



Universidade de A Coruña  
Facultad de Informática y Computación

# **Detección de alcohol en cerveza mediante espectrofotometría**

## **Memoria**

### **Autores:**

Eduardo Juaristi Pérez  
Gabriel Fernández Fernández  
Sergio Marcos Vázquez  
Pedro Pazos Curra

Curso 2021-2022



# Índice

1. Introducción	6
2. Descripción del problema	8
3. Análisis bibliográfico / Estado del arte	12
4. Desarrollo	14
5. Conclusiones finales	56
6. Trabajo futuro	57
7. Bibliografía	58

# Contenidos del documento

A continuación se detallan y explican los temas que van a ser abordados a lo largo del documento.

1. **Introducción:** Descripción del dominio y las técnicas a utilizar.
2. **Descripción del problema:** detalle del problema a resolver, la base de datos utilizada y las restricciones del problema.
3. **Análisis bibliográfico:** resumen de los trabajos recientes que abordan un problema similar, así como de artículos útiles para entender el contexto de técnicas o métodos utilizados.
4. **Desarrollo:** exposición de resultados de nuestra solución en 6 aproximaciones diferentes en las cuales se abordará de una forma sensiblemente distinta la misma tarea y se discutirán los datos obtenidos.
5. **Conclusiones finales:** discusión detallada de todos los resultados obtenidos de forma global y análisis de las diferentes técnicas empleadas, mejores y peores resultados.
6. **Trabajo futuro:** Exploración de posibles ampliaciones o cambios de perspectiva que pudiesen ser ejecutados con nuestro problema como base fundamental.
7. **Bibliografía:** Selección de documentos bibliográficos.

# Glosario de términos

1. **Redes neuronales (RRNNAA):** Esta conocida técnica funciona en base a un conjunto de nodos interconectados cuyos pesos se modifican dinámicamente para resolver un amplio conjunto de problemas de una forma similar a la que lo haría un cerebro humano con una distribución de capas de neuronas del estilo [Capas de entrada, Capas ocultas, Capas de salida].
2. **Máquinas de soporte vectorial (SVM) :** El funcionamiento de una SVM se resume en la búsqueda de hiperplanos que separan linealmente y de forma óptima puntos de dos clases diferentes.
3. **Árboles de decisión:** Un árbol de decisión es un modelo en el que para tomar decisiones se comprueba un conjunto de condiciones (preguntas) de forma incremental para llegar a conclusiones que ayudan a solucionar problemas de diversos campos.
4. **kNN (k Nearest Neighbour):** kNN es un algoritmo de clasificación supervisada en el cual se procede calculando la distancia de un punto a sus vecinos y teniendo en cuenta la clase que es asignada a estos vecinos. Dependiendo del valor de k que se le aporte, este punto será clasificado en una clase u otra, teniendo en cuenta sus vecinos cercanos.
5. **DL (Deep Learning):** DL es un modelo basado en RNA que consiste en la construcción de una red con un gran conjunto de capas ocultas (esencial para que sea “profundo”), lo cual permite una gran capacidad de reconocimiento de patrones en, especialmente, imágenes, pero también en campos como el reconocimiento automático de lenguaje natural o sistemas de recomendación.
6. **Espectroscopia:** Estudio de la absorción y emisión de luz y otras radiaciones en la materia. Implica la separación de la luz en sus constituyentes según la longitud de onda. Se realiza un proceso similar al de la separación de la luz a través de un prisma en los colores del arcoiris.
7. **Espectrometría:** Técnica de medición de la distribución de la luz a lo largo del espectro óptico.
8. **Espectrofotometría:** Rama de la espectroscopia electromagnética dedicada a la medición cuantitativa de la reflexión o propiedades de transmisión de un material como una función de longitud de onda.
9. **Reflectancia:** Eficacia con la que un material refleja la energía radiante. Fracción de potencia electromagnética reflejada al incidir.

10. **Absorbancia:** Medida proporcional al camino óptico de la muestra y a la concentración de la sustancia en ésta. El origen del término proviene de la cuantificación de la luz absorbida por un material, pero su acepción hoy en día es la cuantificación de la luz “perdida” por un sistema detector a través de otros materiales o mecanismos.
11. **Intensidad:** Medida que, en relación con la longitud de onda de la luz, forma el espectro infrarrojo.

# 1. Introducción

La cerveza es una bebida alcohólica creada por la fermentación del almidón presente en ciertos cereales, comúnmente los granos germinados de cebada y malta. Posteriormente puede ser aromatizada por todo tipo de ingredientes, siendo el más conocido el lúpulo, dando lugar al amargor y aroma que es hoy en día reconocido en todo el mundo.

El objetivo de este proyecto es emplear características de la cerveza obtenidas mediante la técnica de espectrofotometría, para discernir entre aquellas que contengan alcohol y las que no, así como estimar la cantidad. Estas propiedades serían extraídas haciendo uso de un espectrofotómetro.

En primera instancia, unos resultados satisfactorios proporcionarán una forma fácil de clasificar cerveza y analizar con cierta agilidad su porcentaje de alcohol para, por ejemplo, poder detectar lotes defectuosos en una fábrica antes de distribuir la mercancía. A mayores, podrían derivar en trabajos similares del mismo ámbito. Como ejemplos prácticos, tendríamos la detección de gluten, estupefacientes externos introducidos en la bebida, etcétera. Por lo tanto, una supuesta solución mediante un sistema de aprendizaje automático permitiría un alto grado de escalabilidad y una base en futuras ampliaciones con un variado abanico de aplicaciones prácticas.

Para el desarrollo de nuestro sistema, utilizaremos teoría específica del dominio del aprendizaje automático, en concreto:

1. Redes neuronales (RRNNAA) : Esta conocida técnica funciona en base a un conjunto de nodos interconectados cuyos pesos se modifican dinámicamente para resolver un amplio conjunto de problemas de una forma similar a la que lo haría un cerebro humano con una distribución de capas de neuronas del estilo [Capas de entrada, Capas ocultas, Capas de salida].
2. Máquinas de soporte vectorial (SVM) : El funcionamiento de una SVM se resume en la búsqueda de hiperplanos que separan linealmente y de forma óptima puntos de dos clases diferentes.
3. Árboles de decisión : Un árbol de decisión es un modelo en el que para tomar decisiones se comprueba un conjunto de condiciones (preguntas) de forma incremental para llegar a conclusiones que ayudan a solucionar problemas de diversos campos.
4. kNN (k Nearest Neighbour) : kNN es un algoritmo de clasificación supervisada en el cual se procede calculando la distancia de un punto a sus vecinos y teniendo en cuenta la clase que es asignada a estos vecinos. Dependiendo del valor de k que se le aporte, este punto será clasificado en una clase u otra, teniendo en cuenta sus vecinos cercanos.

5. DL (Deep Learning): DL es un modelo basado en RNA que consiste en la construcción de una red con un gran conjunto de capas ocultas (esencial para que sea “profundo”), lo cual permite una gran capacidad de reconocimiento de patrones en, especialmente, imágenes, pero también en campos como el reconocimiento automático de lenguaje natural o sistemas de recomendación.

El presente documento está dividido en 6 aproximaciones diferentes al mismo problema, de las cuales hay para cada una dos apartados pertenecientes al desarrollo y a las conclusiones, con además cuatro apartados únicos para la introducción, descripción del problema, estado del arte y bibliografía. Los pormenores de cada una de las aproximaciones serán detallados en sus respectivas secciones.



## 2. Descripción del problema

### 2.1. Introducción

Como se ha indicado anteriormente, nuestro objetivo es hallar un patrón que determine si la cerveza lleva alcohol en base a una serie de características obtenidas mediante un espectrofotómetro “DLP Nirscan Nano” de “Texas Instruments” [1] al estudiar los enlaces de carbono de la misma. Esta técnica, conocida como espectrofotometría, es una técnica no destructiva centrada en la medición de la cantidad de energía lumínica absorbida y refractada por un sistema químico en función de la longitud de onda de la radiación, de esta forma se puede obtener la cantidad de enlaces de carbono de la muestra y, en base a ella, derivar una serie de propiedades [2]. Una de las ventajas de la espectrofotometría (la técnica que hemos decidido usar), es que es una técnica no destructiva, es decir, para el análisis de una muestra (enlaces de carbono o proteínas como tal) no se destruye la muestra, lo cual, favorece su reutilización. Por otro lado, existen otro tipo de técnicas, como es el caso de la espectrometría de masas, que son destructivas, por lo que la muestra sólo puede ser utilizada una vez. Nuestro enfoque, tiene como ventaja la posibilidad de realizar varias mediciones sobre el mismo ejemplar, permitiendo un resultado más preciso.

### 2.2. Base de datos

Se analizaron un total de 22 marcas distintas de cerveza con algunas de sus distintas variantes, cada muestra fue analizada 6 veces y disponemos de 82 tipos distintos de cerveza. Las muestras fueron analizadas obteniendo distintos valores de absorbancia, intensidad, reflectancia, queda registrada también la longitud de onda, que aumenta de forma constante para todas las cervezas, inducida por la configuración del espectrofotómetro (más adelante desglosamos estos datos). La base de datos fue proporcionada por un miembro del equipo docente y fue creada en los laboratorios de la universidad, disponemos de seis archivos con información dispar:

En una de ellas, *BD.x/sx*, disponemos de la marca, nombre, graduación, una foto de la etiqueta en el reverso de la botella (a color y legible) .

En *Cervezas.x/sx*, tenemos marca, nombre, graduación, tipo (botella o lata) y origen geográfico.

Por la parte de los archivos JSON, en *HadaBeer.json* y *HadaBeer2.json* disponemos de graduación, nombre, intensidad, absorbancia, reflectancia, longitud de onda y hora de medición. Respecto a la longitud de onda tenemos 228 valores comunes para todas las cervezas, obtenidos de dividir la onda devuelta por el espectrofotómetro en 228 puntos. Los valores de intensidad, reflectancia y

absorbancia se midieron para cada cerveza en cada uno de dichos puntos, teniendo así 228 valores tanto para intensidad como para reflectancia y absorbancia. En *Cerveza.json* y *Cerveza2.json* tenemos los mismos atributos y además un ID y el tipo de escáner (Column o Hadamard).

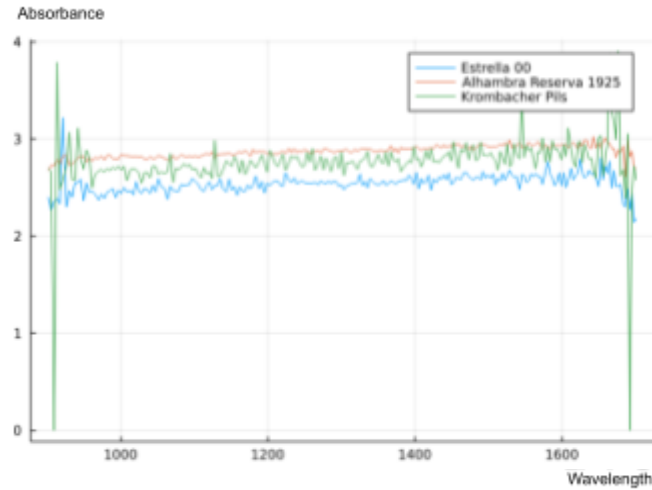


Fig. 1: Función de relación entre Longitud de onda y Absorbancia

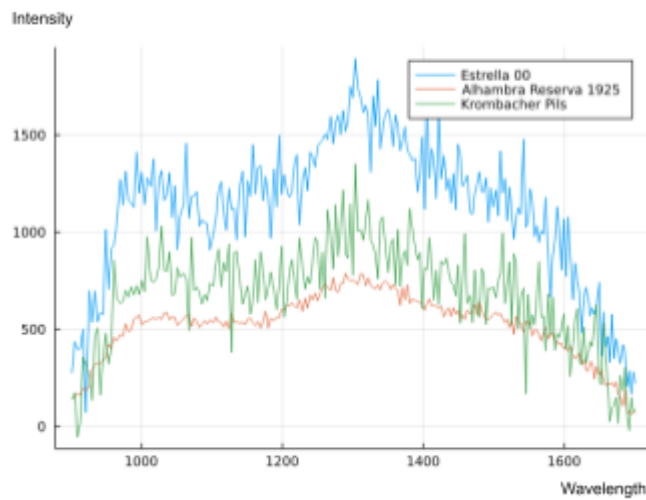


Fig. 2: Función de relación entre Longitud de onda y Intensidad

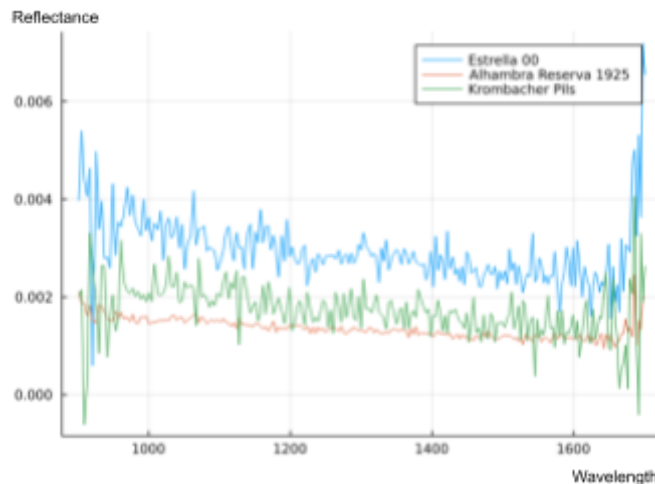


Fig. 3: Función de relación entre Longitud de onda y Reflectancia

Haciendo un análisis de los datos proporcionados podemos observar la relación entre la longitud de onda y los valores de absorbancia (Fig.1), intensidad (Fig.2) y reflectancia (Fig.3) para tres cervezas con distinta graduación de alcohol. Las cervezas escogidas fueron la Estrella Galicia 00 (con 0° de alcohol), la Alhambra Reserva 1925 (con 6.5° de alcohol) y la Krombacher Pils (con 4.8° de alcohol). En las gráficas se puede observar como las cervezas están ordenadas según su graduación, pudiendo relacionar aquellas con más graduación con un menor valor de intensidad y reflectancia, así como con un mayor valor de absorbancia.

### 2.3. Restricciones

Las muestras fueron empleadas recién salidas de la nevera (entre 4 y 5 °C), vertidas en una celda de cuarzo y analizadas a dos milímetros.

Estas condiciones se deben a que a distintas temperaturas se obtienen mediciones distintas, por lo que se debe realizar el análisis en un rango común de temperaturas, y la celda de cuarzo ofrece una contaminación menor de los resultados debido a su estructura cristalina.

Las restricciones para este problema serían varias, como el hecho de disponer de un espectrofotómetro, un aparato de investigación de coste elevado y que requiere de unas buenas condiciones para el análisis de las características ya que, tanto las variaciones en temperatura o la intensidad de la luz en el ambiente en que se toman las muestras, pueden variar en gran medida los resultados. A mayores, se debe de tener en cuenta que el espectrofotómetro usado es una versión de pequeñas dimensiones, lo cual implica una mayor portabilidad, pero también mediciones menos exactas.

## 2.4. Métrica objetiva

Como hemos comentado anteriormente, la determinación de alcohol en una bebida alcohólica es un problema cuya solución requiere un alto grado de sensibilidad y buena precisión, debido a las implicaciones legales a las que está sometida cualquier bebida de cierta graduación; por ejemplo, en una fábrica de cerveza antes de sacar un lote al mercado. Debido a esto, la métrica que buscaremos en nuestra solución será que determine correctamente si una bebida es o no realmente alcohólica, y nos interesa que nuestro sistema final sea realmente sensible y preciso.

### 3. Análisis bibliográfico / Estado del arte

El siguiente punto a mencionar es la existencia de estudios realizados con anterioridad de similar naturaleza, ya sean acerca del análisis del alcohol mediante espectrometría o resolución de problemas de clasificación en base a datos obtenidos también por dicha técnica [8].

Para conocer el estado actual del problema, hemos explorado aproximaciones similares hechas sobre el mismo tema o uno análogo. Mencionaremos en este punto artículos, investigaciones y todo tipo de trabajos sobre la materia para guiar al lector antes de empezar a desglosar el desarrollo de nuestra propia aproximación.

Antes de adentrarnos en los problemas concretos, debemos informarnos de las bases, técnicas y métodos que vamos a utilizar para elaborar nuestro sistema. Es por eso que mencionaremos [10] como una base teórica en la que apoyarnos para informarnos sobre el machine learning. En esta obra, se da una extensa introducción a la materia; con modelos, métodos y estadística aplicada a ellos.

Además, para la realización de este problema, debemos informarnos acerca de la espectrometría, y para ello haremos uso de [2], para obtener información acerca de qué es y los valores que se extraen haciendo uso de la misma.

En primer lugar, mencionaremos el TFM de María Suárez Díaz en el Máster en Biotecnología Alimentaria para la Universidad de Oviedo, [9], que estudia distintos atributos de la cerveza como el amargor y el envejecimiento mediante espectrofotometría y fluorescencia molecular. A pesar de usar la espectrometría para muestras de cerveza, sus conclusiones son mucho más enfocadas a las relaciones entre la fluorescencia y el amargor junto con otros factores como el color o la transparencia. De todas formas, en su desarrollo utiliza los valores de absorbancia de los diferentes tipos de cervezas (Ale, Lager, trigo), arrojando curvas bastante diferenciadas en su gráfica de absorción (con los máximos de absorción respectivamente en ~275 nm, ~275 nm y 266 nm). Esta información nos es de interés, aunque la autora no detalla en la conclusión ninguna inducción que nos sea de utilidad en nuestro problema.

Un interesante estudio de un muy similar problema fue realizado por la Universidad de Tianjin junto con la Universidad Tecnológica de Pekín [3], donde se estudiaron diversas muestras con diferentes concentraciones de alcohol mediante el uso de espectrometría, para así determinar el mejor rango de espectros en los que detectar la concentración de alcohol. El artículo concluye que los rangos de longitud de onda [1100 - 1330 nm] y [2110 - 2400 nm] tienen una buena estabilidad y capacidad predictiva.

Semejante también al artículo [4], que usa la técnica de Laser Raman Spectroscopy, una técnica de espectroscopia basada en la utilización del *scatter*, pero concluye que debe desarrollarse un método más riguroso para eliminar los compuestos fluorescentes de las

muestras de bebidas alcohólicas. De no ser así, el método no arroja datos uniformemente válidos sobre el conjunto de bebidas observadas.

A continuación, mencionaremos el artículo [5], en el cual se estudia mediante la espectroscopia FTIR, espectroscopia FTIR *vapour-phase* y espectroscopia NIR. Se ha seleccionado el rango de onda entre 900 y 1750  $\text{cm}^{-1}$  el metanol y etanol presentes en una muestra para determinar si una bebida alcohólica está adulterada. Concluye en que los datos muestran unos buenos resultados y una predicción fiable del metanol en las muestras utilizadas.

En el estudio para la Universidad de La Rioja [8], se estudia mediante espectrofotometría NIR un problema de clasificación con redes neuronales y máquinas de soporte vectorial de las variedades de uvas con mediciones entre 1600 y 2400 nm, entrenando con 20 variedades de hojas así como muestras específicas de diferentes viñedos. Al igual que algunos de los anteriores, concluye en que es una posibilidad prometedora con unos buenos resultados tanto en el ámbito local como global, permitiendo la discriminación *in situ* de un alto número de variedades de forma portátil y no destructiva.

Por último, hemos escogido [7], donde se lleva a cabo un estudio más genérico que se aplica a todo tipo de bebidas alcohólicas mediante espectrometría NIR derivativa. Los resultados son bastante buenos, siendo capaces de determinar con un error bajo la graduación de un amplio abanico de muestras de 11 bebidas alcohólicas diferentes. De todas formas, es comentado el hecho de que el porcentaje de azúcar (fructosa) influye negativamente en las mediciones, añadiendo errores a los valores.

En conclusión, podemos observar que en los experimentos de espectrofotometría suele haber un rango donde se puede obtener una mayor información y la posibilidad de que ciertos atributos tengan mayor relevancia, permitiéndonos dejar de lado rangos que nos son superfluos. Además, queda en evidencia que el área tiene potencial de experimentación con técnicas de aprendizaje automático que mejoren la capacidad de clasificación que se obtienen.

## 4. Desarrollo

La validación cruzada (Cross Validation) es una técnica estadística para el análisis de conjuntos de datos en la cual se parte el conjunto de datos y se repiten las mediciones para después calcular la media aritmética de las mismas.

Los atributos que, a priori, nos van a interesar para desarrollar nuestro sistema serán intensidad, absorbancia y reflectancia. Determinaremos que la graduación umbral que usaremos para dividir en la primera aproximación serán los 5.5° (5,5% de alcohol). En el futuro, ampliaremos el problema.

En cuanto a la longitud de onda, el conjunto de valores usados son 228 pasos entre [901.2617,1700.8995], conjunto que se mantiene constante para todas las entradas, por lo tanto es probable que no nos interese utilizar este atributo o filtrar antes, ya que sabemos que el rango en el que más alcohol podremos detectar es de 1100 a 1330 nm y de 2110 a 2400 nm [3] (este último rango no nos serviría ya que en nuestro caso las mediciones del espectrofotómetro se realizan entre 900 y 1700 nm), de modo que los datos serán acotados para el rango de longitud de entre 1100 y 1330 nm, lo cual reduce el número de mediciones de cada muestra de 228 a 62. Posteriormente, elaboramos normalización de tipo min/max para situar los valores entre 0 y 1 debido a que no hay en nuestra base de datos valores extremadamente anómalos que redujesen de forma sensible la efectividad de esta técnica.

	media(Intensidad)	media(Reflectancia)	media(Absorbancia)
<u>cerveza1</u>	1252.38	0.00272879	2.56454
<u>cerveza2</u>	2181.78	0.00474995	2.32448
<u>cerveza3</u>	1009.73	0.00219706	2.65856
<u>cerveza4</u>	1551.0	0.00337868	2.47326
<u>cerveza5</u>	1071.0	0.00233491	2.63237
<u>cerveza6</u>	1345.13	0.0029211	2.53706
<u>cerveza7</u>	1259.3	0.00273126	2.56378
<u>cerveza8</u>	1585.13	0.00345567	2.46205
<u>cerveza9</u>	1192.9	0.00259561	2.58615
<u>cerveza10</u>	1066.49	0.0023189	2.63509

Fig.4: Matriz de entrada

	media(Intensidad)	media(Reflectancia)	media(Absorbancia)
<u>cerveza1</u>	0.319237	0.318391	0.55525
<u>cerveza2</u>	0.98043	0.979829	0.0132787
<u>cerveza3</u>	0.146609	0.144379	0.767508
<u>cerveza4</u>	0.531681	0.531069	0.349174
<u>cerveza5</u>	0.190198	0.189488	0.708376
<u>cerveza6</u>	0.385218	0.381323	0.493204
<u>cerveza7</u>	0.32416	0.319199	0.553525
<u>cerveza8</u>	0.55596	0.556266	0.323852
<u>cerveza9</u>	0.276924	0.274804	0.604022
<u>cerveza10</u>	0.186991	0.184249	0.714527

Fig.5: Matriz de entrada normalizada (min/max)

Para las técnicas que vamos a utilizar, los datos de entrada serán diferentes combinaciones entre los valores de absorbancia, reflectancia e intensidad correspondientes al rango anteriormente especificado. Dichas combinaciones son explicadas en la introducción de cada una de las respectivas aproximaciones.

## 4.1. Primera Aproximación

### 4.1.1. Introducción

Como fue comentado en la introducción, elaboramos 4 técnicas diferentes para el desarrollo del sistema. Por lo tanto, dividiremos este apartado por cada uno de los métodos.

Los datos de entrada en esta primera aproximación serán la media de la absorbancia en el rango de longitud de onda 1100-1330, así como las medias en dicho rango de intensidad y de reflectancia.

### 4.1.2. Redes de Neuronas Artificiales

Las 8 arquitecturas diferentes que han sido probadas (todas con 1 capa de entrada y 1 de salida) y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Topología	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
[ 16 ]	0.34	1.0	0.0	0.79	0	0.65	0.65
[ 32 ]	0.29	0.92	0.29	0.8	0.66	0.71	0.70
[ 64 ]	0.26	0.96	0.29	0.82	0.8	0.72	0.73
[ 128 ]	0.26	0.96	0.29	0.82	0.8	0.72	0.73
[ 16 32 ]	0.39	1.0	0.0	0.79	0	0.65	0.65
[ 32 64 ]	0.27	0.92	0.3	0.81	0.6	0.73	0.72
[ 64 64 ]	0.26	0.96	0.29	0.82	0.8	0.72	0.73
[ 64 128 ]	0.29	0.98	0.18	0.81	0.83	0.69	0.70

Tabla 1: Métricas Primera Aprox. RNA

Topología	Matriz de confusión			Topología	Matriz de confusión		
[ 16 ]				[ 16 32 ]			
			Predicción				Predicción
			Neg Pos				Neg Pos



	<table> <tr> <td rowspan="2">Real</td><td>Neg</td><td>0</td><td>27</td></tr> <tr> <td>Pos</td><td>0</td><td>52</td></tr> </table>	Real	Neg	0	27	Pos	0	52		<table> <tr> <td rowspan="2">Real</td><td>Neg</td><td>0</td><td>27</td></tr> <tr> <td>Pos</td><td>0</td><td>52</td></tr> </table>	Real	Neg	0	27	Pos	0	52												
Real	Neg		0	27																									
	Pos	0	52																										
Real	Neg	0	27																										
	Pos	0	52																										
[ 32 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>4</td><td>48</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	4	48	[ 32 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>18</td></tr> <tr> <td>Pos</td><td>4</td><td>49</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	18	Pos	4	49
				Predicción																									
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	4	48																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	18																										
	Pos	4	49																										
[ 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>2</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	2	50	[ 64 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>2</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	2	50
				Predicción																									
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	2	50																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	2	50																										
[ 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>2</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	2	50	[ 64 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>0</td><td>20</td></tr> <tr> <td>Pos</td><td>0</td><td>10</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	0	20	Pos	0	10
				Predicción																									
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	2	50																										
		Predicción																											
		Neg	Pos																										
Real	Neg	0	20																										
	Pos	0	10																										

Tabla 2: Matriz de Confusión Primera Aprox. RNA

### 4.1.3. Máquinas de Soporte Vectorial

Las 8 configuraciones diferentes de hiperparámetros de la SVM que han sido probadas y los resultados obtenidos en las métricas que derivan de la matriz de confusión y sus respectivos valores son los siguientes:

kernel	k. degree	gamma	c	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
"rbf"	3	2	1	0.3	0.1	1.0	0.18	0.68	1.0	0.7

"rbf"	3	2	2	0.3	0.1	1.0	0.18	0.68	1.0	0.7
"rbf"	3	2	3	0.33	0.1	0.95	0.16	0.67	0.5	0.66
"poly"	3	2	1	0.33	0.0	1.0	0.0	0.66	0.0	0.66
"linear"	3	2	1	0.33	0.0	1.0	0.0	0.66	0.0	0.66
"sigmoid"	3	2	0.9	0.34	0.11	0.9	0.16	0.69	0.33	0.65
"sigmoid"	3	2	1	0.36	0.2	0.85	0.26	0.68	0.4	0.63
"sigmoid"	3	2	3	0.43	0.5	0.6	0.43	0.70	0.38	0.56

Tabla 3: Métricas Primera Aprox. SVM

(Kernel; k.degree; gamma; c)	Matriz de confusión				(Kernel; k.degree; gamma; c)	Matriz de confusión			
("rbf"; 3; 2; 1)			Predicción		("linear"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	20	0		Real	Neg	20	0
		Pos	9	1			Pos	10	0
("rbf"; 3; 2; 2)			Predicción		("sigmoid"; 3; 2; 0.9)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	20	0		Real	Neg	18	2
		Pos	9	1			Pos	8	1
("rbf"; 3; 2; 3)			Predicción		("sigmoid"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	19	1		Real	Neg	17	3
		Pos	9	1			Pos	8	2
("poly"; 3; 2; 1)			Predicción		("sigmoid"; 3; 2; 3)			Predicción	
			Neg	Pos				Neg	Pos
		Neg	20	0			Neg	12	8

	Real	Pos	10	0		Real	Pos	5	4
--	------	-----	----	---	--	------	-----	---	---

Tabla 4: Matriz de Confusión Primera Aprox. SVM

#### 4.1.4. Árboles de decisión

Los 6 valores diferentes de profundidad que han sido probados y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Maxdepth	Tasa fallo	Sensibilidad	Especificidad	F1 score	Valor pred neg	Valor pred pos	Valor precisión
4	0.3	0.5	0.8	0.526316	0.761905	0.55556	0.7
2	0.3	0.2	0.95	0.307692	0.703704	0.6667	0.7
6	0.266667	0.5	0.85	0.555556	0.772727	0.625	0.733333
15	0.233333	0.5	0.9	0.588235	0.782609	0.714286	0.766667
1	0.310345	0	1	0	0.666667	0	0.666667
7	0.2	0.5	0.95	0.625	0.791667	0.8333	0.8

Tabla 5: Métricas Primera Aprox. Árboles de Decisión

Maxdepth	Matriz de confusión			Maxdepth	Matriz de confusión		
4				15			
	Real	Neg	16		Real	Neg	18
		Pos	5			Pos	5
2				1			
	Real	Neg	19		Real	Neg	20
		Pos	8			Pos	10
6				7			
	Real	Neg	17		Real	Neg	19
		Pos	5			Pos	5

Tabla 6: Matriz de Confusión Primera Aprox. Árboles de Decisión

### 4.1.5. k Nearest Neighbors

Los 6 valores diferentes de k que han sido probados son los siguientes:

k	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec
3	0.241379	0.444444	0.9	0.533333	0.782609	0.666667	0.758621
4	0.233333	0.4	0.95	0.533333	0.76	0.8	0.766667
5	0.233333	0.5	0.9	0.588235	0.782609	0.714286	0.766667
6	0.233333	0.5	0.9	0.588235	0.782609	0.714286	0.766667
7	0.275862	0.555556	0.8	0.555556	0.8	0.555556	0.724138
10	0.206897	0.33	0.1	0.5	0.769231	1.0	0.793103

Tabla 7: Métricas Primera Aprox. KNN

k	Matriz de confusión			k	Matriz de confusión		
3				6			
	Real		Predicción		Real		Predicción
			Neg Pos				Neg Pos
		Neg	18 2			Neg	18 2
		Pos	5 4			Pos	5 5
4				7			
	Real		Predicción		Real		Predicción
			Neg Pos				Neg Pos
		Neg	19 1			Neg	16 4
		Pos	6 4			Pos	4 5
5				10			
	Real		Predicción		Real		Predicción
			Neg Pos				Neg Pos
		Neg	18 2			Neg	20 0
		Pos	5 5			Pos	6 3

Tabla 8: Métricas Primera Aprox. KNN

#### 4.1.6. Discusión

En RRNN, los resultados son en general buenos. F1 Score ronda los 0.8 (un valor bastante bueno). Las topologías que mejor funcionan han sido [64], [64,64] y [64,128].

Los mejores resultados del SVM han sido obtenidos en kernel = "rbf" y  $c = 1$  y 2, con precisiones del 70%. De todas formas, los resultados en general son muy pobres, pese a haberse probado un conjunto muy variado de configuraciones, puede verse claramente en los bajísimos resultados de F1 Score.

Además, puede observarse que en la técnica de árboles de decisión la profundidad máxima 7 es la que da mejores resultados. Hemos escogido 1 para el valor de una de las iteraciones para que quede constancia de que es un valor de profundidad demasiado pequeño, y efectivamente da unos malos resultados.

En cuanto a kNN, el valor que mejor funciona en nuestro conjunto de datos es  $k = 10$ , con una precisión (accuracy) del 79%, pero F1 no alcanza los 0.60. Cabe destacar, además, que los datos que arrojan  $k = 3, 5$  y  $6$  son idénticos y muy similares a  $k = 2$ .

## 4.2. Segunda Aproximación

### 4.2.1. Introducción

Tras el desarrollo de las 4 técnicas para el problema escogido, ampliaremos el problema en una nueva aproximación: en lugar de clasificar las entradas en mayores y menores que 5,5°, vamos a acercar este umbral a lo que sería denominado una cerveza sin alcohol: 3,3°.

### 4.2.2. Redes de Neuronas Artificiales

Las 8 arquitecturas diferentes que han sido probadas (todas con 1 capa de entrada y 1 de salida) y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Topología	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
[ 16 ]	0.15	0	1.0	0	0.84	0	0.84
[ 32 ]	0.15	0	1.0	0	0.84	0	0.84
[ 64 ]	0.15	0	1.0	0	0.84	0	0.84
[ 128 ]	0.15	0	1.0	0	0.84	0	0.84
[ 16 32 ]	0.15	0	1.0	0	0.84	0	0.84
[ 32 64 ]	0.15	0	1.0	0	0.84	0	0.84
[ 64 64 ]	0.15	0	1.0	0	0.84	0	0.84
[ 64 128 ]	0.15	0	1.0	0	0.84	0	0.84

Tabla 9: Métricas Segunda Aprox. RNA



Topología	Matriz de confusión				Topología	Matriz de confusión			
[ 16 ]			Predicción		[ 16 32 ]			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	67	0		Real	Neg	67	0
		Pos	12	0			Pos	12	0
	[ 32 ]			Predicción		[ 32 64 ]			Predicción
Neg				Pos	Neg				Pos
Real		Neg	67	0	Real		Neg	67	0
		Pos	12	0			Pos	12	0
[ 64 ]				Predicción			[ 64 64 ]		
	Neg			Pos	Neg	Pos			
	Real	Neg	67	0	Real	Neg		67	0
		Pos	12	0		Pos		12	0
	[ 128 ]			Predicción		[ 64 128 ]			
Neg				Pos	Neg		Pos		
Real		Neg	67	0	Real		Neg	67	0
		Pos	12	0			Pos	12	0

Tabla 31: Matriz de Confusión Segunda Aprox. RNA

### 4.2.3. Máquinas de Soporte Vectorial

Las 8 configuraciones diferentes de hiperparámetros de la SVM que han sido probadas y los resultados obtenidos en las métricas que derivan de la matriz de confusión y sus respectivos valores son los siguientes:

kernel	k. degree	gamma	c	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
"rbf"	3	2	1	0.1333	0.0	1.0	0	0.867	0	0.8667
"rbf"	3	2	2	0.1333	0.0	1.0	0	0.867	0	0.8667
"rbf"	3	2	3	0.1333	0.0	1.0	0	0.867	0	0.8667
"poly"	3	2	1	0.1333	0.0	1.0	0	0.867	0	0.8667
"linear"	3	2	1	0.1333	0.0	1.0	0	0.867	0	0.8667
"sigmoid"	3	2	0.9	0.2	0.0	0.923077	0	0.923	0	0.8
"sigmoid"	3	2	1	0.2	0.0	0.923077	0	0.923	0	0.8
"sigmoid"	3	2	3	0.2	0.0	0.923077	0	0.923	0	0.8

Tabla 10: Métricas Segunda Aprox. SVM

(Kernel; k.degree; gamma; c)	Matriz de confusión				(Kernel; k.degree; gamma; c)	Matriz de confusión			
("rbf"; 3; 2; 1)			Predicción		("linear"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	26	0		Real	Neg	26	0
		Pos	4	0			Pos	4	0
	("rbf"; 3; 2; 2)			Predicción		("sigmoid"; 3; 2; 0.9)			Predicción
Neg				Pos	Neg				Pos
Real		Neg	26	0	Real		Neg	24	2
		Pos	4	0			Pos	4	0
("rbf"; 3; 2; 3)				Predicción			("sigmoid"; 3; 2; 1)		
	Neg			Pos	Neg	Pos			
	Real	Neg	26	0	Real	Neg		24	2
		Pos	4	0		Pos		4	0

("poly"; 3; 2; 1)				
			Predicción	
			Neg	Pos
	Real	Neg	26	0
		Pos	4	0

("sigmoid"; 3; 2; 3)				
			Predicción	
			Neg	Pos
	Real	Neg	24	2
		Pos	4	0

Tabla 11: Matriz de Confusión Segunda Aprox. SVM

#### 4.2.4. Árboles de Decisión

Los 6 valores diferentes de profundidad que han sido probados y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Maxdepth	Tasa fallo	Sensibilidad	Especificidad	F1 score	Valor pred neg	Valor pred pos	Valor precisión
4	0.133	0.9615	0.25	0.9259	0.5	0.8928	0.8666
2	0.133	1.0	0.0	0.9285	0.0	0.8666	0.8666
6	0.166	0.9230	0.25	0.90566	0.3333	0.8888	0.8333
15	0.133	0.8846	0.75	0.92	0.5	0.9583	0.8666
1	0.133	1.0	0.0	0.9285	0.5	0.8666	0.8666
7	0.133	0.9615	0.25	0.9259	0.5	0.8928	0.8666

Tabla 12: Métricas Segunda Aprox. Árboles de Decisión

Maxdepth	Matriz de confusión				Maxdepth	Matriz de confusión			
4			Predicción		15			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	3		Real	Neg	3	1
		Pos	1	25			Pos	3	23
2			Predicción		1			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	0	4		Real	Neg	0	4
		Pos	0	26			Pos	0	26
6			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	3		Real	Neg	1	3
		Pos	2	24			Pos	1	25

Tabla 13: Matriz de Confusión Segunda Aprox. Árboles de Decisión

Como se puede observar, todas las métricas han mejorado considerablemente, lo cual no es necesariamente bueno. Si observamos la matriz de confusión, podemos observar una gran desproporción de los patrones entre las 2 clases. Esto tiene como consecuencia, una falta de capacidad de generalización ya que el conjunto de datos no es ni representativo, ni significativo.

#### 4.2.5. k Nearest Neighbors

Los 6 valores diferentes de k que han sido probados son los siguientes:

k	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec
3	0.066	0.75	0.961	0.75	0.961538	0.75	0.933
4	0.133	0.0	1	0.0	0.866	0.0	0.866
5	0.133	0.0	1	0.0	0.866	0.0	0.866
6	0.133	0.0	1	0.0	0.866	0.0	0.866
7	0.133	0.0	1	0.0	0.866	0.0	0.866
10	0.133	0.0	1	0.0	0.866	0.0	0.866

Tabla 14: Métricas Segunda Aprox. KNN

k	Matriz de confusión				k	Matriz de confusión			
3			Predicción		6			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	25	1		Real	Neg	26	0
		Pos	1	3			Pos	4	0
4			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	26	0		Real	Neg	26	0
		Pos	4	0			Pos	4	0
5			Predicción		10			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	26	0		Real	Neg	26	0
		Pos	4	0			Pos	4	0

Tabla 15: Matriz de Confusión Segunda Aprox. KNN

Analizando los datos obtenidos en kNN, se observa una muy considerable mejoría en los resultados obtenidos en la experimentación con  $k = 3$ . Como fue mencionado en el anterior punto, esto no implica necesariamente una buena noticia, pues la matriz de confusión también está muy desbalanceada.

#### 4.2.6. Discusión

Como puede observarse en los resultados obtenidos, esta aproximación tiene en general unos datos peores que en la primera debido a que el rango de alcohol escogido tiene una representación mucho peor en la base de datos (hay muchas más entradas de cervezas con alcohol que sin alcohol), dando lugar a resultados pobres. Ninguna topología de las RNA se salva con un F1 de 0, al igual que las SVM.

Podríamos destacar como kNN con  $k = 3$  como la mejor, con un *F1 score* de 0.75 y una precisión del 93%. Sin embargo, los mejores resultados de la aproximación se encuentran en los árboles de decisión, que se comportan de una forma excelente con un F1 casi siempre en 0.92 y una precisión nunca por debajo de 0.83.

## 4.3. Tercera Aproximación

### 4.3.1. Introducción

Para esta tercera aproximación emplearemos el producto de la desviación típica de la reflectancia e intensidad y la desviación típica de la absorbancia. Además volveremos a situar la graduación umbral en los 5.5°.

### 4.3.2. Redes de Neuronas Artificiales

Las 8 arquitecturas diferentes que han sido probadas (todas con 1 capa de entrada y 1 de salida) y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Topología	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
[ 16 ]	0.34	1.0	0.0	0.79	0	0.65	0.65
[ 32 ]	0.25	0.98	0.29	0.83	0.88	0.72	0.74
[ 64 ]	0.25	0.98	0.29	0.83	0.88	0.72	0.74
[ 128 ]	0.29	0.98	0.18	0.81	0.83	0.69	0.7
[ 16 32 ]	0.25	0.98	0.29	0.83	0.88	0.72	0.74
[ 32 64 ]	0.31	0.88	0.29	0.78	0.57	0.7	0.68
[ 64 64 ]	0.29	0.96	0.22	0.81	0.75	0.7	0.7
[ 64 128 ]	0.3	0.96	0.18	0.8	0.71	0.69	0.69

Tabla 16: Métricas Tercera Aprox. RNA

Topología	Matriz de confusión				Topología	Matriz de confusión			
[ 16 ]					[ 16 32 ]				
[ 16 ]	Real	Neg	0	27	[ 16 32 ]	Real	Neg	8	19
		Pos	0	52			Pos	1	51
	Real	Neg	0	27		Real	Neg	8	19
		Pos	0	52			Pos	1	51

[ 32 ]	<table border="1"> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>1</td><td>51</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	1	51	[ 32 64 ]	<table border="1"> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>6</td><td>46</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	6	46
				Predicción																									
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	1	51																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	6	46																										
[ 64 ]	<table border="1"> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>19</td></tr> <tr> <td>Pos</td><td>1</td><td>51</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	19	Pos	1	51	[ 64 64 ]	<table border="1"> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>6</td><td>21</td></tr> <tr> <td>Pos</td><td>2</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	6	21	Pos	2	50
				Predicción																									
		Neg	Pos																										
Real	Neg	8	19																										
	Pos	1	51																										
		Predicción																											
		Neg	Pos																										
Real	Neg	6	21																										
	Pos	2	50																										
[ 128 ]	<table border="1"> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>5</td><td>22</td></tr> <tr> <td>Pos</td><td>1</td><td>51</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	22	Pos	1	51	[ 64 128 ]	<table border="1"> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>5</td><td>22</td></tr> <tr> <td>Pos</td><td>2</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	22	Pos	2	50
				Predicción																									
		Neg	Pos																										
Real	Neg	5	22																										
	Pos	1	51																										
		Predicción																											
		Neg	Pos																										
Real	Neg	5	22																										
	Pos	2	50																										

Tabla 32: Matriz de Confusión Tercera Aprox. RNA

### 4.3.3. Máquinas de Soporte Vectorial

Las 8 configuraciones diferentes de hiperparámetros de la SVM que han sido probadas y los resultados obtenidos en las métricas que derivan de la matriz de confusión y sus respectivos valores son los siguientes:



kernel	k. degree	gamma	c	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred neg.	Valor pred. pos.	Valor prec.
"rbf"	3	2	1	0.3	1.0	0.1	0.81	1.0	0.689	0.7
"rbf"	3	2	2	0.3	1.0	0.1	0.81	1.0	0.689	0.7
"rbf"	3	2	3	0.3	1.0	0.1	0.81	1.0	0.689	0.7
"poly"	3	2	1	0.33	1.0	0	0.8	0	0.666	0.666
"linear"	3	2	1	0.33	1.0	0	0.8	0	0.666	0.666
"sigmoid"	3	2	0.9	0.33	0.95	0.1	0.79	0.5	0.678	0.666
"sigmoid"	3	2	1	0.33	0.95	0.2	0.79	0.5	0.678	0.666
"sigmoid"	3	2	3	0.46	0.65	0.3	0.65	0.3	0.650	0.533

Tabla 17: Métricas Tercera Aprox. SVM

(Kernel; k.degree; gamma; c)	Matriz de confusión				(Kernel; k.degree; gamma; c)	Matriz de confusión			
("rbf"; 3; 2; 1)			Predicción		("linear"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	0	10
		Pos	0	20			Pos	0	20
("rbf"; 3; 2; 2)			Predicción		("sigmoid"; 3; 2; 0.9)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	1	9
		Pos	0	20			Pos	1	19
("rbf"; 3; 2; 3)			Predicción		("sigmoid"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	1	9
		Pos	0	20			Pos	1	19

("poly"; 3; 2; 1)				Predicción	
				Neg	Pos
	Real	Neg	0	10	
		Pos	0	20	

("sigmoid"; 3; 2; 3)				Predicción	
				Neg	Pos
	Real	Neg	3	7	
		Pos	7	13	

Tabla 18: Matriz de Confusión Tercera Aprox. SVM

#### 4.3.4. Árboles de decisión

Los 6 valores diferentes de profundidad que han sido probados y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Maxdepth	Tasa fallo	Sensibilidad	Especificidad	F1 score	Valor pred neg	Valor pred pos	Valor precisión
4	0.3	0.85	0.4	0.790698	0.57143	0.7	0.73913
2	0.3	0.95	0.2	0.808511	0.66667	0.703704	0.7
6	0.2667	0.9	0.4	0.818182	0.6667	0.75	0.7333
15	0.2	0.95	0.5	0.863636	0.8333	0.791667	0.8
1	0.33333	1.0	0	0.8	0	0.6667	0.66667
7	0.13333	0.95	0.7	0.904762	0.875	0.863636	0.86667

Tabla 19: Métricas Tercera Aprox. Árboles de Decisión

Maxdepth	Matriz de confusión				Maxdepth	Matriz de confusión			
4			Predicción		15			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	4	6		Real	Neg	5	5
		Pos	3	17			Pos	1	19
2			Predicción		1			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	2	8		Real	Neg	0	10
		Pos	1	19			Pos	0	20
6			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	4	6		Real	Neg	7	3
		Pos	2	18			Pos	1	19

Tabla 20: Matriz de Confusión Tercera Aprox. Árboles de Decisión

### 4.3.5. k Nearest Neighbors

Los 6 valores diferentes de k que han sido probados son los siguientes:

k	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec
3	0.3	0.8	0.5	0.780488	0.555556	0.761905	0.7
4	0.266667	0.8	0.6	0.8	0.6	0.8	0.733333
5	0.266667	0.9	0.4	0.818182	0.666667	0.75	0.733333
6	0.233333	0.85	0.6	0.829268	0.666667	0.809524	0.766667
7	0.233333	0.9	0.5	0.837209	0.714286	0.782609	0.766667
10	0.333333	0.7	0.6	0.736842	0.5	0.777778	0.666667

Tabla 21: Métricas Tercera Aprox. KNN

k	Matriz de confusión				k	Matriz de confusión			
3			Predicción		6			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	5	5		Real	Neg	6	4
		Pos	4	16			Pos	3	17
4			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	6	4		Real	Neg	5	5
		Pos	4	16			Pos	2	18
5			Predicción		10			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	4	6		Real	Neg	6	4
		Pos	2	18			Pos	6	14

Tabla 22: Matriz de Confusión Tercera Aprox. KNN

#### 4.3.6. Discusión

Al intentar simplificar las entradas haciendo el producto de dos de ellas podemos observar que obtenemos mejores resultados que los obtenidos en la primera y segunda aproximación (dejando a la vista bastantes soluciones aceptables, como el árbol de decisión con *maxdepth* = 7, con *F1 score* de 0.9).

En las técnicas utilizadas, la de árboles de decisión es la que arroja con poca diferencia mejores estadísticas (expuestas anteriormente), sin ser las restantes del todo pobres: RNA y KNN alcanzan 0.83 de F1 y 0.81 las SVM.

## 4.4. Cuarta Aproximación

### 4.4.1. Introducción

Para esta cuarta aproximación emplearemos la media por la desviación típica de los datos en el rango de longitud de onda [1100 - 1330 nm], manteniendo el umbral entre si hay o no alcohol en 5,5°.

### 4.4.2. Redes de Neuronas Artificiales

Las 8 arquitecturas diferentes que han sido probadas (todas con 1 capa de entrada y 1 de salida) y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Topología	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
[ 16 ]	0.35	1.0	0	0.78	0	0.64	0.64
[ 32 ]	0.27	0.96	0.28	0.81	0.8	0.71	0.72
[ 64 ]	0.30	0.92	0.28	0.79	0.66	0.70	0.69
[ 128 ]	0.31	0.98	0.14	0.8	0.8	0.97	0.68
[ 16 32 ]	0.31	0.92	0.25	0.78	0.63	0.69	0.68
[ 32 64 ]	0.3	0.92	0.28	0.79	0.66	0.70	0.69
[ 64 64 ]	0.31	0.96	0.17	0.79	0.71	0.68	0.68
[ 64 128 ]	0.3	0.98	0.17	0.80	0.83	0.68	0.69

Tabla 23: Métricas Cuarta Aprox. RNA

Topología	Matriz de confusión				Topología	Matriz de confusión			
[ 16 ]			Predicción		[ 16 32 ]			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	0	28		Real	Neg	7	21
		Pos	0	51			Pos	4	47

[ 32 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>20</td></tr> <tr> <td>Pos</td><td>2</td><td>49</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	20	Pos	2	49	[ 32 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>20</td></tr> <tr> <td>Pos</td><td>4</td><td>47</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	20	Pos	4	47
				Predicción																									
		Neg	Pos																										
Real	Neg	8	20																										
	Pos	2	49																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	20																										
	Pos	4	47																										
[ 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>20</td></tr> <tr> <td>Pos</td><td>4</td><td>47</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	20	Pos	4	47	[ 64 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>5</td><td>23</td></tr> <tr> <td>Pos</td><td>2</td><td>49</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	23	Pos	2	49
				Predicción																									
		Neg	Pos																										
Real	Neg	8	20																										
	Pos	4	47																										
		Predicción																											
		Neg	Pos																										
Real	Neg	5	23																										
	Pos	2	49																										
[ 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>4</td><td>24</td></tr> <tr> <td>Pos</td><td>1</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	4	24	Pos	1	50	[ 64 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>5</td><td>23</td></tr> <tr> <td>Pos</td><td>1</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	23	Pos	1	50
				Predicción																									
		Neg	Pos																										
Real	Neg	4	24																										
	Pos	1	50																										
		Predicción																											
		Neg	Pos																										
Real	Neg	5	23																										
	Pos	1	50																										

Tabla 33: Matriz de Confusión Cuarta Aprox. RNA

#### 4.4.3. Máquinas de Soporte Vectorial

Las 8 configuraciones diferentes de hiperparámetros de la SVM que han sido probadas y los resultados obtenidos en las métricas que derivan de la matriz de confusión y sus respectivos valores son los siguientes:

kernel	k. degree	gamma	c	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred neg.	Valor pred. pos.	Valor prec.
"rbf"	3	2	1	0.3	1.0	0.1	0.816 3	1.0	0.6897	0.7
"rbf"	3	2	2	0.3	1.0	0.1	0.816 3	1.0	0.6897	0.7
"rbf"	3	2	3	0.3	1.0	0.1	0.816 3	1.0	0.6897	0.7

"poly"	3	2	1	0.333	1.0	0.0	0.8	0	0.6667	0.6667
"linear"	3	2	1	0.333	1.0	0.0	0.8	0	0.6667	0.6667
"sigmoid"	3	2	0.9	0.366	0.95	0.0	0.77	0	0.6667	0.6333
"sigmoid"	3	2	1	0.366	0.95	0.0	0.77	0	0.6667	0.6333
"sigmoid"	3	2	3	0.433	0.75	0.2	0.69	0.28	0.6558	0.56

Tabla 24: Métricas Cuarta Aprox. SVM

(Kernel; k.degree; gamma; c)	Matriz de confusión				(Kernel; k.degree; gamma; c)	Matriz de confusión			
("rbf"; 3; 2; 1)			Predicción		("linear"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	0	10
		Pos	0	20			Pos	0	20
("rbf"; 3; 2; 2)			Predicción		("sigmoid"; 3; 2; 0.9)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	1	9
		Pos	0	20			Pos	2	18
("rbf"; 3; 2; 3)			Predicción		("sigmoid"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	2	8
		Pos	0	20			Pos	3	17
("poly"; 3; 2; 1)			Predicción		("sigmoid"; 3; 2; 3)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	0	10		Real	Neg	2	8
		Pos	0	20			Pos	5	15



Tabla 25: Matriz de Confusión Cuarta Aprox. SVM

#### 4.4.4. Árboles de decisión

Los 6 valores diferentes de profundidad que han sido probados y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Maxdepth	Tasa fallo	Sensibilidad	Especificidad	F1 score	Valor pred neg	Valor pred pos	Valor precisión
4	0.3	1.0	0.1	0.81	1.0	0.68	0.7
2	0.3	1.0	0.1	0.81	1.0	0.68	0.7
6	0.2	0.9	0.6	0.85	0.75	0.81	0.8
15	0.2333	0.85	0.6	0.82	0.66	0.80	0.7667
1	0.3	1.0	0.1	0.81	1.0	0.7	0.68
7	0.2	0.9	0.6	0.85	0.75	0.81	0.8

Tabla 26: Métricas Cuarta Aprox. Árboles de Decisión

Maxd epth	Matriz de confusión				Maxd epth	Matriz de confusión			
4			Predicción		15			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	6	4
		Pos	0	20			Pos	3	17
2			Predicción		1			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	1	9
		Pos	0	20			Pos	0	20
6			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	6	4		Real	Neg	6	4
		Pos	2	18			Pos	2	18

Tabla 27: Matriz de Confusión Cuarta Aprox. Árboles de Decisión

#### 4.4.5. k Nearest Neighbors

Los 6 valores diferentes de k que han sido probados son los siguientes:

k	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec
3	0.16	0.95	0.6	0.88	0.85	0.82	0.83
4	0.2	0.9	0.6	0.85	0.75	0.81	0.8
5	0.16	0.95	0.6	0.88	0.85	0.82	0.83
6	0.2	0.9	0.6	0.85	0.75	0.81	0.8
7	0.26	0.9	0.4	0.81	0.66	0.75	0.73
10	0.2	1.0	0.4	0.86	1.0	0.76	0.8

Tabla 28: Métricas Cuarta Aprox. KNN

k	Matriz de confusión				k	Matriz de confusión			
3			Predicción		6			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	6	4		Real	Neg	6	4
		Pos	1	19			Pos	2	18
4			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	6	4		Real	Neg	4	6
		Pos	2	18			Pos	2	18
5			Predicción		10			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	6	4		Real	Neg	4	6
		Pos	1	19			Pos	0	20

Tabla 29: Matriz de Confusión Cuarta Aprox. KNN

#### 4.4.6. Discusión

En esta cuarta aproximación, se observan unos resultados en general bastante buenos (F1 sólo baja de 0.8 en algunas configuraciones de SVM, en el resto lo supera). En especial en KNN, donde con  $k = 3$  y  $k = 5$  se alcanza un F1 de 0.88 y una precisión de 0.83 y en la técnica de árboles de decisión, donde se alcanza un F1 de 0.85 y una precisión de 0.8 en Maxdepth = 6 y Maxdepth = 7.

## 4.5. Quinta Aproximación

Esta aproximación, inicialmente reservada para la técnica de Deep Learning, será realizada con las técnicas tradicionales previamente usadas debido a problemas relacionados con el desarrollo de Deep Learning.

### 4.5.1. Introducción

Para esta quinta aproximación dividiremos el rango de longitud de onda [1100 - 1330 nm] en cuatro subrangos y obtendremos de ellos el máximo valor para después emplear la media de los máximos de intensidad por la media de los máximos de reflectancia y la media de los máximos de absorbancia.

Mantenemos el umbral entre si hay o no alcohol en 5,5°.

### 4.5.2. Redes de Neuronas Artificiales

Las 8 arquitecturas diferentes que han sido probadas (todas con 1 capa de entrada y 1 de salida) y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Topología	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
[ 16 ]	0.3671	1.0	0.0	0.7752	0	0.6329	0.6329
[ 32 ]	0.3291	0.92	0.2413	0.7797	0.6364	0.6765	0.6708
[ 64 ]	0.32	0.92	0.24	0.77	0.63	0.67	0.67
[ 128 ]	0.35	0.96	0.10	0.77	0.6	0.64	0.64
[ 16 32 ]	0.31	0.92	0.27	0.78	0.66	0.68	0.68
[ 32 64 ]	0.27	0.94	0.34	0.81	0.76	0.71	0.71
[ 64 64 ]	0.27	0.94	0.34	0.81	0.76	0.72	0.71
[ 64 128 ]	0.34	0.94	0.17	0.77	0.625	0.66	0.65

Tabla 30: Métricas Quinta Aprox. RNA

Topología	Matriz de confusión	Topología	Matriz de confusión																										
[ 16 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>0</td><td>29</td></tr> <tr> <td>Pos</td><td>0</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	0	29	Pos	0	50	[ 16 32 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>21</td></tr> <tr> <td>Pos</td><td>4</td><td>46</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	21	Pos	4	46
				Predicción																									
		Neg	Pos																										
Real	Neg	0	29																										
	Pos	0	50																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	21																										
	Pos	4	46																										
[ 32 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>7</td><td>22</td></tr> <tr> <td>Pos</td><td>4</td><td>46</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	7	22	Pos	4	46	[ 32 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>10</td><td>19</td></tr> <tr> <td>Pos</td><td>3</td><td>47</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	10	19	Pos	3	47
				Predicción																									
		Neg	Pos																										
Real	Neg	7	22																										
	Pos	4	46																										
		Predicción																											
		Neg	Pos																										
Real	Neg	10	19																										
	Pos	3	47																										
[ 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>7</td><td>22</td></tr> <tr> <td>Pos</td><td>4</td><td>46</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	7	22	Pos	4	46	[ 64 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>10</td><td>19</td></tr> <tr> <td>Pos</td><td>3</td><td>47</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	10	19	Pos	3	47
				Predicción																									
		Neg	Pos																										
Real	Neg	7	22																										
	Pos	4	46																										
		Predicción																											
		Neg	Pos																										
Real	Neg	10	19																										
	Pos	3	47																										
[ 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>3</td><td>26</td></tr> <tr> <td>Pos</td><td>2</td><td>48</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	3	26	Pos	2	48	[ 64 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>5</td><td>24</td></tr> <tr> <td>Pos</td><td>3</td><td>47</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	24	Pos	3	47
				Predicción																									
		Neg	Pos																										
Real	Neg	3	26																										
	Pos	2	48																										
		Predicción																											
		Neg	Pos																										
Real	Neg	5	24																										
	Pos	3	47																										

Tabla 31: Matriz de Confusión Quinta Aprox. RNA

### 4.5.3. Máquinas de Soporte Vectorial

Las 8 configuraciones diferentes de hiperparámetros de la SVM que han sido probadas y los resultados obtenidos en las métricas que derivan de la matriz de confusión y sus respectivos valores son los siguientes:

kernel	k. degree	gamma	c	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
"rbf"	3	2	1	0.38	0.93	0.18	0.737	0.67	0.61	0.61
"rbf"	3	2	2	0.38	0.93	0.18	0.737	0.67	0.61	0.61
"rbf"	3	2	3	0.38	0.93	0.18	0.737	0.67	0.61	0.61
"poly"	3	2	1	0.46	0.93	0.0	0.7	0.0	0.56	0.53
"linear"	3	2	1	0.42	1.0	0.0	0.73	0.0	0.57	0.57
"sigmoid"	3	2	0.9	0.42	1.0	0.0	0.73	0.0	0.57	0.57
"sigmoid"	3	2	1	0.42	1.0	0.0	0.73	0.0	0.57	0.57
"sigmoid"	3	2	3	0.34	0.87	0.36	0.74	0.67	0.65	0.65

Tabla 32: Métricas Quinta Aprox. SVM

(Kernel; k.degree; gamma; c)	Matriz de confusión				(Kernel; k.degree; gamma; c)	Matriz de confusión			
("rbf"; 3; 2; 1)			Predicción		("linear"; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	2	9		Real	Neg	0	11
		Pos	1	14			Pos	0	15
("rbf"; 3; 2; 2)			Predicción		("sigmoid" ; 3; 2; 0.9)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	2	9		Real	Neg	0	11
		Pos	1	14			Pos	0	15
("rbf"; 3; 2; 3)			Predicción		("sigmoid" ; 3; 2; 1)			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	2	9		Real	Neg	0	11
		Pos	1	14			Pos	0	15

("poly"; 3; 2; 1)			Predicción	
			Neg	Pos
	Real	Neg	0	11
		Pos	1	14

("sigmoid"; 3; 2; 3)			Predicción	
			Neg	Pos
	Real	Neg	4	7
		Pos	2	13

Tabla 33: Matriz de Confusión Quinta Aprox. SVM

#### 4.5.4. Árboles de decisión

Los 6 valores diferentes de profundidad que han sido probados y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Maxdepth	Tasa fallo	Sensibilidad	Especificidad	F1 score	Valor pred neg	Valor pred pos	Valor precisión
4	0.19	1.0	0.5	0.86	1.0	0.76	0.81
2	0.42	0.875	0.1	0.72	0.33	0.61	0.58
6	0.15	0.81	0.9	0.87	0.75	0.93	0.85
15	0.23	0.75	0.8	0.8	0.67	0.86	0.77
1	0.42	0.81	0.2	0.7	0.4	0.62	0.58
7	0.15	0.875	0.8	0.875	0.8	0.875	0.85

Tabla 35: Métricas Quinta Aprox. Árboles de Decisión

Maxd epth	Matriz de confusión				Maxd epth	Matriz de confusión			
4			Predicción		15			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	5	5		Real	Neg	8	2
		Pos	0	16			Pos	4	12
2			Predicción		1			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	1	9		Real	Neg	2	8
		Pos	2	14			Pos	3	13
6			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	9	1		Real	Neg	8	2
		Pos	3	13			Pos	2	14

Tabla 36: Matriz de Confusión Quinta Aprox. Árboles de Decisión

#### 4.5.5. k Nearest Neighbors

Los 6 valores diferentes de k que han sido probados son los siguientes:

k	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec
3	0.3077	0.8235	0.4444	0.7778	0.5714	0.7368	0.6923
4	0.3461	0.6470	0.6667	0.7097	0.5	0.7857	0.6538
5	0.3077	0.7647	0.5556	0.7647	0.5556	0.7647	0.6923
6	0.3846	0.5882	0.6667	0.6667	0.4615	0.7692	0.6153
7	0.2307	0.8823	0.5556	0.8333	0.7142	0.7895	0.7692
10	0.2307	0.8823	0.5556	0.8333	0.7142	0.7895	0.7692

Tabla 37: Métricas Quinta Aprox. KNN



k	Matriz de confusión				k	Matriz de confusión			
3			Predicción		6			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	4	5		Real	Neg	6	3
		Pos	3	14			Pos	7	10
4			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	6	3		Real	Neg	5	4
		Pos	6	11			Pos	2	15
5			Predicción		10			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	5	4		Real	Neg	5	4
		Pos	4	13			Pos	2	15

Tabla 38: Matriz de Confusión Quinta Aprox. KNN

#### 4.5.6. Discusión

Se puede apreciar en los resultados de esta quinta aproximación unas métricas bastante aceptables y con una arquitectura especialmente sobresaliente: el árbol de decisión con Maxdepth = 7; con un F1 Score de 0.875 y un valor de precisión de 0.85, lo cual son unos datos realmente sorprendentes y los mejores obtenidos a lo largo de todo el desarrollo.

## 4.6. Sexta Aproximación

### 4.6.1. Introducción

En esta última aproximación realizaremos la misma división por subconjuntos que en la anterior, con la diferencia de que esta vez emplearemos los valores mínimos de cada subrango. Al finalizar, haremos el producto de la media de los mínimos de intensidad por la media de los mínimos de reflectancia y usaremos el resultado junto a la media de los mínimos de absorbancia.

### 4.6.2. Redes de Neuronas Artificiales

Las 8 arquitecturas diferentes que han sido probadas (todas con 1 capa de entrada y 1 de salida) y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Topología	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
[ 16 ]	0.36	1.0	0.0	0.77	0.0	0.63	0.63
[ 32 ]	0.31	0.88	0.34	0.77	0.63	0.7	0.68
[ 64 ]	0.35	0.96	0.10	0.77	0.6	0.65	0.65
[ 128 ]	0.35	0.96	0.10	0.77	0.6	0.65	0.65
[ 16 32 ]	0.32	0.92	0.27	0.79	0.67	0.69	0.68
[ 32 64 ]	0.3	0.9	0.34	0.79	0.67	0.7	0.7
[ 64 64 ]	0.29	0.92	0.34	0.8	0.71	0.7	0.71
[ 64 128 ]	0.29	0.96	0.27	0.8	0.8	0.69	0.71

Tabla 39: Métricas Sexta Aprox. RNA

Topología	Matriz de confusión	Topología	Matriz de confusión																										
[ 16 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>0</td><td>29</td></tr> <tr> <td>Pos</td><td>0</td><td>50</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	0	29	Pos	0	50	[ 16 32 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>21</td></tr> <tr> <td>Pos</td><td>4</td><td>46</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	21	Pos	4	46
				Predicción																									
		Neg	Pos																										
Real	Neg	0	29																										
	Pos	0	50																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	21																										
	Pos	4	46																										
[ 32 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>10</td><td>19</td></tr> <tr> <td>Pos</td><td>6</td><td>44</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	10	19	Pos	6	44	[ 32 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>10</td><td>19</td></tr> <tr> <td>Pos</td><td>5</td><td>45</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	10	19	Pos	5	45
				Predicción																									
		Neg	Pos																										
Real	Neg	10	19																										
	Pos	6	44																										
		Predicción																											
		Neg	Pos																										
Real	Neg	10	19																										
	Pos	5	45																										
[ 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>3</td><td>26</td></tr> <tr> <td>Pos</td><td>2</td><td>48</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	3	26	Pos	2	48	[ 64 64 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>10</td><td>19</td></tr> <tr> <td>Pos</td><td>4</td><td>46</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	10	19	Pos	4	46
				Predicción																									
		Neg	Pos																										
Real	Neg	3	26																										
	Pos	2	48																										
		Predicción																											
		Neg	Pos																										
Real	Neg	10	19																										
	Pos	4	46																										
[ 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>3</td><td>26</td></tr> <tr> <td>Pos</td><td>2</td><td>48</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	3	26	Pos	2	48	[ 64 128 ]	<table> <tr> <td colspan="2" rowspan="2"></td><td colspan="2">Predicción</td></tr> <tr> <td>Neg</td><td>Pos</td></tr> <tr> <td rowspan="2">Real</td><td>Neg</td><td>8</td><td>21</td></tr> <tr> <td>Pos</td><td>2</td><td>48</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	8	21	Pos	2	48
				Predicción																									
		Neg	Pos																										
Real	Neg	3	26																										
	Pos	2	48																										
		Predicción																											
		Neg	Pos																										
Real	Neg	8	21																										
	Pos	2	48																										

Tabla 40: Matriz de Confusión Sexta Aprox. RNA

### 4.6.3. Máquinas de Soporte Vectorial

Las 8 configuraciones diferentes de hiperparámetros de la SVM que han sido probadas y los resultados obtenidos en las métricas que derivan de la matriz de confusión y sus respectivos valores son los siguientes:

kernel	k. degree	gamma	c	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec.
"rbf"	3	2	1	0.3846	0.9333	0.1818	0.7368	0.6667	0.6086	0.6154
"rbf"	3	2	2	0.3846	0.9333	0.1818	0.7368	0.6667	0.6086	0.6154
"rbf"	3	2	3	0.3846	0.9333	0.1818	0.7368	0.6667	0.6086	0.6154
"poly"	3	2	1	0.3846	0.9333	0.1818	0.7368	0.6667	0.6086	0.6154
"linear"	3	2	1	0.4231	1.0	0.0	0.7317	0	0.5769	0.5769
"sigmoid"	3	2	0.9	0.4231	1.0	0.0	0.7317	0	0.5769	0.5769
"sigmoid"	3	2	1	0.4231	1.0	0.0	0.7317	0	0.5769	0.5769
"sigmoid"	3	2	3	0.3077	0.8667	0.4545	0.7647	0.7143	0.6842	0.6923

Tabla 41: Métricas Sexta Aprox. SVM

(Kernel; k.degree; gamma; c)	Matriz de confusión	(Kernel; k.degree; gamma; c)	Matriz de confusión																										
("rbf"; 3; 2; 1)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>2</td><td>9</td></tr> <tr> <th>Pos</th><td>1</td><td>14</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	2	9	Pos	1	14	("linear"; 3; 2; 1)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>0</td><td>11</td></tr> <tr> <th>Pos</th><td>0</td><td>15</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	0	11	Pos	0	15
				Predicción																									
		Neg	Pos																										
Real	Neg	2	9																										
	Pos	1	14																										
		Predicción																											
		Neg	Pos																										
Real	Neg	0	11																										
	Pos	0	15																										
("rbf"; 3; 2; 2)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>2</td><td>9</td></tr> <tr> <th>Pos</th><td>1</td><td>14</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	2	9	Pos	1	14	("sigmoid"; ; 3; 2; 0.9)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>0</td><td>11</td></tr> <tr> <th>Pos</th><td>0</td><td>15</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	0	11	Pos	0	15
				Predicción																									
		Neg	Pos																										
Real	Neg	2	9																										
	Pos	1	14																										
		Predicción																											
		Neg	Pos																										
Real	Neg	0	11																										
	Pos	0	15																										
("rbf"; 3; 2; 3)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>2</td><td>9</td></tr> <tr> <th>Pos</th><td>1</td><td>14</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	2	9	Pos	1	14	("sigmoid"; ; 3; 2; 1)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>0</td><td>11</td></tr> <tr> <th>Pos</th><td>0</td><td>15</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	0	11	Pos	0	15
				Predicción																									
		Neg	Pos																										
Real	Neg	2	9																										
	Pos	1	14																										
		Predicción																											
		Neg	Pos																										
Real	Neg	0	11																										
	Pos	0	15																										
("poly"; 3; 2; 1)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>2</td><td>9</td></tr> <tr> <th>Pos</th><td>1</td><td>14</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	2	9	Pos	1	14	("sigmoid"; ; 3; 2; 3)	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>5</td><td>6</td></tr> <tr> <th>Pos</th><td>2</td><td>13</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	6	Pos	2	13
				Predicción																									
		Neg	Pos																										
Real	Neg	2	9																										
	Pos	1	14																										
		Predicción																											
		Neg	Pos																										
Real	Neg	5	6																										
	Pos	2	13																										

Tabla 42: Matriz de Confusión Sexta Aprox. SVM

#### 4.6.4. Árboles de decisión

Los 6 valores diferentes de profundidad que han sido probados y los resultados obtenidos en las métricas que derivan de la matriz de confusión son los siguientes:

Maxdepth	Tasa fallo	Sensibilidad	Especificidad	F1 score	Valor pred neg	Valor pred pos	Valor precisión
4	0.26	0.93	0.4	0.81	0.8	0.71	0.73
2	0.42	0.87	0.1	0.71	0.33	0.6	0.57
6	0.15	0.81	0.9	0.86	0.75	0.92	0.84
15	0.076	0.875	1.0	0.93	0.83	1.0	0.92
1	0.42	0.81	0.2	0.7	0.4	0.61	0.57
7	0.11	0.81	1.0	0.89	0.76	1.0	0.88

Tabla 43: Métricas Sexta Aprox. Árboles de Decisión

Maxd epth	Matriz de confusión				Maxd epth	Matriz de confusión			
4			Predicción		15			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	4	6		Real	Neg	10	0
		Pos	1	15			Pos	2	14
2			Predicción		1			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	4	6		Real	Neg	2	8
		Pos	1	15			Pos	3	13
6			Predicción		7			Predicción	
			Neg	Pos				Neg	Pos
	Real	Neg	9	1		Real	Neg	10	0
		Pos	3	13			Pos	3	13

Tabla 44: Matriz de Confusión Sexta Aprox. Árboles de Decisión

### 4.6.5. k Nearest Neighbors

Los 6 valores diferentes de k que han sido probados son los siguientes:

k	Tasa fallo	Sensibilidad	Especificidad	F1	Valor pred. neg.	Valor pred. pos.	Valor prec
3	0.23	0.88	0.56	0.83	0.71	0.79	0.77
4	0.31	0.65	0.78	0.73	0.54	0.85	0.69
5	0.23	0.82	0.67	0.82	0.67	0.82	0.77
6	0.27	0.76	0.67	0.79	0.6	0.81	0.73
7	0.31	0.88	0.33	0.79	0.6	0.71	0.69
10	0.31	0.88	0.33	0.79	0.6	0.71	0.69

Tabla 45: Métricas Sexta Aprox. KNN

k	Matriz de confusión	k	Matriz de confusión																										
3	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>5</td><td>4</td></tr> <tr> <th>Pos</th><td>2</td><td>15</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	5	4	Pos	2	15	6	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>6</td><td>3</td></tr> <tr> <th>Pos</th><td>4</td><td>13</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	6	3	Pos	4	13
				Predicción																									
		Neg	Pos																										
Real	Neg	5	4																										
	Pos	2	15																										
		Predicción																											
		Neg	Pos																										
Real	Neg	6	3																										
	Pos	4	13																										
4	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>7</td><td>2</td></tr> <tr> <th>Pos</th><td>6</td><td>11</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	7	2	Pos	6	11	7	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>3</td><td>6</td></tr> <tr> <th>Pos</th><td>2</td><td>15</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	3	6	Pos	2	15
				Predicción																									
		Neg	Pos																										
Real	Neg	7	2																										
	Pos	6	11																										
		Predicción																											
		Neg	Pos																										
Real	Neg	3	6																										
	Pos	2	15																										
5	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>6</td><td>3</td></tr> <tr> <th>Pos</th><td>3</td><td>14</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	6	3	Pos	3	14	10	<table> <tr> <th colspan="2" rowspan="2"></th><th colspan="2">Predicción</th></tr> <tr> <th>Neg</th><th>Pos</th></tr> <tr> <th rowspan="2">Real</th><th>Neg</th><td>3</td><td>6</td></tr> <tr> <th>Pos</th><td>2</td><td>15</td></tr> </table>			Predicción		Neg	Pos	Real	Neg	3	6	Pos	2	15
				Predicción																									
		Neg	Pos																										
Real	Neg	6	3																										
	Pos	3	14																										
		Predicción																											
		Neg	Pos																										
Real	Neg	3	6																										
	Pos	2	15																										

Tabla 46: Matriz de Confusión Sexta Aprox. KNN

#### **4.6.6. Discusión**

Se puede apreciar en los resultados de esta sexta aproximación unas métricas bastante aceptables y con una arquitectura especialmente sobresaliente que incluso mejora la obtenida en la aproximación anterior: el árbol de decisión con  $\text{Maxdepth} = 15$ ; con un  $\text{F1\_Score}$  de 0.93 y un valor de precisión de 0.92, lo cual son unos datos realmente sorprendentes y los mejores obtenidos a lo largo de todo el desarrollo.



## 5. Conclusiones finales

Como se ha comentado en el punto 2.3 de este documento, las métricas objetivas que utilizamos para medir la bondad de nuestra solución son la sensibilidad y la precisión. En vista de esto, analizaremos los resultados obtenidos a lo largo de las seis aproximaciones en detalle para extraer unas últimas conclusiones.

Para nuestro problema, los mejores resultados obtenidos son generalmente medidos en los árboles de decisión (P. ej.: 4ª Aprox, Maxdepth = 15; 2ª Aprox, Maxdepth = 2), donde los valores son sobresalientemente buenos. También se obtienen buenos valores en KNN, en concreto cuando el valor de k es 3, aunque lo alcanza el nivel de bondad alcanzado en la primera. Por tanto, podemos inferir que la mejor técnica para nuestro problema de entre las probadas es la de árboles de decisión.

A continuación, en la Tabla 47 mostraremos las métricas que mejor han funcionado en cada aproximación y la técnica con la que se han obtenido.

Aproximación	Configuración	F1_Score	Sensibilidad	Precisión
1ª Aproximación	RNA => [64,64]	0.82	0.96	0.73
2ª Aproximación	KNN => k = 3	0.75	0.75	0.93
3ª Aproximación	Árbol Decisión => Maxdepth = 7	0.90	0.95	0.87
4ª Aproximación	KNN => k = 3	0.88	0.95	0.83
5ª Aproximación	Árbol Decisión => Maxdepth = 7	0.875	0.875	0.85
6ª Aproximación	Árbol Decisión => Maxdepth = 15	0.93	0.875	0.92

Tabla 47: Mejores resultados obtenidos en las distintas aproximaciones

## 6. Trabajo futuro

Como ha sido expuesto, el análisis de bebidas alcohólicas como la cerveza mediante espectrofotometría es un campo con amplias posibilidades en vista a futuros trabajos adyacentes.

Podrían experimentarse diferentes aproximaciones similares para este mismo problema, como, por ejemplo:

- Transformaciones de las características que ya hemos manejado (incluyéndolas o no)
  - El valor de la integral de las funciones obtenidas con las tres características
  - Construir nuevas variables como combinación lineal de las anteriores
- Inclusión de patrones y características nuevos sobre nuevas mediciones replicando las características de la original (color, aroma, tiempo de envejecimiento...)
- Nuevas mediciones con diferentes condiciones o técnicas que las mediciones originales, dando lugar a una nueva base de datos que podría aportar mejores resultados, por ejemplo, mediciones...
  - A mayor temperatura o con temperatura variable
  - Con otras bebidas fermentadas (vino, licor...)
  - Con la técnica de espectrometría de masas

En cuanto a aplicaciones posteriores de nuestra solución, sería posible aplicarlo en, por ejemplo, una fábrica de cerveza o un laboratorio que las analice, para determinar si un lote de cervezas sin alcohol es realmente tal, agilizando mucho el proceso (el tiempo de computación del programa es muy corto en comparación a un análisis tradicional en un laboratorio) y abaratándolo en gran medida (necesidad de menos expertos involucrados en el proceso).

## 7. Bibliografía

1. Espectrofotómetro DLP Nirscan Nano de Texas Instruments:  
<https://www.ti.com/tool/DLPNIRNANOEVIM>
2. Carlos Brunatti, Ana María Martín, Introducción a la Espectroscopía de Absorción Molecular Ultravioleta, Visible e Infrarrojo Cercano:  
<http://materias.fi.uba.ar/6305/download/Espectrofotometria.pdf>
3. Xiaofei Wang, Yanfei Bao, Guili Liu, Gang Li, Ling Lin, Study on the Best Analysis Spectral Section of NIR to Detect Alcohol Concentration Based on SiPLS:  
<https://cyberleninka.org/article/n/548375.pdf>
4. Danielle Cleveland, Matthew Carlson, Evan D. Hudspeth, Lauren E. Quattrochi, Kathleen L. Batchler, Shrimati A. Balram, Seongun Hong & Robert G. Michel - Raman Spectroscopy for the Undergraduate Teaching Laboratory: Quantification of Ethanol Concentration in Consumer Alcoholic Beverages and Qualitative Identification of Marine Diesels Using a Miniature Raman Spectrometer:  
<https://www.tandfonline.com/doi/full/10.1080/00387010701525638>
5. Nagarajan R, Mehrotra R & Bajaj M M - Quantitative analysis of methanol, an adulterant in alcoholic beverages, using attenuated total reflectance spectroscopy:  
<http://nopr.niscair.res.in/handle/123456789/4840>
6. Zhilei Zhou, Wei Ni, ZhongWei Ji, Shuangping Liu, Xiao Han, Xiuting Li & Jian Mao - Development of a Rapid Method for Determination of Main Higher Alcohols in Fermented Alcoholic Beverages Based on Dispersive Liquid-Liquid Microextraction and Gas Chromatography-Mass Spectrometry:  
<https://link.springer.com/article/10.1007/s12161-019-01668-4>
7. Máximo Galignani, Salvador Garrigues & Miguel de la Guardia - Direct determination of ethanol in all types of alcoholic beverages by near-infrared derivative spectrometry:  
<https://pubs.rsc.org/en/content/articlelanding/1993/an/an9931801167/unauth>
8. Salvador Gutiérrez, Javier Tardaguila, Juan Fernández-Novales, María P. Diago - Support Vector Machine and Artificial Neural Network Models for the Classification of Grapevine Varieties Using a Portable NIR Spectrophotometer: [Support Vector](#)

[Machine and Artificial Neural Network Models for the Classification of Grapevine Varieties Using a Portable NIR Spectrophotometer](#)

9. Suárez Díaz, M. (2013). Cerveza: Componentes y Propiedades (Trabajo de fin de máster). [“CERVEZA: COMPONENTES Y PROPIEDADES”](#).
10. Alpaydin, E. (2021). *Machine learning*. MIT Press.  
[https://books.google.es/books?hl=es&lr=&id=2nQJEAAAQBAJ&oi=fnd&pg=PR7&dq=Alpaydin,+E.+\(2021\).+Machine+learning.+MIT+Press.&ots=fH23O8Txmt&sig=Qlso79GxjmU5QKZPUWZbYSVkp4&redir\\_esc=y#v=onepage&q=Alpaydin%2C%20E.%20\(2021\).%20Machine%20learning.%20MIT%20Press.&f=false](https://books.google.es/books?hl=es&lr=&id=2nQJEAAAQBAJ&oi=fnd&pg=PR7&dq=Alpaydin,+E.+(2021).+Machine+learning.+MIT+Press.&ots=fH23O8Txmt&sig=Qlso79GxjmU5QKZPUWZbYSVkp4&redir_esc=y#v=onepage&q=Alpaydin%2C%20E.%20(2021).%20Machine%20learning.%20MIT%20Press.&f=false)