007, con muchos de sus puntos relacionados al diagnóstico y apoyo a la toma de decisiones con el manejo de suelo que resulta ser válido en la agricultura de países desarrollados. También identificaron la capacidad de construir lo

que se requiera en términos de centros de excelencia, diseño de equipo y el desarrollo de sistema de apoyo a la decisión. La importancia de garantizar la calidad de los suelos ha generado que las investigaciones se vuelquen hacia metodologías científicas que nos permitan determinar sus propiedades fisicoquímicas, estas metodologías se aspiran implementar con el fin de hacer frente a las consecuencias de muchos problemas ambientales, específicamente los suelos Landeta-Escamilla et al., 2019. Una de las técnicas más actuales que está siendo usada para la obtención de resultados es el Infrarrojo Cercano (NIR), desafortunadamente para la implementación en suelos se requiere de un conocimiento más específico, debido a ello se implementa la espectroscopia de capacitancia con algunos algoritmos de inteligencia artificial, con ello se pretenden estimar las variables físico-químicas del suelo utilizadas para el cultivo en este caso concretamente de la caña de azúcar. Las ventajas son: errores por debajo de ocho por ciento en la estimación de las variables contra las pruebas de laboratorios, es un sistema portátil, con un bajo coste, que es fácil de usar y que se puede implementar para probar otros tipos de suelos. Por lo cual se considera un resultado interesante para la aportación científica del proyecto propuesto. Con los resultados de esta investigación se pretende retomar la técnica para la valoración del análisis de la calidad del café. La espectroscopia infrarroja por transformada de Fourier (FTIR) se puede utilizar para detectar cambios en grupos funcionales de compuestos, como el café. Sin embargo, la dispersión, la reflexión y la no homogeneidad de la matriz de la muestra pueden causar problemas que resultan en una baja calidad espectral. En el trabajo se ha implementado la fotoacústica para realizar un estudio de café tostado y molido de tres variedades de Coffea arábica cultivado Gordillo-Delgado et al., 2012. La comparación entre los espectros de café obtenidos por espectrometría FTIR-PAS y por espectrofotometria FTIR mostró una mejor resolución del método anterior, que, ayudado por el análisis de componentes principales, permitió la identificación de algunas bandas de absorción que permiten la discriminación entre los diferentes tipos de café. La técnica fotoacústica (PA) en el estudio del café permite un análisis espectroscópico de muestras opacas orgánicas y su caracterización térmica; Por lo tanto, esta técnica proporciona información sobre la pigmentación y los parámetros de modelado de tostado del café, cuando se aplica a muestras de café verde y en polvo, respectivamente Delgado et al., 2008. La caracterización de las muestras a través de los espectros, nos permiten identificar las características físicas en el café de origen. En este trabajo se dan a conocer mediciones espectroscópicas PA del café en polvo y análisis de correlación que permiten proponer un criterio de discriminación. Además, la difusividad térmica de las muestras de café verde se midió mediante la técnica de PA de celda fotoacústica abierta. Los resultados presentados permitirán un posible método para certificar el café orgánico de diferentes regiones y su calidad de cultivo. Dias et al., 2018 La calidad de un café radica en sus granos, los análisis químicos y la evaluación sensorial del café son los métodos más aplicados para el control de calidad del café tostado y molido. En el trabajo se realizó un análisis de espectroscopia fotoacústica infrarroja (FTIR-PAS) en cafe tostado y molido (RG). Se estudió una serie de mezclas de granos saludables con defectuosos de diferentes tipos de café. El análisis de componentes principales permitió predecir la cantidad y la naturaleza de los defectos en las mezclas, mientras que el análisis discriminante de mínimos cuadrados parciales reveló similitudes entre las mezclas. Se obtuvo un modelo predictivo exitoso para el número de pruebas realizadas, demostrando que la técnica puede ser usada para la caracterización y clasificación de

mezclas de café.

Hernández-Aguilar et al., 2019 El objetivo fue el estudio de la espectroscopia fotoacústica (PAS) en la industria alimentaria para evaluar la calidad de los alimentos. Se llevó a cabo una investigación en el desarrollo de tecnologías fototérmicas, las cuales destacan el análisis de alimentos y la determinación de compuestos. Un método con menos desperdicio químico y una cantidad mínima de muestra y una técnica no destructiva. La técnica PAS permite obtener propiedades ópticas de los alimentos, en función de su color, que es el parámetro más útil en la industria agroalimentaria ya que la calidad y los sabores de los alimentos están íntimamente asociados a su color. La técnica PAS aplicada en alimentos y plantas, permite determinar las propiedades, concentraciones o presencia de compuestos químicos y adulterados en los alimentos. En este sentido, para algunos investigadores, el PAS se considera un método analítico. La aplicación del PAS en las diferentes industrias: alimentaria y agroalimentaria son de las que más se verían beneficiadas con el uso de esta tecnología, debido a la obtención de sus espectros de PA en donde se puede observar a través de la banda de absorbancia la variedad de los compuestos de cada producto analizado. La aplicación de la técnica PAS para los granos, es otro estudio que ha mostrado resultados certeros para evaluar la calidad sanitaria de los mismos, permitiendo obtener sus espectros de absorbancia óptica, en función de la longitud de onda, los cuales proporcionan información sobre los procesos de absorbancia óptica que se producen en la muestra. Los resultados obtenidos de las diferentes aplicaciones en la agroindustria con la técnica PAS permite caracterizar los alimentos a través de espectros ópticos de absorbancia, realizando calibraciones y análisis matemático de datos. En esta técnica se pueden tener diferentes configuraciones fotoacústicas que se pueden variar de acuerdo a las áreas del conocimiento, la obtención de los espectros puede ser en función de la longitud de onda o en función de la frecuencia de modulación de la luz. Los espectros pueden ser de amplitud, fase y señales en función de la frecuencia, lo que llevaría a la aplicación de otros métodos para el análisis de las muestras. Cebi et al., 2017 El desarrollo de un método nuevo, rápido, eficaz, no destructivo y rentable para la detección de monohidrato de clorhidrato de sibutramina, que hasta ahora es el adulterante más común en los suplementos dietéticos destinados a la pérdida de peso como el té verde, café verde y té de hierbas mezclado. Otros métodos como los cuales incluyen cromatografía de gases y líquida son también confiables, pero su alto costo y características de operación los dejan fuera del alcance para el logro del objetivo. La espectroscopia FTIR se propone por ser no destructiva y rápida para obtener huellas dactilares bioquímicas proporcionando información sobre la estructura y composición molecular. Esta técnica ha mostrado en productos alimenticios ser una gran herramienta para la detección en la adulteración de los alimentos, la asociación de la espectroscopia FTIR puede tener potencial en la detección de sibutramina en productos herbales naturales, alimentos y suplementos dietéticos. Para este estudio se hizo uso de el espectrómetro Bruker Tensor 27 Bremen-Alemania) con un divisor de haz KBr y un detector DLaTGS. Todas las mediciones se realizaron con un accesorio ATR de diamante de un solo rebote. Las adquisiciones de datos se lograron utilizando OPUS Versión 7.2 para Windows de Bruker Gmbh. En la investigación, se utilizaron dos conjuntos de muestras para probar y validar la metodología desarrollada para determinar la adulteración de sibutramina. Las muestras fueron adulteradas con sibutramina a diferentes niveles. Las cantidades de té y café se mantuvieron en un valor cons-

tante para cada muestra, mientras que la cantidad de adulteración cambió entre 0.375 y 12 mg. El resultado del análisis dio como resultado un grupo de bandas vibratorias significativas asociadas a los componentes químicos de la sibutramina. De la detección de todas las bandas la determinada con numero de onda máximo a  $2698\text{ cm}^{-1}$  en el  $2700 - 2600\text{ cm}^{-1}$  rango espectral, se consideró que era la característica espectral más importante que se utilizó posteriormente para la detección de sibutramina en las muestras adulteradas, ya que es específico de las sales de clorhidrato. El método utilizó las capacidades únicas de huellas dactilares de la técnica de espectroscopía FTIR ya que el método se basa en las características espectrales específicas del clorhidrato de sibutramina monohidrato. Con este estudio se considera la detección de monohidrato de clorhidrato de sibutramina en los suplementos dietéticos; por medio de la técnica FTIR-ATR combinada con la quimiometría.

El uso de herramientas como apoyo para la determinación de características sensoriales es una opción para atender a situaciones como las pandemias, las cuales ponen de manifiesto la necesidad de contar con diferentes sistemas inteligentes para determinar si las características sensoriales cumplen los requisitos establecidos de calidad. Por ello, se han realizado investigaciones para establecer diferentes metodologías para determinar sus características mediante la implementación de algoritmos de Inteligencia Artificial, algunos relacionados con la diferenciación de café orgánico de café no orgánico mediante la determinación de oligoelementos presentes con resultados precisos (98.2 %) con el algoritmo Naïve Ba para ocho minerales Barbosa et al., 2014, local lavado o de exportación de local sin lavar o de exportación en grano de café crudo logrando un 89.1 % de precisión Wallenlign et al., 2019. Otros autores han utilizado el color para determinar defectos Chou et al., 2019 , Oliveira et al., 2016, grado de tueste Leme et al., 2019, y humedad Virgen-Navarro et al., 2016 a través de diferentes algoritmos de Inteligencia Artificial, logrando resultados precisos del 97 % Hendrawan et al., 2019. Las características más evaluadas están relacionadas con atributos sensoriales en café tostado Livio y Hodhod, 2018; Ribeiro et al., 2012, café tostado en escamas bebida esterilizada Goyal y Goyal, 2013, café de civeta Wakhid et al., 2020, Nespresso Viejo et al., 2021 y cafés especiales Chang et al., 2021 mediante la aplicación de diferentes metodologías como: Mínimos cuadrados parciales, regularización bayesiana, sistema experto difuso, árbol de decisión y redes neuronales convolucionales. Zhang et al., 2016 Elaboró un sistema en el que la espectroscopia de transmitancia del infrarrojo medio se combinó con un algoritmo de reconocimiento de patrones para determinar cuatro variedades de café de China, concluyendo que seis modelos de clasificación tenían una precisión del 95 %. En Arboleda, 2018 se combinaron la NIRS y la red neuronal artificial de retropropagación (FFBPANN) para determinar la calidad del café Robusta y el café de civeta, y se obtuvieron resultados con una precisión del 99.98 % para distinguir los granos de café de civeta de los que no lo son. Por último, Berardinis et al., 2020a combinó características objetivas con subjetivas para evaluar las características más óptimas mediante la implementación de tres modelos diferentes, es decir, Random Forest, Support Vector Regression y Multi-Layer Perceptron, concluyendo que Random Forest fue el mejor modelo para el sistema. La implementación de un sistema inteligente para la recomendación de los granos de café, es la herramienta aplicada bajo el esquema de modelos no supervisados para el objetivo de este trabajo. Usando el modelo de bosques aleatorios, se obtuvo un RMSE de 0.3562, así mismo el método de los k vecinos más cercanos para identificar los granos de café más similares. 1340 reseñas digitalizadas

de 1312 granos de café arábica y 28 robusta. Para cada revisión, el conjunto de datos muestra 44 características divididas en medidas de calidad proporcionadas por los revisores en una escala del 1 al 10, metadatos de frijoles y metadatos de granjas. Las 44 características se limpiaron y dividieron en características objetivas y subjetivas. Se adoptaron dos estrategias de selección de características: selección de características univariadas y selección de características basada en árboles para calcular la importancia de las características. Se experimentó con tres modelos diferentes que son adecuados para la regresión de múltiples objetivos, es decir, RF, SVR y un MLP. se obtuvo un RMSE promedio en todos los atributos de 0.3750. Como el RMSE más bajo alcanzado fue para el RF, este fue elegido como el mejor modelo para el sistema Berardinis et al., 2020b. Oliveira et al., 2016 Este estudio presenta un sistema de visión por computadora para analizar y clasificar los granos de café verde con base en técnicas de inteligencia computacional. 12050 g de muestras (30 por color) de los siguientes grupos: verde blanquecino, verde, verde caña y verde azulado. Un sistema de visión por computadora-MLP feed - forward una capa oculta con 5 neuronas y función de activación no lineal y una neurona de salida lineal entrenado con Levenberg-Marquardt para datos de entrenamiento:  $1.2 \pm 1.24$  por ciento y para la validación  $1.125$  por ciento  $1\text{ cm}^{-1} \pm$  por ciento. Los resultados muestran que el sistema desarrollado permite realizar con precisión y determinar objetivamente el color de los granos de café en el color CIE L \* a \* b \* (estándar de medición de color para los alimentos). El sistema desarrollado es capaz de distinguir los colores de los granos de café verde logrando resultados en consonancia con los expertos en café que clasifican los granos de café utilizando inspección visual. El método se puede extender a otras aplicaciones dentro de la caracterización de calidad en los alimentos a escala industrial que permitan una mejor caracterización de los mismos Ribeiro et al., 2012 El uso de modelos de regresión (PLS) para la cata de café utilizando 10 g de café tostado y molido en 100 ml de agua caliente, permitió evaluar la acidez, el amargor, el sabor, la limpieza (taza limpia), el cuerpo (sensación en boca) y la calidad general del mismo. La calidad e intensidad de cada atributo se evaluó simultáneamente utilizando una escala que variaba de 1 a 5, perfiles cromatográficos y características subjetivas (seis atributos sensoriales (acidez, amargor, sabor, limpieza, cuerpo y calidad general)) café tostado "Se seleccionó la variable llevado a cabo por el ordenado. El método de selección de predictores (OPS). Mínimos cuadrados parciales (PLS) fue el método de regresión utilizado para modelado. "Siete compuestos volátiles (6.1 por ciento) son importantes en la construcción de todos los modelos. Otros catorce (12.2 por ciento) se utilizaron en cinco modelos, y 18 compuestos más (15.6 por ciento) fueron importante para la construcción de 4 modelos de regresión. Por lo tanto, el 34 por ciento de los picos se seleccionaron para al menos cuatro modelos de regresión, lo que demuestra que la mayoría de estos compuestos son importantes para proporcionar estas características sensoriales. Los errores de predicción de estos modelos, utilizando 4 o 5 variables latentes, fueron iguales a 0.28, 0.33, 0.35, 0.33, 0.34 y 0.41, para cada uno de los atributos. Goyal y Goyal, 2013 La importancia de la calidad del café, ha dado lugar a la investigación y desarrollo de algoritmos para determinar la calidad sensorial del café tostado. En el estudio de la investigación el análisis de diferentes algoritmos (feedforward, ANN, Algoritmo de Levenberg-Marquardt, algoritmo de gradiente conjugado de actualización de Polak-Ribiére, algoritmo de gradiente conjugado de actualización de Fletcher-Reeves, algoritmo de descenso de gradiente con tasa de aprendizaje adaptativo, algoritmo de

cuasi-Newton BFG, Powell-Beale) fueron determinantes para seleccionar el mejor algoritmo para la predicción sensorial ambos modelos fueron desarrollado para probar la eficiencia de capas dobles ocultas para la predicción de la calidad sensorial del café tostado con sabor. El algoritmo que proporciono mejores resultados fue el de regularización bayesiana. El estudio demostró que las redes feedforward con dos capas ocultas aprenden relaciones complejas más rápidamente que las redes de una sola capa, y son mejores en la predicción de la calidad sensorial de bebida esterilizada con sabor a café tostado. Walleigh et al., 2019 La clasificación de imágenes para determinar la calidad de granos de café fue realizada bajo el modelo CNN (redes neuronales convolucionales). Las imágenes se recopilaron en dos rondas con nueve meses de diferencia utilizando diferentes dispositivos de captura. El primer conjunto de las imágenes se recopiló durante la temporada de cosecha, lo que resultó en un conjunto de datos dominado por muestras de la categoría. Considerando que, el segundo conjunto se recolectó cuando se clasificaron los frijoles de la cosecha del año anterior lo que da como resultado más muestras de la categoría sin lavar. El modelo citado de CNN clasificó los granos de café crudo en 12 grados de calidad y alcanzó el 89.1 por ciento. Barbosa et al., 2014 La determinación de oligoelementos en muestras de café se realizó mediante una espectrometría de masas con plasma acoplado inductivamente ICP-MS (ELAN DRCII, PerkinElmer, CT, EE. UU.) Con argón de alta pureza (99.999 porciento, White Martins, Brasil). Sucharov y Suchara (2006) proporcionan los parámetros instrumentales y las condiciones optimizadas. Se utilizó el aprendizaje automático para clasificar como orgánico o no orgánico dependiendo de los oligoelementos de café orgánico. La clasificación de las muestras se realizó mediante el uso de tres algoritmos de minería de datos, denominados Multilayer Perceptron (MLP), Support Vector Machine (SVM) y Naïve Bayes (NB). En general, se encontraron niveles más altos de elementos químicos en el café convencional en comparación con las muestras orgánicas. Los resultados se consideran precisos con un resultado del 96.3 por ciento cuando todas las variables eran consideradas, sin embargo, se obtuvo mejor resultado por encima del 98 por ciento de precisión cuando solo se realizaba con ocho minerales seleccionados por el algoritmo de selección. Los minerales son Zn, Se, Ba, U, Tl, Th, Dy y Mo. Se determinó que las técnicas de minería de datos se pueden utilizar como una herramienta poderosa y versátil para la evaluación de la autenticidad de muestras de café orgánico. Wakhid et al., 2020 Se obtuvo el rendimiento de varias combinaciones de algoritmos de clasificación y parámetros estadísticos. Los datos sin procesar se obtuvieron de la matriz de sensores en el dispositivo e-nose. Los datos sin procesar estaban en forma de señales digitales, que se extrajeron para obtener los valores de los parámetros estadísticos. Para obtener resultados predictivos, los datos se ingresaron en varios algoritmos de clasificación. Se evaluaron los resultados de la clasificación para determinar la precisión y el rendimiento de los métodos y parámetros utilizados. Este proceso se repitió utilizando diferentes combinaciones de métodos de clasificación y parámetros estadísticos para obtener la combinación de mejor rendimiento. Los métodos utilizados para la clasificación fueron logísticos regresión, máquina de vectores de soporte (SVM), decisión clasificador de árbol (DTC) y clasificador de Bayes ingenuo. El mejor rendimiento se obtuvo mediante la combinación de la clasificación del árbol de decisiones algoritmo con los parámetros de desviación estándar y promedio al comparar 6 clases de café (97 por ciento). Al comparar 2 clases, el nivel de precisión fue del 100 por ciento. El número de clases de comparación influye en el nivel de

precisión del método. Chou et al., 2019 El uso de aprendizaje profundo (GALDAM) fue una de las técnicas usadas para determinar los defectos en café verde en su proceso de producción. Cinco mecanismos clave, incluido el diseño de la red profunda de inspección sensible a defectos (DIDN), el diseño de la red profunda de cambio de frijol (BSDN), el marco basado en GAN para el aumento de datos etiquetados, el optimizador basado en GA para la GAN propuesta marco, y la generación de imágenes de frijol de baja tasa de detección, se presentarán en las siguientes subsecciones. Estos mecanismos funcionan en dos redes profundas críticas, DIDN y BSDN, y un optimizador estructurado GAN, y colaboran en generar un modelo DIDN calificado para la inspección de defectos con imágenes limitadas de frijoles etiquetados por humanos, "defectos verdes" esquema de inspección de frijoles defectuosos basado en aprendizaje profundo (DL-DBIS), junto con datos etiquetados automatizados estructurados por GAN (red generativa-adversaria) método de aumento (GALDAM) para mejorar el esquema propuesto. Los resultados de las pruebas de un estudio de caso revelan que el esquema propuesto puede generar modelos de manera eficiente y efectiva para identificar granos defectuosos con precisión y valores de precisión de hasta el 80 por ciento. Livio y Hodhod, 2018 Se demostró que a través de AI Cupper es posible clasificar el café utilizando sistemas expertos difusos. El sistema se compone de tres capas combinadas que han logrado el uso de términos lingüísticos (bajo, medio, alto y muy alto) que representan las puntuaciones de entrada de los atributos individuales. La segunda capa es el Fuzzy Expert System que es, de hecho, la capa de razonamiento. Esta capa recibe los puntajes traducidos de la interfaz de usuario y los procesa con su motor de inferencia / razonamiento. Luego, pasa este subconjunto al Defuzzifier para traducir el subconjunto en un valor nítido agregado, el resultado del puntaje de calificación de calidad del café. AI Cupper fue probado ingresando setenta y tres tipos de granos de cafés, juzgados por nueve catadores. Los resultados de las pruebas AI Cupper son muy prometedores, ya que hubo un 95 por ciento coincidencia en todas las calificaciones de calidad del café. Hendrawan et al., 2019 El procesamiento de imágenes digitales y el modelo ANN pueden ser una herramienta en la detección del porcentaje de fenol total, pH y pureza del café Luwak. Las capas de entrada incluyen colores y características de textura. La capa de salida expresa el porcentaje de pureza, fenol total y pH en el café Luwak. El diseño de la mejor topología ANN se logra a través del análisis de sensibilidad con una variedad de funciones de aprendizaje; función de activación; tasa de aprendizaje e impulso (0.1, 0.5, 0.9); capa oculta (1, 2); nodo de capa oculta (10, 20, 30, 40) con la validación más baja del parámetro MSE. Este estudio formula las 3 funciones de activación, es decir, purelin, tansig y logsig. Pureza, fenol y pH Café Luwak en grano verde ANN 5 entradas, 30 nodos en la capa oculta 1, 40 en la capa 2 y 3 salidas, la mejor estructura es la 5-30-40-3 con una tasa de aprendizaje de 0.1 y un impulso de 0.5, que produce una validación de MSE de 0.0442. La estructura ANN seleccionada produce un entrenamiento R de 0.99502; R validación de 0.97933 y entrenamiento MSE de 0.0085; y validación de MSE de 0.0442. Virgen-Navarro et al., 2016 El uso de un sistema de monitoreo de color para determinar el contenido de humedad del grano durante el tostado usando un modelo ANFIS. Este modelo permitió estimar el contenido de humedad del frijol que se relaciona con el grado de tueste. El dispositivo construido para este propósito puede sustituir a los instrumentos convencionales para medición de color sensible a altas temperaturas. La metodología se implementó con imágenes digitales mediante el proceso de secado a tres temperaturas diferentes (400, 450 y

500 °C) cada cinco minutos. El modelo ANFIS consiste en: una red de retroalimentación para buscar reglas de decisión difusas que se desempeñen correctamente en una tarea dada, un conjunto de reglas de tipo IF-THEN Takagi-Sugeno, y un sistema de inferencia difusa. El parámetro de las funciones de pertenencia se ajusta mediante un algoritmo de retropropagación y el sistema difuso aprende de la información modelada. Se observó un alto coeficiente de determinación para los tres niveles de temperatura del aire; además, el círculo RMSE revela desviaciones mínimas del comportamiento experimental (RMSE <0.002). A pesar de los buenos resultados de ajuste, algunas mejoras pueden ser significativas. En esta investigación se utilizó la propagación hacia atrás como un algoritmo de aprendizaje; una opción interesante para explorar podría ser otro método adaptativo como Levenberg-Marquardt para acelerar la convergencia y reducir tiempo de entrenamiento. Leme et al., 2019 Se plantea un modelo de transformación de color que convierte el espacio de color RGB -Red Green Blue- al CIE L a,b independiente utilizando redes neuronales artificiales -ANNs-, considerando imágenes de granos de café enteros y molidos. Además, se propone un sistema de reconocimiento de grados capaz de predecir el valor Agtron / SCAA de cada muestra, utilizando modelos de regresión polinomial y ANN para imágenes en color y en escala de grises. Grado de tostado del café -color- Se seleccionaron dos tipos de cafés comerciales, el primero siendo procesado en seco y el segundo procesado en húmedo, fue posible implementar seis tipos de modelos, siendo cinco modelos polinomiales y un modelo ANN. Los modelos polinomiales se ajustaron mediante el estimador de mínimos cuadrados, mientras que el modelo neural se estimó mediante el algoritmo de retropropagación con ajuste de peso mediante el algoritmo de Levenberg Marquardt con regularización bayesiana. Los resultados del modelo de aproximación polinomial en escala de grises -P2G- obtuvieron el mejor ajuste, donde la prueba de Tukey se realizó con un 95 por ciento de confianza. El error absoluto medio encontrado por este modelo fue  $0.9361 \pm 1.3097$ , aunque las redes neuronales artificiales alcanzaron un mejor rendimiento para los granos enteros, un modelo simple como el P2G -modelo polinomial de segundo orden con valores de escala de grises- pudo estimar el grado de tostado de los granos molidos con buena precisión. Esto puede explicarse por el hecho de que las imágenes de las muestras de café molido son mucho más homogéneas que las de las muestras completas. Arboleda, 2018 La combinación de NIRS y ANN fueron un enfoque para la clasificación del café de algalias. Los resultados experimentales con ANN indican que él es viable clasificar civet de no civet café con un grado de precisión de 95 a 100 porciento. Los resultados se pueden interpretar que hay diferentes componentes en el café de algalias que lo hace diferente del café sin algalias. La primera parte es el uso de NIRS para adquirir valores de absorbancia en una longitud de onda elegida de 85 muestras de un solo grano de civeta y 85 muestras de un solo grano de café sin civeta. La segunda parte es la clasificación de los valores de absorbancia por propagación de retroalimentación neuronal artificial red (FFBPANN) calidad del café El primer grupo consistió en granos de Robusta ofrecidos a un gato de algalias enjaulado y el segundo grupo fue el grupo de control o el café sin algalias. Un total de 65 muestras por grupo fueron sometidas a múltiples entrenamientos hasta un nivel general aceptable se alcanzó una precisión superior al 99 por ciento. Se utilizaron muestras de prueba para evaluar si la red puede clasificar qué muestras pertenecen a sus respectivos grupos. La salida de la ANN es el número de clasificaciones correctas. Un total de 130 muestras de entrenamiento, 40 muestras de prueba y 48 muestras de validación se uti-

lizaron en este estudio y se sometieron a FFBBANN, que consta de 4 entradas, 2 capas ocultas y 1 capa de salida hasta que se alcanzó una precisión general del 99.98 por ciento. La red ANN entrenada se validó diez veces utilizando diferentes combinaciones de conjuntos. de 20 muestras por validación de 48 muestras de validación y se logró una precisión que oscila entre el 95 y el 100 por ciento. Aunque estos estudios difieren en cuanto a los instrumentos y el algoritmo utilizado, sus resultados corroboran el resultado de este estudio, que el café civet y no civet son diferentes. Corrales et al., 2014 La detección de la roya en cultivos debido a sus condiciones climáticas, variedad de suelos, cultivos etc, da lugar a proponer una Red neuronal capaz de solucionarlo. 21 atributos que se dividen en 4 categorías: condiciones climáticas (6 atributos), propiedades de fertilidad del suelo (5 atributos), propiedades físicas del cultivo (6 atributos) y manejo del cultivo (4 atributos). Promedio de humedad relativa en los últimos 2 meses. 2 Horas de humedad relativa > 90 por ciento en los últimos meses. 3 Promedio de variación de temperatura en el último mes. 4 Días de lluvia en el último mes. 5 Precipitación acumulada en los últimos 2 meses. 6 Precipitación acumulada nocturna en el último mes. Propiedades de fertilidad del suelo 7 pH. 8 Materia orgánica. 9 K. 10 Ca. 11 Arcilla. Propiedades físicas del cultivo 12 Variedad de café. 13 Densidad de plantas por hectárea. 14 Espaciamiento de plantas. 15 Espaciamiento de surcos. 16 Edad de cultivo. 17 Porcentaje de sombra. Manejo del cultivo 18 Control de la roya del café en el último mes. 19 Control de la roya del café en los últimos 3 meses. 20 Fertilización en los últimos 4 meses. Regresión de vectores de soporte de roya del café, Red neuronal de retropropagación y Árbol de regresión M5 la evaluación de los clasificadores utilizados para DS3 es mejor. Esto se debe a que presenta una correlación positiva más cercana a 1, con respecto a los valores obtenidos para DS1 y DS2, donde SVR ( $r = 0.4705$ ) es el mejor resultado. Viejo et al., 2021 La tecnología de bajo costo, son herramientas que se vuelven necesaria en el área de calidad del café. Se presentan los modelos e-nose y ML de bajo costo desarrollados en este estudio. Máquinas de monodosis de café para evaluar los posibles cambios en los perfiles e intensidades de aroma originales. Predice los aromas del café y la intensidad de tostado utilizando un e-nose portátil (inalámbrico) de bajo costo recientemente desarrollado junto con el modelado de aprendizaje automático (ML). El modelo 1 se desarrolló utilizando los datos de e-nose como entradas para clasificar las muestras en baja, media y alta intensidad de acuerdo con lo informado en su etiqueta. Por otro lado, el Modelo 2 se desarrolló utilizando las salidas de e-nose como entradas para predecir la abundancia relativa basada en el área de pico de 45 aromas medidos usando GC / MS utilizados como objetivos aroma e intensidad de tostado cafés nespresso Un total de 25 ML supervisados clasificadores disponibles en Matlab® R2020b Classification Learner en Statistics and Machine Learning Toolbox 12.0, que constan de árboles de decisión (tres algoritmos), análisis discriminante (dos algoritmos), regresión logística (un algoritmo), Bayes ingenuo (dos algoritmos), vector de soporte máquina (SVM; seis algoritmos), k clasificadores de vecino más cercano (KNN; seis algoritmos) y clasificadores de conjuntos (cinco algoritmos, se probaron (datos no mostrados). Junto con esto, 17 algoritmos de redes neuronales artificiales (ANN), se probaron utilizando un código escrito en Matlab® R2020b para evaluar todos los diferentes algoritmos de entrenamiento en un bucle y encontrar los modelos más precisos basados en el coeficiente de correlación / precisión y rendimiento. y fue del 98 por ciento con etapas de entrenamiento y prueba que resultaron en 100 y 94 por ciento, respectivamente. El modelo 2 tuvo una precisión muy alta en el

entrenamiento, validación y etapas de prueba basadas en el coeficiente de correlación ( $R = 0.99$ ,  $R = 0.98$  y  $R = 0.99$ , respectivamente), con una precisión global  $R = 0.99$ . Para el Modelo 1, los mejores resultados se obtuvieron utilizando ANN de reconocimiento de patrones con los valores medios de diez secciones del segmento más alto y estable de la curva de cada sensor como entradas para clasificar las muestras en i) bajo (3-5), ii) medio (6-9) y iii) alto (10 –13) intensidad. El algoritmo de entrenamiento de Levenberg -Marquardt fue seleccionado como el mejor basado en el coeficiente de correlación más alto ( $R$ ) y el rendimiento sin signos de sobreajuste. Arana et al., 2015 El objetivo principal de la investigación se centra en el diseño de un sistema experto que reconozca los cafés 100 porciento Arábica y su clasificación por grupos, colombianos y otros. Así mismo la obtención de huellas dactilares de alta calidad (método RMN) en el modo de automatización para extractos en metanol no deuterado utilizando un enfoque de presaturación múltiple. Las muestras analizadas fueron proporcionadas por Almacafé SA, Bogotá (Colombia) como material de café finamente molido, en total 192. De las muestras se obtuvieron sus espectros y de acuerdo a la experimentación se analizaron por corrida y repetición para su elección al azar y monitoreo constante, para resultados con valores atípicos se programaron repeticiones. De los 334 espectros resultantes, 20 se eligieron al azar antes de cualquier análisis para constituir un conjunto de validación, mientras que los datos restantes se utilizaron para la construcción del modelo (conjuntos de entrenamiento y prueba). La preparación de las muestras fue mediante extracción a temperatura ambiente de 0.2 g de café en polvo en 1 mL de metanol de calidad cromatográfica. La experimentación se llevó a cabo en un espectrómetro Bruker de 400 MHz equipado con un cabezal de sonda directo con gradientes, un accesorio de ajuste y ajuste automático, un BVT-1000 y BCU-1 para la regulación de la temperatura y un cambiador de muestras Sample Express. La medición de las muestras fue a una temperatura de 300 °K. La experimentación se llevó a cabo por medio de tres métodos, antes de cada ejecución (16 muestras), se calibró la temperatura del imán utilizando una muestra de metanol y la reproducibilidad y calidad de los espectros comprobados con un tubo de referencia. Para la detección de los valores atípicos se usó un análisis PCA. La clasificación de las muestras se llevó a cabo por análisis de quimiometría y por último la validación de los modelos se usaron para clasificar las muestras para el conjunto de validación externa. La determinación de la huella digital del café por la técnica de RMN sirve para discriminar el café de un país con respecto a otros países, incluido un bajo costo para realizarlo.

# CAPÍTULO 2: Estudio de las características físicas y los defectos del café verde que influyen en las notas sensoriales mediante modelos de Machine Learning

En este capítulo se dan a conocer las técnicas que se utilizaron para determinar las variables de importancia y analizar cómo influyen en las notas sensoriales. También se desarrollaron algoritmos automatizados enfocados a la predicción de las notas sensoriales del café utilizando diferentes técnicas de Machine Learning, y cada una de estas técnicas se autoajustaron con hiperparámetros para encontrar los parámetros que dan mayor precisión a cada algoritmo. Como resultados finales, se obtuvo la información de qué algoritmo y sus hiperparámetros tienen mayor precisión para la determinación de determinada nota sensorial en la fragancia, aroma, sabor y resabio.

## 2.1 Introducción

Las características físicas del café verde son esenciales para determinar su calidad. Se considera que existen propiedades físicas tales como volumen, densidad, masa y porosidad en el café verde que constituyen diferencias altamente significativas «Discrimination of defective (Full Black, Full Sour and Immature) and nondefective coffee beans by their physical properties», 2014. La discriminación de café defectuoso y no defectuoso influye significativamente en su calidad. Un producto de mala calidad priva de la venta en el mercado, por lo que la evaluación de los productos antes de ir al mercado requiere una técnica de selección Chandrasekar y Viswanathan, 1999; se presenta una metodología que, utilizando atributos físicos y químicos específicos, permite diferenciar los tipos de granos analizados para detectar defectos y no defectos. Los tipos de café Arábica y Robusta tienen propiedades físicas como el tamaño, la densidad, la masa y la resistencia. Una variante de la técnica aplicada se demostró en la resistencia al aplastamiento de los granos entre un café y otro. Otra característica a destacar fue la humedad, que interfirió en los resultados para ambas variantes de café Sinaga y Julianti, 2021. Las características físicas del café difieren según su tipo; la técnica de separación del color según su tueste sólo fue eficaz para el café robusta, determinando una mejor calidad Wondimkun et al., 2020. Los métodos aplicados en las regiones debido a sus condiciones geográficas y climas son necesarios para controlar la calidad del café. Un estudio demostró que la calidad física de los granos de café verde Gayo obtenidos mediante el método indonesio de procesado semilavado, comparado con el método de beneficiado totalmente lavado (método tradicional) ofrece resultados preliminares para el método de beneficiado, su grado de tueste y la calidad del café ensayado Mendonça et al., 2009. Según los resultados, el proceso de tostado influye directamente en las propiedades físicas del café robusto Sidikalang, y los diferentes colores del café cereza, se determinó que los resultados cambian con respecto a la condición inicial de los granos, incluyendo cambios en el contenido de agua, la masa, el valor de porosidad y la densidad aparente Yusibani et al., 2022. Entre los componentes relevantes para determinar la calidad de los alimentos están las características sensoriales porque estas características, como el sabor, los defectos y la textura, determinan cómo los consumidores perciben los productos Titova et al., 2015. En cuanto

al café, una de las bebidas más consumidas en el mundo, tuvo un consumo global de 166 millones de sacos de 60 kg en 2020 ICO, 2021, de los cuales su calidad ha sido evaluada por expertos catadores que determinan su aspecto, textura, sabor, aroma, etc. Feria-Morales, 2002, El uso de herramientas como apoyo para la determinación de características sensoriales es una opción para atender a situaciones como las pandemias, las cuales ponen de manifiesto la necesidad de contar con diferentes sistemas inteligentes para determinar si las características sensoriales cumplen los requisitos establecidos de calidad. Por ello, se han realizado investigaciones para establecer diferentes metodologías para determinar sus características mediante la implementación de algoritmos de Inteligencia Artificial, algunos relacionados con la diferenciación de café orgánico de café no orgánico mediante la determinación de oligoelementos presentes con resultados precisos (98.2 %) con el algoritmo Naïve Ba para ocho minerales Barbosa et al., 2014, granos de café en bruto lavados localmente o exportados de granos de café en bruto no lavados localmente o exportados, con una precisión del 89.1 % Walleigh et al., 2019, y granos verdes de café Luwak de los no Luwak, concluyendo con una precisión de validación del 97 % Hendrawan et al., 2019.

Otros autores han utilizado el color para determinar defectos Chou et al., 2019 , Oliveira et al., 2016, grado de tostado Leme et al., 2019, y humedad Virgen-Navarro et al., 2016 utilizando diferentes modelos de Inteligencia Artificial, consiguiendo resultados tan precisos como estos estudios. Por último, las características más evaluadas están relacionadas con los atributos sensoriales en café tostado Livio y Hodhod, 2018, Ribeiro et al., 2012, bebida esterilizada con sabor a café tostado Goyal y Goyal, 2013, café de algaria Wakhid et al., 2020, Nespresso Viejo et al., 2021, y cafés especiales Chang et al., 2021 mediante la aplicación de diferentes metodologías como como mínimos cuadrados parciales, regularización bayesiana, sistema experto difuso, árbol de decisión y redes neuronales convolucionales. Otros estudios han incluido más parámetros para describir la variedad y la calidad, por ejemplo, la referencia Zhang et al., 2016 elaboró un sistema en el que la espectroscopía de transmitancia del infrarrojo medio se combinó con un algoritmo de reconocimiento de patrones para determinar cuatro variedades de café de China, concluyendo que seis modelos de clasificación tenían una precisión del 95 %. La referencia Arboleda, 2018 combinó la espectroscopía del infrarrojo cercano (NIRS) y una red neuronal artificial de retropropagación (FFBPANN) para determinar la calidad del café en granos de Robusta y café de civeta, obteniendo resultados del 99 %, con una precisión del 98 % para distinguir los granos de café de civeta de los que no lo son. Por último, referencia Berardinis et al., 2020a combinó características objetivas con subjetivas para evaluar las características óptimas mediante la implementación de tres modelos diferentes, es decir, Random Forest (RFC), Support Vector Regression (SVR), y percepción multicapa (MLP), concluyendo que Random Forest era el mejor modelo para el sistema.

La investigación se centra en los granos de café verde y en cómo, a partir de las propiedades de variedad y tipo, sus propiedades físicas y sus defectos en los granos, puede hacerse una estimación de las propiedades sensoriales del café, que sólo podría obtenerse realizando el proceso de tostado y catación.

## 2.2 Marco Teórico

Los granos de café son producidos por una planta de la familia Rubiaceae y su género Coffea. Dentro de sus especies existen 4 de las cuales 2 son las más destacadas, conocidas como Arábica (*Coffea Arabica*) y Robusta (*Coffea canephora*). Existen diferencias entre cada tipo, las cuales son originadas por su clima, composición del suelo y altitud a la cual son cultivadas. Estos factores influyen de manera directa a la calidad del café.

El café es un producto estratégico en la economía de diversos países, entre ellos México, es considerado un factor determinante para el desarrollo de programas y apoyos al sector cafetalero. El desplome de los precios del café, ha derivado en la generación de estrategias que permitan aumentar la calidad del producto e incentivar sus costos, para ello CEDRSSA, 2018 propone generar un desarrollo óptimo de la producción mediante:

1. Campañas de fitosanidad con el fin de lograr el control de plagas que existen en la actualidad.
2. Intensificar programas de replantación y recepa en parcelas existentes.
3. Adquirir y transferir paquetes tecnológicos que permitan mejorar la capacidad de adaptación al cambio climático en las plantaciones.
4. Impulsar los sistemas de producción sostenibles.
5. Impulsar políticas públicas de la cafeticultura que sean incluyentes, sectorizadas y regionalizadas.
6. Impulsar las inversiones, incentivar la renovación de cafetales.
7. Incentivar la adopción de buenas prácticas agrícolas para lograr el control sostenible de la roya y reducir la incidencia de daños.
8. Impulsar el consumo de la producción nacional de café y aplicar medidas de certidumbre económica, financiera y comercial que garanticen el cumplimiento de los programas productivos agroalimentarios.
9. El apoyo al sector cafetalero a través de los diferentes programas, permitirá el crecimiento del consumo del café en sus diferentes variedades y Tipos.

### 2.2.1 Variedades de café

La variedad de café es determinante para su calidad en taza, estudios fundados en la producción del café en México han dado como resultado de su calidad en 20 variedades de café en Veracruz, México, durante un período de cinco años muestran que Catuai Amarillo, Caturra Rojo, Colombia Brote Café y Colombia Brote Verde tuvieron el mayor rendimiento promedio de frutas, mientras que Pluma Hidalgo 177 tuvo el mayor rendimiento agroindustrial de café cereza/pergamino. En la Tabla 2.1 se muestran las características de algunas variedades de café, ellas enfocadas a la morfología de su planta, su producción promedio y su perfil de taza. Pacamara tuvo el rendimiento de frutas más bajo, pero la proporción más alta de granos para preparar café estilo europeo y

Tabla 2.1: Variedades de Café arábica de portes bajo y alto evaluadas en Zentla, Veracruz.

\*Brote terminal con hojas verdes; †Brote terminal con hojas cafés; §Con tres ramas plagiotrópicas por nudo, en lugar de dos.

Porte bajo	Porte alto
1. Caturra Rojo 1	1. Typica 947
2. Catuai Amarillo 1	2. Bourbon Salvadoreño
3. Caturra Amarillo	13. Batie
4. Pacamara	14. Blue Mountain
5. Garnica F5	15. Typica Xhantocarpa
6. Garnica Iquimite	16. Pluma Hidalgo 177
7. Colombia Brote Verde (BV)*	17. Dessie
8. Colombia Brote Café (BC)†	18. Garnica (TCPA)§
9. Garnica (TCPB)§	19. Typica Tres Cruces (TC)§
10. Oro Azteca	20. Bourbon Tres Cruces (TC)§

los mejores atributos en olor, acidez y cuerpo. La mayoría de las variedades de café en Veracruz tienen un bajo rendimiento promedio de frutas en comparación con otros países.

El estudio también destaca la necesidad de evaluar nuevas variedades en diferentes sistemas de producción y condiciones agroecológicas. Además, se discuten varios aspectos de la producción de café, incluyendo los efectos de la altitud, la exposición a la pendiente y el rendimiento en la calidad del café, así como el valor económico de los recursos genéticos del café. Como parte también de los resultados se incluye una evaluación sensorial de 20 variedades de café en Veracruz, México García, s.f.

## 2.2.2 Tipos de Café

Actualmente los tipos de café son clasificados por su tipo de grano, hasta el momento se agrupan en cuatro categorías: Arábiga, Robusta, Libérica y Excelsa

### 2.2.2.1 Arábica

El café tipo arábica es uno de los más cultivadas a nivel mundial, tiene la característica de ser suave y agradable al gusto. Así como asociar aromas silvestres y frutos secos, complementado con un toque dulce y ácido a su vez. Su sabor es mucho más complejo y rico que el robusta. La complejidad del cultivo del café arábica es un factor a considerar ya que necesita sombra y un clima frío, mucha agua y una gran altitud para su crecimiento correcto, sus plantas suelen cultivarse a una altura que va de 500 hasta 2400 metros de altura.

### **2.2.2.2 Robusta**

El café tipo robusta es característico por ser más amargo que acido, se asocia con matices de frutos secos y madera, Siendo mucho menos afrutado que el arábica. Además, es poco perfumado, y junto con una textura más áspera y cremosa, lo hacen poco atractivo para los gustos más finos. Su cultivo es principalmente en África, Indonesia y Vietnam. Dentro de sus características físicas destacan sus granos más grandes y redondeados, sus plantas son más grandes que las del grano Arábica, oscilan entre 4 y 6 metros. Su crecimiento oscila entre los 500 y 1200 metros de altura y climas cálidos.

### **2.2.2.3 Libérica**

El café tipo libérica tiene su origen en África central, occidental y Liberia. Se caracteriza por su aroma floral picante y una gran variedad de sabores ahumados. Este suele definirse por la variedad de mezclas con otras variantes, con la finalidad de añadir cuerpo y complejidad. Sus cultivos se dan en lugares cálidos y húmedos, esta característica le permite crecer en lugares de baja altitud. Su crecimiento es superior al de Arábica y Robusta y su altura puede llegar a medir hasta 18 metros.

### **2.2.2.4 Excelsa**

El café Excelsa se considera una variante del café Libérica, su crecimiento está limitado al sudeste asiático, la planta tiene una altura que puede llegar a medir hasta 10 metros. Una característica exclusiva del café excelsa es su sabor el cual es considerado único, ya que combina notas ácidas y tonos afrutados. Son ligeros en aroma y cafeína y son difíciles de conseguir fuera de Asia.

### **2.2.3 Café verde**

Los granos de café verde son el fruto de la planta de la familia de las Rubiaceae y del género Coffea (Arábica y Robusta). Son conocidos como grano o café verde por ser frutos no tostados.

En la figura 2.1 se puede apreciar una muestra de café verde de Coatepec Veracruz, el cual es de tipo pergamino y cuya variedad es Typica-Costarica 95.

### **2.2.4 Composición Química**

El café, químicamente se compone de agua y materia seca. La materia seca de los granos del café almendra está constituida por minerales y por sustancias orgánicas que son los carbohidratos, lípidos, proteínas, alcaloides, como la cafeína y la trigonelina, así como, por ácidos carboxílicos y fenólicos, y por compuestos volátiles que dan el aroma a la almendra Quintero, 2011 . En la tabla 2.2 se muestran los componentes químicos de dos tipos de café (arábica y robusta) con porcentajes en base seca.



Figura 2.1: Muestra de café verde Coatepec-Veracruz, Tipo Pergamino, Variedad:Typica - Costa Rica 95

La composición química del grano de café verde (almendra) varía de acuerdo a su cultivo (tipo y variedad).

### 2.2.5 Propiedades Físicas

Las propiedades físicas del café nos permiten tener una valoración a través de la vista, permitiendo conocer sus defectos, así como su aspecto general.

1. Color: El color es la primera impresión que se tiene del café verde ya que su uniformidad de color en el grano da una pauta para el muestreo y su clasificación. La tonalidad de la muestra tiene una variedad de coloraciones, dentro de las que se encuentran; Azul-verde, Azulado-verde, Verde, Verdoso, Verde amarillento, Amarillento-pálido, Amarillento y Café.
2. Olor: Se refiere a la impresión olfativa de los granos tostados (olor limpio, olor extraño). Se valora de acuerdo a la tonalidad que presenta la muestra.
3. Humedad: Este factor es de gran importancia para la evaluación de un buen café, su rango ideal de humedad se encuentra entre 10 a 12 por ciento, se considera que esta característica es muy importante durante el proceso de tueste.
4. Tamaño: El tamaño del grano es uno es un indicativo de la calidad del café. Para el muestreo se hace uso de las zarandas que permiten ver el tamaño del grano, esta característica se puede también percibir de manera visual.

Tabla 2.2: Componentes químicos de café Arábica y Robusta

Componente químico	Arábica (%)	Robusta (%)
Polisacáridos	50.8	56.4
Sacarosa	8	4
Azúcares reductores	0.1	0.4
Proteínas	9.8	9.5
Aminoácidos	0.5	0.8
Cafeína	1.2	2.2
Trigonalina	1	0.7
Lípidos	16.2	10
Ácidos alifáticos	1.1	1.2
Ácidos clorogénicos	6.9	10.4
Minerales	4.2	4.4
Compuestos aromáticos	trazas	trazas

5. Imperfecciones: Las imperfecciones son apreciadas de manera visual y dentro de ellas se encuentra su color, brillo, tamaño y forma.
6. Muestreo: El muestreo se basa en una porción de 350 gramos de café verde que es analizada de manera estandarizada por normas de calidad y se enfoca en analizar todas sus propiedades desde el grano verde hasta la taza.

### 2.2.6 Catación

Es un método el cual es usado para conocer las propiedades sensoriales del café, también conocido como evaluación sensorial, a través de este se identifican los defectos del café, así como sus características sensoriales: acidez, aroma, dulzor, sabor y calidad global. Las personas encargadas de evaluar el café (catadores) también se conocen como panelistas, degustadores y jueces analíticos.

Para realizar una evaluación completa del café, los catadores se enfocan en tres aspectos relevantes: el olor (aroma, fragancia y sabor residual), el gusto (sabor, acidez y dulzura) y la sensación táctil o textura (cuerpo).

La cata consta básicamente de los siguientes pasos basados en la norma Mexicana, [S.F.](#)

Se colocan en la taza 8.25 gramos de café tostado molido (cuando mucho 5 min antes) por cada 150 ml de agua. Inmediatamente aspirar los gases sueltos (aroma seco o fragancia) Se coloca agua caliente a 93 °C de temperatura dentro de cada taza, inhalando inmediatamente los vapores (aroma húmedo). Deje reposar la infusión de tres a cinco minutos para permitir la extracción y dilución adecuadas. se forma una capa o costra

en la superficie de la taza que permite medir el carácter aromático. La capa o costra que se forma se rompe con una cuchara redonda inhalando profundamente los vapores que salen de la taza. Toda la espuma y las partículas se limpian y eliminan de la superficie. Después de 10-12 minutos, se coloca una cucharada de la bebida cerca de la boca y se aspira. La aspiración introduce vapor en la cavidad nasal y distribuye el líquido uniformemente por toda la lengua. La bebida debe mantenerse en la boca de tres a cinco segundos para percibir la intensidad y calidad de las características gustativas: sabor, acidez, dulzura, limpieza y equilibrio. Seguidamente se expulsa la bebida transcurrido este tiempo, en un recipiente destinado a tal efecto, evaluando la sensación que queda en boca tras la degustación para determinar el sabor residual. La lengua se desliza suavemente por el paladar para determinar la textura, el contenido de grasa y la intensidad. La evaluación del sabor debe realizarse en tres etapas: caliente, templada y fría, con el fin de evaluar la consistencia y uniformidad de la bebida. La evaluación total se refleja de manera individual para cada atributo de acuerdo al formulario de catación de acuerdo a la norma.

#### **2.2.6.1 Aroma**

“Es la percepción de las sustancias olorosas y aromáticas después de haberse puesto en la boca. El aroma es el principal componente del sabor de los alimentos”Sensorial y Cafe, 2015

#### **2.2.6.2 Sabor**

“Esta propiedad en el café es muy compleja, ya que combina tres elementos: olor, aroma y gusto. El sabor es lo que diferencia un alimento de otro ”Sensorial y Cafe, 2015.

#### **2.2.6.3 Acidez**

“Se describe como aquella sensación en la lengua que hace salivar. A menudo se describe como “brillante” cuando es favorable o “agria” cuando es desfavorable”Sensorial y Cafe, 2015.

Las muestras de café verde fueron proporcionadas por CAFECOL (Centro Agroecológico del Café A.C.). La vinculación fue un factor importante para la obtención de las mismas, ya que el tipo de variedad que se uso nos permitió obtener resultados de alto impacto a nivel de la investigación.

### **2.3 Materiales y métodos**

La investigación tiene dos temas principales de estudio: El análisis y determinación de las variables de importancia del café, y la estimación de las notas sensoriales a partir de las físicas y defectos de los granos de café verde. La Figura 2.2 muestra un diagrama esquemático que indica cómo, mediante el uso de algoritmos de Machine Learning (ML) y datos como las propiedades físicas de los granos de café verde, como la humedad, la densidad, los defectos de los granos, el tamaño y la forma, es posible determinar diferentes propiedades. Los

resultados de las catas de estas muestras de café también se utilizaron para realizar un análisis de cómo influyen estas variables en las notas sensoriales del café, así como la creación de un modelo basado en algoritmos de aprendizaje automático para estimar las propiedades sensoriales del café a partir de sus propiedades físicas y las deformaciones de los granos de café.

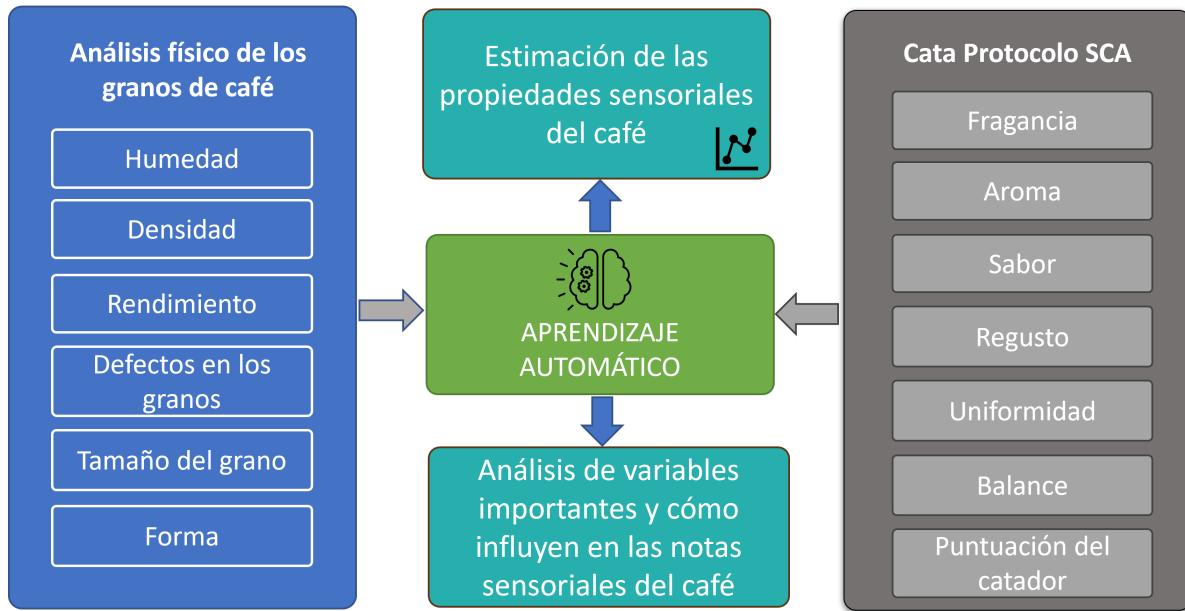


Figura 2.2: Diagrama esquemático de la metodología aplicada.

### 2.3.1 Muestras de café

Las variables físicas del café, como tipo, variedad, propiedades físicas y defectos, e información sensorial del café, como fragancia, aroma, sabor y regusto, se recolectaron de 185 muestras de diferentes variedades de café de especialidad y localidades de México. Las variedades, los tipos y la información sobre el perfil sensorial se obtuvieron de expertos certificados de una organización especializada en café (Centro Agroecológico del Café A.C. (CAFECOL)) y se describen a continuación:

1. Tipo: pergamino, verde, bola, miel;
2. Variedad: Tekisik, marago, sudan rome, mundo novo, sarchimor, cr95, criollo, garnica, typica, bourbon, costa rica, catuai, arabica, colombia, arainema, catura, catimor, marellesa, geisha, pacamara, garnica, oro azteca, jilotepec, miel negra, catuai rojo, amarillo catuai;
3. Propiedades físicas: natural, desmucilaginado, miel, lavado, húmedo, secado, mezcla, tamizado, planilla, secador guardiola, humedad, densidad, negra, amarga, cereza seca;
4. Defectos: hongos, materias extrañas, broca severa, parcial negro, parcial amargo, pergamino, flotante, inmadura, arrugada, cáscara, partida, cascarilla, ligera broca;

5. Calibre: Tamiz 19, tamiz 18, tamiz 17, tamiz 16, tamiz 15, tamiz bajo 15;
6. Forma: Frijol plano, peaberry, triángulo, monstruo, cáscara, bola, rendimiento, mancha;
7. Evaluación de fragancia, aroma, sabor y regusto: Floral, afrutado, herbal, frutos secos, caramelo, chocolate, especiado, resinoso, pirolítico, terroso, fermentado;
8. Resultados de la evaluación: Fragancia, sabor, residual, acidez, cuerpo, equilibrio, puntuación del barista y puntuación total.

La figura 2.3 muestra nueve variedades de café verde cultivadas en las altas montañas de Veracruz, México. Cada una de estas variedades tiene propiedades físicas, defectos y tamaños diferentes.



Figura 2.3: Variedades de café verde. (a) Bourbon, (b) Typica, (c) Colombia, (d) Caturra, (e) Costa Rica, (f) mundo novo, (g) Garnica, (h) Catuai, y (i) Marseillaise.

### 2.3.2 Metodología de la catación del café

La catación es un proceso estandarizado para evaluar el aroma, el sabor y la textura de una muestra de café.

Para realizar una evaluación completa del café, los catadores se centran en tres aspectos relevantes: olor (aroma, fragancia y sabor residual), gusto (sabor, acidez y dulzor) y sensación táctil o textura (cuerpo). La figura 2.4 muestra el proceso de preparación de una taza de café.



Figura 2.4: (a) Análisis de defectos en granos de café verdes, (b) preparación de la muestra de granos de café tostados para la cata, y (c) preparación de la taza de café y proceso de cata.

La cata consta de los siguientes pasos basados en la norma mexicana NMX-F-177- SCFI-2009 Mexicana, s.f.:

1. Se coloca una cantidad de 8.25 g de café tostado molido (máximo 5 min antes) en la taza por cada 150 mL de agua; inmediatamente se aspiran los gases sueltos (aroma seco o fragancia).
2. Aspirar inmediatamente los gases sueltos (aroma o fragancia seca).
3. Se coloca agua caliente a una temperatura de 92 °C dentro de cada taza; aspirar inmediatamente los vapores (aroma húmedo).
4. Dejar reposar la infusión de 3 a 5 min para permitir una extracción y dilución adecuadas.
5. En la superficie de la taza se forma una capa o costra que permite medir el carácter aromático.
6. La capa o costra que se forma se rompe con una cuchara redonda inhalando profundamente los vapores que salen de la taza.
7. Se limpia y se retira toda la espuma y las partículas de la superficie. Transcurridos ocho minutos se coloca una cucharada de la bebida cerca de la boca y se aspira.

8. La aspiración introduce vapor en la cavidad nasal y reparte uniformemente el líquido sobre toda la lengua.
9. Se debe mantener la bebida en la boca de tres a cinco segundos para percibir la intensidad y calidad de las características gustativas: sabor, acidez, dulzor, limpieza y equilibrio.
10. Transcurrido este tiempo, se expulsa la bebida en un recipiente destinado a tal efecto, evaluando la sensación que queda en la boca tras la degustación para determinar el sabor residual. La lengua se desliza suavemente por el paladar para determinar la textura, el contenido en grasa, e intensidad.
11. La evaluación del sabor debe realizarse en tres etapas: caliente, templada y fría, para evaluar la consistencia y uniformidad de la bebida.

### 2.3.3 Metodología de aprendizaje automático

Se presenta la metodología implementada para obtener los parámetros sensoriales del café. La sección de entrenamiento del modelo se centra en generar un modelo que se entrena mediante una metodología de aprendizaje supervisado utilizando las variables de tipo, variedad, propiedades físicas y defectos para estimar cada nota sensorial de aroma, fragancia, sabor y regusto, como se muestra en la Figura 2.5.

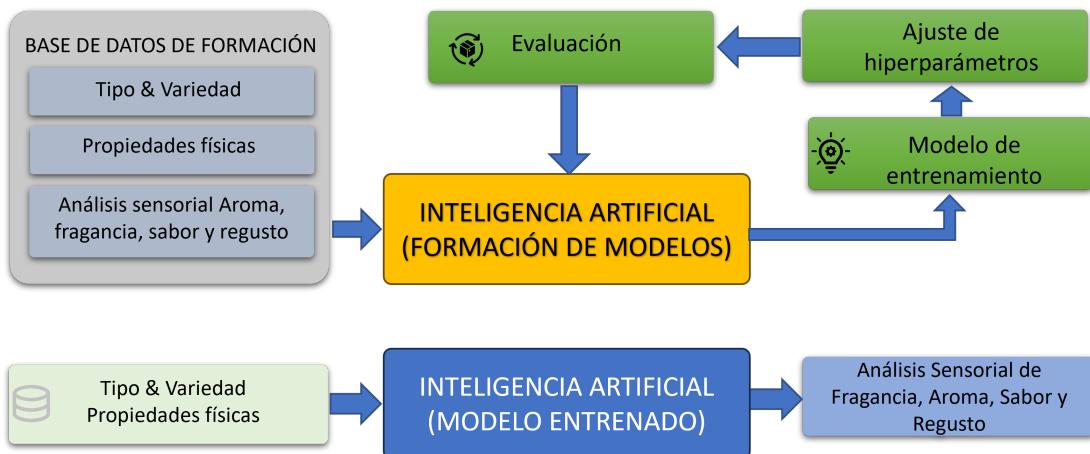


Figura 2.5: Diagrama esquemático de la metodología de Machine Learning implementada y utilizada para el análisis de variables importantes y la estimación de propiedades sensoriales.

El objetivo de la investigación es obtener información sobre qué variables de entrada en la base de datos afectan a la variable objetivo seleccionada. Por lo tanto, los modelos basados en que pueden lograr la estimación de parámetros sensoriales y de calidad del café. La investigación se dividió en la creación de la base de datos, los datos del preprocesamiento, la selección de variables, la validación cruzada, la importancia de las características, los modelos de aprendizaje automático y el ajuste de hiperparámetros.

1. Creación de bases de datos: La información que se utilizará incluye datos sobre las características físicas y sensoriales obtenidos de la evaluación del café. Se utilizó la librería Pandas en Python para gestionar todo lo relacionado con la carga, análisis y almacenamiento de la información en la base de datos.
2. Pre-procesamiento de la base de datos: La base de datos se analizó utilizando un algoritmo que encuentra datos que podrían estar corruptos, como campos vacíos y caracteres especiales. Cuando el algoritmo encuentra datos que presentan problemas, se realiza una eliminación automática para garantizar que los modelos de aprendizaje automático no tengan problemas durante su ejecución debido a estos tipos de datos.
3. Selección de variables: El usuario selecciona las variables a analizar, es decir, determina las variables que deben estar en la entrada y en el destino. El programa permite al usuario escribir los nombres de las variables a analizar para que el sistema automáticamente almacene, analice y agrupe la información en la búsqueda. Un segmento fundamental para obtener resultados lo más parecidos a los reales se consigue garantizando que los datos implementados para entrenar y probar son correctos reduciendo la posibilidad de sobreentrenar el modelo. Por ello, se implementa la técnica de validación cruzada de forma que el algoritmo implemente diferentes pruebas a través de diferentes combinaciones de nK pliegues por el usuario. La figura 2.6 muestra el esquema del segmento de validación cruzada.
4. Importancia de las características: El objetivo era obtener las variables relevantes que impactan positiva o negativas sobre la variable objetivo a analizar. Para ello, se utilizaron los siguientes modelos: Recursive Feature Elimination (RFE), CHI Square, least absoluto (LASSO), CatBoost (CBC), Árbol de Decisión (DTC), Random Forest (DTC), Random Forest (RFC), k-nearest neighbors (KNN), Regresión lineal (LR) y regresión logística (LogR). Todos ellos proporcionan las variables más importantes que influyen en la variable objetivo seleccionada. Para reducir el sobreentrenamiento de los modelos y garantizar que los datos obtenidos sean lo más precisos posible, se utilizó la técnica de validación cruzada.
5. Modelos de aprendizaje automático: Este segmento utiliza los datos de la sección de base de datos y ejecuta el clasificador de vectores de soporte (SVC), K-Nearest Neighbors (KNN), Linear lineal, árbol de decisión, bosque aleatorio, CatBoost, Naïve Bayes (NB) y modelos de regresión logística para predecir la calidad del café y la calidad de la bebida. para predecir la calidad del café y las variables sensoriales. Se realizó una validación cruzada y se ajustaron los hiperparámetros para obtener la mejor precisión posible en la figura 2.7 se muestra el diagrama representativo de la metodología de aprendizaje automático .

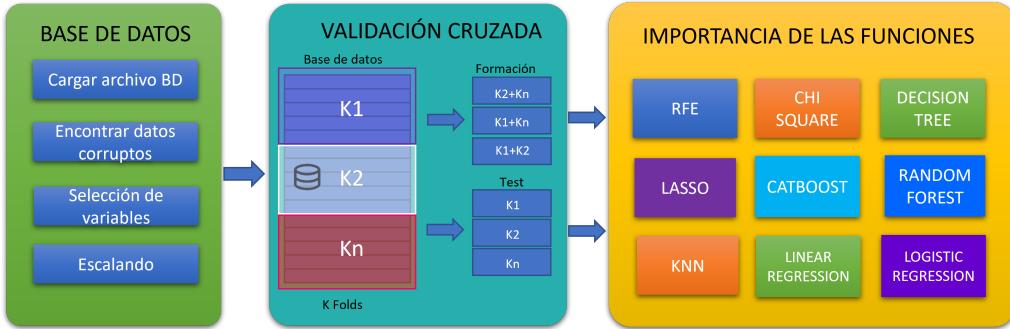


Figura 2.6: Diagrama esquemático que representa la metodología de aprendizaje automático mediante validación cruzada para la determinación de variables importantes.

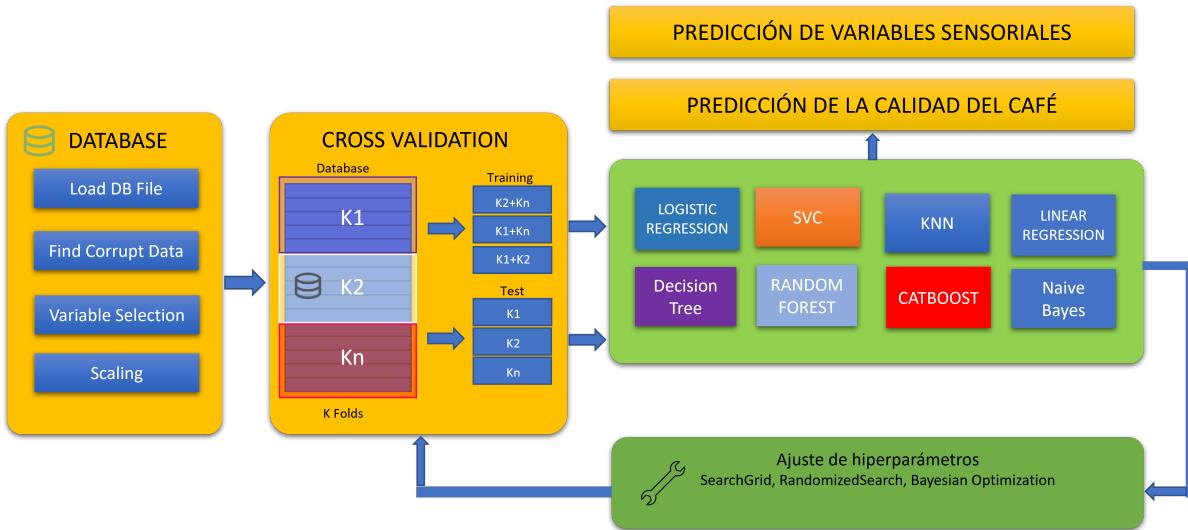


Figura 2.7: Diagrama representativo de la metodología de aprendizaje automático para la generación de modelos de predicción de las notas sensoriales del café.

## 2.4 Resultados

Se realizó un estudio en profundidad para determinar las variables del café que influyen en las variables de las notas sensoriales relacionadas con la fragancia, el aroma, el sabor y el regusto; los experimentos realizados fueron los siguientes.

1. Determinación de las variables más importantes relacionadas con las notas sensoriales del café verde a través de la variedad y el tipo de grano de café verde a través de la variedad y el tipo de granos de café. Se utilizaron ciento ochenta y cinco granos de café verde de la zona de alta montaña de Veracruz, México, cuya información contiene el análisis sensorial de la cata de café con respecto a la variedad y tipo de café. Por lo tanto, se construyó una base de datos con estas características para su análisis. El objetivo de

este experimento se centra en la obtención de datos para analizar la relación existente entre las variables sensoriales de la catación y su estrecha relación con la variedad y tipo de café analizado, con los que identificar las diferentes notas sensoriales de aroma, fragancia sabor y retrogusto que se ven influenciadas por el tipo y variedad de café.

2. Determinación de las variables más importantes relacionadas con las notas sensoriales del café verde a través de las propiedades físicas (humedad, densidad, defectos, tamaño y forma). Se utilizaron ciento ochenta y cinco granos de café verde de la sierra de Veracruz, cuya información incluye el análisis sensorial de la cata de café relativa a las propiedades físicas y defectos de los granos de café, que inciden directamente en las notas sensoriales de aroma, fragancia, sabor y retrogusto.

#### **2.4.1 Predicción de las notas sensoriales**

La predicción de las notas sensoriales del café en aroma, fragancia, sabor y regusto se realizó utilizando información sobre la variedad, el tipo, las propiedades físicas y los defectos de granos de café verde. En este experimento se utilizaron 185 muestras de café verde. Se utilizaron siete modelos de aprendizaje automático para predecir notas sensoriales como florales, afrutadas, herbáceas, a nuez, caramelo, chocolate, picante, resinoso, pirolítico, terroso y fermentado en el aroma, fragancia, sabor y regusto. Durante la cata de café, el catador experto asigna una puntuación que va de 0 a 5 a cada una de las notas sensoriales detectadas en las cuatro etapas de análisis sensorial (aroma, fragancia, sabor y regusto). Esta parte de la investigación se centra en el desarrollo de un algoritmo que pueda predecir las notas sensoriales del café mediante modelos de aprendizaje automático con aprendizaje supervisado configurado para la clasificación utilizando las variables de variedad y tipo, propiedades físicas y defectos de los granos de café verde. El resultado son valores categorizados de notas sensoriales.

#### **2.4.2 Resultados del análisis físico**

A continuación se presentan los resultados del análisis físico de algunas de las variedades como Bourbon, Typica, Colombia, Caturra, Costa Rica y Geisha. Los resultados indican el número medio de granos de café verde con algún defecto en el grano. Se puede observar en la Figura 2.8 que la rajadura es uno de los defectos más comunes en todas las variedades. El café Typica fue una de las muestras que contenía los defectos más característicos, que son inmaduro, arrugado, partido, broca y amargo parcial. Bourbon concentra sus defectos en Parcial negro, hongo, inmaduro y arrugado.

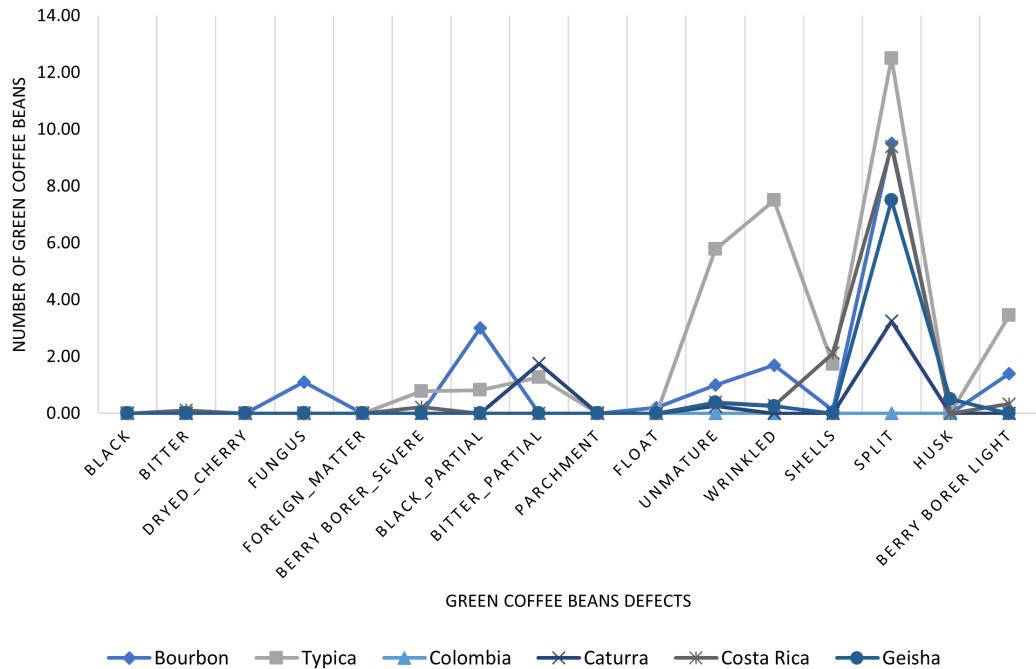


Figura 2.8: Resultados que indican el número de granos de café verde según su tipo de defecto.

La figura 2.9 muestra los resultados medios de las variedades de café Bourbon, Typica, Colombia, Caturra, Costa Rica y Geisha, indicando el número de granos de café verde según su tamaño que contiene cada variedad. El tamaño del grano es un factor importante en el proceso de tueste, ya que el objetivo es garantizar un tueste homogéneo que puede verse afectado si hay muchos granos de diferentes tamaños.

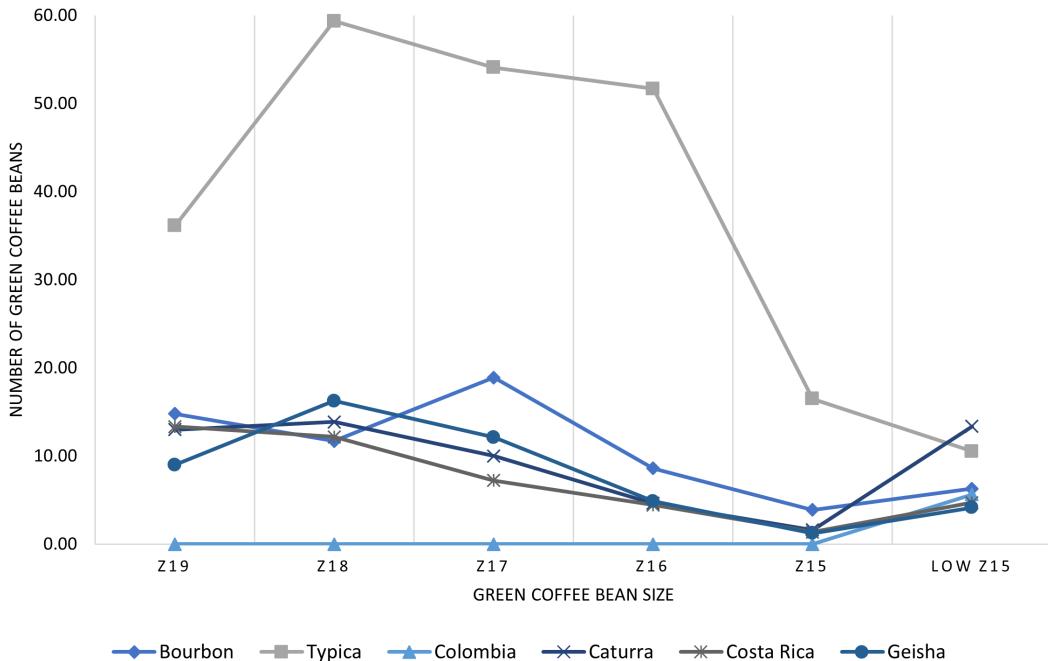


Figura 2.9: Resultados que indican el número de granos de café verde según su tamaño.

En Sunarharum et al., 2018 se describe un método para correlacionar las propiedades físicas de los granos de café verde Java Arabica de los granos de café verde Arábica de Java con sus defectos físicos mediante el uso de gráficos de caja para determinar la distribución de los defectos físicos bajo diferentes métodos de postcosecha y secado. Se concluye que se observan menos defectos en el proceso húmedo en comparación con el proceso seco, y el secado mecánico proporciona una mejor calidad de los granos de café verde y minimiza las pérdidas.

#### 2.4.3 Resultados de la cata de café

Los resultados obtenidos de la cata son los siguientes: notas sensoriales como afrutado herbáceos, frutos secos, caramelo, chocolate, especiados, resinosos, pirolíticos, terrosos, fermentados y fenólicos se analizan. Los resultados de la cata de café se muestran en las Figuras 2.10 y 2.11, donde se puede observar mediante mapas de calor el nivel de presencia de una determinada nota sensorial (azul-baja presencia, rojo-alta presencia). Los resultados se dividen en cuatro secciones: Aroma, fragancia, sabor y regusto. Según los resultados de la cata de aromas (Figura 2.10), las notas sensoriales de picante chocolate, caramelo, nuez, afrutado y floral son las notas más presentes en Marseillaise, Caturra, Arábica, Colombia, Costa Rica, Bourbon y Typica. Las notas herbáceas están muy presentes, especialmente en los cafés Typica, Bourbon, Costa Rica y Caturra. En cuanto a los resultados del aroma (Figura 2.10), las notas afrutadas, de nuez, caramelo y chocolate fueron las más características de los cafés, y chocolate fueron las más características de los cafés estudiados. Las notas resinosas, pirolíticas, terrosas y notas resinosas, pirolíticas, terrosas y fermentadas en los cafés Typica, Bourbon y Caturra.

Los resultados gustativos Figura 2.11 mostraron que las notas afrutadas, de nuez, caramelo, chocolate y especias fueron las más características de los cafés estudiados. Las notas pirolíticas, resinosas y terrosas estuvieron más presentes en los cafés Typica, Bourbon, Colombia y Caturra. Los resultados del retrogusto (Figura 2.11) muestran que todos los cafés tienen notas afrutadas, herbáceas, de frutos secos, caramelo, chocolate y especias. Las notas terrosas y fermentadas están presentes en Typica, Bourbon, Colombia y Caturra. En Sunarharum et al., 2018 se utiliza un gráfico de araña para determinar la influencia del proceso de secado en el promedio obtenido de la evaluación de cinco jueces sobre las notas sensoriales del aroma de granos de café verde Java Arábica. Encontraron que las muestras secadas mecánicamente son más de nuez, tierra y hierba que las muestras secadas al sol. En lugar de utilizar gráficos de araña, los cuales nos ayudan a determinar de forma rápida y coherente la presencia de notas sensoriales en siete variedades de café en los cuatro casos presentados anteriormente. Esta tarea sería inviable utilizando gráficos de araña.

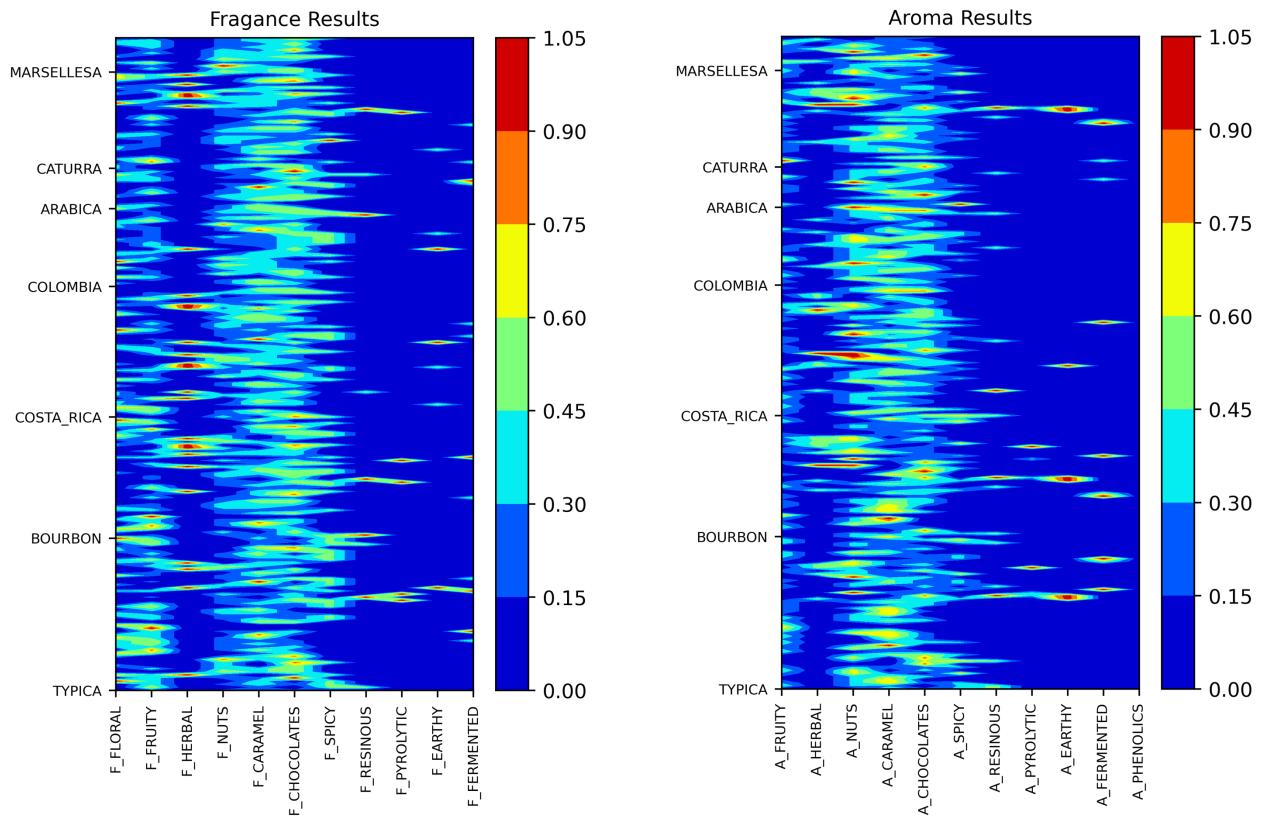


Figura 2.10: Resultados de las notas sensoriales de fragancia y aroma de la cata de café.

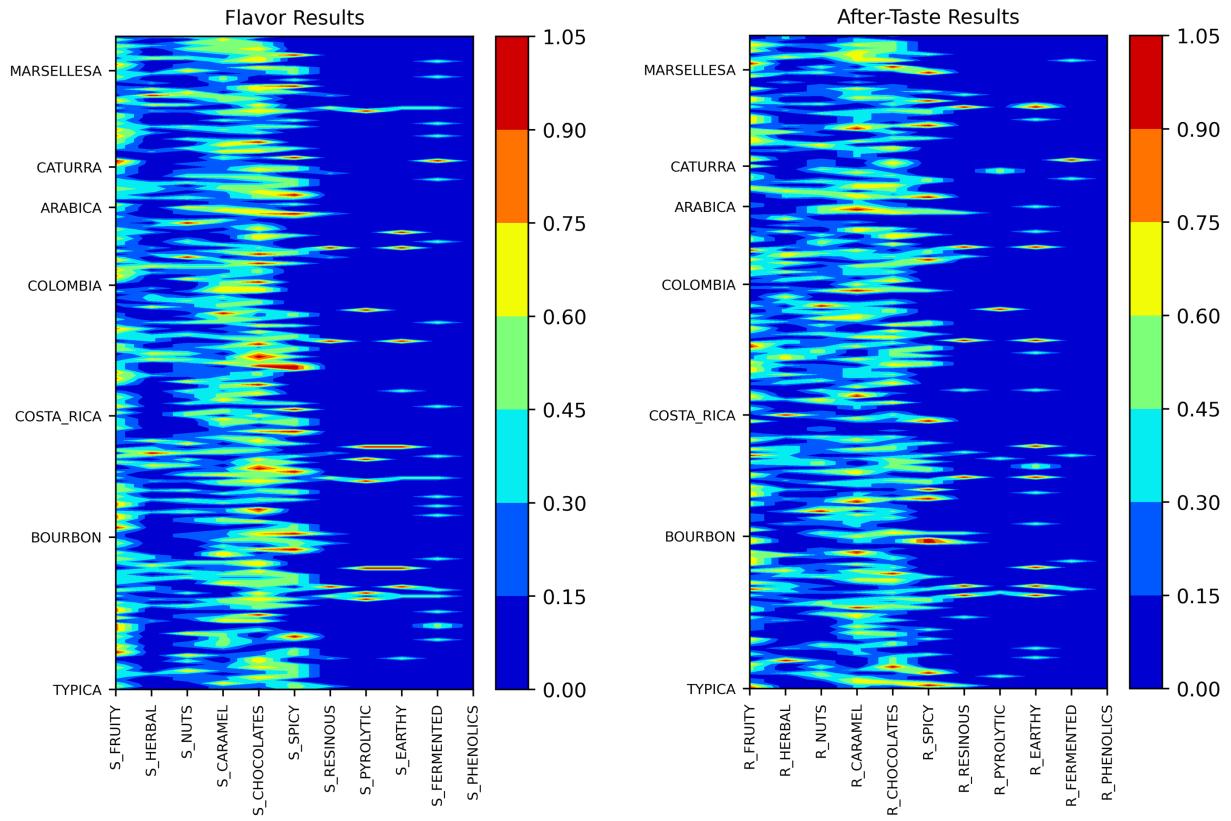


Figura 2.11: Resultados de las notas sensoriales del sabor y el regusto de la cata de café.

#### 2.4.4 Resultados de la importancia de las características del análisis sensorial

La figura 2.12 muestra el número de concordancias de las variables de importancia obtenidas por los métodos utilizados (RFE, LASSO, KNN, CHI2, CATBOOST, DECISION TREE y BOSQUE ALEATORIO). El resultado indica la relación e importancia que existe entre las variedades de café y sus respectivas notas sensoriales, como floral, afrutado, herbal caramelo, chocolate, especiado, resinoso, pirolítico, terroso, fermentado y fenólico, que se analizaron en el aroma, la fragancia, el sabor y el regusto. Las variedades y tipos que estaban más presentes en todas las notas sensoriales de las catas de aroma, fragancia, sabor y regusto fueron las variedades Bourbon, Costa Rica, Colombia, Typica y Maragón.

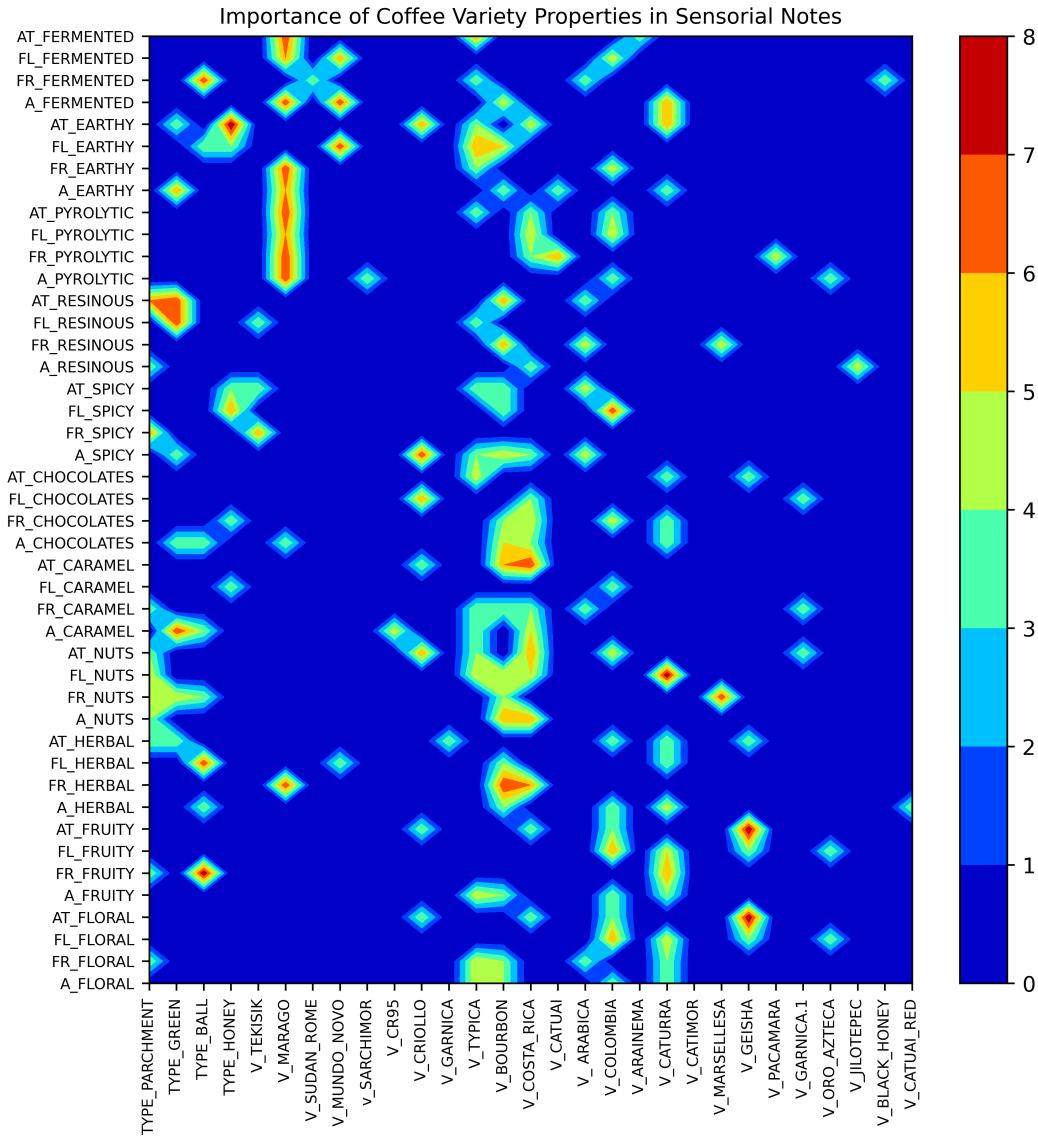


Figura 2.12: Número de concordancias de las variables de importancia obtenidas por los métodos utilizados café en las notas sensoriales e acuerdo a su tipo y variedad.

En la tabla 2.3 se pueden observar los datos reflejados en la figura 2.12

Específicamente en la cata de aromas, Costa Rica, Caturra y Maragon fueron las variedades con el impacto más significativo en todas las notas sensoriales del aroma. En la fragancia, las variedades Bourbon, Costa Rica y Typica tuvieron la influencia más significativa en las notas sensoriales de la fragancia. En el sabor, las variedades Colombia, Caturra y Bourbon fueron las más relevantes. Por último, en el retrogusto, las variedades Costa Rica, Typica, Criollo y Geisha obtuvieron el mayor número de concordancias. La figura 2.13 muestra el número de concordancias de las variables de importancia de las variables físicas, defectos y tamaño obtenidas por los métodos utilizados (RFE, LASSO, KNN CHI2, CATBOOST, DTC y RFC). El resultado indica la relación y la

Tabla 2.3: Variables físicas y defectos del grano que afectan a cada nota sensorial del café en su aroma, fragancia sabor y resabio.

Notas Sensoriales	TIPO Y VARIEDAD DE GRANOS DE CAFÉ
SPICY	Bourbon, Tekisik, Arabica, Typica, Criollo, Colombia, Type Parchment, green and honey
RESINOUS	Type green, Bourbon, Type Parchment, Arabica, Marellesa, Jilotepec, Tekisik, Typica and Costa Rica
PYROLITIC	Marago, Costa Rica, Colombia, Catuai, Pacamara, Sarchimor, Typica, Oro Azteca
NUTS	Marago, Mundo Novo, Sudan Rome, Tekisik, Arabica, Caturra, Type Ball, Costa Rica, Garnica, Colombia, Geisha
HERBAL	Bourbon, Caturra, Type Ball, Colombia, Marago, Costa Rica, Type Parchment, Type Green, Mondo Novo, Garnica, Geisha
FLORAL	Colombia, Caturra, Geisha, Typica, Bourbon, Type Parchment, Criollo, Costa Rica, Arabica, Oro Azteca
FERMENTED	Marago, Mundo Novo, Typica, Type Ball, Caturra, Bourbon, Colombia, Sudan Rome, Arabica
EARTHY	Typica, Marago, Type Honey, Type Green, Caturra, Bourgon, Mondo novo, Criollo, Costa Rica, Colombia, Type Ball
CHOCOLATE	Costa Rica, Caturra, Bourbon, Criollo, Typica, Colombia, Type Green, Type Ball, Type Honey
FRUTAL	Caturra, Colombia, Geisha, Type_Ball, Typica, Type Parchment, Criollo, Bourbon, Costa Rica, Oro Azteca
CARAMEL	Costa Rica, Bourbon, Typica, Type Green, Type Parchment, Type Ball, Type Honey, Criollo, Arabica, Colombia

importancia que existen entre las variables físicas, defectos y tamaño del grano de café verde con respecto a las notas sensoriales como floral, afrutado, herbáceo, nuez, caramelo, chocolate, picante, resinosas, pirolíticas, terrosas, fermentadas y fenólicas, que se analizaron en el aroma, fragancia, sabor y regusto.

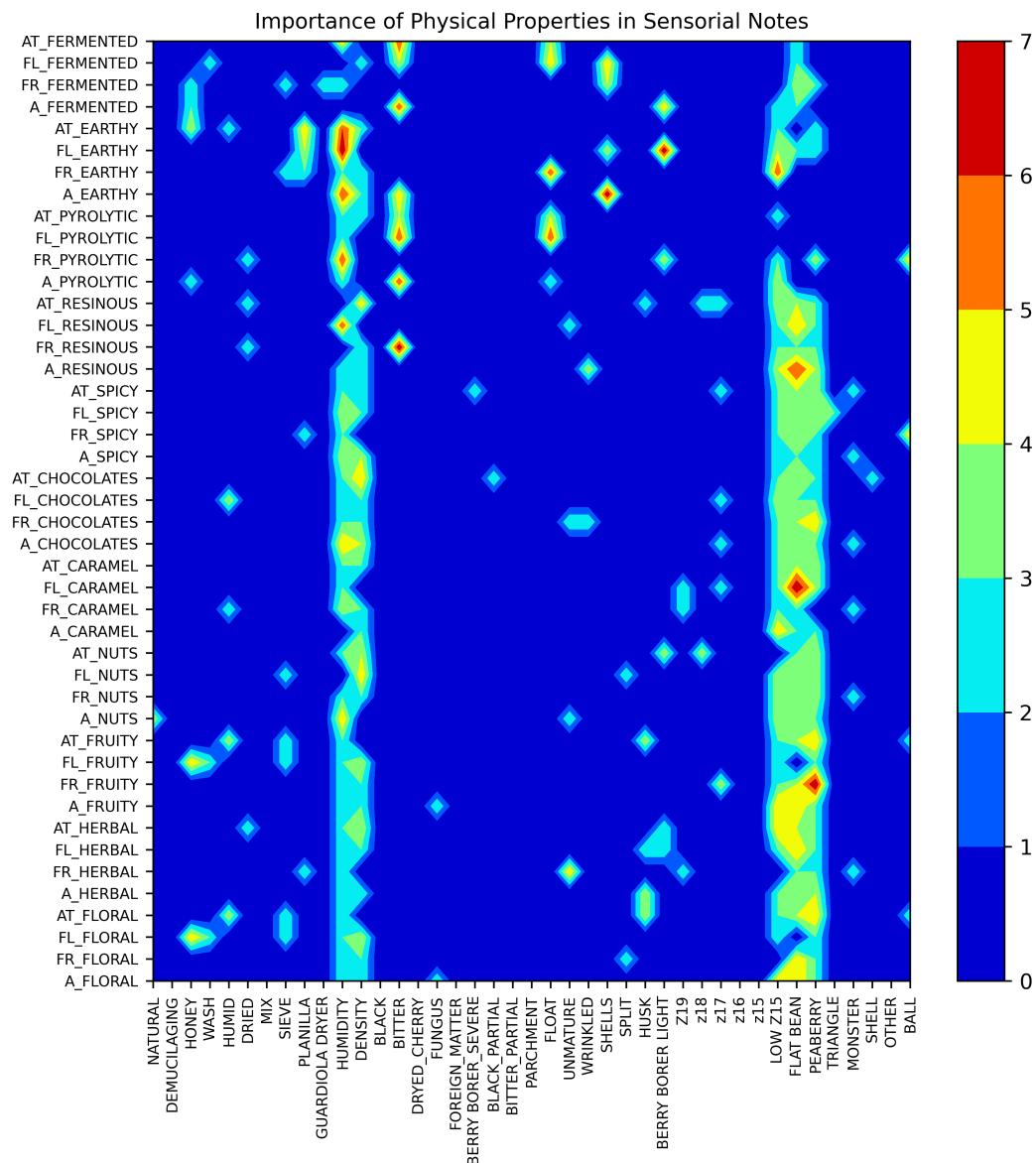


Figura 2.13: Importancia de las propiedades físicas en las notas sensoriales.

Tabla 2.4: Variables físicas y defectos del grano que afectan a cada nota sensorial del café en su aroma, fragancia sabor y resabio.

Notas Sensoriales	Variable física y defectos de los granos de café
SPICY	Monster, ball, planilla, berry borer, z17, triangle
RESINOUS	Bitter, dried, wrinkled, unmature, husk, z18 y z17
PYROLYTIC	Bitter, float, low z15, ball, berry borer, honey, dried, density
NUTS	Natural, berry borer, z19, sieve, unmature, split, monster
HERBAL	Husk, berry borer, unmature, dried, planilla, z19, monster
FLORAL	Sieve, honey, humid, husk, wash, fungus, split, ball.
FERMENTED	Bitter, float, shells, honey, berry borer, wash, sieve, guardiola drier
EARTHY	Planilla, shells, density, berry borer, float, bitter, honey, humid, sieve, flat bean.
CHOCOLATES	Z17, humid, black partial, unmature, wrinkled, monster, shell
FRUTAL	Sieve, honey, humid, husk, z17, wash, fungus, ball

En la tabla 2.4 se pueden observar las variables físicas, los defectos y el tamaño que estuvieron más presentes en todas las notas sensoriales en las catas de aroma, fragancia, sabor y regusto fueron humedad, densidad, baja Z15, aspecto plano, peaberry y amargo. Específicamente en la cata de aroma, las características físicas características físicas fueron miel, humedad, densidad, amargo, hongo, cáscara, Z15, grano plano y peaberry. Para la cata de fragancias, fueron cáscaras, humedad, densidad, amargo, inmaduro, z19, bajo z15, judía plana, y peaberry y bola. Para el sabor, fue humedad, densidad, amargo, float, cáscaras, z19, bajo z15, frijol plano, peaberry, y monstruo. Para el sabor, era la humedad, la densidad, amargo, flotante, conchas, z19, bajo z15, judía plana, peaberry y monstruo. En cuanto al regusto, fue humedad, densidad, amargo, cáscara, z18, z17, bajo z15, judía plana, y peaberry. En la literatura, es posible encontrar varios estudios de investigación en los que las propiedades físicas y sensoriales de los granos de café se han correlacionado mediante correlaciones de Pearson para encontrar si existe significación estadística entre las correlaciones entre pares de propiedades físicas y sensoriales Wondim-kun et al., 2020, Jane et al., 2020. Las conclusiones de esta investigación podrían indicar que existe una coincidencia entre las variables entre propiedades físicas y notas sensoriales. Por lo tanto, en esta investigación se utilizaron siete métodos para seleccionar las propiedades; cada método encuentra qué características físicas influyen más en cada nota sensorial. No todos los métodos determinan las mismas propiedades, por lo que sumando el número de veces que cada propiedad física fue importante para todos los métodos, se llegó a un consenso más sólido que con las correlaciones de Pearson. Además, gracias a los mapas de calor, determinar las propiedades físicas más importantes de cada nota sensorial es más fácil. de cada nota sensorial es más fácil de visualizar.

## 2.4.5 Resultados de la predicción de las notas sensoriales

Uno de los principales retos del aprendizaje automático es la selección del algoritmo para la aplicación deseada. La aplicación deseada; en la mayoría de los casos, la selección del algoritmo se realiza según la experiencia del programador, lo que limita la eficacia del aprendizaje automático. El aprendizaje automático mejora la selección de algoritmos y reduce la tendencia del programador a utilizar un algoritmo o una técnica determinada. La evaluación de las características sensoriales del café es subjetiva, aunque la realicen catadores certificados; como ejemplo de evaluación subjetiva, podemos fijarnos en la evaluación individual de una muestra o de un catador. La media de las puntuaciones de varios catadores evita la subjetividad de la subjetividad de la cata de café, aunque es difícil obtener las puntuaciones de varios catadores muestra a escala comercial. Los resultados muestran que la combinación de distintos modelos de aprendizaje automático permite obtener parámetros sensoriales del café con una precisión comparable a la obtenida combinando las puntuaciones de varios expertos. Esta metodología reduce la subjetividad de los catadores. Los resultados de los siete modelos de aprendizaje automático utilizados para predecir aromas florales, afrutados, herbáceos, nuez, caramelo, chocolate, especiado, resinoso, pirolítico, terroso, fermentado en aroma, fragancia, sabor y regusto. Las figuras 2.14 a 2.17 muestran los resultados de las notas sensoriales basados en la precisión de la estimación de los resultados de aroma, fragancia, sabor y regusto, respectivamente. Los resultados muestran que el RFC, CB, y SVC han obtenido los mejores resultados, donde el resinoso, pirolítico, terroso, y fermentado fueron las notas sensoriales con mejor precisión.

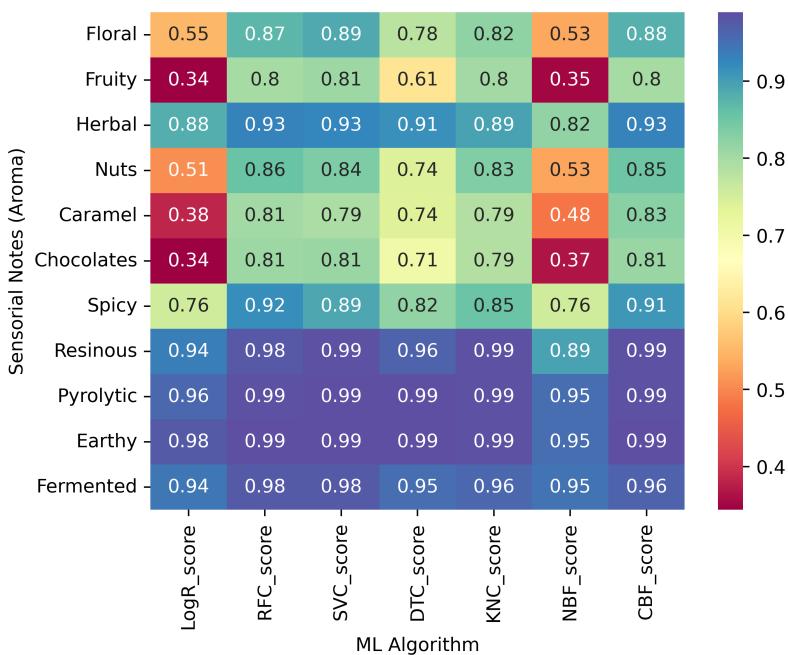


Figura 2.14: Resultados de precisión de los modelos ML utilizados en la predicción de notas sensoriales de aroma.

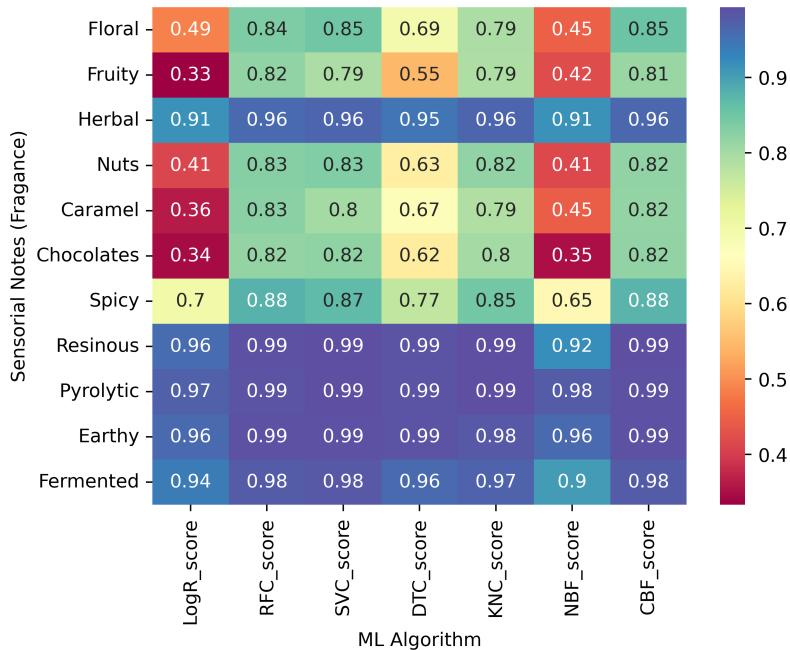


Figura 2.15: Resultados de precisión de los modelos ML utilizados en la predicción de notas sensoriales de fragancias.

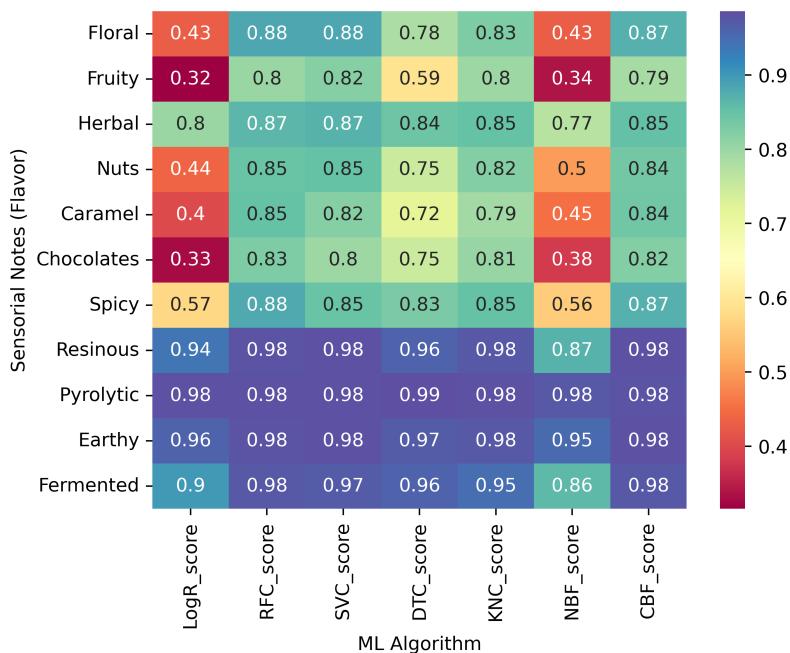


Figura 2.16: Resultados de precisión del algoritmo ML utilizado en la predicción de notas sensoriales de sabor.

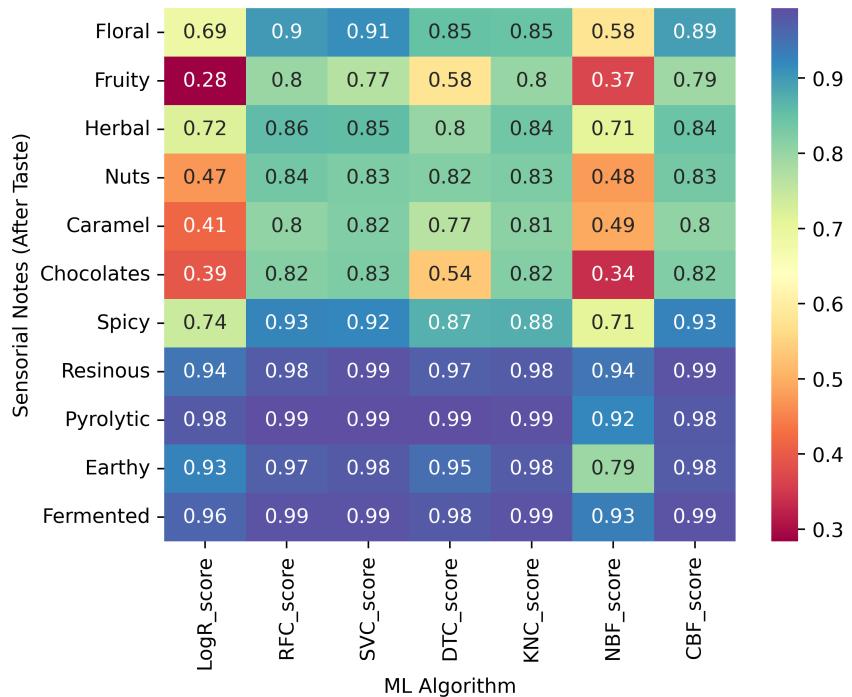


Figura 2.17: Resultados de precisión del algoritmo ML utilizado en la predicción de las notas sensoriales de regusto.

Las tablas 2.5 y 2.6 muestran los resultados de las métricas de evaluación utilizadas para predecir cada una de las notas sensoriales. Los resultados mostrados pertenecen a los modelos de Machine Learning que obtuvieron los mejores resultados en su predicción. La precisión se utilizó para medir el número de predicciones correctas predicciones positivas correctas. Sensibilidad se utilizó para medir la sensibilidad, que se obtiene mediante el clasificador que predijo correctamente entre todos los casos positivos de los datos. También se utilizó la puntuación F1 porque combina la precisión y la recuperación calculando una media armónica de ambas. La macromedia se calcula utilizando la media aritmética de todas las puntuaciones F1 por clase. Y, por último, la media ponderada calcula la media de las puntuaciones de clase teniendo en cuenta el número real de apariciones de la clase en los datos.

Tabla 2.5: Métricas de evaluación utilizadas en la determinación de las notas sensoriales de aroma y fragancia notas sensoriales

Nota sensorial	Modelo ML	Precisión	Sensibilidad	f1-puntuación
Flavor - Floral	SVC	Accuracy		0.8
		Macro avg	0.86	0.93
		Weighted avg	0.89	0.88

**Tabla 2.5 continuación de la página previa**

Nota sensorial	Modelo ML	Precisión	Sensibilidad	f1-puntuación
Flavor - Fruity	SVC	Accuracy		0.82
		Macro avg	0.73	0.64
		Weighted avg	0.81	0.8
Flavor - Herbal	SVC	Accuracy		0.87
		Macro avg	0.62	0.52
		Weighted avg	0.86	0.86
Flavor - Nuts	RFC	Accuracy		0.85
		Macro avg	0.69	0.61
		Weighted avg	0.84	0.84
Flavor - Caramel	RFC	Accuracy		0.85
		Macro avg	0.9	0.87
		Weighted avg	0.85	0.84
Flavor - Chocolates	RFC	Accuracy		0.83
		Macro avg	0.82	0.71
		Weighted avg	0.83	0.82
Flavor - Spicy	RFC	Accuracy		0.88
		Macro avg	0.91	0.91
		Weighted avg	0.88	0.88
Aftertaste - Floral	SVC	Accuracy		0.88
		Macro avg	0.87	0.91
		Weighted avg	0.9	0.89
Aftertaste - Fruity	SVC	Accuracy		0.82
		Macro avg	0.85	0.81
		Weighted avg	0.87	0.82
Aftertaste - Herbal	SVC	Accuracy		0.87
		Macro avg	0.87	0.66
		Weighted avg	0.87	0.86
Aftertaste - Nuts	SVC	Accuracy		0.85
		Macro avg	0.86	0.89
		Weighted avg	0.85	0.85

**Tabla 2.5 continuación de la página previa**

Nota sensorial	Modelo ML	Precisión	Sensibilidad	f1-puntuación
Aftertaste - Caramel	RFC	Accuracy		0.85
		Macro avg	0.84	0.77
		Weighted avg	0.84	0.84
Aftertaste—Chocolates	RFC	Accuracy		0.83
		Macro avg	0.68	0.7
		Weighted avg	0.81	0.81
Aftertaste—Spicy	CB	Accuracy		0.88
		Macro avg	0.88	0.74
		Weighted avg	0.89	0.88

Tabla 2.6: Métricas de evaluación utilizadas en la determinación de las notas sensoriales de aroma y fragancia

notas sensoriales

Nota sensorial	Modelo ML	Precisión	Sensibilidad	f1-puntuación
Aroma - Floral	SVC	Accuracy		0.9
		Macro avg	0.78	0.78
		Weighted avg	0.9	0.9
Aroma - Fruity	SVC	Accuracy		0.81
		Macro avg	0.87	0.86
		Weighted avg	0.82	0.81
Aroma - Herbal	SVC	Accuracy		0.93
		Macro avg	0.58	0.39
		Weighted avg	0.92	0.92
Aroma - Nuts	RFC	Accuracy		0.86
		Macro avg	0.81	0.7
		Weighted avg	0.86	0.86
Aroma - Caramel	CB	Accuracy		0.83
		Macro avg	0.85	0.81
		Weighted avg	0.85	0.83
Aroma - Chocolates	CB	Accuracy		0.81
		Macro avg	0.87	0.81

**Tabla 2.6 continuación de la página previa**

Nota sensorial	Modelo ML		Precisión	Sensibilidad	f1-puntuación
		Weighted avg	0.82	0.81	0.81
Aroma - Spicy	RFC	Accuracy			0.92
		Macro avg	0.61	0.59	0.6
		Weighted avg	0.9	0.92	0.91
Fragance - Floral	SVC	Accuracy			0.85
		Macro avg	0.87	0.88	0.88
		Weighted avg	0.85	0.86	0.85
Fragance - Fruity	RFC	Accuracy			0.82
		Macro avg	0.75	0.72	0.73
		Weighted avg	0.84	0.81	0.82
Fragance - Herbal	SVC	Accuracy			0.96
		Macro avg	0.81	0.74	0.77
		Weighted avg	0.94	0.95	0.94
Fragance - Nuts	SVC	Accuracy			0.83
		Macro avg	0.86	0.78	0.81
		Weighted avg	0.83	0.82	0.82
Fragance - Caramel	RFC	Accuracy			0.83
		Macro avg	0.86	0.86	0.86
		Weighted avg	0.84	0.84	0.84
Fragrance—Chocolates	CB	Accuracy			0.82
		Macro avg	0.86	0.76	0.81
		Weighted avg	0.81	0.77	0.78
Fragrance—Spicy	CB	Accuracy			0.88
		Macro avg	0.89	0.82	0.85
		Weighted avg	0.89	0.88	0.88

En la Tabla 2.7 se muestra el resultado de los mejores algoritmos de Machine Learning con hiperparametros para la estimacion de las notas sensoriales de aroma y en la tabla 2.8 para las notas de fragancia.

Tabla 2.7: Los mejores algoritmos ML hiperparámetros para la estimación de notas sensoriales de Aroma

<b>Notas Sensorial</b>	<b>Hiperparámetros Algoritmos ML</b>
A_FLORAL	SVC "kernel": rbf, "gamma": 1, C": 1000 SVC "kernel": rbf, "gamma": 1, C": 1000
A_HERBAL	RFC "bootstrap": true, criterion": "gini", "max_features": 0.2634559502785016, "min_samples_leaf": 1, "min_samples_split": 0.013696664233304495, "n_estimators": 170. SVC "kernel": rbf, "gamma": 1, C": 100 CBF "max_depth": 6, "learning_rate": 0.05
A_NUTS	RFC "bootstrap": true, criterion": "gini", "max_features": 0.30474046083262707, , "min_samples_leaf": 1, "min_samples_split": 0.016934696827956865, "n_estimators": 171.
A_CARAMEL	CBF "max_depth": 6, "learning_rate": 0.05
A_CHOCOLATES	CBF "max_depth": 6, "learning_rate": 0.05
A_SPICY	CBF "max_depth": 6, "learning_rate": 0.05 SVC "kernel": rbf, "gamma": 1, C": 1000 CBF "max_depth": 6, "learning_rate": 0.05

Tabla 2.8: Los mejores algoritmos ML hiperparámetros para la estimación de notas sensoriales de Aroma

<b>Notas Sensoriales</b>	<b>Hiperparámetros Algoritmos ML</b>
F_FLORAL	SVC "kernel": rbf, "gamma": 1, C": 1000 CBF "max_depth": 6, "learning_rate": 0.2
F_FRUITY	RFC "bootstrap": true, criterion": "gini", "max_features": 0.2606530421228427, "min_samples_leaf": 1, "min_samples_split": 0.013956146634119322, "n_estimators": 105
F_HERBAL	RFC "bootstrap": true, criterion": "gini", "max_features": 0.2611486134479855, "min_samples_leaf": 1, "min_samples_split": 0.013854024616189118, "n_estimators": 87
F_NUTS	SVC "kernel": "linear", "gamma": 1, C": 1000 CBF "max_depth": 2, "learning_rate": 0.05 SVC "kernel": rbf, "gamma": 1, C": 1000

**Table 2.8 continued from previous page**

<b>Notas Sensoriales</b>	<b>Hiperparámetros Algoritmos ML</b>
F_CARAMEL	RFC "bootstrap": true, , "criterion": "gini", "max_features": 0.2606530421228427 "min_samples_leaf": 1, "min_samples_split": 0.013956146634119322, false
F_CHOCOLATES	RFC "bootstrap": true, "criterion": "gini", "max_features": 0.2606530421228427, "min_samples_leaf": 1, "min_samples_split": 0.013956146634119322, "n_estimators": 105 SVC "kernel": rbf, "gamma": 1, "C": 1000 CBF "max_depth": 6, "learning_rate": 0.05
F_SPICY	RFC "bootstrap": true, "criterion": "gini", "max_features": 0.2634559502785016, "min_samples_leaf": 1, "min_samples_split": 0.013696664233304495, "n_estimators": 170 CBF "max_depth": 6, "learning_rate": 0.05

## 2.5 Conclusiones

Una de las mayores contribuciones de este trabajo se basa en los resultados de los experimentos realizados, que indican que es posible conocer cómo las variables físicas y las deformaciones de los granos de café verde impactan en cada una de sus notas sensoriales. Por lo tanto, fue posible obtener un mapa detallado con esta información.

El desarrollo de un modelo automatizado de aprendizaje automático permite determinar las notas sensoriales haciendo posible compararlas con las obtenidas por expertos y, al mismo tiempo, abre la posibilidad de desarrollar algoritmos automatizados más complejos. Mediante la etapa de ajuste de hiperparámetros, se mejora la precisión de la predicción de las propiedades sensoriales, que se aproxima a la obtenida por las puntuaciones combinadas de cataadores entrenados.

Además, al obtener una estimación del perfil sensorial de los granos de café verde con el modelo de Machine Learning propuesto, los productores y comerciantes podrán analizar los datos antes del proceso de tostado del café, dando lugar a un modelo de comercio justo. Al comprender la relación entre las propiedades físicas y el perfil sensorial, los productores podrán mejorar la calidad del producto.

Los resultados de la investigación demuestran que combinando diferentes modelos de aprendizaje automático se puede alcanzar una precisión análoga a la de los expertos en catación; por lo tanto, se reduce la posibilidad de errores inducidos por cuestiones humanas como el cansancio o la subjetividad.

# CAPÍTULO 3: Estudios de espectroscopía NIR en granos de café verde para la determinación de las notas sensoriales utilizando aprendizaje profundo

## 3.1 Introducción

Este capítulo se muestra el estudio realizado a los granos del café verde utilizando técnicas de espectroscopía NIR, con lo cuál a través de algoritmos de inteligencia artificial fue posible determinar las variables de importancia relacionadas a la longitud de onda que impactan en la determinación de las notas sensoriales del café. También se muestran los resultados del entrenamiento de un sistema inteligente basado en aprendizaje profundo que sea capaz de estimar las notas sensoriales del café verde a partir de los resultados de espectroscopía NIR.

## 3.2 Marco Teórico

### 3.2.1 Espectrofotometría

La espectrofotometría es la técnica analítica que mide la intensidad de luz que absorbe una sustancia química, para determinar la concentración presente en una muestra o conocer la cantidad de un producto químico conocido en una sustancia. En términos generales la muestra a analizar dependiendo de sus componentes químicos o compuestos orgánicos tienen una respuesta de absorción diferente en cada longitud de onda electromagnética. Sin embargo a pesar de obtener una relación entre las longitudes de onda y su absorbancia, resulta muy complejo determinar composiciones como las notas sensoriales del café a partir únicamente de estos datos de absorbancia. La técnica de espectrofotometría tiene su sustento en la Ley de Lambert la cual puede "determinar que la iluminación producida por una fuente luminosa es proporcional a la intensidad de la luz de la fuente y al coseno del ángulo que forma la normal a la superficie", Ley de Lambert-Beer la cual afirma que "a mayor concentración, mayor absorbancia" y la Ley de Bouguer-Beer-Lambert la cual es una combinación de las dos.

### 3.2.2 Instrumentos de medición

Durante los trabajos de experimentación, se hace uso de los siguientes equipos de medición para la obtención de datos requeridos para ser usados con los algoritmos de inteligencia artificial.

**Espectrómetro Thermo Nicolet iS5:** El espectrómetro infrarrojo por transformada de Fourier (FTIR) permitió identificar los componentes del café mediante muestras y la colección de datos en el rango del espectro IR-medio usando transmisión, reflexión total atenuada (ATR), y reflexión especular. La espectroscopia infrarroja de transformada de Fourier esta soportada en el principio de interferencia de radiación entre dos haces de luz para formar un interferograma. El producto es una señal producida como función del cambio de paso de luz entre los dos haces. Los dos dominios de distancia y frecuencia son interconvertibles mediante el método matemático

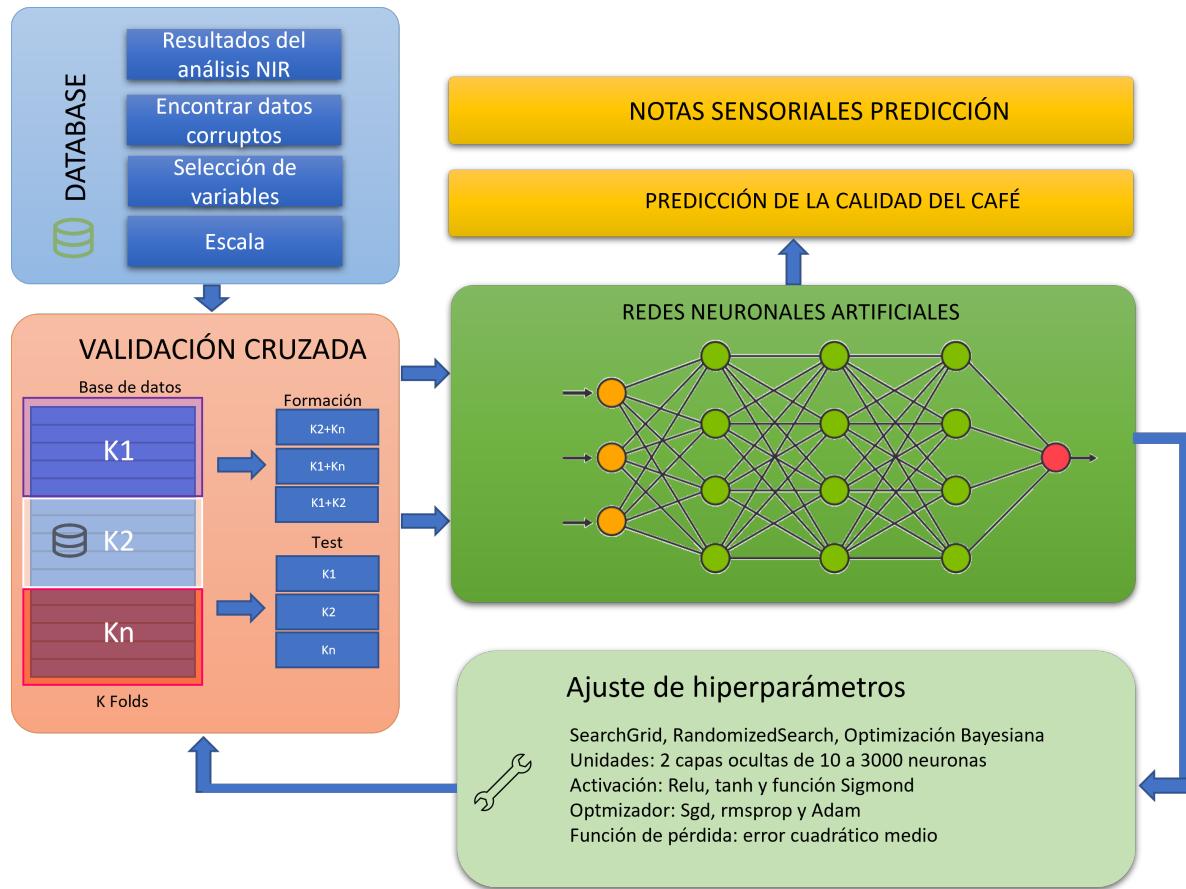


Figura 3.1: Metodología para la determinación de las notas sensoriales del café utilizando datos de espectroscopía NIR a través de Redes Neuronales Artificiales.

de Transformada de Fourier.

**La bascula analítica OHAUS:** Es una herramienta de precisión, la cual es usada para obtener las cantidades requeridas para la preparación de mediciones de muestras en laboratorios.

**Bromuro de potasio :** El bromuro de potasio (KBr) es una sal de color blanco, transparente de 40000 - 385  $cm^{-1}$ , es más higroscópico que el NaCl y tiene buena resistencia mecánica a los golpes la cual es utilizada en la mezcla con la muestra de café verde.

**Prensa y pastillero :** La prensa es una herramienta que está conformada de una barra con soporte de presión y una cavidad en donde se deposita la muestra para la obtención de la pastilla que es utilizada en el espectrómetro su análisis.

### **3.3 Metodología**

En la figura 3.1 se muestra la metodología implementada para el desarrollo de esta investigación. Se puede apreciar que existen cuatro bloques principales. El primer bloque está conformado por la base de datos de espectroscopía NIR, el segundo bloque está enfocado a la validación cruzada de los datos para garantizar que los resultados en el entrenamiento y precisión sean adecuados, el tercer bloque está enfocado a los algoritmos de aprendizaje profundo (Deep Learning) que estuvieron destinados a generar un sistema inteligente capaz de estimar las notas sensoriales de café a partir de los datos obtenidos por espectroscopia NIR. Finalmente el último bloque está enfocado a la autosintonización de hiperparámetros el cual fue requerido para encontrar los valores de los hiperparámetros que obtuvieran una mayor precisión.

#### **3.3.1 Metodología Espectrofotometría NIR**

#### **3.3.2 Obtención de las muestras.**

Las muestras analizadas se obtuvieron a través del centro de investigación CAFECOL las cuales fueron clasificadas de acuerdo a su región, tipo, variedad, aroma, proceso de lavado, secado, selección, defectos etc.

##### **3.3.2.1 Preparación de la muestra.**

Los equipos y materiales utilizados fueron: el Espectrómetro Thermo Nicolet iS5 para realizar los estudios de espectrometría de las muestras, las muestras de café verde para realizar los estudios, una báscula de precisión para pesar cada muestra de café así como también para pesar el Bromuro de potasio que es añadido a la muestra.

Para la cuantificación de las muestras se prepararon pastillas por el método de disco con bromuro de potasio. Para realizar las pastillas se utilizaron 0.08 mg de la muestra de café con 0.8 mg de polvo de KBr. Primero se pulverizo la muestra sólida en un molino de café , después se pasó por el método de tamizado para obtener la muestra homogénea. En un mortero de ágata, se realizó la mezcla de café con bromuro de potasio, considerando que la muestra se encontrara libre de humedad, una vez obtenida la mezcla, se procedió a ponerla en la cavidad formada en el pastillero, enseguida se tapa con el otro tornillo y se produce el prensado inicial. Después de aplicar vacío, se aprieta fuertemente. Una vez transcurrido el tiempo necesario se retiran los dos pernos, se sitúa la porta pastilla sobre una superficie y se obtiene la muestra. Al final se obtuvo un disco de aproximadamente 8 mm de diámetro. Una de las técnicas más popular de manipulación de las muestras sólidas es la formación de pastillas de KBr. Las sales de haluros tienen la propiedad de flujo en frío por lo cual cuando se presiona suficientemente, este material finamente pulverizado presenta propiedades transparentes o translúcidas como el vidrio.

La mezcla para la formación de las pastillas se llevó a cabo de acuerdo a la relación de producto indicado en el manual de usuario del Espectroscopio Thermo Nicolet iS5. Con el objetivo de obtener un conjunto representativo de espectros de las muestras de café en función a su absorbancia, se realizaron 4 corridas de cada muestra,

produciendo un total de 836.

### **3.3.2.2 Valoración de la muestra**

La valoración por el método de Espectroscopia Infrarroja por Transformada de Fourier (FTIR) se realizó usando el equipo Thermo Nicolet iS5, que utiliza un accesorio de reflectancia total atenuada (ATR) de ZnSe acoplado al FTIR. El análisis de la señal obtenida es generado por el Software OMNIC™ Specta, el cual crea una base de datos espectrales para identificar las propiedades fisicoquímicas de las muestras de café analizadas. Para cada muestra de café se obtuvieron 4 espectros en la zona de infrarrojo medio ( $500\text{-}4000\text{ }cm^{-1}$ ), trabajando en modo de absorbancia.

### **3.3.2.3 Sistema Inteligente**

Se utilizó la biblioteca de TensorFlow para la realización de las Redes Neuronales Artificiales, y la biblioteca de kera\_tuner para la sintonización de hiperparámetros con la finalidad de tener un sistema capaz de estimar las notas sensoriales del café de grano verde utilizando señales de espectroscopía. Para lograr obtener el modelo con la mayor precisión posible se realizaron varias pruebas variando los hiperparámetros de la red neuronal. Estos hiperparametros fueron la cantidad de neuronas en la primera y segunda capa, el bías, la función de activación en ambas capaz y el optimizador utilizado.

## **3.4 Experimentación**

### **3.4.1 Obtención de las señales**

Una vez obtenida la pastilla se procede a realizar la medición de la señal. Primero se recolecta el fondo (como referencia de cada muestra), para cada corrida se recolectaron 16 barridos por espectro. Obtenido el fondo se iniciaron las recogidas de las muestras. Se colocó la muestra sobre la superficie del cristal de diamante del ATR.

Como resultados se observan 4 gráficas como muestra de los espectros obtenidos de absorción de diferentes tipos de granos de café. Para finalizar se giró el dispositivo de presión en posición de limpieza, se retiró la muestra de la punta y del cristal de diamante con papel y/o espátulas plásticas, se limpió la punta y el cristal de diamante con alcohol isopropílico y papel toalla (también se puede hacer uso de: papel óptico, algodón o aplicadores de algodón). Todo lo señalado se realizó para cada colección de muestra y evitar interferencias y errores en las lecturas.

El espectro obtenido representa los picos de absorción que corresponden a las frecuencias de las vibraciones entre los lazos de los átomos que componen el material. Debido a que cada material diferente es una combinación única de átomos, no hay dos compuestos que produzcan exactamente el mismo espectro infrarrojo. El tamaño de los picos en el espectro es una indicación directa del tipo y cantidad de compuesto químico presente.

### 3.4.2 Cuantificación de las muestras

Para la cuantificación de las muestras se usó el software IRPal, el cual nos permitió identificar los compuestos orgánicos de las diferentes muestras de café en los rangos de 500-4000  $cm^{-1}$ . Se procedió a cuantificar los elementos químicos de las muestras Tabla (3.1), adquiriendo la información de los espectros FTIR de las mismas. Gracias a la opción de cuantificación se registra la concentración de los diferentes elementos. Los datos obtenidos de las 174 muestras de café se clasificaron en una base de datos, en la que se describe el número de muestra, la posición e intensidad en la huella, sus rangos de frecuencia y sus propiedades fisicoquímicas.

Tabla 3.1: Rango de longitudes de onda con su correspondiente estructura orgánica utilizados en la espectroscopía NIR

Rango Longitud de Onda	Estructura
1695-1705	RCONHR'5-ring
1695-1705	ArCHO
1705-1715	RCO-OH
1710-1720	R2CO
1715-1725	C=C-CO-OH
1715-1725	C=C-CO-OR_ArCOOR
1720-1730	RCHO
1730-1740	RCOOR'
1740-1750	RCONHR'4_ring
1740-1750	R2CO_5-ring
1755-1765	RCO-OH
1755-1765	R-CO-O-CO-R'
1765-1775	RCOOR'5-ring
1815-1825	R-CO-O-CO-R'
1990-2140	R-N=C=S
2100-2140	RC#CH
2100-2360	Si-H_silane
2110-2165	R-N-C
2190-2260	RC#CR
2225-2235	C#N(nitriles)
2245-2255	C#N(nitriles)
2280-2440	P-H_phosphine
2550-2700	(O=)PO-H

**Tabla 3.1 continuación de la página previa**

Rango Longitud de Onda	Estructura
2550-2600	S-H thiols
2695-2830	RHCO
2715-2725	All_aldehydes
2800-3400	RCO-OH
2815-2825	All_aldehydes
2845-2855	CH2
2850-3000	RCH2CH3
2865-2875 -	RCH2CH3
2920-2930 -	CH2-
2955-2965 -	CH3
3000-3100	ArH
3270-3330	RC#C-H
3395-3405	RNH2
3310-3350	R2NH
3310-3335	R2NH
3015-3025	transRCH=CHR_cisRCH
3040-3140	RCH=CH2
3080-3140	R2C=CH2
3200-3500	ArO-H_bonded
3325-3335	RCONHR'
3395-3405	RCONH2
3395-3405	RCO-OH_C=C-CO-OH
3400-3600	RCH2OH
3435-3445	RCONHR'
3445-3455	Ar2NH
3495-3505	RCONH2
3495-3505	RNH2
3550-3600	NOH oxime
3610-3640	O-H_free
3610-3640	ArO-H_free

A continuación en las figuras [3.2](#), [3.3](#), [3.4](#), [3.5](#), [3.6](#), [3.7](#), [3.8](#), [3.9](#), [3.10](#) y [3.11](#) se muestran los análisis obtenidos.

dos relacionados a la cuantificación de diferentes tipos y variedades de café. A través del análisis de los picos ubicados en posiciones específicas de la longitud de onda se puede determinar la presencia de determinadas estructuras orgánicas en la muestra. Dicha información será utilizada para los modelos de aprendizaje máquina y aprendizaje profundo con la finalidad de realizar un análisis de como las variables relacionadas a la longitud de onda y su estructura orgánica impactan en la determinación de las notas sensoriales del café.

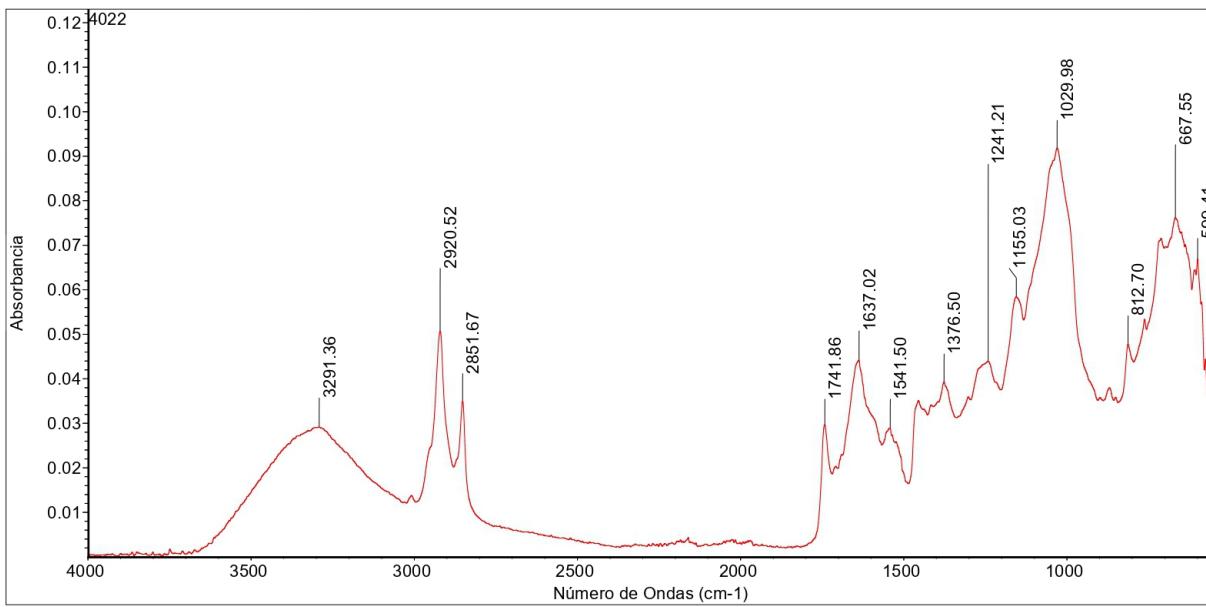


Figura 3.2: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo pergamo.

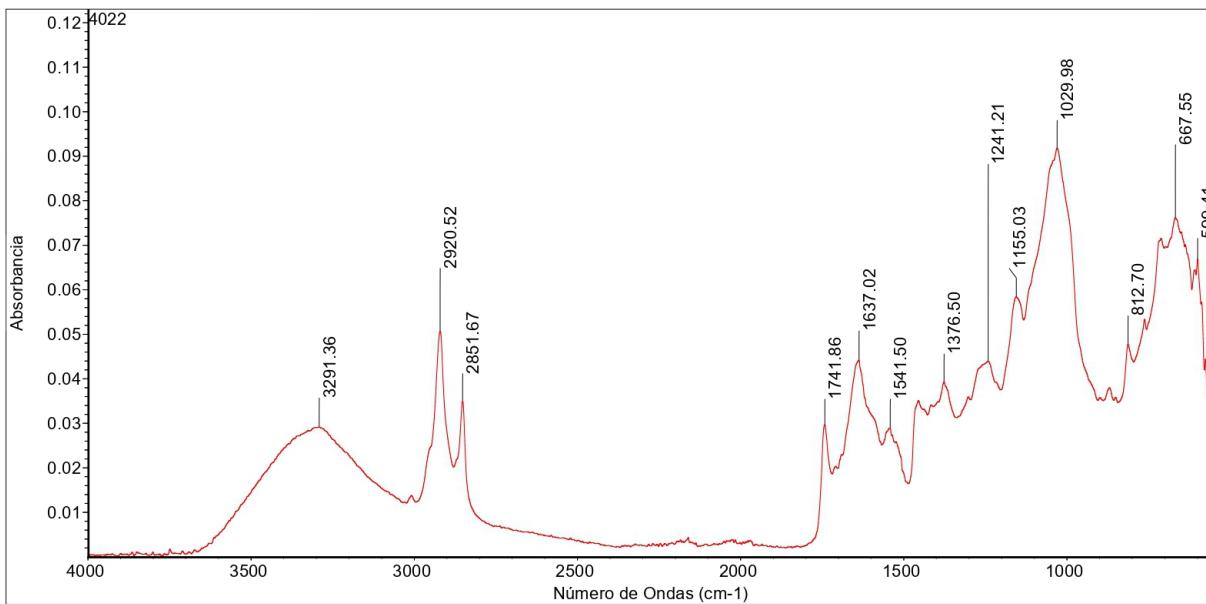


Figura 3.3: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo CR95.

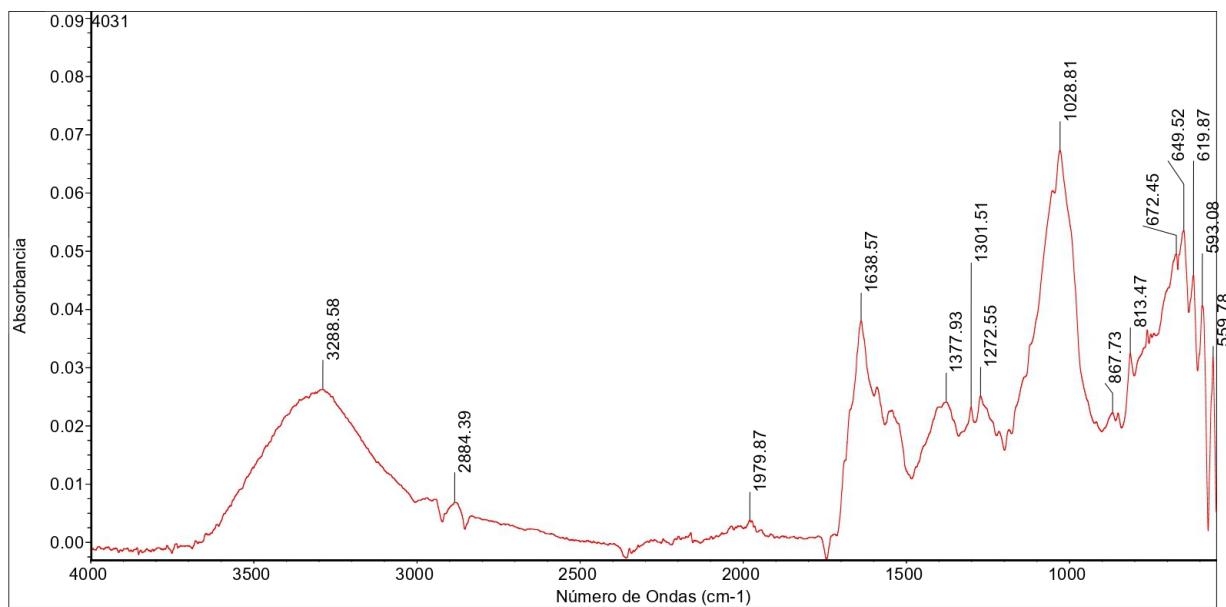


Figura 3.4: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Borbon.

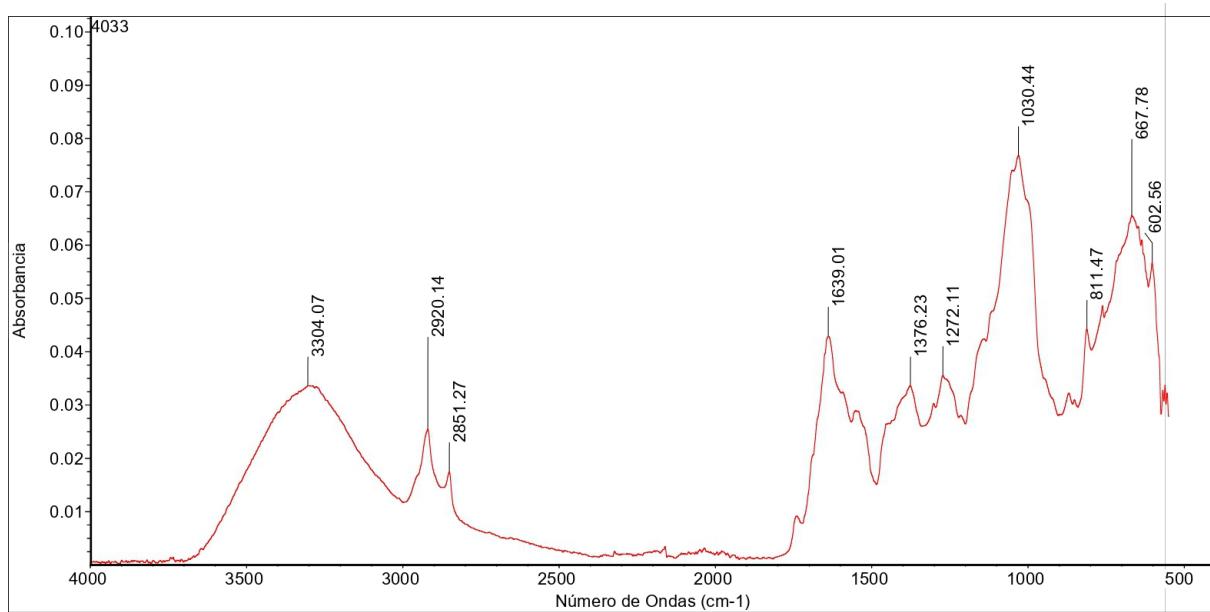


Figura 3.5: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Geisha.

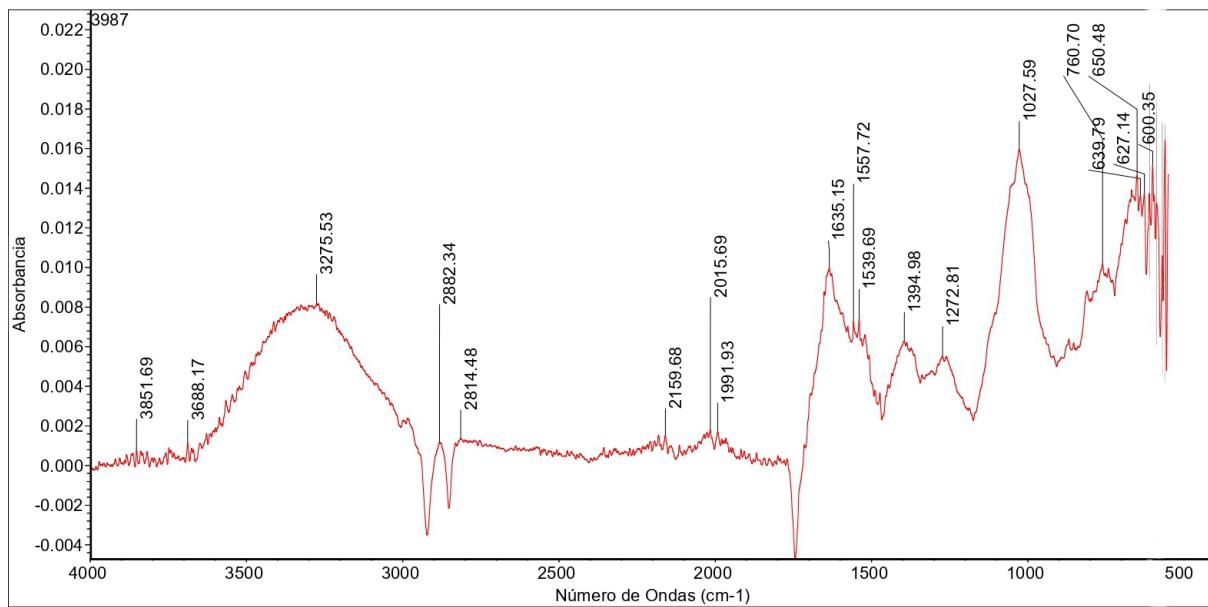


Figura 3.6: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Marago.

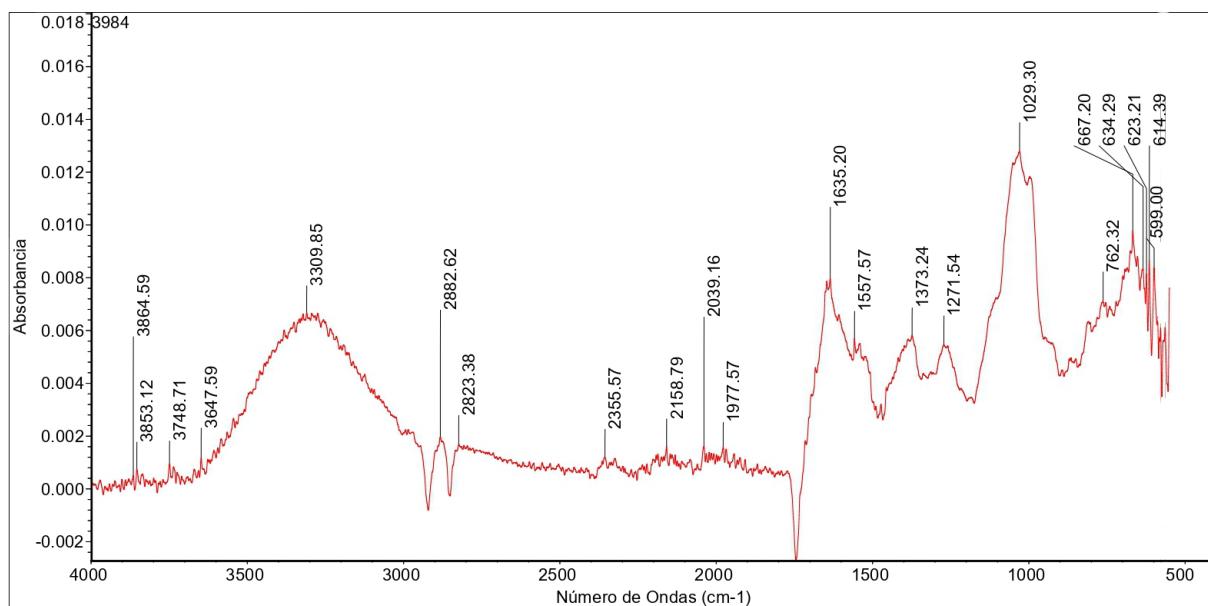


Figura 3.7: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Marellesa.

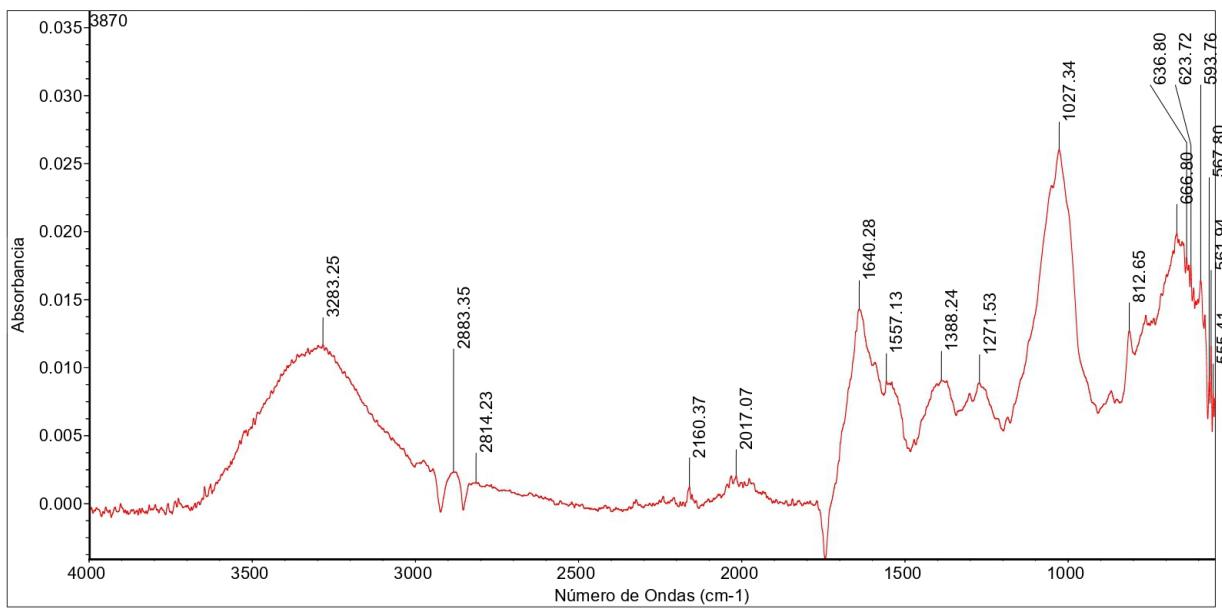


Figura 3.8: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Mundo Novo.

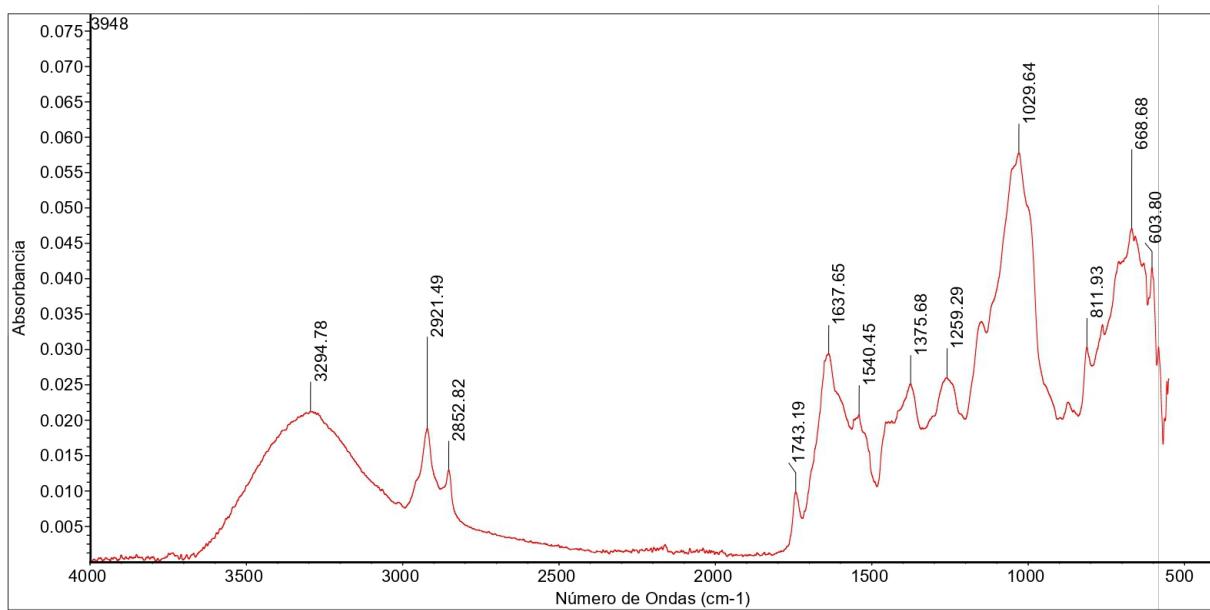


Figura 3.9: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Oro Azteca.

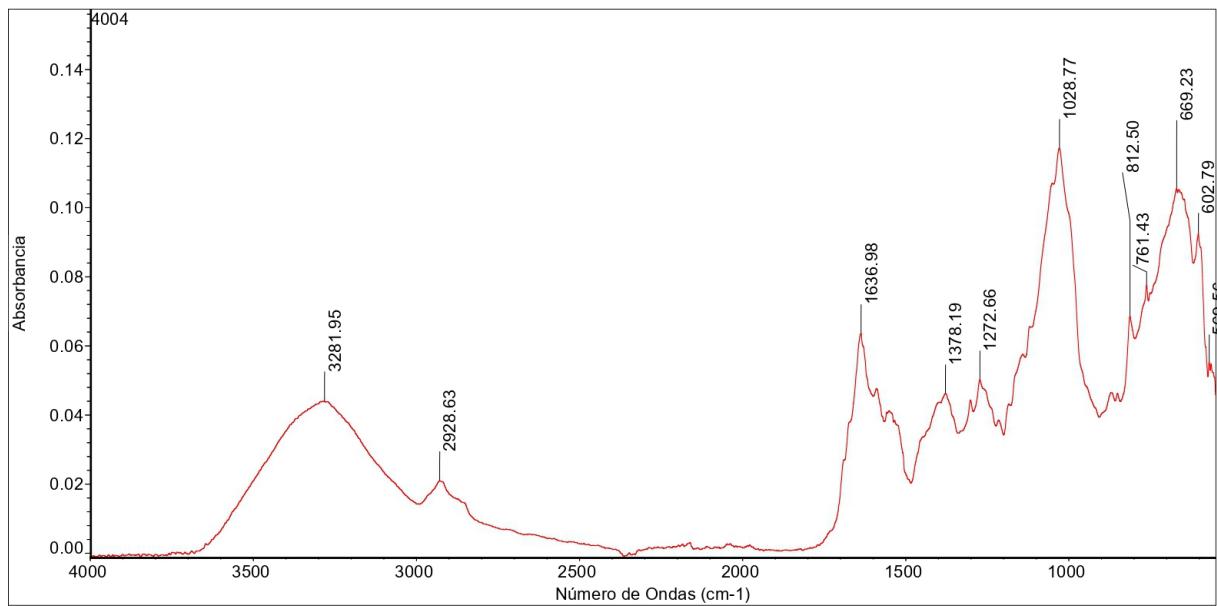


Figura 3.10: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Pakamara.

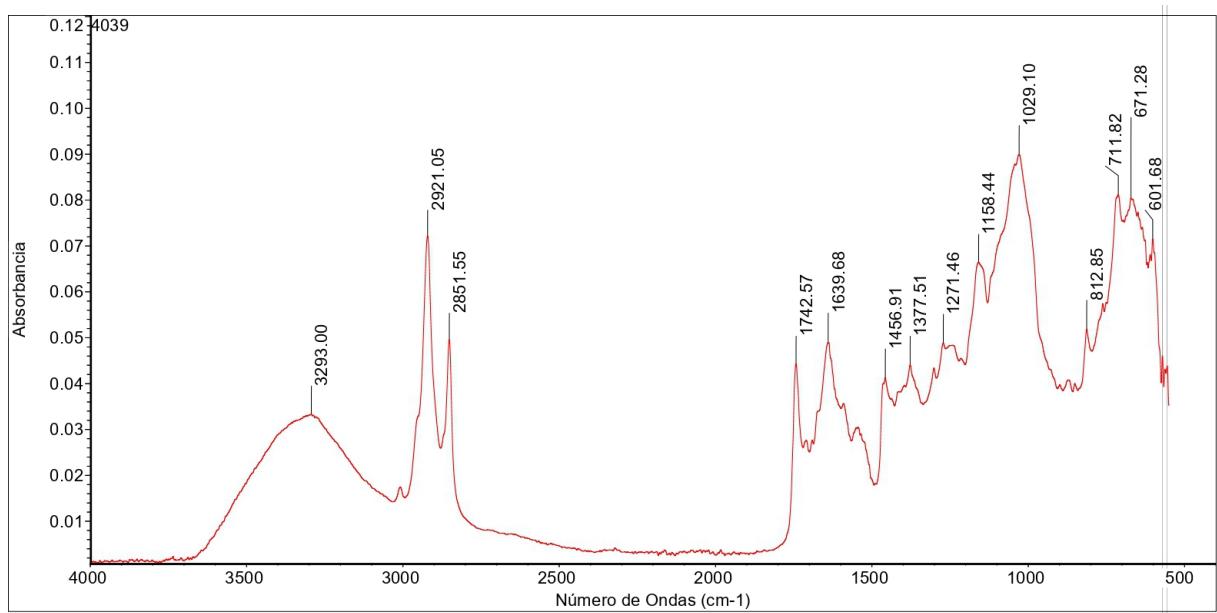


Figura 3.11: Gráfica con los resultados obtenidos en espectroscopía NIR de muestras de café tipo Typica.

Las figuras anteriores muestran la presencia de diferentes estructuras orgánicas a través de la detección de picos de la longitud de onda de los análisis de espectroscopia NIR en cada una de las muestras estudiadas.

A continuación en la figura 3.12 se muestra la frecuencia con la que aparecen diferentes longitudes de onda relacionadas a los análisis de espectroscopía NIR en las muestras de grano de café verde. Es importante remarcar

que las longitudes de onda que aparecen con una mayor frecuencia podrían indicar las estructuras orgánicas con las que está compuesta el café. En este caso las longitudes de onda que se presentan más frecuentemente en todas las muestras de café analizadas son: 665-910nm RNH2\_R2NH, 750-850nm R-Cl, 1000-1350nm R-F, 1030-1230nm RNH2\_R2NH, 1150-1300nm CH2X, 1560-1640nm RNH2, 1655-1665nm R2C=CHR, 2280-2440nm P-H\_phosphine, 3015-3025nm R2NH, 3395-3405nm RCONH2, 3495-3505 RNH2.

Sin embargo es difícil con esta información conocer cómo alguna o diversas combinaciones de longitudes de ondas impactan en alguna nota sensorial. Por esta razón se utilizaron técnicas de Machine Learning y Deep learning para la extracción y obtención de información relevante que pueda determinar el impacto y la relación que tienen las variables entre ellas.

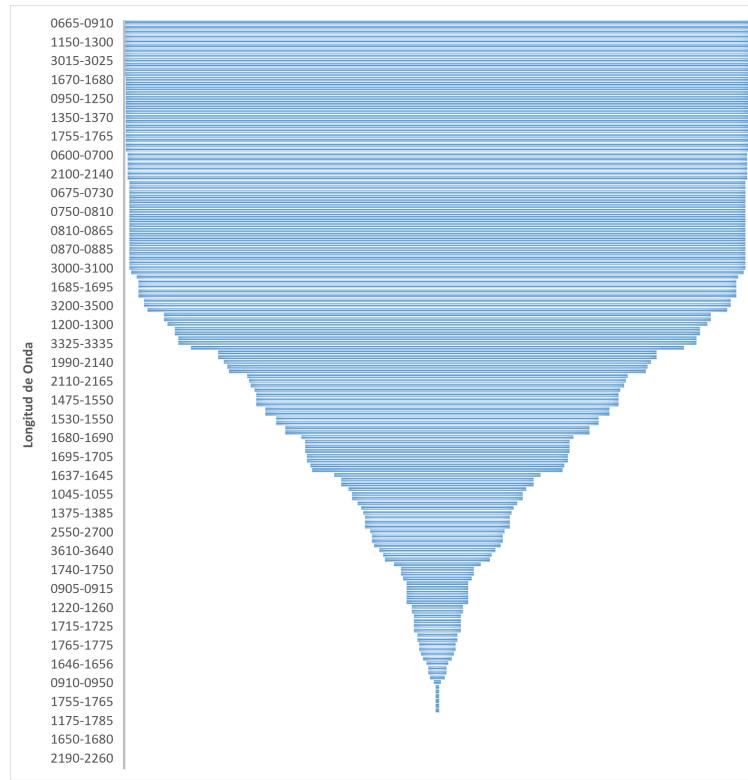


Figura 3.12: Frecuencia de los datos en cada longitud de onda

### 3.4.3 Determinación de Características Importantes

Una de las aportaciones científicas de esta investigación se encuentra en la determinación de las variables de importancia que impactan directamente en las notas sensoriales del café. La metodología para la determinación de estas variables de importancia está descrita en la figura 3.13. Como primer paso de la metodología se encuentra la preparación de las muestras de café verde, posteriormente se realiza el análisis de espectrometría con lo cual es posible obtener los picos en la gráfica de absorbancia-Longitud de onda. Esa información en conjunto con los resultados de los estudios de catación del café en el cual se tiene la información de las notas sensoriales

del café en el aroma, fragancia, sabor y resabio, son enviados a los algoritmos de Machine Learning los cuales utilizan métodos como (RFE, LASSO, KNN, CHI Square, CATBOOST, Random Forest y Logistic Regression) para determinar cómo el comportamiento de la absorbancia respecto a diferentes longitudes de onda influyen en las notas sensoriales del café. Tomando en cuenta que los picos de absorbancia en diferentes longitudes de onda representan la presencia de diferentes estructuras orgánicas, se puede deducir que la combinación de dichas estructuras pueden determinar aromas y sabores requeridos para la obtención de las notas sensoriales.

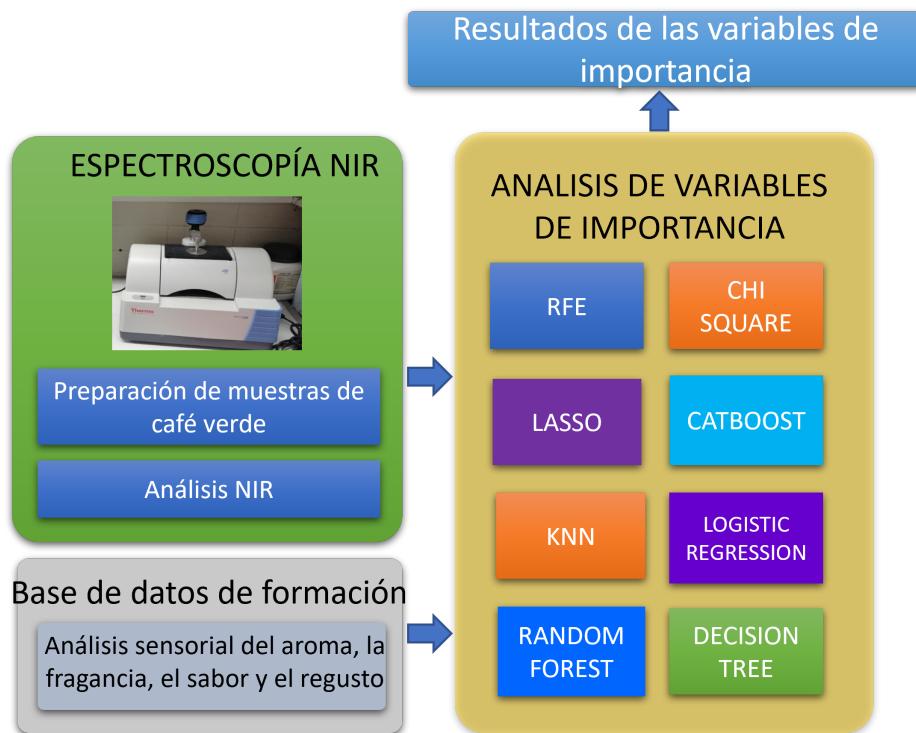


Figura 3.13: Metodología de Variables de Importancia relacionadas a los niveles de absorbancia de diferentes longitudes de onda respecto a las notas sensoriales del café

La figura 3.14 muestra los resultados del análisis de variables de importancia entre las longitudes de onda de 500 a 1200 nm respecto a las notas sensoriales. Se puede apreciar que las longitudes de onda 790-840nm (R<sub>2</sub>C=CHR), 700-900nm (S-OR esters), 1050-1200nm C=S (thiocarbonyl) , 1200-1300nm (N-O amine oxide), 1030-1060nm (S=O Sulfoxide) son lo que tienen un mayor impacto en la determinación de todas las notas sensoriales del café en el fragancia, aroma, sabor y resabio.

La figura 3.15 muestra los resultados del análisis de variables de importancia entre las longitudes de onda de 1200 a 1695 nm respecto a las notas sensoriales. Se puede apreciar que las longitudes de onda de 1250-1360 nm (Ar<sub>2</sub>NH), 1500-1600nm (N=O nitroso), 1637-1645nm (3 ring), 1645-1655nm (RCONHR' \_RCONH<sub>2</sub>) son lo que tienen un mayor impacto en la determinación de todas las notas sensoriales del café en el fragancia, aroma, sabor y resabio.

La figura 3.16 muestra los resultados del análisis de variables de importancia entre las longitudes de onda de 1600 a 3600 nm respecto a las notas sensoriales. Se puede apreciar que las longitudes de onda de 1695-1705nm (ArCHO), 2110-2165nm (R-N-C), 2100-2360nm (Si-H silane), 2850-3000 (RCH<sub>2</sub>CH<sub>3</sub>) son lo que tienen un mayor relevancia en la determinación de todas las notas sensoriales del café en el fragancia, aroma, sabor y resabio.

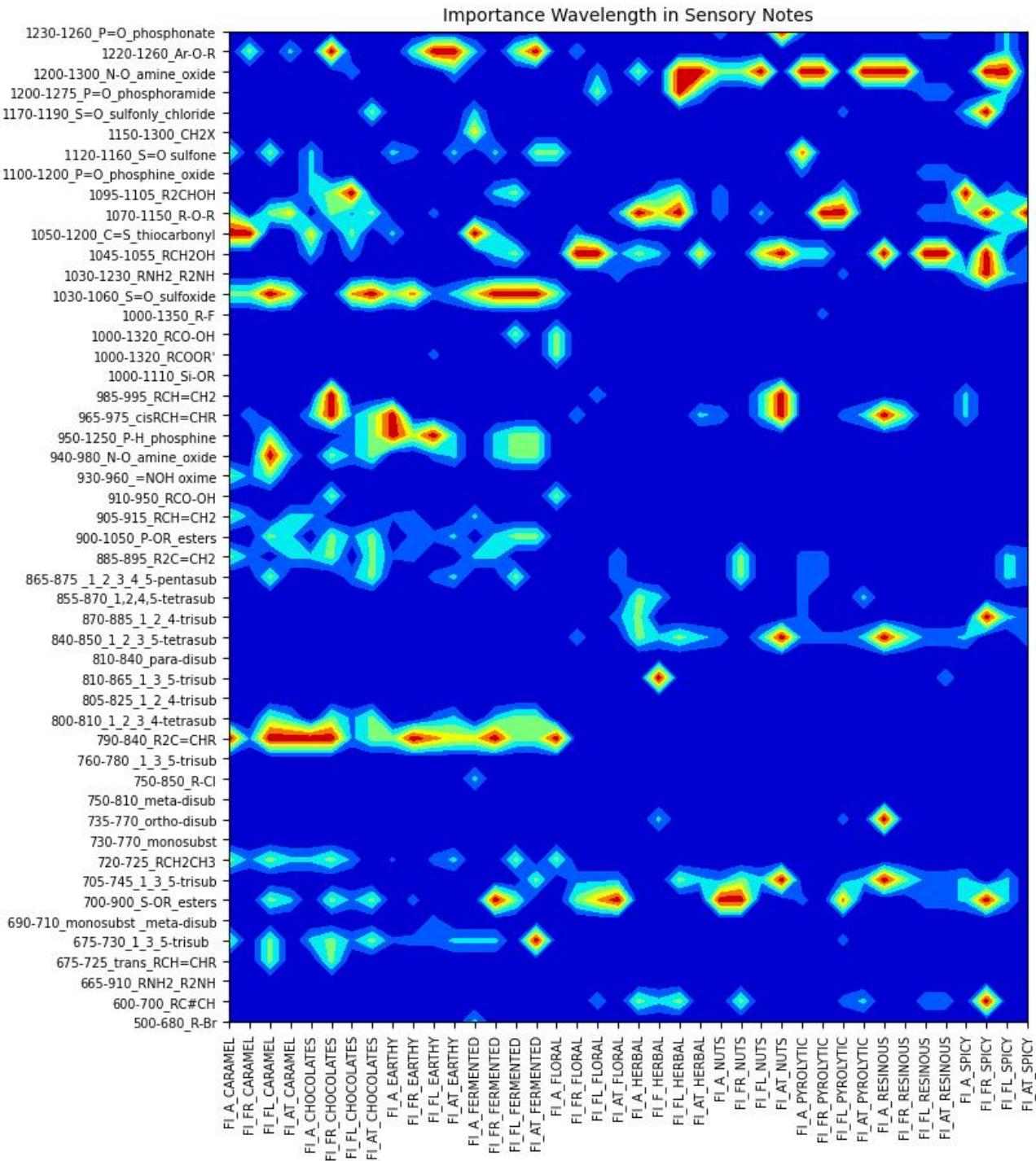


Figura 3.14: Variables de Importancia en las longitudes de onda de 500nm a 1200nm respecto a las notas sensoriales de aroma, fragancia, sabor y resabio. Donde FI\_A es aroma, FI\_FR es fragancia, FI\_FL es sabor y FI\_AT es resabio.

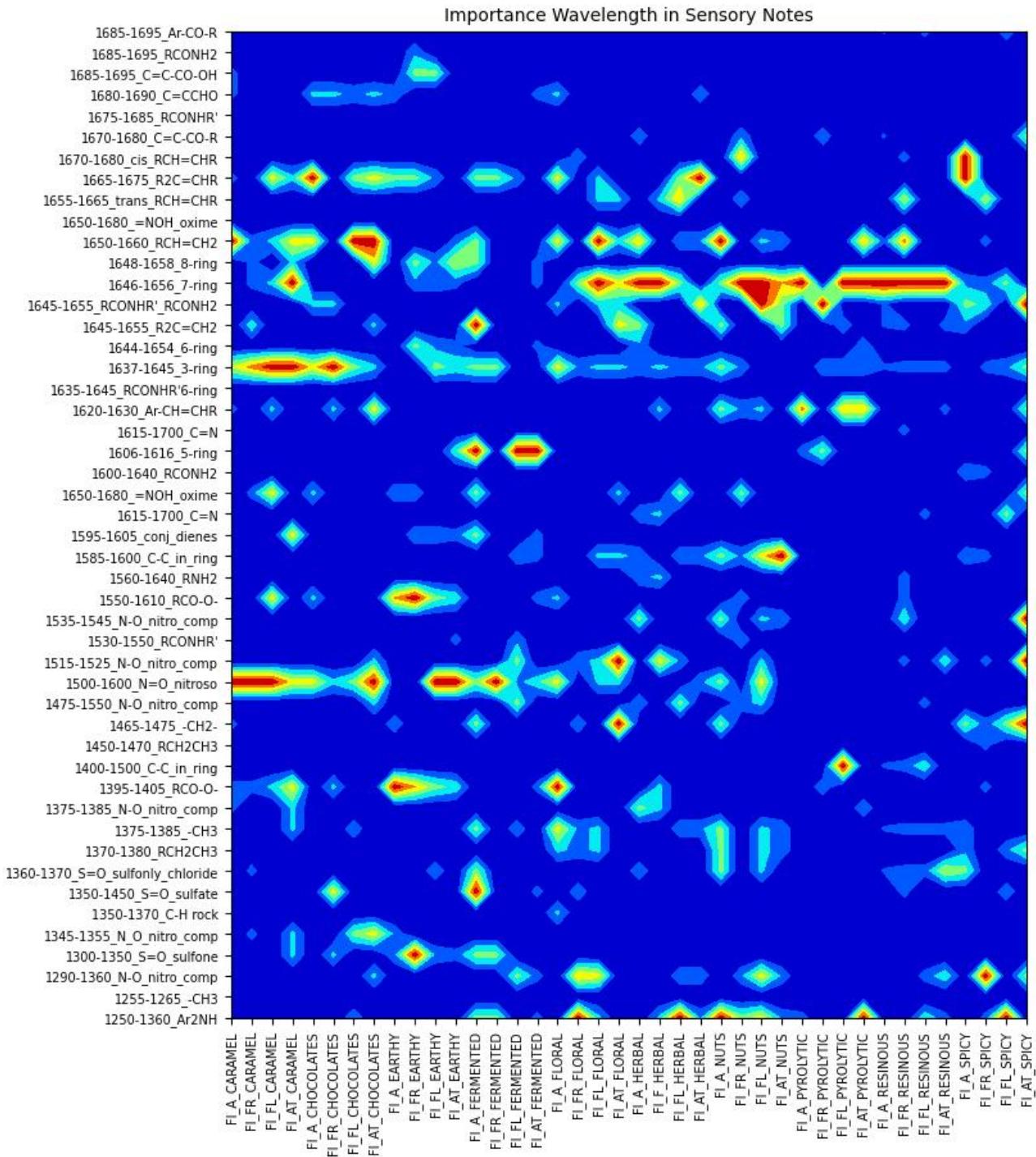


Figura 3.15: Variables de Importancia en las longitudes de onda de 1200nm a 1695nm respecto a las notas sensoriales de aroma, fragancia, sabor y resabio. Donde FI\_A es aroma, FI\_FR es fragancia, FI\_FL es sabor y FI\_AT es resabio.

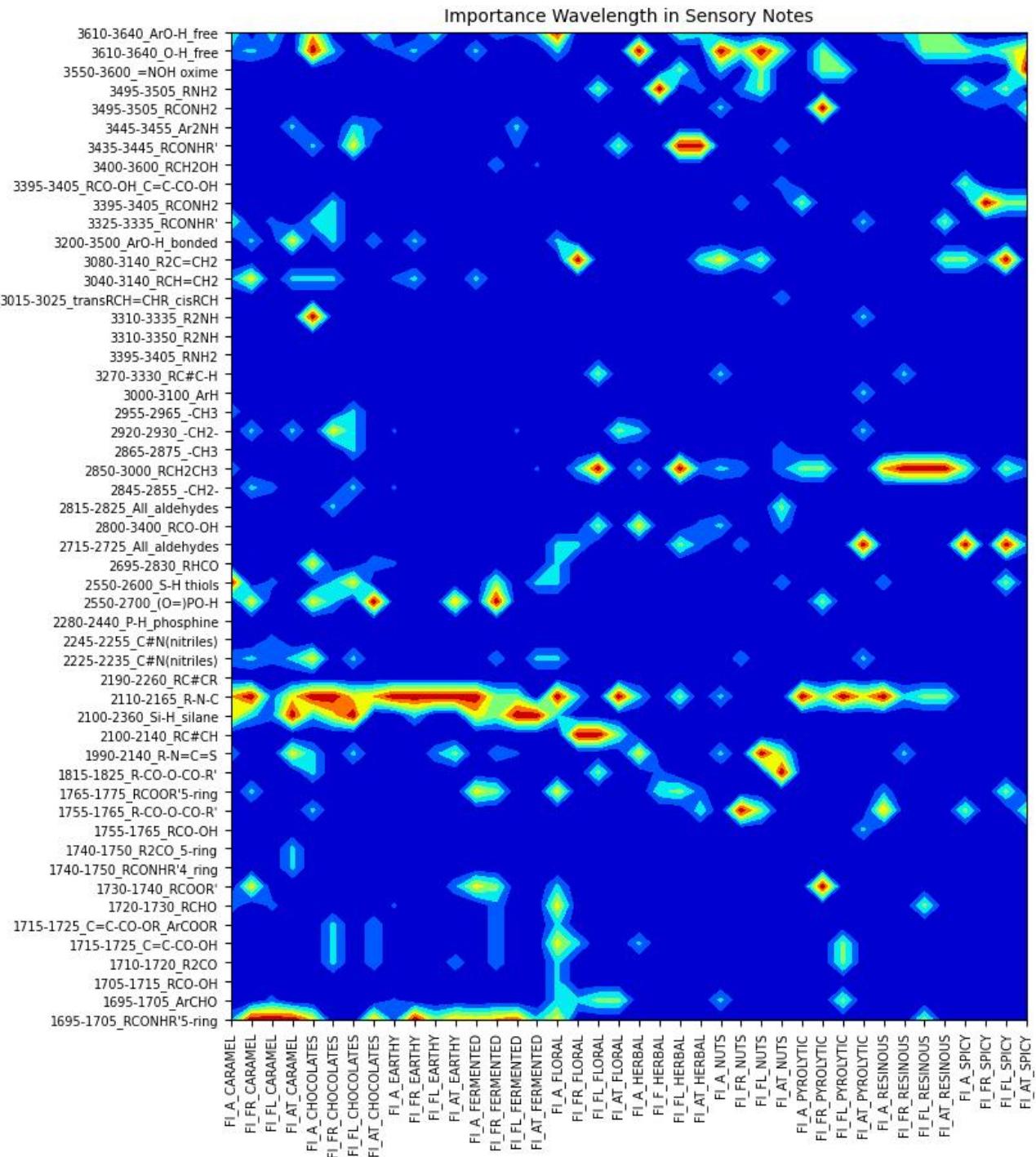


Figura 3.16: Variables de Importancia en las longitudes de onda de 1600nm a 3600nm respecto a las notas sensoriales de aroma, fragancia, sabor y resabio. Donde FI\_A es aroma, FI\_FR es fragancia, FI\_FL es sabor y FI\_AT es resabio.

Se realizó un análisis de las variables de importancia para conocer el impacto que tienen las diferentes

longitudes de onda de todas las notas sensoriales de fragancia, aroma, sabor y resabio.

En la figura 3.17 se muestran los resultados ordenados por importancia de todas las notas sensoriales del **aroma** respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro. Se puede apreciar que las longitudes de onda 2110-2165 (R-N-C), 1646-1656 (7-ring), 1650-1660 (RCH=CH2), 3610-3640 (O-H free) y 1665-1675 (R2C=CHR) son las variables con mayor impacto en la determinación de las notas sensoriales del aroma.

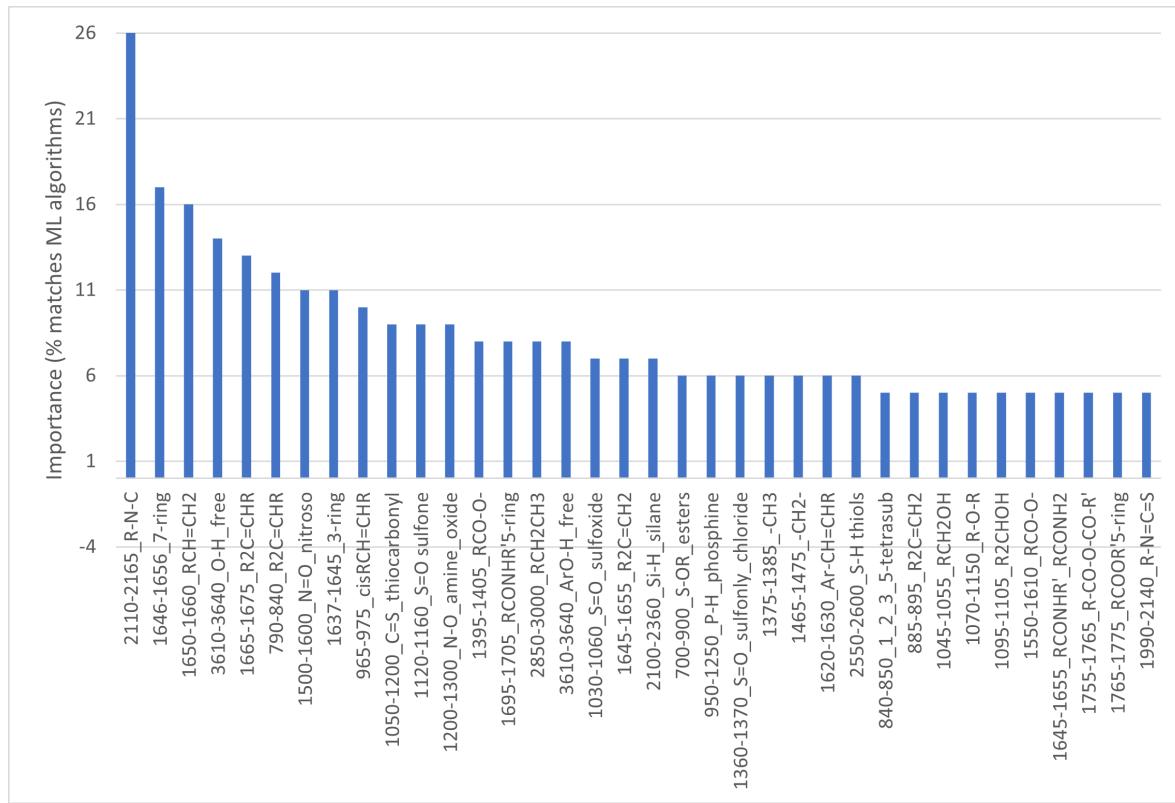


Figura 3.17: Resultados ordenados por importancia de las variables de todas las notas sensoriales del Aroma respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro.

En la figura 3.18 se muestran los resultados ordenados por importancia de todas las notas sensoriales de la **fragancia** respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro. Se puede apreciar que las longitudes de onda 1646-1656 (7 ring), 1637-1645 (3 ring), 2110-2165 (R-N-C), 790-840 (R2c=CHR) y 700-900 (S-OR esters) son las variables con mayor impacto en la determinación de las notas sensoriales del fragancia.

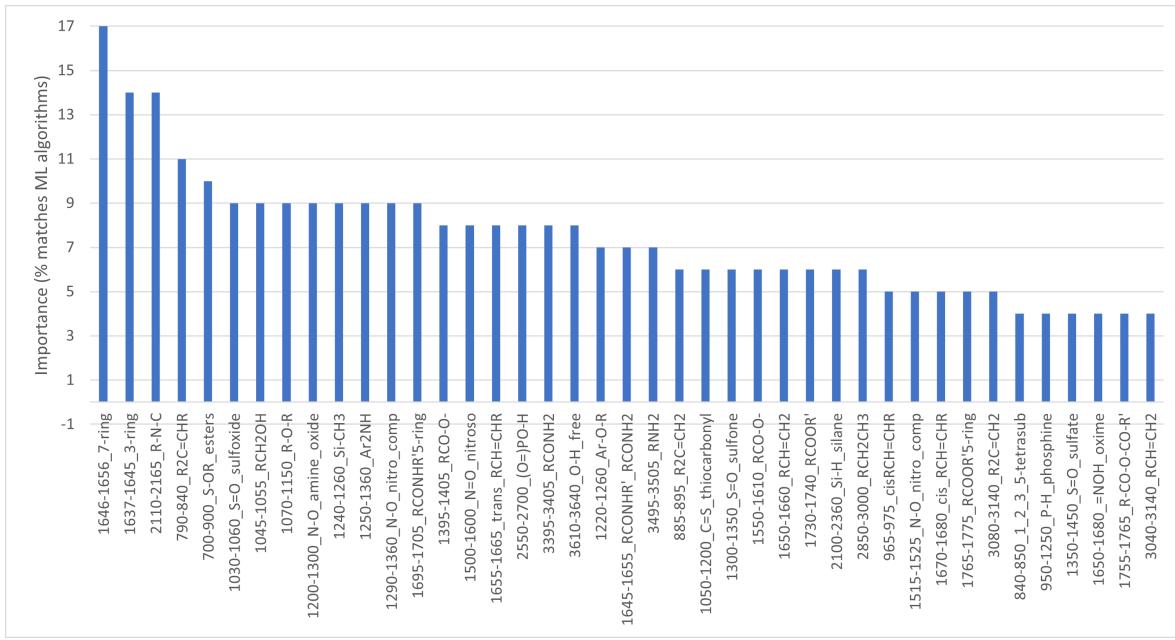


Figura 3.18: Resultados ordenados por importancia de las variables de todas las notas sensoriales del Fragancia respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro.

En la figura 3.19 se muestran los resultados ordenados por importancia de todas las notas sensoriales del **sabor** respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro. Se puede apreciar que las longitudes de onda 1646-1656 (7 ring), 1250-1360 (Ar2NH), 2110-2165 (R-N-C), 1070-1150 (R-O-R) y 1500-1600 (N=O\_nitroso) son las variables con mayor impacto en la determinación de las notas sensoriales del sabor.

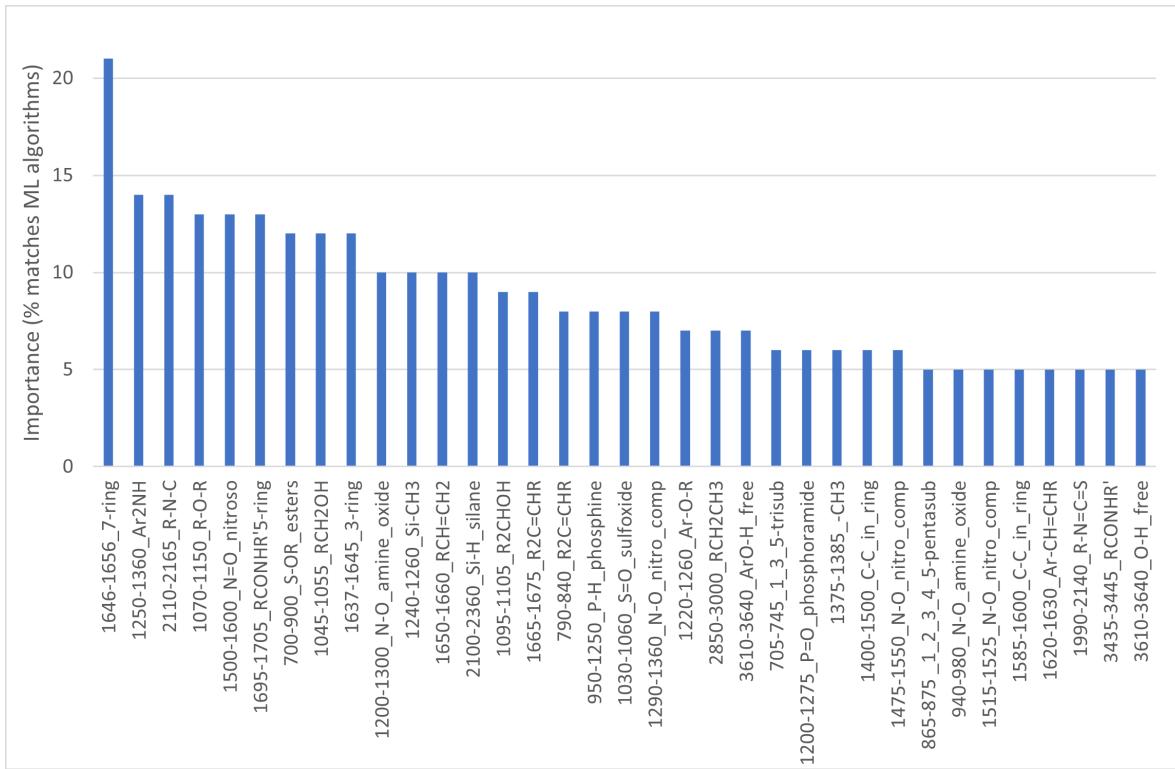


Figura 3.19: Resultados ordenados por importancia de las variables de todas las notas sensoriales del Sabor respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro

En la figura 3.20 se muestran los resultados ordenados por importancia de todas las notas sensoriales del **resabio** respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro. Se puede apreciar que las longitudes de onda 1646-1656 (7 ring), 1500-1600 (N=O\_nitroso), 2110-2165 (R-N-C), 1650-160 (RCH=CH2), 1695-1705 (RCONHR5-ring), 1637-1645 (3 ring) y 2100-2360 Si-H silane son las variables con mayor impacto en la determinación de las notas sensoriales del resabio.

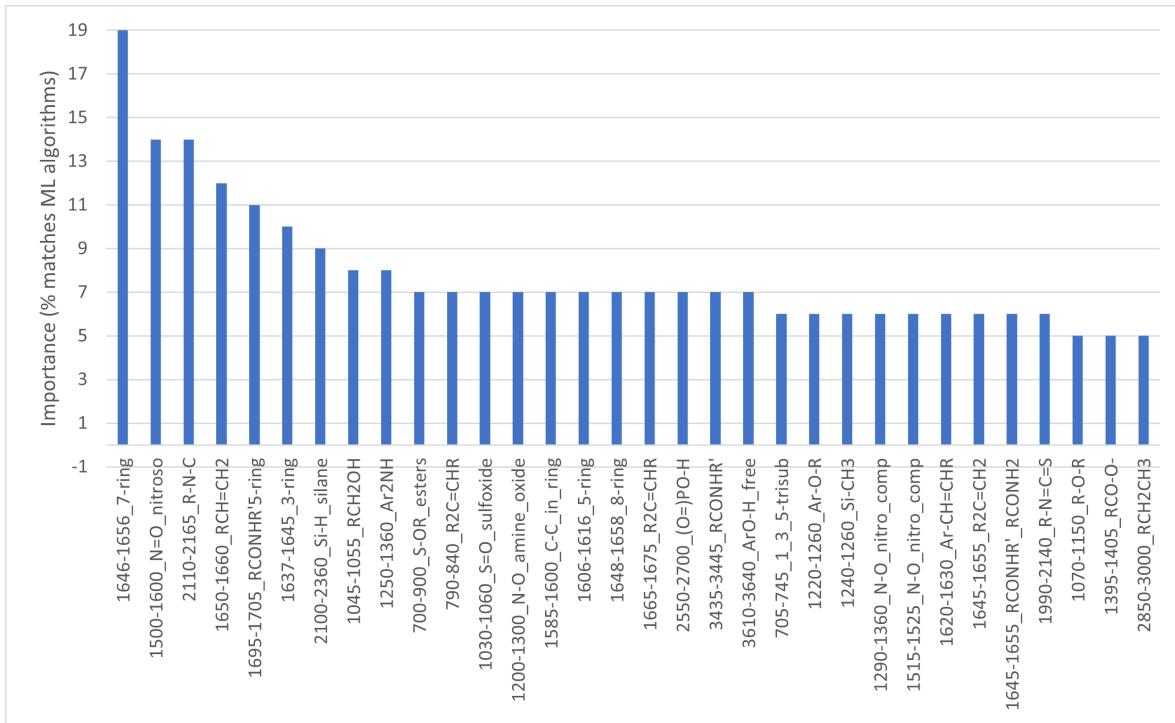


Figura 3.20: Resultados ordenados por importancia de las variables de todas las notas sensoriales del Resabio respecto a las estructuras orgánicas con su longitud de onda del espectrofotómetro

En las tablas 3.2 y 3.3 se muestran las variables de importancia del aroma, fragancia, sabor y resabio que impactan en las notas sensoriales (caramelo, chocolate, terroso, fermentado, floral, herbal, nueces, pirolítico, resinoso, especies). Se puede apreciar claramente como distintas estructuras orgánicas inciden en el comportamiento de cada una de las notas sensoriales del café.

Tabla 3.2: variables de importancia del aroma, fragancia, sabor y resabio que impactan en las notas sensoriales (caramelo, chocolate, terroso, fermentado, floral)

Caramelo	Chocolates	Earthy	Fermented	Floral
1500-1600 (N=O_nitros)	2110-2165 (R-N-C)	2110-2165 (R-N-C)	2100-2360 (Si-H_silane)	700-900 (S-OR_esters)
1637-1645 (3-ring)	1650-1660 (RCH=CH2)	950-1250 (P-H_phosphine)	1695-1705 (RCONHR'5-ring)	1045-1055 (RCH2OH)
1695-1705 (RCONHR'5-ring)	1500-1600 (N=O_nitroso)	790-840 (R2C=CHR)	1606-1616 (5-ring)	1646-1656 (7-ring)
1650-1660 (RCH=CH2)	1637-1645 (3-ring)	1395-1405 (RCO-O-)	1030-1060 (S=O_sulfoxide)	2110-2165 (R-N-C)
2110-2165 (R-N-C)	790-840 (R2C=CHR)	1550-1610 (RCO-O-)	2110-2165 (R-N-C)	1650-1660 (RCH=CH2)
		1220-1260 (Ar-O-R)		
		1500-1600 (N=O_nitroso)		

Tabla 3.3: variables de importancia del aroma, fragancia, sabor y resabio que impactan en las notas sensoriales (herbal, nueces, pirolítico, resinoso, especies)

Herbal	Nuts	Pyrolytic	Resinous	Spicy
1646-1656 (7-ring)	1646-1656 (7-ring)	1646-1656 (7-ring)	1646-1656 (7-ring)	3395-3405 (RCONH2)
3435-3445 (RCONHR')	3610-3640 (O-H_free)	1200-1300 (N-O_amine_oxide)	2850-3000 (RCH2CH3)	1070-1150 (R-O-R)
1250-1360 (Ar2NH)	1585-1600 (C-C_in_ring)	1620-1630 (Ar-CH=CHR)	1045-1055 (RCH2OH)	3610-3640 (O-H_free)
1665-1675 (R2C=CHR)	1250-1360 (Ar2NH)	1070-1150 (R-O-R)	2110-2165(R-N-C)	1465-1475 (-CH2-)
1070-1150 (R-O-R)	1200-1300 (N-O_amine_oxide)	2110-2165 (R-N-C)	1360-1370 (S=O_sulfonly_chloride)	1290-1360 (N-O_nitro_comp)
	700-900 (S-OR_esters)			1645-1655 (RCONHR'_RCONH2)

### 3.4.4 Modelos de Inteligencia Artificial para la predicción de Propiedades Sensoriales y calidad del café utilizando datos de espectroscopía NIR

Utilizando el algoritmo automatizado que se programó para la generación y entrenamiento de modelos de Redes Neuronales Artificiales se realizaron diversos experimentos con la finalidad de saber con qué precisión se podrían estimar las variables de las notas sensoriales del café a partir de datos de espectroscopía NIR.

La figura 3.21 muestran los resultados de la regresión lineal obtenidas entre el valor deseado y el valor estimado de 4 muestras de café.

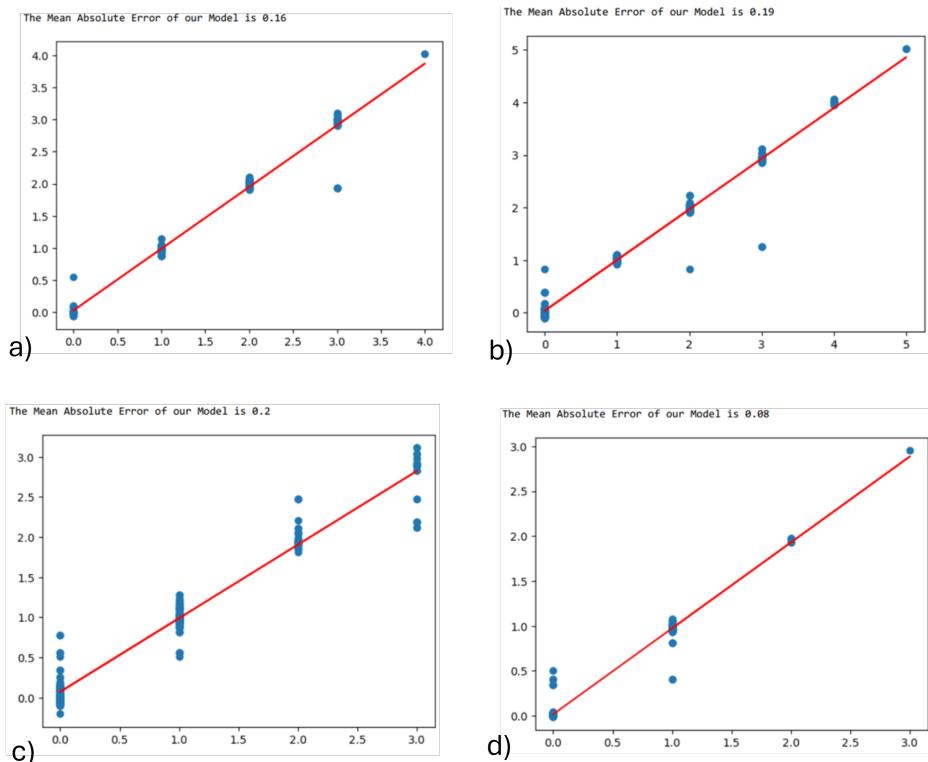


Figura 3.21: Precisión obtenida por las redes neuronales. Valores estimados vs valores deseados. a) Aroma Floral, b) Aroma chocolate, c) Aroma nueces, d) Aroma floral.

Las tablas 3.4, 3.5, 3.6 y ?? muestran los resultados de cada una de las notas sensoriales del café en el aroma, fragancia, sabor y resabio. Las tablas muestran los mejores resultados obtenidos de la experimentación con diferentes hiperparámetros de las redes neuronales. Indican el número de neuronas (units\_l1), el bias (bias\_l1) y la función de activación de la primera capa (act\_l1), el número de neuronas de la segunda capa (units\_l2), el bias (bias\_l2) de y la función de activación (act\_l2) de la segunda capa, el método de optimización utilizado (optimizer), y los resultados de la precisión obtenida por el entrenamiento (Train), prueba (Test), la precisión de la regresión lineal (Reg\_Lin) y el MAE.

Tabla 3.4: Resultados de la precisión para la estimación del Aroma del café utilizando redes neuronales con diferentes configuraciones de hiperparámetros

CARACTERISTICA	units_l1	bias_l1	act_l1	units_l2	bias_l2	act_l2	optimizer	Train	Test	Reg_Lin	MAE
A_FLORAL	378	False	relu	1898	False	tanh	rmsprop	0.026	0.056	0.98	0.24
A_FRUITY	1754	False	relu	2810	False	tanh	adam	0.061	0.101	0.99	0.32
A_HERBAL	2058	False	relu	826	True	tanh	rmsprop	0.001	0.002	0.99	0.04
A_NUTS	2778	True	tanh	2362	False	relu	adam	0.021	0.038	0.99	0.2
A_Caramel	2474	True	relu	138	False	relu	adam	0.006	0.025	0.99	0.16
A_CHOCOLATES	2570	True	relu	2618	True	relu	adam	0.045	0.037	0.99	0.19
A_SPICY	2410	True	relu	138	False	relu	rmsprop	0.011	0.007	0.99	0.08
A_RESINOUS	1274	False	relu	26	False	relu	adam	0	0	0.99	0.01
A_PYROLYTIC	1578	True	relu	1146	True	relu	rmsprop	0	0	0.99	0.01
A_EARTHY	10	False	relu	298	False	relu	rmsprop	0	0	0.99	0
A_FERMENTED	250	True	tanh	362	False	relu	adam	0.002	0.006	0.99	0.08
A_PHENOLICS	266	False	relu	1274	False	tanh	rmsprop	0	0	0	0

Tabla 3.5: Resultados de la precisión para la estimación de la Fragancia del café utilizando redes neuronales con diferentes configuraciones de los hiperparámetros

CARACTERISTICA	units_I1	bias_I1	act_I1	units_I2	bias_I2	act_I2	optimizer	Train	Test	Reg_Lin	MAE
CARACTERISTICA	units_I1	bias_I1	act_I1	units_I2	bias_I2	act_I2	optimizer	Train	Test	Reg_Lin	MAE
F_FLORAL	2810	False	relu	266	True	relu	rmsprop	0.017	0.028	0.99	0.17
F_FRUITY	938	True	relu	858	False	relu	rmsprop	0.018	0.033	0.99	0.18
F_HERBAL	2842	True	relu	1130	True	relu	rmsprop	0.004	0.003	0.99	0.05
F_NUTS	554	False	relu	2938	False	tanh	adam	0.018	0.015	0.99	0.12
F_CARAMEL	458	False	relu	586	True	relu	adam	0.03	0.035	0.99	0.19
F_CHOCOLATES	202	True	tanh	1706	False	relu	adam	0.033	0.036	0.99	0.19
F_SPICY	986	False	tanh	1018	False	relu	adam	0.007	0.015	0.99	0.12
F_RESINOUS	474	False	tanh	1098	False	relu	adam	0	0	0.99	0
F_PYROLYTIC	842	False	relu	378	True	tanh	rmsprop	0	0	0.99	0.01
F_EARTHY	2266	True	tanh	426	False	relu	adam	0	0	0.99	0.01
F_FERMENTED	362	False	tanh	1610	False	relu	adam	0	0	0.99	0

Tabla 3.6: Resultados de la precisión para la estimación del sabor del café utilizando redes neuronales con diferentes configuraciones de hiperparámetros

CARACTERISTICA	units_l1	bias_l1	act_l1	units_l2	bias_l2	act_l2	optimizer	Train	Test	Reg_Lin	MAE
S_FLORAL	1002	True	relu	1546	True	tanh	rmsprop	0.027	0.035	0.99	0.19
S_FRUITY	1898	False	relu	234	False	tanh	adam	0.062	0.066	0.99	0.26
S_HERBAL	2170	False	relu	2026	True	relu	rmsprop	0.013	0.013	0.99	0.11
S_NUTS	2970	False	relu	810	False	relu	adam	0.003	0.006	0.99	0.08
S_Caramel	2634	False	relu	2250	False	relu	rmsprop	0.008	0.008	0.99	0.09
S_CHOCOLATES	538	True	relu	2634	False	relu	adam	0.011	0.01	0.99	0.1
S_SPICY	2298	True	relu	2026	True	relu	rmsprop	0.019	0.025	0.99	0.16
S_RESINOUS	538	True	relu	2634	False	relu	adam	0	0	0.99	0
S_PYROLYTIC	2378	True	relu	2810	True	tanh	rmsprop	0	0	0.99	0
S_EARTHY	410	True	tanh	1914	True	relu	adam	0	0	0.99	0
S_FERMENTED	1402	False	relu	810	True	tanh	rmsprop	0	0	0.99	0.02
S_PHENOLICS	138	False	relu	1274	True	tanh	rmsprop	0	0	0	0

Tabla 3.7: Resultados de la precisión para la estimación del resabio del café utilizando redes neuronales con diferentes configuraciones de hiperparámetros

CARACTERISTICA	units_l1	bias_l1	act_l1	units_l2	bias_l2	act_l2	optimizer	Train	Test	Reg_Lin	MAE
R_FLORAL	1114	False	relu	2778	True	relu	rmsprop	0.031	0.032	0.99	0.18
R_FRUITY	250	True	tanh	938	True	relu	rmsprop	0.022	0.034	0.99	0.19
R_HERBAL	746	False	relu	314	False	relu	adam	0.013	0.008	0.99	0.09
R_NUTS	138	True	relu	2794	True	relu	adam	0.02	0.018	0.99	0.13
R_Caramel	186	False	relu	890	False	tanh	rmsprop	0.036	0.032	0.99	0.18
R_CHOCOLATES	74	False	tanh	2730	False	relu	adam	0.009	0.014	0.99	0.12
R_SPICY	1690	False	relu	506	False	relu	rmsprop	0.004	0.001	0.99	0.04
R_RESINOUS	266	True	relu	906	True	relu	adam	0	0	0.99	0
R_PYROLYTIC	2890	False	relu	2522	False	relu	rmsprop	0	0	0.99	0
R_EARTHY	970	False	relu	26	False	tanh	rmsprop	0	0	0.99	0.01
R_FERMENTED	1658	False	relu	1978	False	tanh	rmsprop	0.003	0.004	0.99	0.06
R_PHENOLICS	362	False	relu	1834	True	tanh	rmsprop	0	0	0	0

### **3.5 Conclusiones**

En este capítulo se generaron resultados que permitieron conocer la relación que existe entre ciertas combinaciones de estructuras orgánicas del café verde, obtenidas a través de la detección de picos relacionados a la absorbancia utilizando espectroscopía NIR, respecto a las notas sensoriales del café en aroma, fragancia, sabor y resabio. Los resultados presentan una aportación al estado del arte ya que fue posible determinar cuales fueron las longitudes de onda en los estudios de espectroscopia que impactan en cada una de las notas sensoriales del café como lo son afrutado, herbal, nueces, caramelo, chocolates, especies, resinoso, pirolítico, fermentado y fenólico. Este conocimiento generado nos permite conocer a profundidad la relación existente entre el perfil sensorial y los resultados de la espectroscopía en muestras de café verde.

Otro aportación que se realizó en este capítulo fue la generación de un sistema inteligente basado en Redes Neuronales Artificiales que es capaz de estimar las notas sensoriales del café utilizando los datos del la espectroscopía NIR en el café verde. Este tema también representa una aportación al conocimiento ya que aunque existen algunas investigaciones que han logrado estimar algunos parámetros sensoriales del café utilizando datos de espectroscopía, estos análisis los han desarrollado con café tostado y se han concentrado en estimar solo ciertas características que son más generales en el perfil sensorial del café. En comparación con los resultados obtenidos en esta sección se puede mencionar que el sistema fue capaz de realizar estimaciones de las notas sensoriales del café en el aroma, fragancia, sabor y resabio utilizando los resultados obtenidos por espectroscopía NIR en muestras de café verde obteniendo una precisión promedio del 99 %.

## CONCLUSIONES

Los resultados de este trabajo han demostrado que es factible comprender cómo las variables físicas y las deformaciones de los granos de café verde influyen en sus notas sensoriales, lo que ha permitido la creación de un detallado mapa analizando dicha relación.

La implementación de un modelo automatizado de aprendizaje automático ha posibilitado la determinación precisa de las notas sensoriales, lo que permite compararlas con las evaluaciones de expertos y desarrollar algoritmos más sofisticados. El ajuste de hiperparámetros ha contribuido a mejorar la precisión de la predicción de las propiedades sensoriales, aproximándose a las puntuaciones de catadores entrenados.

La estimación del perfil sensorial de los granos de café verde antes del proceso de tostado, a través del modelo de Machine Learning propuesto, ofrece a productores y comerciantes la oportunidad de analizar los datos y mejorar la calidad del producto. Esto también puede conducir a un modelo de comercio más justo y transparente.

La combinación de diferentes modelos de aprendizaje automático ha demostrado ser efectiva para alcanzar una precisión similar a la de los expertos en catación, reduciendo la posibilidad de errores inducidos por factores humanos como el cansancio o la subjetividad.

Otro punto importante ha sido la generación de información relevante sobre la relación entre las estructuras orgánicas del café verde detectadas mediante espectroscopía NIR y sus notas sensoriales, en aspectos como aroma, fragancia, sabor y resabio. Este conocimiento contribuye significativamente al estado del arte en el campo.

Se ha desarrollado un sistema inteligente basado en Redes Neuronales Artificiales que puede estimar las notas sensoriales del café utilizando datos de espectroscopía NIR en café verde, logrando una precisión promedio del 99 %. Este avance representa una contribución importante, ya que amplía el alcance de la estimación sensorial a partir de datos espectroscópicos y proporciona resultados altamente precisos.

## RECOMENDACIONES

Optimización de algoritmos de aprendizaje automático: Aunque el modelo actual ha demostrado una alta precisión en la estimación de las notas sensoriales del café verde, aún podría investigarse la optimización de los algoritmos de aprendizaje automático para mejorar aún más esta precisión. Se podrían explorar diferentes técnicas de aprendizaje automático, como redes neuronales profundas, para ver si se pueden obtener resultados aún más precisos.

Impacto del procesamiento del café verde en las notas sensoriales: Aunque el estudio se centró en el café verde, sería interesante investigar cómo el proceso de tostado y otros métodos de procesamiento del café afectan las notas sensoriales. Esto podría implicar la aplicación del mismo modelo de aprendizaje automático a diferentes etapas del procesamiento del café y comparar los resultados para comprender mejor cómo estos procesos influyen en las características sensoriales finales.

Validación del modelo en diferentes regiones y variedades de café: Para asegurar la generalización y aplicabilidad del modelo desarrollado, sería importante validar su rendimiento en diferentes regiones productoras de café y con diversas variedades de café. Esto permitiría determinar si el modelo es robusto y puede aplicarse ampliamente en la industria del café.

Exploración de otras técnicas de espectroscopía para caracterizar el café verde: Aunque el estudio se centró en la espectroscopía NIR, podría ser útil investigar otras técnicas espectroscópicas para caracterizar el café verde y su relación con las notas sensoriales. Esto podría incluir técnicas como la espectroscopía de resonancia magnética nuclear (RMN) o la espectroscopía de masas, entre otras.

Aplicación del modelo en otros productos agrícolas: Dado el éxito del modelo en la estimación de las notas sensoriales del café verde, podría investigarse su aplicabilidad en otros productos agrícolas. Esto podría incluir productos como el cacao, el té, las especias u otros productos alimenticios donde las características sensoriales son importantes para la calidad del producto.

## Bibliografía

- Agrícola Nacional CAFÉ.* (2017). SAGARPA. [https://www.gob.mx/cms/uploads/attachment/file/256426/B\\_sico-Caf\\_.pdf](https://www.gob.mx/cms/uploads/attachment/file/256426/B_sico-Caf_.pdf)
- Aitkenhead, M. J., Gaskin, G. J., Lafouge, N., & Hawes, C. (2017). Phylis: A low-cost portable visible range spectrometer for soil and plants. *Sensors (Switzerland)*, 17. <https://doi.org/10.3390/s17010099>
- Alessandrini, L., Romani, S., Pinnavaia, G., & Rosa, M. D. (2008). Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Analytica Chimica Acta*, 625, 95-102. <https://doi.org/10.1016/j.aca.2008.07.013>
- Arana, V. A., Medina, J., Alarcon, R., Moreno, E., Heintz, L., Schäfer, H., & Wist, J. (2015). Coffee's country of origin determined by NMR: The Colombian case. *Food Chemistry*, 175, 500-506. <https://doi.org/10.1016/j.foodchem.2014.11.160>
- Arboleda, E. R. (2018). Discrimination of civet coffee using near infrared spectroscopy and artificial neural network. *International Journal of Advanced Computer Research*, 8, 324-334. <https://doi.org/10.19101/IJACR.2018.839007>
- Barbin, D. F., de Souza Madureira Felicio, A. L., Sun, D. W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, 61, 23-32. <https://doi.org/10.1016/j.foodres.2014.01.005>
- Barbosa, R. M., Batista, B. L., Varrique, R. M., Coelho, V. A., Campiglia, A. D., & Barbosa, F. (2014). The use of advanced chemometric techniques and trace element levels for controlling the authenticity of organic coffee. *Food Research International*, 61, 246-251. <https://doi.org/10.1016/j.foodres.2013.07.060>
- Berardinis, J. D., Pizzuto, G., Lanza, F., Chella, A., Meira, J., & Cangelosi, A. (2020a). At Your Service: Coffee Beans Recommendation from a Robot Assistant. *HAI 2020 - Proceedings of the 8th International Conference on Human-Agent Interaction*, 257-259. <https://doi.org/10.1145/3406499.3418765>
- Berardinis, J. D., Pizzuto, G., Lanza, F., Chella, A., Meira, J., & Cangelosi, A. (2020b). At Your Service: Coffee Beans Recommendation from a Robot Assistant. *HAI 2020 - Proceedings of the 8th International Conference on Human-Agent Interaction*, 257-259. <https://doi.org/10.1145/3406499.3418765>
- Bhawani, S. A., Fong, S. S., & Ibrahim, M. N. M. (2015). Spectrophotometric Analysis of Caffeine. *International Journal of Analytical Chemistry*, 2015. <https://doi.org/10.1155/2015/170239>
- Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Discrimination of Arabica and Robusta in Instant Coffee by Fourier Transform Infrared Spectroscopy and Chemometrics. *Journal of Agricultural and Food Chemistry*, 44, 170-174. <https://doi.org/10.1021/jf950305a>
- Cebi, N., Yilmaz, M. T., & Sagdic, O. (2017). A rapid ATR-FTIR spectroscopic method for detection of sibutramine adulteration in tea and coffee based on hierarchical cluster and principal component analyses. *Food Chemistry*, 229, 517-526. <https://doi.org/10.1016/j.foodchem.2017.02.072>

- CEDRSSA. (2018). Reporte del café en México: diagnóstico y perspectiva. *Reporte del café en México: diagnóstico y perspectiva*, 33.
- Chandrasekar, V., & Viswanathan, ; R. (1999). *Physical and Thermal Properties of Coffee*. <http://www.idealibrary.com>
- Chang, Y. T., Hsueh, M. C., Hung, S. P., Lu, J. M., Peng, J. H., & Chen, S. F. (2021). Prediction of specialty coffee flavors based on near-infrared spectra using machine and deep-learning methods. *Journal of the Science of Food and Agriculture*, 101, 4705-4714. <https://doi.org/10.1002/jsfa.11116>
- Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation. (2016). *European Food Research and Technology*, 242, 1403-1409. <https://doi.org/10.1007/s00217-016-2643-y>
- Chou, Y. C., Kuo, C. J., Chen, T. T., Horng, G. J., Pai, M. Y., Wu, M. E., Lin, Y. C., Hung, M. H., Su, W. T., Chen, Y. C., Wang, D. C., & Chen, C. C. (2019). Deep-learning-based defective bean inspection with GAN-structured automated labeled data augmentation in coffee industry. *Applied Sciences (Switzerland)*, 9. <https://doi.org/10.3390/app9194166>
- Cirillo, M. Â., Ramos, M. F., Borém, F. M., De, F. M., Ribeiro, D. E., & Menezes, F. S. D. (2019). Acta Scientiarum of coffee with different qualities using the distribution of the extremes of the highest scores. <https://doi.org/10.4025/actasciagron.v41i1.39323>
- Corrales, D. C., Ledezma, A., Q., A. J. P., Hoyos, J., Figueroa, A., & Corrales, J. C. (2014). A new dataset for coffee rust detection in Colombian crops base on classifiers. *Sistemas y Telemática*, 12, 9. <https://doi.org/10.18046/syt.v12i29.1802>
- da Silva Araújo, C., Macedo, L. L., Vimercati, W. C., Ferreira, A., Prezotti, L. C., & Saraiva, S. H. (2020). Determination of pH and acidity in green coffee using near-infrared spectroscopy and multivariate regression. *Journal of the Science of Food and Agriculture*, 55. <https://doi.org/10.1002/jsfa.10270>
- Delgado, F. G., Rincón, F. Z., & Vela, J. A. O. (2008). Coffee certification criterion using the photoacoustic technique. *Microelectronics Journal*, 39, 1331-1332. <https://doi.org/10.1016/j.mejo.2008.01.072>
- Dias, R. C. E., Valderrama, P., Março, P. H., dos Santos Scholz, M. B., Edelmann, M., & Yeretzian, C. (2018). Quantitative assessment of specific defects in roasted ground coffee via infrared-photoacoustic spectroscopy. *Food Chemistry*, 255, 132-138. <https://doi.org/10.1016/j.foodchem.2018.02.076>
- Discrimination of defective (Full Black, Full Sour and Immature) and nondefective coffee beans by their physical properties. (2014). *Journal of Food Process Engineering*, 37, 524-532. <https://doi.org/10.1111/jfpe.12113>
- Feria-Morales, A. M. (2002). Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. *Food Quality and Preference*, 13(6), 355-367. [https://doi.org/https://doi.org/10.1016/S0950-3293\(02\)00028-9](https://doi.org/https://doi.org/10.1016/S0950-3293(02)00028-9)

- Ferreira, H. A., Liska, G. R., Cirillo, M. A., Borém, F. M., Ribeiro, D. E., Cortez, R. M., & Guiraldeli, C. H. C. (2016). Sensory Analysis Of Specialty Coffees Performed With Consumer. *IEEE Latin America Transactions*, 14, 1507-1512.
- Franca, A. S., Mendonça, J. C., & Oliveira, S. D. (2005). Composition of green and roasted coffees of different cup qualities. *LWT - Food Science and Technology*, 38, 709-715. <https://doi.org/10.1016/j.lwt.2004.08.014>
- Garcia, P. (s.f.). PRODUCCIÓN Y CALIDAD EN VARIEDADES DE CAFÉ (Coffea arabica L.)
- Gordillo-Delgado, F., Marín, E., Cortés-Hernández, D. M., Mejía-Morales, C., & García-Salcedo, A. J. (2012). Discrimination of organic coffee via Fourier transform infrared-photoacoustic spectroscopy. *Journal of the Science of Food and Agriculture*, 92, 2316-2319. <https://doi.org/10.1002/jsfa.5628>
- Goyal, S., & Goyal, G. K. (2013). Machine Learning ANN Models for Predicting Sensory Quality of Roasted Coffee Flavoured Sterilized Drink. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 2, 09-13. <https://doi.org/10.14201/adcaij201426913>
- Hendrawan, Y., Widyaningtyas, S., & Sucipto, S. (2019). Computer vision for purity, phenol, and pH detection of Luwak coffee green bean. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17, 3073-3085. <https://doi.org/10.12928/TELKOMNIKA.v17i6.12689>
- Hernández-Aguilar, C., Domínguez-Pacheco, A., Cruz-Orea, A., & Ivanov, R. (2019). Photoacoustic spectroscopy in the optical characterization of foodstuff: A review. *Journal of Spectroscopy*, 2019. <https://doi.org/10.1155/2019/5920948>
- ICO. (2021). Coffee prices rallied in April 2021 amid an anticipated decline in production and signs of economic recovery from covid-19 pandemic, 1-10.
- Jane, J. C., Kahiu, N., James, W. M., & Chrispine, O. O. (2020). Assessment of Arabusta coffee hybrids [Coffea arabica L. X Tetraploid Robusta (Coffea canephora)] for green bean physical properties and cup quality. *African Journal of Food Science*, 14, 119-127. <https://doi.org/10.5897/ajfs2019.1852>
- Landeta-Escamilla, O., Sandoval-Gonzalez, O., Martínez-Sibaja, A., de Jesús Agustín Flores-Cuautle, J., Posada-Gómez, R., & Alvarado-Lassman, A. (2019). Intelligent spectroscopy system used for physicochemical variables estimation in sugar cane soils. *Sensors (Switzerland)*, 19. <https://doi.org/10.3390/s19020240>
- Leme, D. S., da Silva, S. A., Barbosa, B. H. G., Borém, F. M., & Pereira, R. G. F. A. (2019). Recognition of coffee roasting degree using a computer vision system. *Computers and Electronics in Agriculture*, 156, 312-317. <https://doi.org/10.1016/j.compag.2018.11.029>
- Livio, J., & Hodhod, R. (2018). AI Cupper: A Fuzzy Expert System for Sensorial Evaluation of Coffee Bean Attributes to Derive Quality Scoring. *IEEE Transactions on Fuzzy Systems*, 26, 3418-3427. <https://doi.org/10.1109/TFUZZ.2018.2832611>
- Mendonça, J. C., Franca, A. S., & Oliveira, L. S. (2009). Physical characterization of non-defective and defective Arabica and Robusta coffees before and after roasting. *Journal of Food Engineering*, 92, 474-479. <https://doi.org/10.1016/j.jfoodeng.2008.12.023>

Mexicana, N. (s.f.). *SECRETARÍA DE ECONOMÍA CAFÉ VERDE DE ESPECIALIDAD-ESPECIFICACIONES, CLASIFICACIÓN Y EVALUACIÓN SENSORIAL SPECIALITY GREEN COFFEE-SPECIFICATIONS, PREPARATIONS AND SENSORY EVALUATION.*

Oliveira, E. M. D., Leme, D. S., Barbosa, B. H. G., Rodarte, M. P., & Pereira, R. G. F. A. (2016). A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food Engineering*, 171, 22-27. <https://doi.org/10.1016/j.jfoodeng.2015.10.009>

Orduña, A. D., Rodríguez, G. Á., Hernández, G. L., Gonzalez, O. O. S., Escamilla, O. L., Sanchez, B. E. G., & de Jesús Agustín Flores Cuautle, J. (2016). Development and performance evaluation of moisture range system for green coffee grains. *Cogent Food & Agriculture*, 2, 1-7. <https://doi.org/10.1080/23311932.2016.1210067>

Quintero, G. I. P. (2011). COMPOSICIÓN QUÍMICA DE UNA TAZA DE CAFÉ El. *Cenicafé*, 414, 12. <https://doi.org/10.1049/iet-map:20060301>

Revista fitotecnia mexicana publicada por la Sociedad Mexicana de Fitogenética. (2016). *Revista fitotecnia mexicana*, 39, 297-304. [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0187-73802016000300297#B6](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-73802016000300297#B6)

Ribeiro, J. S., Augusto, F., Salva, T. J., & Ferreira, M. M. (2012). Prediction models for Arabica coffee beverage quality based on aroma analyses and chemometrics. *Talanta*, 101, 253-260. <https://doi.org/10.1016/j.talanta.2012.09.022>

Ribeiro, J. S., Ferreira, M. M., & Salva, T. J. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, 83, 1352-1358. <https://doi.org/10.1016/j.talanta.2010.11.001>

Sensorial, E., & Cafe, D. (2015). *Evaluación sensorial del café*. “Creación de Capacidades en Asistencia Técnica a Productores de Café en Guatemala” Financiado. <http://scanprogram.org/wp-content/uploads/2012/08/20151026-Evaluacion-sensorial-del-cafe.pdf>

Shepherd, K. D., & Walsh, M. G. (2007). Review: Infrared spectroscopy - Enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy*, 15, 1-19. <https://doi.org/10.1255/jnirs.716>

Sinaga, S. H., & Julianti, E. (2021). Physical characteristics of Gayo arabica coffee with semi-washed processing. *IOP Conference Series: Earth and Environmental Science*, 782. <https://doi.org/10.1088/1755-1315/782/3/032093>

Sunarharum, W. B., Yuwono, S. S., Pangestu, N. B., & Nadhiroh, H. (2018). Physical and sensory quality of Java Arabica green coffee beans. *IOP Conference Series: Earth and Environmental Science*, 131. <https://doi.org/10.1088/1755-1315/131/1/012018>

Titova, T., Nachev, V., & Damyanov, C. (2015). FOOD QUALITY EVALUATION ACCORDING TO THEIR COLOR CHARACTERISTICS. 14, 1-10.

- Viejo, C. G., Tongson, E., & Fuentes, S. (2021). Integrating a low-cost electronic nose and machine learning modelling to assess coffee aroma profile and intensity. *Sensors*, 21, 1-16. <https://doi.org/10.3390/s21062016>
- Virgen-Navarro, L., Herrera-López, E. J., Corona-González, R. I., Arriola-Guevara, E., & Guatemala-Morales, G. M. (2016). Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed. *Expert Systems with Applications*, 54, 162-169. <https://doi.org/10.1016/j.eswa.2016.01.027>
- Wakhid, S., Sarno, R., Sabilla, S. I., & Maghfira, D. B. (2020). Detection and classification of indonesian civet and non-civet coffee based on statistical analysis comparison using E-Nose. *International Journal of Intelligent Engineering and Systems*, 13, 56-65. <https://doi.org/10.22266/IJIES2020.0831.06>
- Walleigh, S., Polceanu, M., Jemal, T., & Buche, C. (2019). Coffee grading with convolutional neural networks using small datasets with high variance. *Journal of WSCG*, 27, 113-120. <https://doi.org/10.24132/JWSCG.2019.27.2.4>
- Weldegebreal, B., Redi-Abshire, M., & Chandravanshi, B. S. (2017). Development of new analytical methods for the determination of caffeine content in aqueous solution of green coffee beans. *Chemistry Central Journal*, 11, 1-9. <https://doi.org/10.1186/s13065-017-0356-3>
- Wondimkun, Y. W., Emire, S. A., & Esho, T. B. (2020). Investigation of Physical and Sensory Properties of Ethiopian Specialty Dry Processed Green Coffee Beans. *Acta Universitatis Cibiniensis. Series E: Food Technology*, 24, 39-48. <https://doi.org/10.2478/aucft-2020-0004>
- Yusibani, E., Putra, R. I., Rahwanto, A., Surbakti, M. S., Rajibussalim & Rahmi. (2022). Physical properties of Sidikalang robusta coffee beans medium roasted from various colors of coffee cherries. *Journal of Physics: Conference Series*, 2243. <https://doi.org/10.1088/1742-6596/2243/1/012046>
- Zhang, C., Wang, C., Liu, F., & He, Y. (2016). Mid-infrared spectroscopy for coffee variety identification: Comparison of pattern recognition methods. *Journal of Spectroscopy*, 2016. <https://doi.org/10.1155/2016/7927286>

# PRODUCTOS ACADÉMICOS

## PRODUCTOS GENERADOS

### **1. Escritura del artículo JCR.**

Como parte de los resultados; se documentó la investigación a través de un artículo, el cual fue publicado en la revista Processes 2024, 12, 18. <https://doi.org/10.3390/pr12010018>.

### **2. Registro de Software.**

Otro resultado fue el registro de obra ante INDAUTOR, el cual a través de un programa basado en algoritmos de aprendizaje máquina, son capaces de estimar y predecir las notas sensoriales del café. La estimación se realizó utilizando diversos algoritmos como Regresión logística, SVC, KNN, Decisión Tree, Random Forest, CatBoost, Naive Bayes.

### **3. Proyecto Financiado COVEICYDET.**

A través del financiamiento del proyecto 15 2243/2021, se lograron obtener resultados en la experimentación, con técnicas de espectroscopía para la determinación de la calidad del café con la ayuda de algoritmos de inteligencia artificial, logrando una estimación cuantitativa y cualitativa de las propiedades fisicoquímicas del café, permitiendo realizar una aportación al estado del arte.

## ANEXOS

**Determinación de Variables de Importancia:** El código está enfocado a la determinación de las variables de importancia de las variables de variedad, tipo y defecto del grano del café que influyen en la determinación de las notas sensoriales de la fragancia, aroma, sabor y resabio.

```
1
2 import pandas as pd
3 from tensorflow.keras.models import Sequential
4 from tensorflow.keras.layers import Dropout
5 from tensorflow.keras.layers import Dense
6 from tensorflow.keras.optimizers import SGD
7 from scikeras.wrappers import KerasClassifier #pip install scikeras==0.1.8 --user
8 from sklearn.model_selection import cross_val_score
9 from sklearn.preprocessing import LabelEncoder
10 from sklearn.model_selection import StratifiedKFold
11 from sklearn.preprocessing import StandardScaler
12 from sklearn.pipeline import Pipeline
13
14 DB_NIR=pd.read_csv("2021_Nov_DB_Total_en.csv")
15 DB_NIR.head()
16 DB_NIR.info
17
18 #Seleccion de las variables a Utilizar
19
20 from sklearn import preprocessing
21
22 DB_NIR=DB_NIR.fillna(1)
23 X=DB_NIR.iloc[ValMIni:ValMEnd,ValIni:ValEnd].values
24 yi=DB_NIR[VAR_OUTPUT].values
25 y=yi[ValMIni:ValMEnd]
26
27 min_max_scaler = preprocessing.MinMaxScaler()
28 Xs=min_max_scaler.fit_transform(X)
29
30 # Use train_test_split() Method.
31 from sklearn.model_selection import train_test_split
32 Xtrain, Xtest, ytrain,ytest= train_test_split(Xs,y, test_size=0.4)
33
34 from tensorflow.keras.models import Sequential
35 from tensorflow.keras import layers
36 from tensorflow.keras.optimizers import SGD
37
38 def build_model(hyperparams):
39     model = Sequential()
```

```

40     model.add(layers.Input(shape=(Xs.shape[1],)))
41     model.add(layers.Dense(units=hyperparams.Int("units_l1", 10, 2000, step=16),
42                            use_bias=hyperparams.Boolean("bias_l1"),
43                            activation=hyperparams.Choice("act_l1", ["relu", "tanh"]))
44                            ))
45     #model.add(Dropout(0.05))
46     model.add(layers.Dense(units=hyperparams.Int("units_l2", 10, 2000, step=16),
47                            use_bias=hyperparams.Boolean("bias_l2"),
48                            activation=hyperparams.Choice("act_l2", ["relu", "tanh"]))
49                            ))
50     #model.add(Dropout(0.05))
51     model.add(layers.Dense(1))
52
53     optim=hyperparams.Choice("optimizer",["sgd","rmsprop","adam"])
54     model.compile(optim, loss="mean_squared_error", metrics=["mean_squared_error"])
55
56     return model
57
58     #pip install keras-tuner --upgrade
59 from keras_tuner import RandomSearch
60 from keras_tuner import Objective
61
62 tuner1 = RandomSearch(hypermodel=build_model,
63                       objective="val_mean_squared_error",
64                       max_trials=10,
65                       #seed=123,
66                       project_name="Regression",
67                       overwrite=True
68                       )
69
70 tuner1.search(Xtrain, ytrain, batch_size=32, epochs=150, validation_data=(Xtest,ytest))
71
72 best_params = tuner1.get_best_hyperparameters()
73
74 best_params[0].values
75
76 best_model = tuner1.get_best_models()[0]
77
78 best_model.summary()
79
80 Y_test_reg_preds = best_model.predict(Xtest)
81
82 ytest[:5], Y_test_reg_preds[:5]

```

Listing 3.1: Código para la Determinación de Variables de Importancia

**Estimación de propiedades sensoriales del café utilizando la variedad, tipo y defectos en grano:** El código está enfocado en la generación de un sistema inteligente que sea capaz de realizar estimaciones de las notas sensoriales del aroma, fragancia, sabor y resabio del café verde utilizando como parámetros de entrada su variedad, tipo y defectos en granos de café.

```
1
2 import numpy as np
3 import pandas as pd
4 import matplotlib.pyplot as plt
5 import warnings
6 warnings.filterwarnings("ignore")
7
8 from scipy.stats import uniform, truncnorm, randint
9 import pandas as pd
10 import numpy as np
11 from sklearn.preprocessing import StandardScaler
12 from scipy.stats import uniform, truncnorm, randint
13
14 from scipy.stats import loguniform
15 from sklearn.linear_model import LogisticRegression
16 from sklearn.model_selection import RepeatedStratifiedKFold
17 from sklearn.model_selection import RandomizedSearchCV
18
19 from sklearn.ensemble import RandomForestClassifier
20 from sklearn.svm import SVC
21 from sklearn.tree import DecisionTreeClassifier
22 from sklearn.neighbors import KNeighborsClassifier
23 from sklearn.naive_bayes import GaussianNB
24 from catboost import CatBoostClassifier
25
26 import json
27
28 import warnings
29 warnings.filterwarnings("ignore")
30
31 cv=4
32
33
34 ValIni=1
35 ValEnd=20
36
37 #varouput=['Florales', 'Frutales', 'Herbales', 'Nueces', 'Caramelos', 'Chocolates', '
38 #           Especiados', 'Resinosos',
39 #           'Piroliticos','Fermentados']
```

```

40 varouput=['F_FLORAL','F_FRUITY','F_HERBAL','F_NUTS','F_CARAMEL','F_CHOCOLATES','F_SPICY','
41 F_RESINOUS','F_PYROLYTIC','F_EARTHY','F_FERMENTED']

42

43 #DataF1=pd.read_csv("CAFECOL_FISICO-SENSORIAL_OLFATIVO_4.csv")
44 #DataF1=pd.read_csv("CAFECOL_FISICO-SENSORIAL_CATALACION_3.csv")
45 DataF1=pd.read_csv("DataBase/2021_Nov_DB_Total_en_Fragance.csv")

46

47

48 columns_name=DataF1.columns[ValIni:ValEnd].values

49

50 #Seleccion de las variables a Utilizar
51 X=DataF1.iloc[:,ValIni:ValEnd].values
52 y=DataF1[varouput[0]].values

53

54 # Escalamiento de Datos
55 scaler=StandardScaler()
56 scaler.fit(X)
57 Xs=scaler.transform(X)
58 Xs.shape

59

60 #Logistic Regression
61

62 def LR(Xs,y):
63     # define model
64     model = LogisticRegression()
65     # define search space
66     model_params = {
67         'solver': ['newton-cg', 'lbfgs', 'liblinear'],
68         'penalty': ['l2', 'elasticnet'],
69         'C': loguniform(1e-5, 10)}
70
71
72     search = RandomizedSearchCV(model, model_params, n_iter=10, scoring='accuracy', n_jobs
73     ==-1, random_state=1, cv=cv)
74     # execute search
75     result = search.fit(Xs, y)
76     # summarize result
77
78     return json.dumps(result.best_params_), result.best_score_
79
80     #RandomForestClassifier
81
82 def RFC(Xs,y):
83

```

```

84     model_params = {
85         # randomly sample numbers from 4 to 204 estimators
86         'n_estimators': randint(4,200),
87         # normally distributed max_features, with mean .25 stddev 0.1, bounded between 0 and
88         1
89         'max_features': truncnorm(a=0, b=1, loc=0.25, scale=0.1),
90         # uniform distribution from 0.01 to 0.2 (0.01 + 0.199)
91         'min_samples_split': uniform(0.01, 0.199)}
92
93
94     #Define and Train the Model with Random Search
95
96
97     # create random forest classifier model
98     rf_model = RandomForestClassifier()
99
100
101    # set up random search meta-estimator
102    # this will train 100 models over 5 folds of cross validation (500 models total)
103    clf = RandomizedSearchCV(rf_model, model_params, n_iter=100, cv=cv, n_jobs=-1,
104                             random_state=1)
105
106
107    # train the random search meta-estimator to find the best model out of 100 candidates
108    model = clf.fit(Xs, y)
109
110
111    # print winning set of hyperparameters
112    model_best=model.best_estimator_.get_params()
113
114
115    return json.dumps(model_best), model.best_score_
116
117
118    # train the model on train set
119 def SVCFun(Xs,y):
120
121     model = SVC()
122
123     # defining parameter range
124     param_grid = {'C': [0.1, 1, 10, 100, 1000],
125                  'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
126                  'kernel': ['rbf','linear']}
127
128     # define search
129     search = RandomizedSearchCV(model, param_grid, n_iter=500, scoring='accuracy', n_jobs
130                                 =-1, cv=cv, random_state=1)
131
132
133     # execute search
134     result = search.fit(Xs, y)
135
136     # summarize result
137     return json.dumps(result.best_params_),result.best_score_
138
139
140    #Decision Tree Classifier
141 def DTC(Xs,y):
142
143     model=DecisionTreeClassifier()

```

```

127
128     param_grid=[{"criterion":['gini','entropy'],
129                  "max_depth":range(1,10),
130                  "min_samples_split":range(1,10),
131                  "min_samples_leaf":range(1,5)
132                 }
133
134     # define search
135     search = RandomizedSearchCV(model, param_grid, n_iter=500, scoring='accuracy', n_jobs
136     =-1, cv=cv, random_state=1)
137     # execute search
138     result = search.fit(Xs, y)
139     return json.dumps(result.best_params_), result.best_score_
140
141     #KNEIGHBORS CLASSIFIERS
142
143 def KNC(Xs,y):
144
145     model=KNeighborsClassifier()
146
147     param_grid={"leaf_size":range(1,50),
148                "n_neighbors":range(1,30),
149                "p": [1,2]
150               }
151
152     # define search
153     search = RandomizedSearchCV(model, param_grid, n_iter=500, scoring='accuracy', n_jobs
154     =-1, cv=cv, random_state=1)
155     result=search.fit(Xs,y)
156
157     return json.dumps(result.best_params_),result.best_score_
158
159     #NAIVE BAYES
160
161 def NBF(Xs,y):
162     model=GaussianNB()
163
164     param_grid = {
165         'var_smoothing': np.logspace(0,-9, num=100)}
166
167     # define search
168     search = RandomizedSearchCV(model, param_grid, n_iter=500, scoring='accuracy', n_jobs
169     =-1, cv=cv, random_state=1)
170     result=search.fit(Xs,y)
171
172     return json.dumps(result.best_params_),result.best_score_
173
174     #CATBOOST
175
176 def CBF(Xs,y):

```

```

170
171     model = CatBoostClassifier()
172
173     param_grid = { 'learning_rate':[0.05,0.2] ,
174                   'max_depth': [2,6]}
175
176     # define search
177     search = RandomizedSearchCV(model, param_grid, scoring='accuracy', n_jobs=-1, cv=cv)
178     result=search.fit(Xs,y)
179
180     return json.dumps(result.best_params_),result.best_score_
181
182
183 #name_file_csv_out=['ML_TV_Florales.csv', 'ML_TV_Frutales.csv', 'ML_TV_Herbales.csv', ' '
184 #                      'ML_TV_Nueces.csv', 'ML_TV_Caramelos.csv', 'ML_TV_Chocolates.csv',
185 #                      'ML_TV_Especiados.csv', 'ML_TV_Resinosos.csv', 'ML_TV_Piroliticos.csv',
186 #                      'ML_TV_Fermentados.csv']
187
188
189
190 name_file_csv_out=['FI_DB_FRAGANCIA_1.csv','FI_DB_SABOR_1.csv','FI_DB_RESIDUAL_1.csv',' '
191 #                      'FI_DB_ACIDEZ_1.csv',
192 #                      'FI_DB_CUERPO_1.csv','FI_DB_BALANCE_1.csv','FI_DB_PUNTAJE_CATADOR_1.csv',
193 #                      'FI_DB_TOTAL_1.csv']
194
195
196 name_file_csv_out=['F_FLORAL.csv','F_FRUITY.csv','F_HERBAL.csv','F_NUTS.csv','F_CARAMEL.csv'
197 #                      , 'F_CHOCOLATES.csv',
198 #                      'F_SPICY.csv','F_RESINOUS.csv','F_PYROLYTIC.csv','F_EARTHY.csv', ' '
199 #                      F_FERMENTED.csv']
200
201
202
203 LR_best_score=np.zeros(len(name_file_csv_out))
204 RFC_best_score=np.zeros(len(name_file_csv_out))
205 SVC_best_score=np.zeros(len(name_file_csv_out))
206 DTC_best_score=np.zeros(len(name_file_csv_out))
207 KNC_best_score=np.zeros(len(name_file_csv_out))
208 NBF_best_score=np.zeros(len(name_file_csv_out))
209 CBF_best_score=np.zeros(len(name_file_csv_out))
210
211
212 LR_best_parameters =["" for x in range(len(name_file_csv_out))]
213 RFC_best_parameters=["" for x in range(len(name_file_csv_out))]
214 SVC_best_parameters=["" for x in range(len(name_file_csv_out))]
215 DTC_best_parameters=["" for x in range(len(name_file_csv_out))]
216 KNC_best_parameters=["" for x in range(len(name_file_csv_out))]
217 NBF_best_parameters=["" for x in range(len(name_file_csv_out))]
218 CBF_best_parameters=["" for x in range(len(name_file_csv_out))]
```

```

210
211
212 for x in range(len(name_file_csv_out)):
213
214     y=DataF1[varouput[x]].values
215
216     LR_best_parameters[x],LR_best_score[x]=LR(Xs,y)
217     print("LR_best_score{0}, score:{1}".format(x,LR_best_score[x]))
218
219     RFC_best_parameters[x], RFC_best_score[x]=RFC(Xs,y)
220     print("RFC_best_score{0}, score:{1}".format(x,RFC_best_score[x]))
221
222     SVC_best_parameters[x],SVC_best_score[x]=SVCFun(Xs,y)
223     print("SVC_best_score{0}, score:{1}".format(x,SVC_best_score[x]))
224
225     DTC_best_parameters[x],DTC_best_score[x]=DTC(Xs,y)
226     print("DTC_best_score{0}, score:{1}".format(x,DTC_best_score[x]))
227
228     KNC_best_parameters[x],KNC_best_score[x]=KNC(Xs,y)
229     print("KNC_best_score{0}, score:{1}".format(x,KNC_best_score[x]))
230
231     NBF_best_parameters[x],NBF_best_score[x]=NBF(Xs,y)
232     print("NBF_best_score{0}, score:{1}".format(x,NBF_best_score[x]))
233
234     CBF_best_parameters[x],CBF_best_score[x]=CBF(Xs,y)
235     print("NBF_best_score{}:{}".format(x,CBF_best_score[x]))
236
237     #Se crea una matriz que contiene la suma de todos los elementos importantes encontrados
238     #por todoslos métodos
239
240 d1 = {'LogR_parameters': LR_best_parameters, 'RFC_parameters': RFC_best_parameters, 'SVC_parameters':SVC_best_parameters, 'DTC_parameters':DTC_best_parameters, 'KNC_parameters': KNC_best_parameters, 'NBF_parameters':NBF_best_parameters, 'CBF_parameters':CBF_best_parameters}
241
242 d = {'LogR_score': LR_best_score, 'RFC_score': RFC_best_score, 'SVC_score':SVC_best_score, 'DTC_score':DTC_best_score, 'KNC_score': KNC_best_score, 'NBF_score':NBF_best_score, 'CBF_score':CBF_best_score}
243
244 print(varouput)
245 df = pd.DataFrame(data=d, index=varouput)
246 df1= pd.DataFrame(data=d1, index=varouput)
247
248 df.to_csv("ML_all_Best_Score.csv",index = True, header=True)
249 df1.to_csv("ML_all_Hiperp.csv",index = True, header=True)
250

```

```

251 import matplotlib.pyplot as plt
252 df.plot.bar(rot=0,figsize=(20,12),title='Results', ylabel='Accuracy',ylim=(0.4,1))
253 plt.savefig("ML_all_TotalScore.jpg",dpi=300)

```

Listing 3.2: Código para la estimación de notas sensoriales del café

**Estimación de las notas sensoriales utilizando espectroscopía NIR y Redes Neuronales Artificiales** El código está enfocado a la determinación de las notas sensoriales del café en el aroma, fragancia, sabor y resabio, utilizando los resultados obtenidos en los análisis de espectroscopía NIR como variables de entrada utilizadas para la estimación de dichas notas a través del entrenamiento de Redes Neuronales Artificiales.

```

1
2 import pandas as pd
3 from tensorflow.keras.models import Sequential
4 from tensorflow.keras.layers import Dropout
5 from tensorflow.keras.layers import Dense
6 from tensorflow.keras.optimizers import SGD
7 from scikeras.wrappers import KerasClassifier #pip install scikeras==0.1.8 --user
8 from sklearn.model_selection import cross_val_score
9 from sklearn.preprocessing import LabelEncoder
10 from sklearn.model_selection import StratifiedKFold
11 from sklearn.preprocessing import StandardScaler
12 from sklearn.pipeline import Pipeline
13
14
15 #####
16 #CONFIGURACIÓN DE PARÁMETROS
17 #####
18
19 #'R_FLORAL','R_FRUITY','R_HERBAL','R_NUTS','R_Caramel','R_CHOCOLATES','R_SPICY','R_RESINOUS'
20     , 'R_PYROLYTIC','R_EARTHY','R_FERMENTED'
21 VAR_OUTPUT="F_HERBAL";
22
23 #Selección de Muestras de Café
24 ValMIni=0
25 ValMEnd=869
26
27 #Selección de los parametros fisicos
28 ValIni=1
29 ValEnd=72
30
31 DB_NIR=pd.read_csv("2021_Nov_DB_Total_en.csv")
32 DB_NIR.head()
33 DB_NIR.info

```

```

34
35 #Seleccion de las variables a Utilizar
36
37 from sklearn import preprocessing
38
39 DB_NIR=DB_NIR.fillna(1)
40 X=DB_NIR.iloc[ValMIni:ValMEnd,ValINI:ValEnd].values
41 yi=DB_NIR[VAR_OUTPUT].values
42 y=yi[ValMIni:ValMEnd]
43
44 min_max_scaler = preprocessing.MinMaxScaler()
45 Xs=min_max_scaler.fit_transform(X)
46
47 # Use train_test_split() Method.
48 from sklearn.model_selection import train_test_split
49 Xtrain, Xtest, ytrain,ytest= train_test_split(Xs,y, test_size=0.4)
50
51 from tensorflow.keras.models import Sequential
52 from tensorflow.keras import layers
53 from tensorflow.keras.optimizers import SGD
54
55 def build_model(hyperparams):
56     model = Sequential()
57     model.add(layers.Input(shape=(Xs.shape[1],)))
58     model.add(layers.Dense(units=hyperparams.Int("units_l1", 10, 2000, step=16),
59                           use_bias=hyperparams.Boolean("bias_l1"),
60                           activation=hyperparams.Choice("act_l1", ["relu", "tanh"]))
61                           ))
62     #model.add(Dropout(0.05))
63     model.add(layers.Dense(units=hyperparams.Int("units_l2", 10, 2000, step=16),
64                           use_bias=hyperparams.Boolean("bias_l2"),
65                           activation=hyperparams.Choice("act_l2", ["relu", "tanh"]))
66                           ))
67     #model.add(Dropout(0.05))
68     model.add(layers.Dense(1))
69
70     optim=hyperparams.Choice("optimizer",["sgd","rmsprop","adam"])
71     model.compile(optim, loss="mean_squared_error", metrics=["mean_squared_error"])
72
73     return model
74 print(Xs)
75
76 from keras_tuner import RandomSearch
77 from keras_tuner import Objective
78
79 tuner1 = RandomSearch(hypermodel=build_model,

```

```

80         objective="val_mean_squared_error",
81         max_trials=10,
82         #seed=123,
83         project_name="Regression",
84         overwrite=True
85     )
86
87 tuner1.search(Xtrain, ytrain, batch_size=32, epochs=150, validation_data=(Xtest,ytest))
88
89 best_params = tuner1.get_best_hyperparameters()
90
91 best_params[0].values
92
93 best_model = tuner1.get_best_models()[0]
94
95 best_model.summary()
96
97 Y_test_reg_preds = best_model.predict(Xtest)
98
99 ytest[:5], Y_test_reg_preds[:5]
100
101 plt.plot(ytest)
102 plt.plot(Y_test_reg_preds)
103
104 import matplotlib.pyplot as plt
105
106 plt.scatter(ytest, Y_test_reg_preds)
107 plt.savefig(VAR_OUTPUT+'.png')
108
109 _,train_acc=best_model.evaluate(Xtrain,ytrain,verbose=1)
110 _,test_acc=best_model.evaluate(Xtest,ytest,verbose=1)
111
112
113 print('Train: %.3f, Test: %.3f' % (train_acc, test_acc))
114
115 import tensorflow as tf
116 # Guardar el Modelo
117 best_model.save(VAR_OUTPUT+'.h5')
118
119 # Recrea exactamente el mismo modelo solo desde el archivo
120 new_model1 = tf.keras.models.load_model(VAR_OUTPUT+'.h5')
121
122 _,train_acc=best_model.evaluate(Xtrain,ytrain,verbose=1)
123 _,test_acc=best_model.evaluate(Xtest,ytest,verbose=1)
124
125

```

```
126 print('Train: %.3f, Test: %.3f' % (train_acc, test_acc))
```

Listing 3.3: Estimación de las notas sensoriales utilizando espectroscopía NIR y Redes Neuronales Artificiales