



**UNIVERSIDAD  
DEL AZUAY**

**Facultad de Ciencia y Tecnología  
Escuela de Ingeniería en Alimentos**

**“Desarrollo de un método analítico FTIR para determinar la  
composición de la mayonesa”**

**Trabajo de graduación previo a la obtención del título de:  
INGENIERA EN ALIMENTOS**

**Autora:**

**ANA SOFIA GUAYAS VIDAL**

**Director:**

**PIERCOSIMO TRIPALDI CAPPELLETTI**

**CUENCA - ECUADOR**

**2022**

## **DEDICATORIA**

Esta tesis va dedicada a mis padres Carlos Guayas y Ana Vidal por permitirme aprovechar cada uno de los sacrificios que han hecho para ser la persona que soy ahora, por permitirme seguir adelante y cumplir cada una de mis metas, por las enseñanzas durante toda esta etapa de mi vida y sobre todo por ser mi mayor fuente de inspiración; a mis hermanos Karla y David quienes con su apoyo incondicional y su alegría no me han dejado rendirme. Todos mis logros son dedicados a mi familia, quienes han estado siempre.

## **AGRADECIMIENTOS**

Quiero agradecer a Dios porque ha sido guía durante toda esta trayectoria de mi vida, a mis padres, porque sin su apoyo nada de esto hubiera sido posible, gracias por su amor incondicional, por enseñarme a ser fuerte y nunca darme por vencida, por inculcarme valores que me han servido a lo largo de mi vida y por ser mi mayor ejemplo a seguir.

A mis hermanos, por sus consejos, sus palabras de aliento y por estar siempre en cada momento de mi vida.

A cada uno de los docentes que fueron parte de este trabajo Dr. Piercósimo Tripaldi y Mgst. Andrés Pérez, por brindarme sus conocimientos y su sabiduría. Además, a todos los que hicieron parte de mi formación universitaria.

A mis compañeros con los que pude compartir experiencias y que día a día me enseñaban algo nuevo, gracias por los buenos y malos momentos. También, quiero agradecer a mi compañera Mónica González por su apoyo incondicional en este trabajo.

A toda mi familia que se ha hecho presente en cada logro demostrando su cariño sincero

Y a la Universidad del Azuay por ser mi segundo hogar.

## DESARROLLO DE UN MÉTODO ANALÍTICO FTIR PARA DETERMINAR LA COMPOSICIÓN DE LA MAYONESA

### RESUMEN

En el presente trabajo se calcularon diferentes modelos matemáticos que permitan determinar la composición de la mayonesa (proteínas, grasas y carbohidratos) mediante espectros FTIR. Para ello se realizó un diseño experimental que sirvió para desarrollar muestras patrón o unos modelos estándar, los cuales fueron medidos en el espectrofotómetro de infrarrojo. De acuerdo a los modelos aplicados se pudo determinar que mínimos cuadrados ordinales (OLS) con selección de variables de algoritmos genéticos, no es factible para la determinación de composición de la mayonesa. Los mejores resultados se obtuvieron del espectro original sin procesar. La selección de variables óptimas fue la de reemplazo secuencial renovado (RSR) utilizando el método de mínimos cuadrados parciales (PLS) para proteína y carbohidratos; y mínimos cuadrados ordinales (OLS) para grasa.

**Palabras clave:** Espectroscopia infrarroja FTIR, Mayonesa, Reemplazo Secuencial Renovado, Mínimos cuadrados ordinales, Mínimos cuadrados parciales. Diseño experimental.



---

Piercósimo Tripaldi Cappelletti

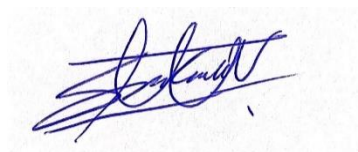
**Director de Tesis**



---

María Fernanda Rosales Medina

**Directora de escuela**



---

Ana Sofía Guayas Vidal

**Autora**

## DESARROLLO DE UN MÉTODO ANALÍTICO FTIR PARA DETERMINAR LA COMPOSICIÓN DE LA MAYONESA

### ABSTRACT

In this study, different mathematical models were calculated that led to determine the composition of mayonnaise, (proteins, fats, and carbs) by the utilization of FTIR spectrums. For this, an experimental design was used to develop pattern samples or standard models, which were measured with an infrared spectrophotometer. According to the applied models, it was determined the ordinal least squares (OLS) with a variable selection of genetic algorithms, was not applicable for the determination of the mayonnaise composition. The best results were obtained from the original-not processed spectrum. The selection of optimal variables was the reshaped sequential replacement (RSR) using the partial least squares method (PLS) for proteins and carbs and OLS for fats.

**Key words:** Infrared spectroscopy (FTIR), Mayonnaise, Reshaped Sequential Replacement, Ordinal Least Squares, Partial Least Squares, Experimental design.




---

Piercósimo Tripaldi Cappelletti

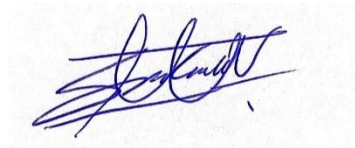
**THESIS DIRECTOR**




---

María Fernanda Rosales Medina

**FACULTY DIRECTOR**




---

Ana Sofía Guayas Vidal

**AUTHOR**

Translated by



Ana Sofía Guayas Vidal



## INDICE DE CONTENIDOS

DEDICATORIA .....	ii
AGRADECIMIENTOS .....	iii
RESUMEN.....	iv
ABSTRACT .....	v
INDICE DE CONTENIDOS.....	vi
ÍNDICE DE TABLAS .....	vii
ÍNDICE DE FIGURAS.....	viii
1. INTRODUCCIÓN .....	1
2. CAPITULO 1: MATERIALES Y METODOS .....	4
2.1. Espectroscopia infrarroja.....	4
2.1.1. Espectroscopia infrarroja por transformada de Fourier (FTIR) en alimentos.....	4
2.1.2. Infrarrojo cercano (NIR).....	5
2.2. Diseño experimental.....	5
2.2.1. Diseño experimental factorial.....	6
2.3. Métodos quimiométricos .....	7
2.3.1. Calibración multivariante para métodos cuantitativos .....	7
2.3.2. Modelos multivariantes de regresión .....	7
2.3.2.1. Mínimos cuadrados ordinarios (OLS) .....	7
2.3.2.2. Mínimos cuadrados parciales (PLS).....	8
2.3.3. Métodos de selección de variables.....	9
2.3.3.1. Algoritmos genéticos.....	9
2.3.3.2. Reemplazo Secuencial Renovado (RSR).....	10
2.4. Parámetros de calidad del modelo.....	10
2.4.1. Coeficiente de determinación ( <b>R<sup>2</sup></b> ).....	10
2.4.2. <b>R<sup>2</sup></b> Cross Validation.....	11
2.4.3. Coeficiente de correlación cuadrático predictivo <b>Q<sup>2</sup></b> .....	11
3. RESULTADOS Y DISCUSIÓN .....	11
4. CONCLUSIONES .....	26
5. REFERENCIAS.....	27

**ÍNDICE DE TABLAS**

Tabla 1. Diseño experimental factorial	13
Tabla 2. Composición por experimento	13
Tabla 3. Abreviaciones utilizadas para presentar los resultados	21
Tabla 4. Resultados de los modelos aplicados para proteínas	21
Tabla 5. Resultados de los modelos aplicados para grasa	22
Tabla 6. Resultados de los modelos aplicados para carbohidratos	22
Tabla 7. Mejores resultados obtenidos de proteína y carbohidratos para la posterior fusión de los datos	22
Tabla 8. Resultado de los datos fusionados para proteínas.	23
Tabla 9. Resultado de los datos fusionados para carbohidratos	23
Tabla 10. Mejores modelos para proteína, grasa y carbohidratos.	24

**ÍNDICE DE FIGURAS**

Figura 1. Esquema de trabajo	12
Figura 2. Corrección de la Línea de base del espectro infrarrojo de la muestra de mayonesa.	15
Figura 3. Corrección de la Línea de base, suavización del espectro infrarrojo de la mayonesa.	16
Figura 4. Corrección de la Línea de base, suavización y normalización de la escala del espectro infrarrojo de la mayonesa	17
Figura 5. Espectro de la mayonesa junto a su primera y segunda derivadas	18



## 1. INTRODUCCIÓN

Existen varios métodos de determinación de composición de las mayonesas, los cuales pueden ser muy costosos ya sea por los reactivos que se utilizan, de laboratorio como tal y de los distintos equipos utilizados en la determinación, incluso el proceso puede resultar demasiado largo y los resultados que se obtienen no siempre son precisos, esto puede deberse a diferentes factores como son: malas mediciones de los reactivos y un control inadecuado de los equipos. (Greenfield & Southgate, 2006).

Estos métodos son:

### **Determinación de la concentración total de proteínas**

Se utiliza el método Kjeldahl, este está dividido en tres pasos: digestión, neutralización y titulación. Es el más utilizado para la determinación de proteínas, se basa en que un alimento se conduce con un ácido fuerte para la liberación de nitrógeno el cual puede determinarse mediante una adecuada técnica de titulación. Seguido de eso, se calcula la cantidad de proteína presente a partir de la concentración de nitrógeno del alimento. Es por eso que el mismo enfoque básico se usa hoy en día, incluso se han ido implementando mejoras de manera que se acelere el proceso y se puedan obtener mediciones sin tanto margen de error en los resultados (McClements, 2003). Este método no da el contenido de proteína directamente, en ese caso, se necesita un factor de conversión denominado (F) el cual sirve para convertir la concentración de nitrógeno medida en una concentración de proteína. El valor de este factor de conversión (F) es de 6,25 (equivaliendo a 0,16g de nitrógeno por gramo de proteína), este solo es un valor promedio y cada proteína contiene un factor de conversión distinto según la composición de aminoácidos. (AOAC, 2005).

### **Determinación del contenido total de grasa**

La hidrólisis ácida se utiliza para muestras complejas, es decir, muestras que contienen proteínas, carbohidratos, etc. Consiste en liberar los lípidos ocluidos o unidos de manera que la grasa quede expuesta para su posterior análisis.

### **Método de soxleth**

Para este método se utiliza éter de petróleo, este será el reactivo que extraerá el extracto etéreo de la muestra. Se utilizan 2 g de muestra, seguido de eso se colocan en un papel filtro bien cerrado de manera que no se pierda alimento en el proceso. Se coloca el papel dentro de un cartucho de extracción, y se ponen ambos dentro del equipo. Debajo de la muestra tiene que ir un matraz aforado con el éter de petróleo y una vez realizado eso se pone a trabajar el equipo entre 4 y 8 h. Luego de que ha transcurrido entre este tiempo siempre controlado que sea el necesario, se evaporará el éter, quedando así la grasa del alimento en el matraz. La cantidad de grasa se obtiene mediante una diferencia de peso del matraz antes y después del proceso (AOAC, 2005).

### **Determinación del contenido total de carbohidratos**

Una vez que se miden todos los demás componentes de la muestra se realiza el cálculo del contenido de carbohidratos mediante la siguiente fórmula:

$$\% \text{ Carbohidratos} = 100 - \% \text{ humedad} - \% \text{ proteína} - \% \text{ lípidos} - \% \text{ minerales}$$

Este método puede tener resultados erróneos debido a que al realizar la cuantificación de los demás componentes siempre existe un margen de error. (Salvatierra, 2015)

Estos procesos suelen ser largos y costosos, es por eso que se pretende hacer el uso del espectrofotómetro de infrarrojo, de manera que pueda mejorar estas limitantes que existen en los métodos estándares.

La mayonesa es una salsa muy utilizada tanto a nivel industrial como a nivel casero, debido a su facilidad de obtención y a sus grandes cantidades de producción, por lo que se vuelve fundamental tener un control sobre estas y sobre los ingredientes que se utilizan en la elaboración de sus productos. Los componentes que se encuentran en mayor cantidad en los alimentos, son: agua, proteínas, grasas y carbohidratos. Siendo la composición y la calidad de la mayonesa, grandes influyentes en el consumo de la misma, debido a las actuales tendencias de consumo. Lo que se busca es poder analizarlos de una manera mucho más rápida y sin tanto costo, obteniendo además resultados precisos que permitan tomar decisiones factibles. Siendo así la espectrometría de infrarrojo (FTIR) una gran alternativa para lograr estos objetivos (Pérez, 2013).

La espectroscopia de infrarroja se ha aplicado a varios alimentos con la finalidad de cumplir diferentes objetivos como han sido la cuantificación de pectinas, proteínas,

grasas, carbohidratos, etc.; control de calidad, analizar cambios en estructura molecular de algún alimento e incluso es un método aplicable para bebidas (Flores, 2017; Mejía, 2010; Rosas & Fernández, 2012).

Al ser la espectroscopia de infrarrojo una buena alternativa para mejorar los procesos de determinación de la composición en mayonesa, se establecen algunos objetivos:

- Aplicar un diseño experimental para desarrollar muestras patrón y medir en el espectrofotómetro de infrarrojo.
- Utilizar técnicas quimiométricas para relacionar la composición de las muestras del diseño experimental con su espectro FTIR.
- Determinar si es un proceso aplicable a la determinación de la composición de mayonesas.

Para ello se calculan diferentes modelos matemáticos, los cuales permiten determinar la composición de la mayonesa mediante espectroscopia FTIR.

## 2. CAPITULO 1: MATERIALES Y METODOS

### 2.1. Espectroscopia infrarroja

La espectroscopia IR es una técnica que se basa fundamentalmente en las vibraciones de los átomos de una molécula, cuando esta pasa por una radiación infrarroja mediante la muestra que está siendo analizada. Los picos en un espectro de absorción aparecen a una longitud de onda determinada y esto corresponde a una frecuencia de vibración de una parte de la molécula de muestra. Se transmite una luz infrarroja a través de la muestra y queda registrada la cantidad de energía absorbida lo que permite en un rango de longitudes de onda registrar estas ondas en unas gráficas, las cuales posteriormente pueden ser analizadas. Aquí se pueden detectar algunas vibraciones de las moléculas, incluidos algunos grupos funcionales como C-H, O-H, N-H, C=O, etc (Dufour, 2009).

#### 2.1.1. Espectroscopia infrarroja por transformada de Fourier (FTIR) en alimentos

La espectrometría FTIR es un método analítico rápido y sobre todo accesible, el cual nos brinda una gran cantidad de información tanto cualitativa como cuantitativa acerca de la composición de la muestra que se esté analizando (Torres, 2016). El equipo donde se lleva a cabo este proceso se lo conoce como “Espectrómetro de infrarrojo” en el cual se pueden colocar muestras tanto solidas como liquidas, garantizando procesos estables, bajo control y con especificaciones de los productos finales (Mondragón, 2020).

Tradicionalmente, el contenido de proteínas, grasas y carbohidratos en distintos alimentos se estima de acuerdo a procedimientos volumétricos, mediciones del índice de refracción, entre otras. Las cuales pueden tardar mucho tiempo, suelen ser costosas y sus resultados no siempre suelen ser precisos (Greenfield & Southgate, 2006). Por lo tanto, la espectroscopia FTIR se ha vuelto un método alternativo para poder analizar estos componentes en muestras de diferentes alimentos, siendo mucho menos costosa o rentable y más rápida que los métodos tradicionales. La espectroscopia infrarroja FTIR se ha utilizado cada vez más, junto con la quimiometría, para así estudiar una variedad de muestras de alimentos y, en particular, para estudiar alimentos líquidos como bebidas. (Leopold et al., 2011).

Dentro de la espectroscopía infrarroja la zona espectral del infrarrojo medio (MIR) ha despertado interés debido a su gran capacidad de determinar la estructura de las moléculas (Lozano, 2013). Esta región se extiende desde  $4000\text{ cm}^{-1}$  hasta  $400\text{ cm}^{-1}$ , proporcionando información química, estructural y de composición cuantitativa sobre las moléculas constituyentes en las fases gaseosa, líquida y sólida. Cuando la materia es atravesada por la radiación electromagnética, los enlaces de las moléculas realizan movimientos (vibración y rotación), los cuales implican una absorción evidente de la energía proporcionada. Sobre la base de la energía suministrada y la cantidad absorbida por la muestra irradiada, y usando tratamientos matemáticos de espectro, es posible determinar la composición química de la muestra y los compuestos correlacionados (de Marchi et al., 2014).

### **2.1.2. Infrarrojo cercano (NIR)**

Esta región NIR se basa en la absorción, transmisión o reflexión de la luz sobre una muestra, que en función de su naturaleza absorberá más o menos cantidad de energía, esta zona cubre la región entre 780 y 2500nm, y basado en la absorción, transmisión o reflexión de la luz, esta región ofrece una técnica rápida y no destructiva para la determinación de la composición de varios componentes en una muestra. Los espectros NIR de alimentos comprenden bandas anchas correspondientes a sobretonos y combinaciones de modos vibracionales que involucran enlaces químicos C–H, O–H y N–H, dándonos así una cantidad de información que, con la aplicación de la quimiometría, puede resultar muy factible para la clasificación y cuantificación de varios parámetros (Armenta et al., 2010; Valenciaga & Simoes, 2006).

### **2.2. Diseño experimental**

Un diseño experimental es la aplicación y planificación de experimentos, de manera que se espera obtener resultados óptimos, con un mínimo tiempo y con mínimo costo global (Hidalgo et al., 2018). Además, es importante tener en cuenta que, la calidad de la información que se obtiene depende directamente de la ejecución de todos estos experimentos, es decir, de nuestro diseño experimental. Este consiste en analizar que pruebas se pueden hacer y la manera en las que se las puede ejecutar, de manera que se puedan obtener datos que al ser analizados estadísticamente brinden evidencia objetiva que nos permitan responder las interrogantes que se han planteado previamente, incluso se puede manipular para que este nos brinde información que se requiere para su mejoría (Gutiérrez & Salazar, 2008).

Los diseños experimentales se han vuelto un pilar fundamental para garantizar estudios exitosos, debido a que se puede tener un control en la variabilidad de los resultados que se obtengan de la experimentación. Para realizar un diseño experimental es importante analizar todas las fuentes que puedan generar variabilidad en los datos, es decir, nuestras variables tanto dependientes como independientes, de manera que se puedan obtener conclusiones válidas con su análisis estadístico previo (Reyes, 2009).

De acuerdo a la naturaleza de los datos se hace una elección del diseño experimental más adecuado, teniendo en cuenta todas las características que pueden influir en el resultado del experimento. En este caso el diseño experimental se aplicó como base del estudio con el fin tener una generación de datos ordenada que nos permita desarrollar muestras patrón o muestras estándares para su posterior proceso.

### **2.2.1. Diseño experimental factorial**

Se representan como diseño factorial  $2^k$  los diseños en donde cada uno de sus factores que son denominados “k” pueden adoptar únicamente dos niveles que pueden ser tanto cualitativos como cuantitativos. Se pueden aplicar estos modelos ya sea con replicas o sin ellas, de manera tal que se puede obtener únicamente una respuesta o más de una respuesta para cada una de las combinaciones, además no solo es aplicable en casos de optimización, sino que también es importante para espectros FTIR. (Fernández, 2020; Medina & López, 2011)

El objetivo principal de este diseño es probar combinaciones de los niveles de los factores en el que el desempeño del proceso sea mejor y de acuerdo a esto, observar cómo se comporta la respuesta en función de estas diferentes combinaciones (Gutiérrez & Salazar, 2008). Todas estas combinaciones posibles son realizadas a través de una previa realización de la matriz de diseño, en donde se utilizan los signos – y + para ambos niveles de un factor (Fernández, 2020).

Según (Pérez & López, 2017) a menudo existen variables experimentales que se cree que tienen influencia en el resultado experimental y se quiere saber cómo logran esto, para ello será necesario determinar la influencia directa de las variables experimentales, así como de sus posibles efectos de interacción, por ende, nos permitirán reducir la cantidad total de experimentos, de manera que se ahorre así tiempos, costos y esfuerzo.

### **2.3. Métodos quimiométricos**

La quimiometría es muy utilizada para realizar procesos de selección de variables, para diseñar métodos analíticos y para procesamiento e interpretación de datos que se han medido. Para ello, la quimiometría usa métodos matemáticos y estadísticos que permiten diseñar o seleccionar experimentos o procesos de medición óptimos y además proporcionar la mayor cantidad de información química a través del análisis de los datos químicos (Vandeginste, 1987).

#### **2.3.1. Calibración multivariante para métodos cuantitativos**

La calibración multivariante tiene como objetivo o se la utiliza con el fin de relacionar dos tipos de medidas sobre una muestra: una medida rápida, además de ser económica y de fácil obtención como puede ser un espectro infrarrojo, y otra medida que puede requerir mucho tiempo para obtener resultados pudiendo esta otra medición ser más complicada y cara, como puede ser la determinación de proteína o grasas de un alimento específico (Ferré, 2006). Aquí se utiliza más de una longitud de onda, permitiendo así la corrección de ruidos espectrales y algunos efectos de la matriz como pueden ser: las interacciones químicas, las cuales dan paso a daños o variaciones en los espectros no lineales con relación a la concentración (Kalivas, 2009).

Los modelos de calibración multivariante son muy útiles para análisis cuantitativo utilizando técnicas espectroscópicas debido a que medir un espectro es una manera rápida de que se pueda crear muchas medidas, además de que puede resultar un tanto simple. Incluso la calibración multivariante sirve para poder cuantificar a partir de ellas y restablecer la velocidad de análisis (Dufour, 2009).

#### **2.3.2. Modelos multivariantes de regresión**

##### **2.3.2.1. Mínimos cuadrados ordinarios (OLS)**

Este es el método de regresión más simple y multiparamétrica, del cual se obtiene un modelo del tipo lineal. El problema de regresión de mínimos cuadrados ordinarios se representa a través del siguiente modelo:

$$y = Xb + e$$

Donde:

$b$ : vector de coeficientes a estimar

Para determinar el vector de los coeficientes  $b$  de la fórmula, se utiliza la siguiente fórmula:

$$b = (X^T X)^{-1} X^T y$$

Una vez que se determina los coeficientes del modelo, se puede determinar el vector de respuesta (Ildikó & Todeschini, 1994)

$$\hat{y} = Xb$$

### 2.3.2.2. Mínimos cuadrados parciales (PLS)

Se desarrolla para evitar el efecto de la multicolinealidad en las estimaciones de los parámetros de una regresión. El problema de la multicolinealidad, o relación de dependencia extrema entre variables explicativas en un modelo de regresión, produce situaciones de inestabilidad de los coeficientes de regresión. Este método trabaja con la selección de variables, permitiéndonos reducir el número de variables y por lo tanto la reducción de la dimensión de cálculo, siempre y cuando esto sea lo más óptimo para la respuesta. Existen casos en los que hay una gran cantidad de variables explicativas y pocas variables latentes que recogen la mayor variabilidad de la variable respuesta. El objetivo de este método de regresión es extraer estas variables latentes, recogiendo a su vez la mayor variación de las variables explicativas ya que así funcionan para modelar la variable respuesta de la mejor manera (Gaviria et al., 2016).

La regresión PLS transforma la matriz de variables explicativas  $X$  de orden  $n \times p$ , teniendo en cuenta el vector respuesta  $Y$  de orden  $n \times 1$ , en una matriz de variables latentes  $T$  y  $U$ . La técnica se basa en extraer sucesivamente factores de ambas matrices  $X$  e  $Y$  de manera que la covarianza entre dichos factores es máxima (Gaviria, 2015).

Donde:

$$t = Xw \quad y \quad u = Yc$$

$$\max t^T u = (Xw)^T (Yc) = w^T X^T Yc$$

$$t = Xw = 1$$



$$u=Yc=1$$

Este método relaciona la matriz de las variables de entrada X y la matriz de las variables de salida Y a través de las variables latentes de cada una (Rodríguez, 2015).

$$X = TP^T + E$$

$$Y = UC^T + F$$

Donde:

T y U son las matrices compuestas por los vectores latentes.

P y C son las matrices compuestas por los vectores de peso de las respuestas.

E y F son las matrices de error

### **2.3.3. Métodos de selección de variables**

Los métodos de selección de variables son una serie de métodos que se aplican con la finalidad de poder conseguir la mayor cantidad de información posible, de un sistema integrado por muchas variables, usando pocas variables del sistema, es decir, seleccionando un subconjunto de variables más relevantes para la respuesta, estos métodos se pueden aplicar a casos de regresión y clasificación (Gallegos, 2021).

#### **2.3.3.1. Algoritmos genéticos**

Los algoritmos genéticos son procesos inspirados en la teoría de la evolución del hombre, es decir, los individuos evolucionan de acuerdo a la selección y supervivencia de los que más se adapten. Para poder imitar estos principios, los algoritmos genéticos se basan en una población, donde cada individuo representara una solución candidata para un problema de optimización. Las mejores soluciones tendrán más oportunidad de ser seleccionadas y reproducirse, esta solución se representa por un “cromosoma”, luego se seleccionan de manera probabilística los cromosomas para la siguiente generación de acuerdo con sus valores de aptitud (Cuevas et al., 2021)

Dentro de los algoritmos genéticos se tienen algunos pasos: codificación de variables, creación de la población de partida, evaluación de la respuesta, cruce y mutación. En

algunos casos los cromosomas seleccionados producen descendencia debido a que se aparean de manera aleatoria, cuando sucede esta descendencia, el cruce y la mutación también se darán de manera aleatoria. Esto sucede ya que los cromosomas con valores altos de aptitudes tienen grandes posibilidades de ser seleccionados. El proceso de evolución se repetirá las veces necesarias hasta que se satisfagan algunas condiciones finales que son: llegar a la mejor respuesta posible y cumplir un número determinado de generaciones (Kumar et al., 2010; Niazi & Leardi, 2012).

### 2.3.3.2. Reemplazo Secuencial Renovado (RSR)

Este método de reemplazo secuencial SR fue propuesto por Miller, donde la idea principal fue que, a partir de un modelo generado de manera aleatoria (semilla) ver si se puede obtener un mejor modelo. De esta manera, el mejor modelo que se encuentre se convertirá en la nueva semilla para de esta encontrar mejores modelos, este procedimiento se continua hasta que no se encuentra un mejor modelo (Grisoni et al., 2014)

A este modelo se le han implementado algunas funciones, las cuales son:

- Disminuir el tiempo de cálculo
- Identificar modelos con diferentes patologías como pueden ser: sobreajuste, correlación casual, redundancia de variables y colinealidad de predictores
- Aumento de la probabilidad de convergencia sobre el modelo

## 2.4. Parámetros de calidad del modelo

### 2.4.1. Coeficiente de determinación ( $R^2$ )

El coeficiente de determinación  $R^2$  o también conocido como estadística R-cuadrado en los modelos de regresión lineal, son una medida de que tan bien explica el modelo de regresión los datos observados. Esto se expresa por la ecuación de regresión, como la proporción de varianza en la variable dependiente.

$$R^2 = \frac{\text{variación explicada}}{\text{variación total}}$$

Donde:

$R^2$  nos indica la proporción de la variación total en los n valores observados de la variable dependiente que explica el modelo de regresión lineal simple (Vilca, 2011).

Cuando este coeficiente de determinación  $R^2$  es +1, se observa una relación lineal perfecta entre  $x$  e  $y$ , esto quiere decir que el 100 % del cambio en  $y$  se explica por el cambio en  $x$ . Por otro lado, cuando este coeficiente es mayor a 0 y menor a 1, tenemos una relación lineal más débil entre  $x$  e  $y$ , lo que significa que no toda la variación de  $y$  se explica por la variación de  $x$  (Palma, 2022)

#### **2.4.2. $R^2$ Cross Validation**

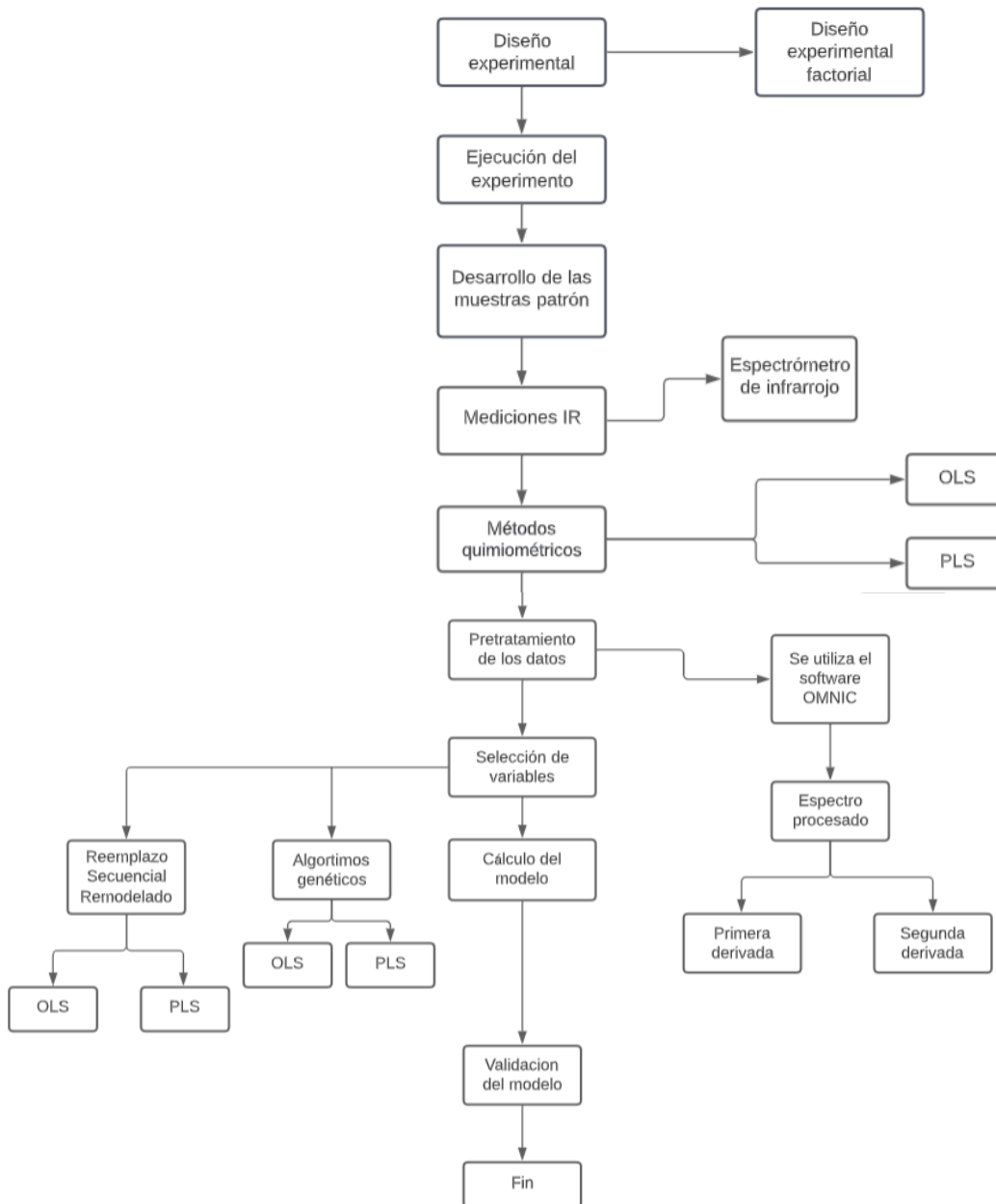
La validación cruzada se refiere a un grupo de métodos que nos permitan medir el rendimiento de un modelo predictivo dado en nuevos grupos de datos considerados como prueba. Básicamente esta consiste en dividir los datos en dos grupos:

El grupo de entrenamiento, utilizado para construir el modelo; y el conjunto de validación del modelo como tal estimando el error de predicción. Además, a esta también se la conoce como método de remuestreo, esto debido a que, implica ajustar el mismo método estadístico algunas veces usando distintos subgrupos de datos (Kassambara, 2018).

#### **2.4.3. Coeficiente de correlación cuadrático predictivo $Q^2$**

Es un indicador de la capacidad predictiva del modelo, cuando este es =1 nos indica una predictividad prácticamente perfecta. Si bien es cierto los parámetros estadísticos deben ser lo más cercano a 1 posible (maximizado) en el mejor modelo. Sin embargo, el valor de  $Q^2$  normalmente aumenta solo cuando los predictores agregados son útiles para predecir los compuestos omitidos (Gramatica, 2007).

### **3. RESULTADOS Y DISCUSIÓN**



**Figura 1.** Esquema de trabajo

**Condiciones del diseño experimental**

Para el diseño experimental se utilizó un diseño factorial 2k de 4 factores, donde se obtuvo un diseño en variables codificadas, debido a que se trabajan con diferentes variables se estandarizan las escalas de variación de todos los componentes es por eso que se utiliza +1 para el nivel más alto y -1 para el nivel bajo, como se puede observar

en la tabla 1. Haciendo esto se transforman los valores originales y se logra que todas las variables que tengan la misma influencia, sin importar la calidad de estas tal como se muestra en la tabla 2. Seguido de esto, se ejecutó el diseño experimental obteniendo así las muestras patrón, para así medirlas en el espectrómetro de infrarrojo.

Para obtener los datos de la tabla 2, se realizó una muestra de mayonesa previa a la elaboración del diseño experimental, donde se utilizaron diferentes cantidades de los ingredientes hasta obtener una mayonesa con buena consistencia, de manera que sirvió como un patrón para obtener así poder realizar el diseño factorial. Los ingredientes utilizados fueron aceite girasol, huevo moreno de gallina, vinagre blanco y goma xanthan.

**Tabla 1.** Diseño experimental factorial

<b>Experimento</b>	<b>Huevo</b>	<b>Aceite</b>	<b>Goma</b>	<b>Vinagre</b>
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	-1	1	-1	-1
4	1	1	-1	-1
5	-1	-1	1	-1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	-1
9	-1	-1	-1	1
10	1	-1	-1	1
11	-1	1	-1	1
12	1	1	-1	1
13	-1	-1	1	1
14	1	-1	1	1
15	-1	1	1	1
16	1	1	1	1
17	0	0	0	0
18	0	0	0	0
19	0	0	0	0

**Tabla 2.** Composición por experimento

Experimento	Huevo		Vinagre (ml)	Goma Xantana (gr)	Aceite (gr)
	Clara (gr)	Yema (gr)			
1	29.7	19.8	27	0.27	126.9
2	36.3	24.2	27	0.27	126.9
3	29.7	19.8	27	0.27	155.1
4	36.3	24.2	27	0.27	155.1
5	29.7	19.8	27	0.33	126.9
6	36.3	24.2	27	0.33	126.9
7	29.7	19.8	27	0.33	155.1
8	36.3	24.2	27	0.33	155.1
9	29.7	19.8	33	0.27	126.9
10	36.3	24.2	33	0.27	126.9
11	29.7	19.8	33	0.27	155.1
12	36.3	24.2	33	0.27	155.1
13	29.7	19.8	33	0.33	126.9
14	36.3	24.2	33	0.33	126.9
15	29.7	19.8	33	0.33	155.1
16	36.3	24.2	33	0.33	155.1
17	33	22	30	0.3	141
18	33	22	30	0.3	141
19	33	22	30	0.3	141

## Mediciones IR

### -Espectrometría de infrarroja por transformada de Fourier

Los espectros infrarrojos de las muestras de mayonesa, se obtuvieron en el espectrofotómetro infrarrojo FTIR de la marca Thermo Scientific, modelo Nicolet Summit Pro, con una resolución espectral de 0,45 cm<sup>-1</sup> para ayudar a distinguir varios componentes en pequeñas cantidades (Thermo Fisher Scientific, 2021). Las muestras se realizaron a partir del diseño experimental factorial, dándonos así un total de 19 experimentos como se puede ver en la tabla 1, los cuales se midieron por triplicado de manera que se pueda realizar un posterior promedio de los mismos.

## Tratamiento de los datos

### -Pretratamientos

La lectura de los espectros obtenidos se llevó a cabo en un software Thermo Scientific OMNIC Paradigm (Thermo Electron Corporation). Donde se utilizaron los siguientes

menús: Automatic Baseline Correct, Automatic Smooth, Normalize Scale y Save. Seguidamente, se realizó la primera y segunda derivada de los espectros ya procesados. Una vez que se obtuvieron los datos se hizo un promedio de estos, debido a que las mediciones de los espectros fueron por triplicado.

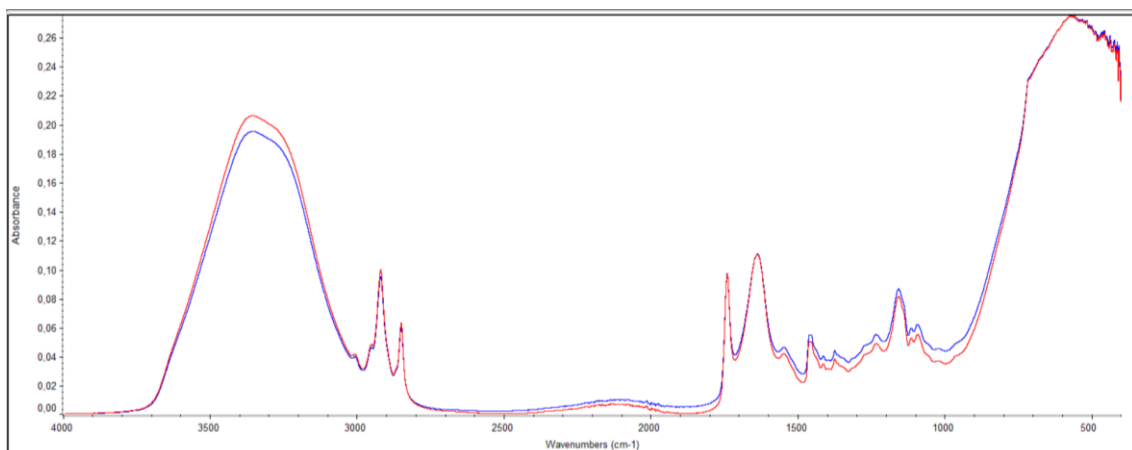
Después, se analizaron los rangos tanto de proteínas como de grasas y carbohidratos, debido a que la mayor cantidad de información no siempre se encuentra en los mismos datos para todos los componentes. Los rangos utilizados fueron: 1500-1700  $\text{cm}^{-1}$  para las proteínas; 2800-3000  $\text{cm}^{-1}$  para los carbohidratos y 900-1900  $\text{cm}^{-1}$  para las grasas.

**Automatic Baseline Correct:** Este menú se aplica para corregir de manera automática las líneas de base que están inclinadas o curvas (Thermo Fisher Scientific, 2021).

Consiste en analizar un ajuste cuadrático del espectro  $Y(x)$ , usando mínimos cuadrados. (Thermo Electron Corporation)

$$Y(x) \sim Y'(x) = ax^2 + bx + c$$

En la Figura 2, se muestra el espectro de la mayonesa, con su respectiva corrección de línea de base. La gráfica roja corresponde al espectro original, y la gráfica azul al espectro corregido su línea de base.



**Figura 2.** Corrección de la Línea de base del espectro infrarrojo de la muestra de mayonesa.

**Automatic Smooth:** Se usa para mejorar la apariencia de los espectros, generalmente para los picos que se generan por el ruido. (Thermo Electron Corporation).

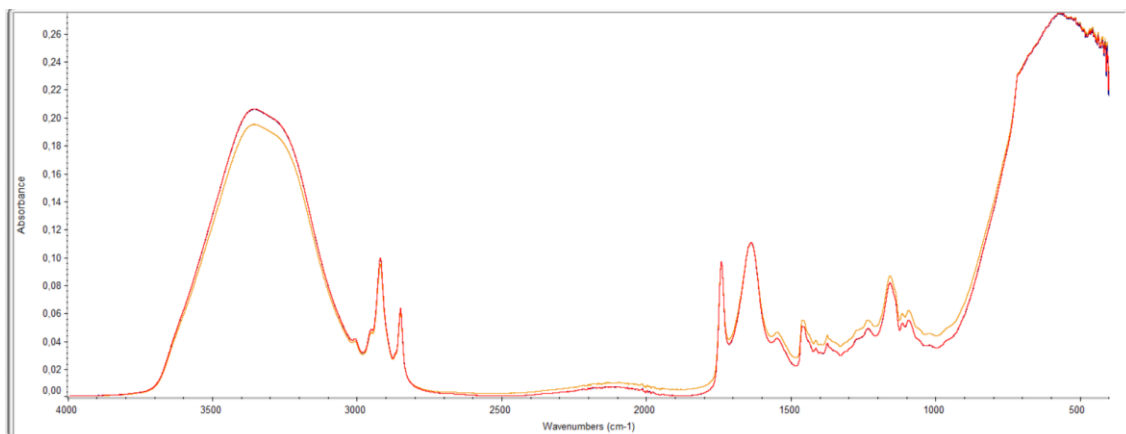
Para este tratamiento, se utiliza el método de Savitzky – Golay, donde se suavizan las señales, calculando así una regresión polinomial de grado n, usando n+1 puntos en el espectro respectivo. Dando como resultado nuevos puntos (Vélez, 2022).

$$SF = \frac{K}{\frac{\%T * 2.302585}{100}}$$

Donde:

K: frecuencia de la ecuación

En la figura 3, se muestra un espectro al cual, se ha corregido su línea de base y se ha suavizado. La grafica roja corresponde al espectro de la muestra de mayonesa y al suavizado de la señal, y la gráfica amarilla a la corrección de línea de base.



**Figura 3.** Corrección de la Línea de base, suavización del espectro infrarrojo de la mayonesa.

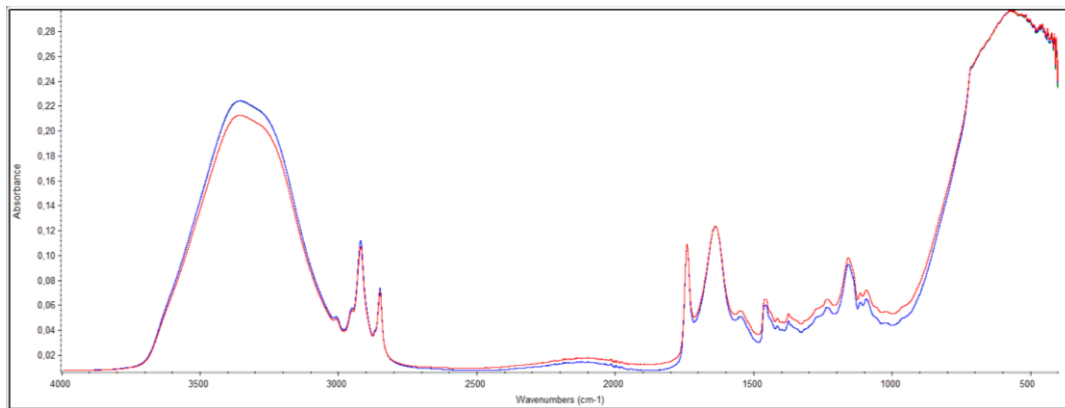
**Normalize scale:** Se utiliza para cambiar la escala del eje Y de los espectros seleccionados a una escala "normal" en la que los valores Y de los puntos de datos van desde 0 unidades de absorbancia para el punto más bajo hasta 1 unidad de absorbancia para el pico más alto (Thermo Electron Corporation).

Aquí se enlaza la absorbancia a una determinada longitud de onda, los espectros se encuentran en un rango de 0-1 siendo, 0 el punto más bajo y 1 el punto más alto. (Vélez, 2022).

$$\text{Normalizado} = \frac{(Abs - AbsMin)}{AbsMax - AbsMin}$$



En la figura 4, se observa un espectro al cual se han aplicado, corrección de la línea de base, suavización y normalización de la escala. La grafica roja corresponde al espectro original y la gráfica azul corresponde al espectro ya tratado



**Figura 4.** Corrección de la Línea de base, suavización y normalización de la escala del espectro infrarrojo de la mayonesa

**Primera derivada:** Este es útil para revelar picos que aparecen como hombros en los espectros originales. Es decir, los picos se vuelven más estrechos y son más fáciles de ver. Se obtiene derivando la absorbancia en relación al número de onda

$$D_i = \frac{Y_{i-1} - Y_{i+1}}{X_{i-1} - X_{i+1}}$$

Donde:

$D_i$ : derivada en cada punto  $i$

$Y_i$ : es la absorbancia en cada punto  $i$

$X_i$ : es la ubicación de cada punto  $i$ . (Thermo Electron Corporation).

**Segunda derivada:** Esta nos sirve para encontrar la ubicación del centro de los hombros en los espectros originales. Esta curva es más compleja que la primera derivada, con bandas significativamente más estrechas. Se obtiene aplicando la siguiente ecuación

$$D'_i = \frac{Y_{i-2} - 2 * Y_i + Y_{i+2}}{(X_{i-1} - X_{i+1})^2}$$

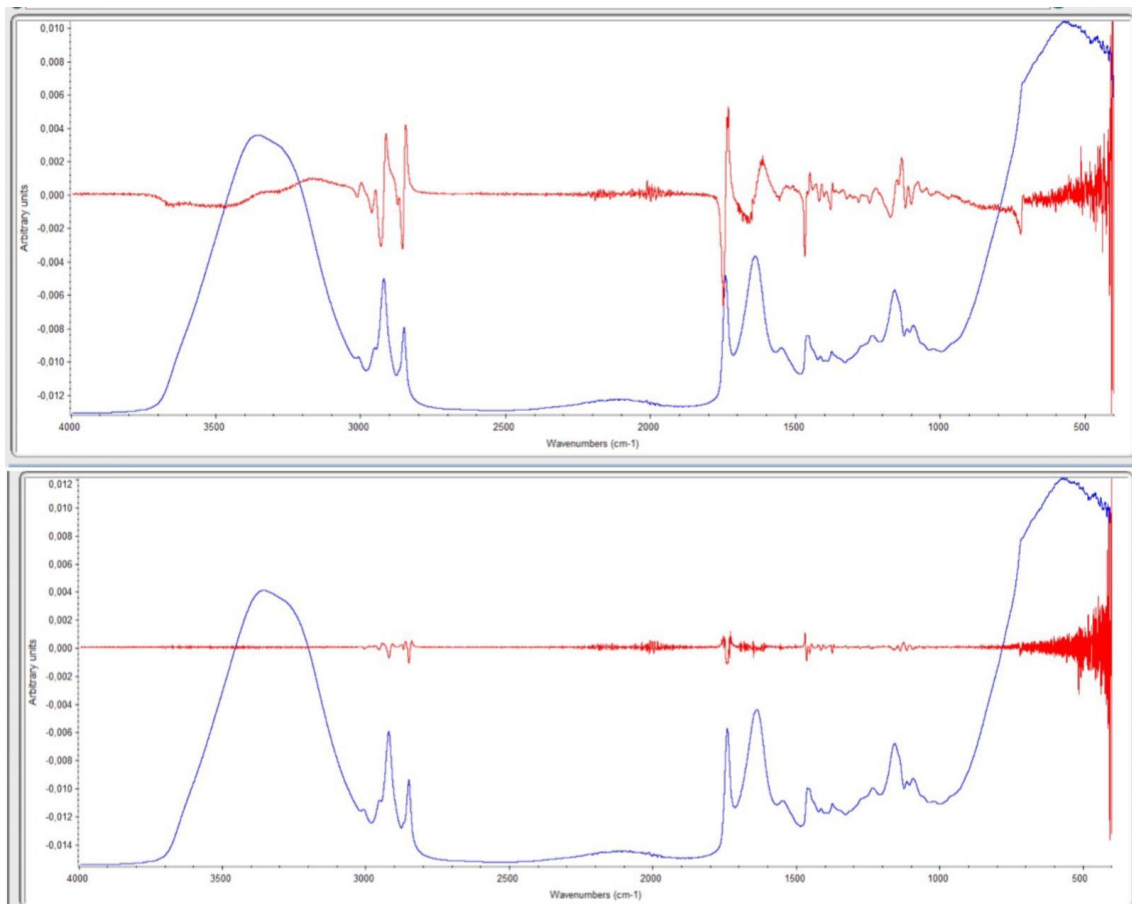
Donde:

$D'_i$ : segunda derivada en cada punto  $i$

$Y_i$ : es la absorbancia en cada punto  $i$

$X_i$ : es la ubicación de cada punto  $i$ . (Thermo Electron Corporation)

En la figura 5, se presenta el espectro de una muestra de mayonesa, junto con su primera y segunda derivada.



**Figura 5.** Espectro de la mayonesa junto a su primera y segunda derivada.

### **Selección de variables**

Para realizar la selección de variables se utilizó algoritmos genéticos y reemplazo secuencial renovado, donde se hace una selección de los predictores más relevantes para la validación del modelo. Los parámetros utilizados tanto para algoritmos genéticos como para reemplazo secuencial renovado fueron OLS y PLS.

### **Partición de datos en training y test**

Para realizar esta partición de datos, se dividió la base de datos de manera aleatoria. La primera parte representa el 70% de los datos, los cuales serían utilizados como training set, en donde se desarrolla la selección de variables y los modelos. El 30% se utilizó como test set, en la etapa de desarrollo de modelos, para validar, los modelos resultantes del training set.

### **Cálculo de modelo**

Se realizaron los cálculos mediante el software “MATLAB” desarrollado por MathWorks en donde se utilizaron los siguientes toolbox: ga-toolbox 4.4, ga-model.m, regression toolbox 1.3 y Reshaped Sequential Replacement (RSR) desarrollado por Milano Chemometrics Research Group de la Universidad de la Bicocca de Milan Italia. Se fueron registrando los datos de los diferentes componentes de la muestra y finalmente se calculó el modelo por OLS y PLS para ambos casos de selección de variables.

### **Validación del modelo**

Para validar los modelos de regresión se evalúa su capacidad predictiva, utilizando un punto de referencia:

TSS:  $\sum(y - \bar{y})^2$  Suma total de cuadrados

Este valor se refiere al valor medio de la respuesta  $y$ . Los modelos de regresión son mejores cuando la suma residual de cuadrados (RSS) es menor.

RSS:  $\sum(y - \hat{y})^2$  Suma residual de cuadrados

#### **-Coeficiente de determinación $R^2$**

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$$

#### **-Coeficiente cuadrático predictivo $Q^2$**

$$Q^2 = 1 - \frac{PRESS}{TSS}$$

PRESS:  $\sum(y - \hat{y})^2$  Suma de cuadrados del error residual predictivo

$$Q^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$$

Si bien es cierto, los cálculos para estos parámetros son casi idénticos. La diferencia está en que RSS se calcula a través de los datos en los que el algoritmo se entrena, mientras que PRESS se calcula a través de los datos retenidos.

Según los valores externos utilizados en el ajuste del modelo son independientes entre sí. Lo que significa que el valor de  $Q^2$  derivado del conjunto de datos externos, es decir,  $Q_{EXT}^2$  y el promedio de los valores de  $Q^2$  que se obtengan, tomando de manera independiente cada dato externo deben coincidir (Consonni et al., 2009).

$$Q_{EXT}^2 = \frac{\sum_{i=1}^{n_{EXT}} Q_i^2}{n_{EXT}}$$

Donde:

$Q_i^2$  es el  $Q_{EXT}^2$  calculado, tomando en cuenta únicamente el  $i$ -ésimo valor del conjunto de valores de prueba.

$n_{EXT}$ : número total de datos externos

En las tablas 4,5 y 6, se puede observar los resultados obtenidos de cada uno de los modelos aplicados para cada componente de la mayonesa. Los resultados se presentan de manera que, D0 indica los resultados en el espectro procesado sin derivar, D1 indica los resultados de los modelos aplicados en la primera derivada y D2 los resultados de los modelos aplicados en la segunda derivada.

En la tabla 7 se presentan los mejores resultados que se obtuvieron tanto para proteínas como para carbohidratos, con el fin de usarlos para realizar la fusión. En el caso de las grasas no se pudo mejorar el modelo debido a que los resultados fueron bajos.

En la tabla 8 y 9 se encuentran los datos fusionados para proteína y carbohidratos respectivamente, esto con la finalidad de analizar si el modelo mejora o no. La fusión se realizó con los datos de los espectros procesados sin derivar, tanto para proteínas como para carbohidratos. En el caso de la grasa, el modelo no se puede mejorar.

En la tabla 10, están los mejores resultados obtenidos para cada uno de los componentes de la mayonesa, cada uno con su respectivo modelo aplicado.

En la tabla 3, tenemos todas las abreviaciones que se utilizan en las tablas para la presentación de los resultados respectivos.

**Tabla 3.** Abreviaciones utilizadas para presentar los resultados

Abreviación	Significado
D0	Espectro procesado sin derivar
D1	Primera Derivada
D2	Segunda Derivada
GA	Algoritmos Genéticos
RSR	Reemplazo Secuencial Renovado
OLS	Mínimos Cuadrados Ordinales
PLS	Mínimos Cuadrados Parciales
$Q^2$	Coefficiente de Correlación Cross Validado "Leave one out" externo
$R^2$	Coefficiente de determinación

**Tabla 4.** Resultados de los modelos aplicados para proteínas

Pretratamiento	D0				D1				D2			
Selección de variables	GA	GA	RSR	RSR	GA	GA	RSR	RSR	GA	GA	RSR	RSR
Método de regresión	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS
$R^2$	No hay modelo	0.745	0.994	0.997	No hay modelo	0.197	0.697	0.786	No hay modelo	No hay modelo	0.682	0.652
$Q^2$		0.592	0.603	0.426		0.058	-4.121	-5.938			0.591	0.618

**Tabla 5.** Resultados de los modelos aplicados para grasa

<b>Pretratamiento</b>	<b>D0</b>				<b>D1</b>				<b>D2</b>			
<b>Selección de variables</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>
<b>Método de regresión</b>	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS
$R^2$	No hay modelo	0.272	0.996	1	No hay modelo	No hay modelo	1	0.988	No hay modelo	No hay modelo	0.899	0.942
$Q^2$		0.075	-0.678	0.534			-499955.8	-11.2			-0.683	-1.063

**Tabla 6.** Resultados de los modelos aplicados para carbohidratos

<b>Pretratamiento</b>	<b>D0</b>				<b>D1</b>				<b>D2</b>			
<b>Selección de variables</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>
<b>Método de regresión</b>	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS	OLS	PLS
$R^2$	No hay modelo	0.281	0.987	0.989	No hay modelo	No hay modelo	0.966	0.975	No hay modelo	No hay modelo	0.669	0.502
$Q^2$		0.595	0.878	0.574			0.129	0.495			-0.044	-0.844

**Tabla 7.** Mejores resultados obtenidos de proteína y carbohidratos para la posterior fusión de los datos

Componente	Proteína			Carbohidratos		
Pretratamiento	D0			D0		
Selección de variables	GA	RSR	RSR	GA	RSR	RSR
Método de regresión	PLS	OLS	PLS	PLS	OLS	PLS
$R^2$	0.745	0.994	0.997	0.281	0.987	0.989
$Q^2$	0.592	0.603	0.426	0.595	0.878	0.574

*Tabla 8.* Resultado de los datos fusionados para proteínas.

Pretratamiento	D0			
Selección de variables	GA	GA	RSR	RSR
Método de regresión	OLS	PLS	OLS	PLS
$R^2$	No hay modelo	0.97	0.979	0.99
$Q^2$		0.327	0.333	0.124

*Tabla 9.* Resultado de los datos fusionados para carbohidratos.



<b>Pretratamiento</b>	<b>D0</b>			
<b>Selección de variables</b>	<b>GA</b>	<b>GA</b>	<b>RSR</b>	<b>RSR</b>
<b>Método de regresión</b>	<b>OLS</b>	<b>PLS</b>	<b>OLS</b>	<b>PLS</b>
$R^2$	No hay modelo	0.987	0.979	0.979
$Q^2$		0.656	0.448	0.415

*Tabla 10.* Mejores modelos para proteína, grasa y carbohidratos.

<b>Componente</b>	<b>Pretratamiento</b>	<b>Selección de variables</b>	<b>Método de regresión</b>	$R^2$	$Q^2$	<b>Variables (Longitudes de onda)</b>
<b>Proteína</b>	D0	RSR	OLS	0.994	0.603	<b>1500.347, 1505.2, 1549, 1554.3, 1576, 1676.8</b>
<b>Grasa</b>	D0	RSR	PLS	1	0.534	<b>1161.4, 1178.3, 1201, 1446.4, 1473.3, 1854.2</b>
<b>Carbohidratos</b>	D0	RSR	OLS	0.987	0.878	<b>2815.08, 2865.702, 2902.343, 2907.164, 2948.144, 2955.375</b>

#### 4. CONCLUSIONES

La selección de variables con algoritmos genéticos, utilizando el método de mínimos cuadrados ordinales (OLS), no es aplicable para la determinación de la composición en mayonesas debido a que no se genera el modelo en la determinación de proteínas, carbohidratos y grasa.

En el caso de las proteínas, grasas y carbohidratos se obtuvo el mejor modelo en el espectro sin procesar; además el algoritmo de selección de variables que dio un mejor resultado, aplicado en señales infrarrojas en muestras de mayonesa, fue el reemplazo secuencial renovado, utilizando el método PLS para grasa y OLS para proteínas y carbohidratos, con un  $R^2 = 0.994$  y  $Q^2 = 0.603$  para las proteínas;  $R^2 = 1$  y  $Q^2 = 0.534$  para las grasas y  $R^2 = 0.987$  y  $Q^2 = 0.878$  para los carbohidratos. Las longitudes de onda que se utilizaron fueron: en el caso de las proteínas 1500.347, 1505.2, 1549, 1554.3, 1576, 1676.8; para la grasa 1161.4, 1178.3, 1201, 1446.4, 1473.3, 1854.2 y para carbohidratos 2815.08, 2865.702, 2902.343, 2907.164, 2948.144, 2955.375.

Los resultados son buenos pero la validación es baja para el caso de la grasa y proteínas, esto se debe a que el modelo se ha construido con pocos datos y por tanto la validación externa no tendrá muchos datos, haciendo que se vuelva más crítico a la hora de obtener un resultado.

La espectroscopia infrarroja puede ayudar a minimizar tanto tiempo como costos, debido a que en los métodos estándares, se utilizan reactivos que son de alto costo y además son controlados, por ejemplo: el éter de petróleo. Mientras que la espectrometría infrarroja no utiliza ningún reactivo, únicamente es el equipo, la muestra y sus programas necesarios para obtener un resultado.

A nivel industrial con el método de la espectrometría infrarroja se gana tiempo de producción, por lo tanto, costos. Debido a que, al ser utilizada la mayonesa como materia prima, esta debe pasar por un respectivo control de calidad el cual puede tomar mucho tiempo y ser perjudicial para la empresa. Con este método en cambio se obtienen los resultados mucho más precisos de una manera rápida, a través del equipo y sin utilizar reactivos controlados que son de alto costo.

## 5. REFERENCIAS

- AOAC. (2005). *Official Methods of Analysis of AOAC International*. In W. Horwitz & Jr. George W. Latimer (Eds), 18th Edition.
- Armenta, S., Moros, J., Garrigues, S., & de La Guardia Cirugeda, M. (2010). Determination of Olive Oil Parameters by Near Infrared Spectrometry. In V. Preedy & R. Watson (Eds.), *Olives and Olive Oil in Health and Disease Prevention* (pp. 533–544). <https://doi.org/10.1016/B978-0-12-374420-3.00058-9>
- Consonni, V., Ballabio, D., & Todeschini, R. (2009). Evaluation of model predictive ability by external validation techniques. *Journal of Chemometrics*, 194–201. <https://doi.org/10.1002/cem.1290>
- Cuevas, E., Fausto, F., Gálvez, J., & Rodríguez, A. (2021). Algoritmo Genético (GA). In MATLAB (Ed.), *Matlab: computación metaheurística y bio-inspirada* (Primera Edición, pp. 35–59). Alpha Editorial.
- de Marchi, M., Toffanin, V., Cassandro, M., & Penasa, M. (2014). Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits<sup>1</sup>. *Journal of Dairy Science*, 97(3), pp. 1171–1186. <https://doi.org/10.3168/jds.2013-6799>
- Dufour, É. (2009). Principles of Infrared Spectroscopy. In D.-W. Sun (Ed.), *Infrared Spectroscopy for Food Quality Analysis and Control* (pp. 3–25). Elsevier.
- Fernández, S. (2020). Diseño de Experimentos: Diseño Factorial. [Trabajo de fin de master, Politécnica de Catalunya BarcelonaTech,] UPCommons. [https://upcommons.upc.edu/bitstream/handle/2117/339723/TFM\\_Fernandez\\_Bao\\_Sheila.pdf?sequence=1&isAllowed=y](https://upcommons.upc.edu/bitstream/handle/2117/339723/TFM_Fernandez_Bao_Sheila.pdf?sequence=1&isAllowed=y)
- Ferré, J. (2006). Calibración multivariante en análisis cuantitativo. El modelo inverso. *Técnicas de Laboratorio*, 310, pp. 262–266. [http://www.quimica.urv.es/quimio/general/AD\\_TL\\_Calibracisn%20Multivariante.%20El%20modelo%20inverso.pdf](http://www.quimica.urv.es/quimio/general/AD_TL_Calibracisn%20Multivariante.%20El%20modelo%20inverso.pdf)
- Flores, O. (2017). *Estudio de la composición nutricional de un alimento por espectroscopía de infrarrojo cercano con transformada de fourier (ft-nirs) y quimiometría* [Trabajo de fin de máster, Universidad Autónoma de Baja California]. Repositorio UABC.

<https://repositorioinstitucional.uabc.mx/bitstream/20.500.12930/2551/1/MXL120651.pdf>

Gallegos, S. (2021). *Aplicación de técnicas quimiométricas en la interpretación de espectros Raman de diversos tipos de cerveza* [Trabajo de fin de grado, Universidad del Azuay] Repositorio UDA. <https://dspace.uazuay.edu.ec/bitstream/datos/10571/1/16160.pdf>

Gaviria, C. (2015). *Regresión por Mínimos Cuadrados Parciales PLS Aplicada a Datos Variedad Valuados* [Trabajo de fin de Master, Universidad Nacional de Colombia] Repositorio UNAL. <https://repositorio.unal.edu.co/bitstream/handle/unal/57110/8105560.2016.pdf?sequence=1&isAllowed=y>

Gaviria, C. A., Pérez R. A., & Puerta-Yepes, M. E. (2016). Regresión por mínimos cuadrados parciales PLS con datos de intervalo. *Revista de La Facultad de Ciencias*, 5(1), pp. 148-159. <https://doi.org/10.15446/rev.fac.cienc.v5n1.54616>

Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, 26(5), pp. 694–701. <https://doi.org/10.1002/qsar.200610151>

Greenfield, H., & Southgate, D. a. T. (2006). Examen de los métodos de análisis. In FAO (Ed.), *Datos de composición de alimentos “Obtención, gestión y utilización”* (2nd ed, p. 122). <https://www.fao.org/3/y4705s/Y4705S.pdf>

Grisoni, F., Cassotti, M., & Todeschini, R. (2014). Reshaped sequential replacement for variable selection in QSPR: Comparison with other reference methods. *Journal of Chemometrics*, pp. 249–259. <https://doi.org/10.1002/cem.2603>

Gutiérrez, H., & Salazar, R. (2008). Introducción al diseño de experimentos. In P. Roig (Ed.), *Análisis y diseño de experimentos* (2nd ed, pp. 2–14). Mc Graw Hill. [https://gc.scalahed.com/recursos/files/r161r/w19537w/analisis\\_y\\_diseno\\_experimentos.pdf](https://gc.scalahed.com/recursos/files/r161r/w19537w/analisis_y_diseno_experimentos.pdf)

Hidalgo, J., Motato, J., & Ramírez, J. (2018). Diseños experimentales aplicados en Investigación y Desarrollo de dulces de leche. In E. Ordóñez (Ed.), *Leches concentradas azucaradas: de la tradición a la ciencia* (pp. 273–293). USC. <https://libros.usc.edu.co/index.php/usc/catalog/view/74/79/1255>

- Ildikó, F., & Todeschini, R. (1994). *The data analysis handbook*. Elsevier
- Kalivas, J. (2009). Calibration Methodologies. In S. Brown, R. Tauler, & B. Walczak (Eds.), *Comprehensive Chemometrics Chemical and Biochemical Data Analysis* (pp. 1–32). Elsevier.
- Kassambara, A. (2018, March 11). *Cross-Validation Essentials in R*. STHDA. <http://www.sthda.com/english/articles/38-regression-model-validation/157-cross-validation-essentials-in-r/>
- Kumar, M., Husain, M., Upreti, N., & Gupta, D. (2010). Genetic Algorithm: Review and Application. *SSRN Electronic Journal*, 451–454. <https://doi.org/10.2139/ssrn.3529843>
- Leopold, L., Leopold, N., Diehl, H., & Socaciu, C. (2011). Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. *Spectroscopy*, 26(2), 93–104. <https://doi.org/10.3233/SPE-2011-0529>
- Lozano, M. (2013). *Caracterización de la carne procedente de diferentes especies animales mediante espectroscopia mir* [Trabajo de Fin de carrera, Universidad Pública de Navarra] Repositorio Academica-e - UPNA. <https://academica-e.unavarra.es/xmlui/bitstream/handle/2454/7789/578141.pdf?sequence=1&isAllowed=y>
- McClements, J. (2003). Analysis of Proteins. *Analysis of Food Products Food Science*. <https://people.umass.edu/~mcclemen/581Proteins.html>
- Medina, P., & López, Á. (2011). Análisis crítico del diseño factorial 2K sobre casos aplicados. *Scientia Et Technica*, XVII(47), 101–106. <https://www.redalyc.org/articulo.oa?id=84921327018>
- Mejía, C. (2010). *Cuantificación, mediante infrarrojo, de la pectina residual del proceso de biotecnología de clarificación de néctar y jugo de manzana* [Trabajo de graduación, Universidad del Azuay] Repositorio UDA. <https://dspace.uazuay.edu.ec/bitstream/datos/6284/1/08095.pdf>
- Mondragón, P. (2020). El espectrómetro de infrarrojo. In CIATEJ (Ed.), *Principios y aplicaciones de la espectroscopia de infrarrojo en el análisis de alimentos y*

- bebidas*, 7(1), pp. 41–50.  
[https://ciatej.mx/files/divulgacion/divulgacion\\_5f89fd7801268.pdf](https://ciatej.mx/files/divulgacion/divulgacion_5f89fd7801268.pdf)
- Niazi, A., & Leardi, R. (2012). Genetic algorithms in chemometrics. *Journal of Chemometrics*, 26(6), 345–351. doi: 10.1002/cem.2426
- Palma, P. (2022). Critical analysis of the coefficient of determination (R<sup>2</sup>), as an indicator of the quality of linear and non-linear models. *ESPOL-FCNM*, 20(2), 1–12. <http://www.revistas.espol.edu.ec/index.php/matematica/article/view/1037/917>
- Pérez, W., & López, J. (2017). Aplicación de un diseño factorial 2<sup>4</sup> en la remoción de turbiedad del Río Rímac mediante la coagulación y floculación usando goma de tara. *Revista de Investigación Ciencia, Tecnología y Desarrollo*, 3(1), 18–28. <https://doi.org/10.17162/rictd.v1i1.895>
- Pérez, Y. (2013). *Análisis comparativo de propiedades, textura y estabilidad de mayonesas comerciales* [Trabajo fin de master, Universidad de Oviedo] Repositorio RUO. <https://digibuo.uniovi.es/dspace/bitstream/handle/10651/19089/TFM%20Yarima%20Perez%20Chiquito.pdf?sequence=3&isAllowed=y>
- Reyes, M. (2009). *Aplicación del diseño experimental en el desarrollo de las prácticas internas, en el área de operaciones unitarias* [Trabajo de Graduación, Universidad de San Carlos de Guatemala] Biblioteca USAC. [http://biblioteca.usac.edu.gt/tesis/08/08\\_1138\\_Q.pdf](http://biblioteca.usac.edu.gt/tesis/08/08_1138_Q.pdf)
- Rodríguez, E. (2015). Modelo de regresión lineal: mínimos cuadrados parciales (PLS) con método Cuasi-Newton. *REVECITEC*, 6(1), 21–37. <http://ojs.urbe.edu/index.php/revecitec/article/view/1523/1480>
- Rosas, M., & Fernández, J. (2012). FTIR aplicada durante la deshidratación osmótica de mango Ataulfo (*Magnífera indica* L.). *Superficies y Vacío*. 25(1), 8–13. <https://superficiesyvacio.smctsm.org.mx/index.php/SyV/article/view/221/152>
- Salvatierra, D. (2015). *Determinación de la composición química proximal, carbohidratos totales, azúcares libres y fructanos del tipo Inulina - Fructooligosacáridos del Yacón* [Tesis de grado, Universidad Peruana Cayetano Heredia]. In *Repositorio de la Universidad Peruana Cayetano Heredia*. <https://repositorio.upch.edu.pe/bitstream/handle/20.500.12866/668/Determinaci%C>

3%B3n%20de%20la%20composici%C3%B3n%20qu%C3%ADmica%20proximal%20%20carbohidratos%20totales%20%20az%C3%BAcares%20libres%20y%20fructanos%20del%20tipo%20inulina%20-%20fructooligosac%C3%A1ridos%20del%20yac%C3%B3n%20%28Smallenthus%20sonchifolius%20%28Poepp.%20et%20Endl.%29%20H.%20Robinson%29.pdf?sequence=1&isAllowed=y

Thermo Electron Corporation. (s.f.). *Using OMNIC Algorithms*. Madison: Thermo Electron Corporation.  
<https://mmrc.caltech.edu/FTIR/Nicolet/Nicolet%20manuals/Omnic%20Users%20Manual%207.3.pdf>

Thermo Fisher Scientific. (2021). *OMNIC Paradigm Spectroscopy Software*.  
<https://assets.thermofisher.com/TFS-Assets/MSD/Product-Information/OMNIC-Paradigm-A-UG.pdf>

Torres, C. (2016). *Uso de espectroscopía FTIR para la determinación de azúcares simples y ácidos en zumos y néctares de fruta* [Trabajo de Fin de grado, Universidad Politécnica de Valencia] RIUNET Repositorio UPV.  
<https://riUNET.upv.es/handle/10251/71829>

Valenciaga, D., & Simoes, D. O. (2006). La espectroscopia de reflectancia en el infrarrojo cercano (NIR) y sus potencialidades para la evaluación de forrajes. *Cuban Journal of Agricultural Science*, 40(3), 259–267.  
<https://www.redalyc.org/pdf/1930/193017723001.pdf>

Vandeginste, B. (1987). Chemometrics and Species Identification. In W.-Y. Wong & M. Olivucci (Eds.), *Topics in Current Chemistry* (Vol. 141, pp. 2–7). Springer.

Vélez, B. (2022). *Análisis de la transformación de datos espectrales obtenidos por espectroscopía de infrarrojo por transformada de Fourier para la cuantificación de azúcares* [Tesis de grado, Universidad Politécnica Salesiana] Repositorio UPS.  
<https://dspace.ups.edu.ec/bitstream/123456789/22652/4/UPS-CT009802.pdf>

Vilca, A. (2011). *Métodos Predictivos*. [Diapositivas de Power point]. Departamento de Ingeniería de Sistemas, Universidad nacional de la Amazonía Peruana.  
<https://es.slideshare.net/WolfHunterRose/metodos-predictivos>

