



**INSTITUTO TECNOLÓGICO  
SUPERIOR DE GUANAJUATO**

**“Sistema Experto para la Determinación  
de la Vida de Anaquel de Concentrados  
de Fruta”**

**Opción I: Tesis**

**Que para obtener el grado de:  
INGENIERO EN SISTEMAS  
COMPUTACIONALES**

**Presenta:  
Larisa Hernández Aquino**

**Con la asesoría de:  
Dr. Luis Armando García de la Rosa**

**Guanajuato, Gto. 2019**

## **Agradecimientos.**

Doy gracias a dos personas que fueron decisivos para la realización y finalización de mi tesis, en primer lugar a mi asesor, porque la mayor parte del tiempo no somos lo suficientemente capaces de entender todo el sacrificio que ellos tienen que hacer por nosotros, desde el momento en que preparan la clase, y no obstante a eso tenerla que adaptar a cada uno de nosotros para que todos podamos adquirir el nuevo conocimiento y en ocasiones, brindarnos de sus tiempo libre para aclarar dudas, por momentos se nos olvida ser agradecidos y entender, que como estudiantes el papel que nos corresponde es tomar cada tarea que se nos asigna con responsabilidad, no verla con desprecio y asociarla como castigo y pérdida de tiempo, saber valorar la importancia de un maestro, es comprender bien nuestro papel de estudiantes, porque en realidad estas personas son sumamente importantes en nuestro desarrollo como personas y profesionistas. Especialmente en mi caso, mi asesor fue crucial para la realización de esta tesis.

Quiero agradecerle a él por cada detalle y momento dedicado para aclarar cualquier tipo de duda que me surgiera, agradecerle por la claridad y exactitud con la que enseñó cada clase.

Así mismo, doy las gracias a mi pareja por entenderme en todo, gracias a ella porque en todo momento fue un apoyo incondicional en mi vida, fue la felicidad encajada en una sola persona y nunca dudo de que lograría ser toda una ingeniería.

Mis sinceras gracias a mi pareja y a mi asesor, nunca podré terminar de agradecerles por su apoyo incondicional.

## Resumen.

Es importante destacar que, en cada área de ingeniería, industrial, comercial, tecnológica, etc., se requieren realizar procesos para mejorar la calidad de un producto o de un objeto, siendo el caso de este proyecto, que apoyará a determinar la vida de anaquel de concentrados de fruta y mermeladas con un costo relativamente bajo en comparación con los métodos existentes (pruebas de colorimetría), evaluando la pérdida de color como parámetro de deterioro fisicoquímico a través de capturas de imágenes, empleando cómputo móvil. Es decir, un sistema clasificador que aplique visión por computadora y algoritmos de inteligencia artificial para determinar cuántos días útiles le queda al alimento antes de perder propiedades sensoriales tales como color, sabor, olor, entre otros. Estos productos perecederos son elaborados a base de frutas y azúcar y la concentración de esta determina si se trata de una conserva (<65% p/p<sup>1</sup>) o de una mermelada (65-68% p/p), pueden mantenerse hasta dos años en condiciones óptimas para consumo. Se realizará un estudio sobre cuál de las tres técnicas de clasificación propuestas: *Naïve-Bayes* con enfoque gaussiano), *K-Vecinos Más Cercanos* o *KNN* (*K-Nearest-Neighbor*), y *Redes Neuronales Artificiales* (*Pattern Recognition Neural Network*), es la mejor para el problema planteado.

Este sistema experto, podrá ayudar a productores estatales de conservas de fruta del estado de Guanajuato, asesorándolo para que este conozca con exactitud dicha variable, obteniendo con esto, la fecha (aproximada) en que el producto aún conservará su estado inocuo y la mayoría de sus propiedades nutricionales y organolépticas, en otras palabras, el producto se encontrará dentro del periodo de tiempo en el cual conservara los atributos esperados por el consumidor (seguirá manteniendo color, sabor, etc.) y será el momento adecuado para mantenerlo comercializado, provocando con esto que los productores puedan ser más competitivos con sus productos e incrementen sus volúmenes de producción.

---

<sup>1</sup> Porcentaje Peso a Peso (% P/P), es una medida de la concentración que indica el peso de soluto por cada 100 unidades de peso de la solución.

# Índice.

Agradecimientos.....	i
Resumen.....	ii
Capítulo 1. Generalidades del proyecto.....	1
1.1 Planteamiento del problema.....	2
1.2 Justificación.....	2
1.3 Viabilidad.....	4
1.4 Hipótesis.....	5
1.5 Objetivos.....	5
1.5.1 General.....	5
1.5.2 Específicos.....	5
Capítulo 2. Marco Teórico.....	6
2.1 Fundamento Teórico.....	7
2.2 Antecedentes.....	17
Capítulo 3. Metodología.....	22
3.1 Materiales y Métodos.....	23
Capítulo 4. Discusión y análisis de resultados.....	25
4.1. Resultados.....	26
4.1.1. Clasificador de <i>Naïve-Bayes</i> .....	26
4.1.2. Programación de KNN ( <i>K-Nearest-Neighbor</i> ).....	28
4.2. Importancia o trascendencia.....	33
4.3 Conclusiones y recomendaciones.....	33
Datos de la empresa.....	35
a) Organigrama.....	35
b) Misión.....	35
c) Política de calidad.....	36
d) Visión 2023.....	36
e) Objetivos.....	36
Clasificador <i>Naïve-Bayes</i> .....	38

<b>Clasificador KNN.....</b>	<b>41</b>
<b>Red Neuronal.....</b>	<b>43</b>
<b>Referencias.....</b>	<b>44</b>

## Índice de Figuras.

Figura 1: Sistema Biométrico. ....	7
Figura 2: Ejemplificación de algoritmo de detector de Spam. ....	8
Figura 3: Estructura de un sistema de visión industrial típico. ....	9
Figura 4: Toma de rayos X de la columna Lumbar. ....	9
Figura 5: Diagrama de flujo del aprendizaje no supervisado. ....	10
Figura 6: Diagrama de flujo del aprendizaje supervisado. ....	11
Figura 7: Ejemplificación de cómo trabaja el algoritmo KNN. ....	14
Figura 8: Diagrama de un perceptrón simple con cinco señales de entrada. ....	15
Figura 9: Arquitectura unidireccional con tres capas de neuronas: una capa de entrada, una capa oculta.....	16
Figura 10: Siete principios HACCP.....	18
Figura 11: Descripción del proceso de clasificación del mango.....	19
Figura 12: Descripción del proceso de clasificación del mango.....	19
Figura 13: Modelado de datos para clase 97, con media y desviación estándar.....	26
Figura 14: Histograma $L, a^*, b^*$ .....	27
Figura 15: Grafica Gaussiana.....	28
Figura 16: Probabilidad de posterior y clasificación errónea.....	28
Figura 17: Gráfica con los conjuntos de entrenamiento y prueba.....	30
Figura 18: Red Neuronal.....	31
Figura 19: Rendimiento de la red neuronal. ....	32
Figura 20: Entrenamiento de red neuronal.....	32
Figura 21: Organigrama del ITESG.....	35

## **Introducción.**

El presente trabajo tiene como objetivo crear un sistema experto que apoye a los productores locales del estado de Guanajuato, a contar con una herramienta que ayude a obtener de una manera aproximada la vida de anaquel de sus concentrados de fruta (fresa), aplicando inteligencia artificial en el proceso.

En la actualidad se cuenta con un amplio desarrollo de nuevas tecnologías y productos alimenticios en el mercado, es por lo que, las modificaciones en las condiciones de operación en las industrias procesadoras de alimentos, la competitividad en los mercados, las exigencias del consumidor frente a la calidad de los alimentos, exigen la mejora de las técnicas de conservación y medición del tiempo durante el cual el alimento se conserva apto para el consumo humano, desde el periodo de la compra hasta su consumo, en otras palabras se mantiene inocuo (libre de agentes patógenos).

Es esencial identificar y medir estos atributos críticos del alimento en relación a su sabor [1], textura, color y otras características sensoriales, así como las variables que producen el deterioro de estos atributos como la rancidez (o enranciamiento), decoloración o mal olor y bajo qué circunstancias (tiempo-temperatura) se producen estos efectos. Esto ayuda a determinar su perfil de calidad, y en base a éste, medir su deterioro durante su vida de anaquel, hasta el punto que éste no sea aceptable por el consumidor.

En específico, los cambios de color en alimentos procesados se deben a las reacciones de oscurecimiento y decoloración, las cuales son un tipo de reacciones de deterioro que se presentan en los productos alimentarios y se van evidenciando conforme va avanzando la vida útil [4].

La pérdida de color en alimentos envasados como conservas y mermeladas se puede atribuir a diferentes causas:

1. Cocción prolongada. Causa la caramelización del azúcar o afecta los pigmentos (carotenos, antocianinas, betalainas, clorofilas etc.), que se vuelve pardos.

2. Insuficiente enfriamiento después del envasado. Se observa frecuentemente cuando se llenan envases grandes en una sola operación.
3. Pulpa descolorida. Se observa con frecuencia cuando se utiliza pulpa de fresas mal limpiadas. El anhídrido sulfuroso, usualmente enmascara el verdadero color de la pulpa, y la pérdida de color solamente se pone de manifiesto después de la cocción.
4. Contaminación con metales. Los fosfatos de magnesio y potasio, los oxalatos u otras sales insolubles de estos metales producen enturbiamiento. El estaño y el hierro y sus sales pueden originar un aspecto lechoso u oscurecimiento.
5. Causas biológicas. Los daños mecánicos o una madurez excesiva causan el pardeamiento de un gran número de variedades de fruta. [4, 6].

Los atributos como el color, se miden por medio de pruebas, que en ocasiones son destructivas o provocan que el alimento pierda inocuidad. Por lo que, pruebas libres de contacto son adecuadas, ya que permite que los productos empacados o envasados sigan en condiciones para su comercialización. Entre estas pruebas existen las pruebas ópticas, las cuales por medio de Procesamiento Digital de Imágenes pueden obtener información que se relaciona con los parámetros que determinan la vida de anaquel. Impactando favorablemente en la calidad del producto y garantizando que el consumidor final consuma producto en las óptimas condiciones.

# **Capítulo 1. Generalidades del proyecto.**



## 1.1 Planteamiento del problema.

En los últimos años, es evidente el crecimiento del sector agrícola en el estado de Guanajuato, especialmente el cultivo de fresas [28] para la elaboración de concentrados de frutas (mermeladas) u otros productos.

La industria de las mermeladas tiene un origen relativamente reciente en el estado de Guanajuato [26], en virtud de que la mermelada sólo se produce a nivel casero o artesanal en la mayoría de los casos, ya que cumplir con todas las normas y reglas que existen para su elaboración industrial requiere de pruebas físico-químicas, asociada a la falta de herramientas para conocer con exactitud la vida de anaquel del producto final, tiene como resultado pérdida de competitividad, ya que por desconocimiento de la vida útil exacta, no se producen grandes lotes, por lo que, se requiere una herramienta o sistema que ayude al emprendedor a proyectar sus volúmenes de producción de manera óptima.

## 1.2 Justificación.

La industria de alimentos es rigurosa con respecto al control de la calidad, debido a los riesgos sanitarios asociados con el incorrecto manejo de los alimentos [4]. Prueba de esto es la múltiple cantidad de normas (nacionales e internacionales) que deben cumplir los productos para su venta al público o su exportación, entre estas se encuentra: la Norma Oficial Mexicana **NOM-251-SSA1-2009, Prácticas de higiene para el proceso de alimentos, bebidas o suplementos alimenticios** y la **ISO 22000 Gestión de la inocuidad alimentaria**. [4, 7-9].

De igual manera, existen diferentes sistemas que ayudan a la clasificación de frutas usando visión por computadora, los cuales trabajan con diversos métodos entre los cuales se encuentra el algoritmo de K-Vecinos Más Cercanos o *KNN (K-Nearest-Neighbor)*, redes neuronales artificiales, entre otros, dichos sistemas no toman en cuenta la degradación de las frutas, solo se basan en su color o textura y son aplicados en las bandas transportadoras.

Sin embargo, una forma de medir la vida de anaquel está asociada a la evaluación sensorial, a través de escalas hedónicas de aceptabilidad. Con ello se miden características como color, sabor, o el enranciamiento (en caso de grasas), entre otros. Este tipo de estudios se han realizado primordialmente por medio de instrumentos sensibles a estas características, los cuales resultan onerosos, tanto por el precio, como por el uso constante de reactivos. En este sentido, se han propuesto sistemas computacionales para clasificar frutas y hortalizas [10-12], otros reúnen mediciones y obtienen información en tiempo real *in-situ* [14], todos basados en visión por computadora. Sin embargo, no existe un sistema capaz de determinar la vida de anaquel de concentrados de fruta por medio de procesamiento de imágenes, con lo que se cuenta actualmente para la realización del proyecto. Lo que indica que hay una oportunidad de realizar un trabajo innovador. Considerando que el aseguramiento de la calidad alimentaria es una actividad prioritaria en el sector. Esto debido a que el Consejo Estatal Agroalimentario de Guanajuato A.C. reportó que en el 2017 fueron cosechadas 1,027 hectáreas de fresa, con una producción de 32 mil 540 toneladas, lo que colocó a la entidad en la tercera posición a nivel nacional en producción y procesamiento de esta fruta. Cabe señalar que en 2017 el valor de dicha producción alcanzó los \$287 millones 110 mil pesos y generó 2 mil 723 empleos directos de los cuales el 56.64% corresponde a procesadores artesanales [33].

Así mismo, el sistema inteligente para la determinación de vida de anaquel, contribuye al cumplimiento de las metas del Objetivo 2 del Desarrollo Sostenible: Hambre Cero para el 2030 [25], de acuerdo a la Organización de las Naciones Unidas, en el cual, es necesario duplicar la productividad agrícola y los ingresos de los productores de alimentos en pequeña escala, respetando el medio ambiente y la biodiversidad de cada región (Meta 2.3) y asegurar la sostenibilidad de los sistemas de producción de alimentos y aplicando prácticas agrícolas resilientes que aumenten la productividad y la producción, contribuyendo al mantenimiento de los ecosistemas (Meta 2.4).

### **1.3 Viabilidad.**

El Proyecto reúne características, condiciones técnicas y operativas que aseguran el cumplimiento de sus metas y objetivos, dado que recoge los conocimientos y experiencias de profesionales del área que trabajan en su ejecución, además contara con una inversión inicial de 100,000.00 pesos, financiados por el Tecnológico Nacional de México (notificado en el oficio M00.2/598/2019), para la adquisición de los insumos requeridos para llegar a cabo la investigación (objeto de la presente tesis).

Es importante destacar que en la actualidad el mundo del consumismo incrementa constantemente, conocer el tiempo de durabilidad de los alimentos a cobrado gran importancia, esto debido al interés de los consumidores por el cuidado de su salud, lo que los lleva a tomar precauciones para minimizar riesgos de contraer enfermedades por el consumo de alimentos contaminados, o de alimentos procesados.

Por ello el desarrollo de este proyecto beneficiará a los pequeños talleres de la región de Irapuato, puesto que, una vez finalizado y validado se podrá ofertar como servicio a los productores locales y estatales de concentrados de fruta (y mermeladas), poniendo a su disposición una herramienta que los apoye a determinar con certeza la vida útil de sus productos, para que puedan proyectar sus volúmenes de producción, considerando cuánto tiempo puede estar en óptimas condiciones para su consumo. Lo que al final puede acarrear un beneficio social ya que estos pequeños productores son el sustento de múltiples familias (generan alrededor de 100,000 empleos directos). Esto por medio de la Fundación Guanajuato Produce A.C., la cual precisamente sirve para vincular todos los organismos en pro del sector agroalimentario.

Posteriormente las ganancias del estado se verán directamente afectadas, trayendo con ello un incremento en el cultivo de esta fruta, logrando mejorar e impulsar cada vez más el sector agrícola.

## **1.4 Hipótesis.**

“El Sistema experto, podrá ayudar a productores estatales de conservas de fruta, para determinar la vida de anaquel de sus productos, en forma simple y sin necesidad de grandes inversiones.

## **1.5 Objetivos.**

### **1.5.1 General:**

Diseñar y desarrollar un sistema experto para la determinación de la vida de anaquel de concentrados de fruta (fresas), que apoye a la toma de decisiones con base a las características de color y textura, usando visión por computadora.

### **1.5.2 Específicos:**

1. Programar tres clasificadores de inteligencia artificial para la clasificación de las muestras de mermeladas:
  - I. Naïve-Bayes.
  - II. KNN.
  - III. Redes Neuronales
2. Identificar el clasificador que mejor se ajuste al problema, para su implementación.
3. Realizar una prueba piloto con mermeladas artesanales para identificar el desempeño del clasificador ante una situación real.
  - I. Modificar el clasificador si no se obtienen resultados superiores a 95%.
4. Documentar el desarrollo de los clasificadores para su posterior uso en el sistema clasificador.

## **Capítulo 2. Marco Teórico.**

## 2.1 Fundamento Teórico.

Clasificar, es la acción que consiste en asignar un objeto (instancia, dato) a una clase (categoría) [15]. Por ejemplo, podemos catalogar una figura como paisaje, retrato, entre otros. Otro ejemplo es atribuir palabras a categorías gramaticales: sustantivo, verbo, adjetivo, etc. La clasificación encuentra aplicaciones, tales como:

- Sistemas biométricos: donde se adjudica una imagen de una huella, el iris o inclusive el rostro a la persona correspondiente, en sistemas de control de acceso (seguridad) [16], como se puede observar en la **Fig. 1**.



*Figura 1: Sistema Biométrico.*

- Lectores de correo electrónico: los cuales filtran mensajes que sean “spam” (no solicitados), “pishing” (suplantación de identidad), entre otros [17].

En la **Fig. 2** se puede observar un algoritmo detector de spam de redes neuronales artificiales usando la regla de aprendizaje perceptrón adaptados.

```

Require: Update interval: a time interval after which the system will
update its input layer [Defined by user]
Input layer: identify number of layer used for defining spam.
Update time: Current time + Update_interval
Start Training (Algorithm 2)
  While Perceptron System is running do
    If message is received then
      Start Application (Algorithm 3)
    end if
    if current time > update time then Or
      if number of message received > number of message for update
then
      Star Learning (Algorithm 4)
end if
end while

Require Message <- spam or non-spam message (Training corpus)
Innate_Input_Layer <- table (may be empty)
Spam corpus
For each token in the spam message corpus do
  If layer is already exist in in Innate_Input_Layer then
    Innate_Input_Layer.msg_matched <- Innate_Input_Layer.msg_matched +1
    Innate_Input_Layer.spam_matched <- Innate_Input_Layer.spam_matched + spam_increment
  else
    Add token to Innate_Input_Layer
    Innate_Input_Layer.msg_matched <- Innate_Input_Layer.msg_matched + 1
    Innate_Input_Layer.spam_matched <- Innate_Input_Layer.spam_matched + spam_increment
  end if
end for

Ham corpus
For each token in the spam message corpus do
  If token is already exist in Innate_Input_Layer then
    Innate_Input_Layer.msg_matched <- Innate_Input_Layer.msg_matched +1
  end if
end for
token. Weight = Innate_Input_Layer.spam_matched/ Innate_Input_Layer.msg_matched +1

End

```

Figura 2: Ejemplificación de algoritmo de detector de Spam.

- Control de calidad en la industria: para clasificar una pieza o producto como correcta o defectuosa, en forma rápida, exacta, precisa y relativamente poca incertidumbre [18], ver **Fig. 3**.

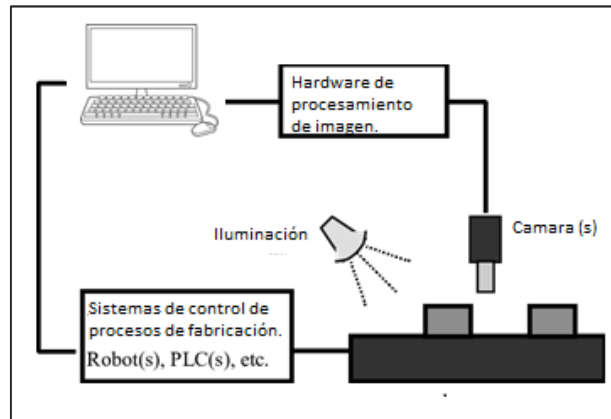


Figura 3: Estructura de un sistema de visión industrial típico.

- Sistemas expertos para asistencia médica: los cuales ayudan a los especialistas médicos a diagnosticar apropiadamente enfermedades, muchos de estos se basan en procesamiento digital de imágenes [19]. Ver **Fig. 4**.

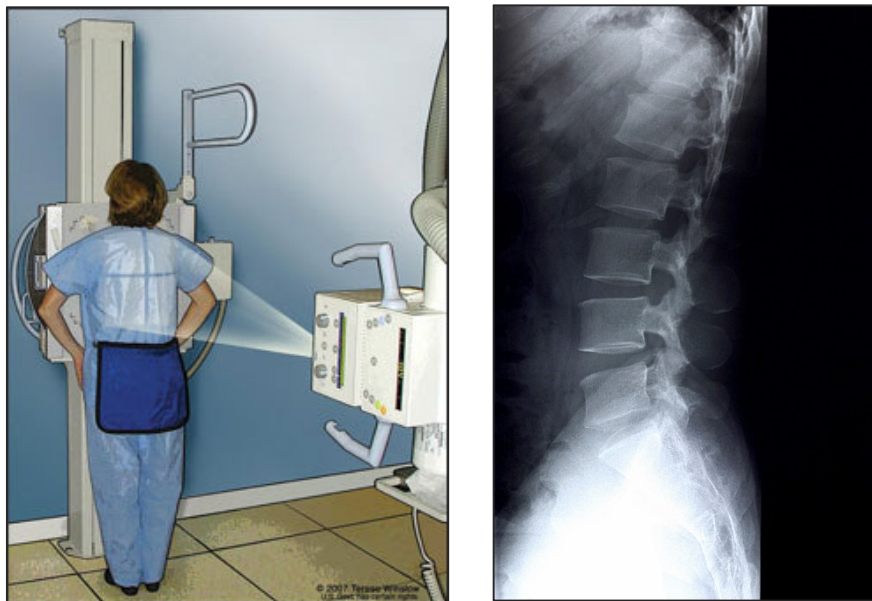


Figura 4: Toma de rayos X de la columna Lumbar.



Para los humanos la acción de clasificar es natural, ya que los sentidos permiten abstraer información del entorno y catalogarla, ayudando a la toma de decisiones. Sin embargo, en los sistemas computacionales, no es una acción tan sencilla, ya que la información deberá de representarse en términos que puedan ser manipulados computacionalmente y entonces proceder a la clasificación. Esto indica que en la clasificación por computadora existen dos problemas, por un lado, la representación de la información (la cual se puede originar de diversas fuentes como sensores, archivos, etc.), y por el otro lado, la clasificación en sí [15].

Se reconoce dos tipos de clasificadores fundamentalmente:

- a) No supervisado o agrupamiento, donde un conjunto de  $n$  objetos se debe dividir (clasificar) en  $k$  clases desconocidas, de tal forma que objetos similares se les asigna la misma clase. Ver **Fig. 5**. Este tipo de clasificadores forman parte de lo que se conoce como *machine learning*.

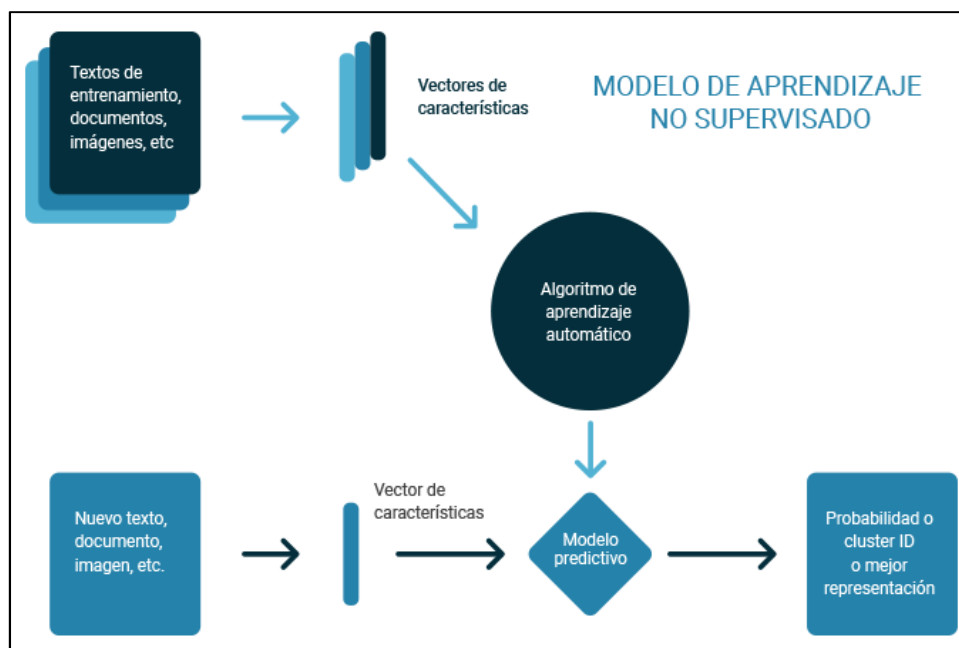


Figura 5: Diagrama de flujo del aprendizaje no supervisado.

b) Supervisado, donde se conocen *a priori* las clases, y el problema se reduce a encontrar una función que asigne (o a cada objeto su clase correspondiente). Ver **Fig. 6**. Este tipo de clasificadores forman parte de lo que se conoce como *deep learning*.

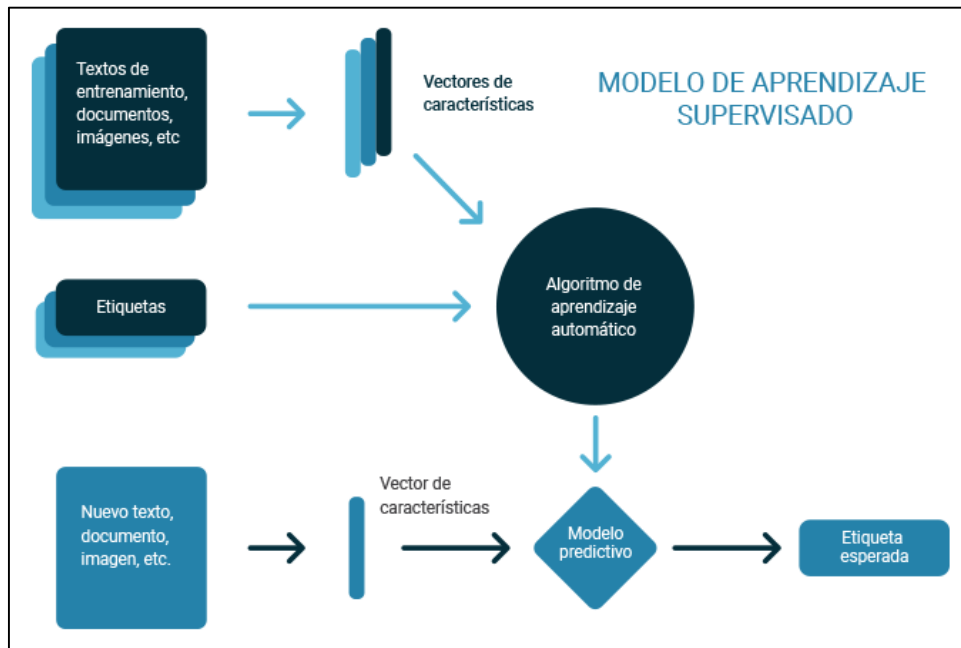


Figura 6: Diagrama de flujo del aprendizaje supervisado.

Los clasificadores no supervisados son muy útiles cuando no se cuenta con la información de salida del problema. Por el contrario, cuando se dispone de esta información (de salida) o metadatos, un clasificador supervisado es más adecuado. Entre los algoritmos de clasificación supervisada, se encuentran: *Naïve-Bayes*, K vecinos más próximos o KNN, lógica difusa y redes neuronales artificiales. [15, 20, 21].

### Clasificador *Naïve-Bayes*

Desde un enfoque bayesiano, el problema de clasificación supervisada consiste en asignar a un objeto descrito por un conjunto de atributos o características,  $X_1, X_2, \dots, X_n$ , a una de  $m$  clases posibles,  $c_1, c_2, \dots, c_m$ , tal que la probabilidad de la clase dados los atributos se maximiza:

$$\boxed{\phantom{\text{Equation}}}$$

$$\text{arg}_C \max [P(C|X_1, X_2, \dots, X_n)] \quad \text{Ec. 1}$$

Donde  $P(C|X_1, X_2, \dots, X_n)$ , es la regla de Bayes o probabilidad *a posteriori*, la cual denota la probabilidad de pertenencia a una clase dado un conjunto de características. La regla de Bayes, también puede escribirse como:

$$P(C|X_1, X_2, \dots, X_n) = \frac{P(C) P(X_1, X_2, \dots, X_n|C)}{P(X_1, X_2, \dots, X_n)} \quad \text{Ec. 2}$$

Donde  $P(C)$  es la probabilidad *a priori* de la clase  $C$  (la probabilidad de pertenecer a la clase  $C$ );  $P(X_1, X_2, \dots, X_n|C)$  es la probabilidad de encontrar del conjunto de características en la clase  $C$  (también conocido como verosimilitud); finalmente  $P(X_1, X_2, \dots, X_n)$ , se conoce como evidencia.

El clasificador bayesiano simple o ingenuo o de *Naïve-Bayes*, se basa en la suposición de que los atributos o características  $X_i$  son condicionalmente independientes entre sí, dada una clase  $C$ . Por lo cual la ecuación 2 se puede reescribir como:

$$P(C|X_1, X_2, \dots, X_n) = P(C)P(X_1|C)P(X_2|C) \dots P(X_n|C) / P(\mathbf{X}) \quad \text{Ec. 3}$$

Donde  $P(\mathbf{X}) = P(X_1, X_2, \dots, X_n)$  y se puede considerar como constante de normalización.

Para clasificación se modifica levemente la ecuación 3 de la siguiente forma:

$$P(C_i|X_1, X_2, \dots, X_n) = P(C_i)P(X_1|C_i)P(X_2|C_i) \dots P(X_n|C_i) / P(\mathbf{X}) \quad \text{Ec. 4}$$

Donde  $C_i$  es  $i$ -ésima clase con  $i = 1, 2, \dots, m$ .

Para determinar en qué clase  $C_i$  pertenece un elemento  $X$  (recordando se trata de un vector de  $1 \times n$ ), solo se debe calcular cada probabilidad por medio de la ecuación 4 y determinar es la mayor.

El clasificador de *Naïve-Bayes* reduce drásticamente la complejidad en espacio y

tiempo de cálculo. En cuanto a espacio de memoria, se requiere la probabilidad previa de las  $m$  clases (vector de  $1 \times m$ ), y las  $k$  probabilidades condicionales de cada atributo dada la clase (si suponemos que los atributos son discretos con  $n$  posibles valores, esto implica  $k$  matrices de  $m \times n$ ). Básicamente el espacio requerido aumenta linealmente con el número de atributos. También el cálculo de la probabilidad posterior se vuelve muy eficiente, ya que se requieren del orden de  $n$  multiplicaciones para calcularla por cada clase dados los atributos (lo que significa complejidad es lineal). [15, 20, 21].

### **K Vecinos Más Cercanos.**

La regla del vecino más cercano o NN (*Nearest Neighbors*), que consiste en clasificar un objeto desconocido en la clase de su vecino más próximo según la disimilitud o distancia [15]. Explorando todo el conocimiento almacenado en el conjunto de entrenamiento para determinar cuál será la clase a la que pertenece una nueva muestra, pero únicamente tiene en cuenta el vecino más próximo a ella, por lo que es lógico pensar que es posible que no se esté aprovechando de forma eficiente toda la información que se podría extraer del conjunto de entrenamiento. Con el objetivo de resolver esta posible deficiencia, surge la regla de los K Vecinos Más Cercanos o KNN (*K-Nearest Neighbors*) [20]. Como extensión de la regla NN, en la que se utiliza la información suministrada por los  $k$  prototipos del conjunto de entrenamiento más cercanos de una nueva muestra para su clasificación.

La idea en la que se basa, reside en comparar un elemento con sus  $k$  vecinos, de modo que éste será clasificado del tipo del que haya más elementos entre estos  $k$  vecinos. Se trata de un algoritmo sencillo y eficaz especialmente cuando existen un gran número de categorías diferentes. El parámetro  $k$  deberá ser decidido *a priori*, y en función de la decisión que se tome se obtendrán unos resultados u otros.

Este tipo de clasificadores es usado cuando existe una posibilidad de traslapes entre los elementos de las clases y, por tanto, la hipótesis determinista no se cumple. Existen varias versiones de este clasificador. Una de ellas es la siguiente: se escoge un entero  $k$ , enseguida se calcula la distancia del elemento a clasificar a los  $k$  elementos más cercanos de las  $N$  clases.

El patrón es clasificado en la clase  $C_i$  ( $i = 1, 2, \dots, N$ ), para la cual la clase  $i$  tenga más votos al momento de la clasificación. La forma de determinar la distancia puede ser muy distinta, se puede emplear la simple distancia euclídea u otras más complejas como la distancia de Mahalanobis. Ahora bien, para cuantificar los votos de cada clase se puede simplemente generar un histograma por clase por cada patrón a clasificar, la categoría con mayor incidencia en el histograma será a la que pertenezca el elemento. Un ejemplo de esto se observa en la **Fig. 7**.

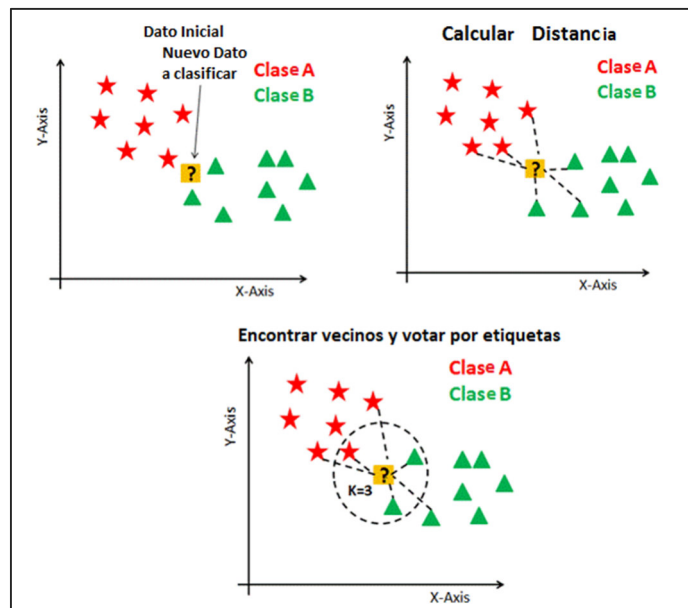


Figura 7: Ejemplificación de cómo trabaja el algoritmo KNN.

## Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) o Artificial Neural Network (ANN), es un modelo computacional que se basa en el funcionamiento del cerebro humano, específicamente en la forma en que se interconectan las neuronas para procesar la información de entrada. En 1957 Frank Rosenblatt propuso el perceptrón simple, basándose en el modelo de lógica umbral de McCulloch y Pitts [20], un perceptrón se compone de varias entradas (de 2 a  $n$ ), las cuales se conectan a un elemento sumador, estas entradas se ven afectadas por pesos que las modifican antes de las suma, finalmente la salida del sumador se conecta a la llamada función de activación (que es donde se lleva a cabo la umbralización), esta función convierte los valores de

entrada en los términos en que se debe manejar la salida. Teniendo en cuenta definición del perceptrón simple o neurona, es evidente que se pueden conectar varios (perceptrones) en forma paralela, para volver más eficiente el procesamiento o aún mejor, interconectándolas entre sí en una o múltiples capas, generando una red de neuronas o red neuronal (las capas intermedias entre la entrada y la salida, se les conoce comúnmente como capas ocultas). Como se aprecia en la **Fig. 8**.

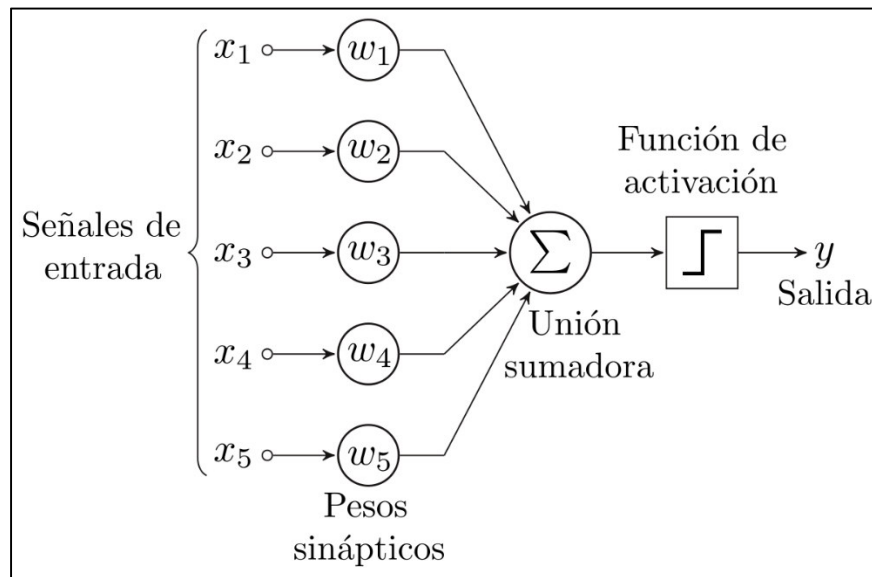


Figura 8: Diagrama de un perceptrón simple con cinco señales de entrada.

Es evidente que una selección aleatoria de los pesos, poco garantiza la correcta aplicación del paradigma del perceptrón, por lo que se requiere de un procedimiento para definir los pesos correctos y convertir (casi) todo valor de entrada, en la salida deseada, dicho proceso se le conoce como entrenamiento. Para realizar a cabo el mencionado entrenamiento, se han propuesto diversas soluciones y una de las mejores es el algoritmo de propagación hacia atrás o *back propagation* (por parte de Werbos en 1975 [20]). La simplicidad del planteamiento de las RNA, la robustez de las soluciones y las múltiples técnicas de entrenamiento las han vuelto en una poderosa herramienta para la resolución de múltiples problemas, como la visión por computadora, la clasificación de datos, filtros digitales, etc.

Existen diversas formas de realizar la clasificación por medio de Redes Neuronales

Artificiales, una de ellas es definir una arquitectura de 3 capas, es decir una capa de  $n$  entradas, otra oculta y finalmente una de  $m$  salidas. Donde  $n$  corresponde al número de muestras que se pasaran a la RNA y  $m$  corresponde a la cantidad de clases. Después se debe alimentar a la red con los datos de entrada (de  $n \times k$ ) y el vector (o matriz) objetivo o *target* (de  $m \times n$ ), y finalmente se aplica el algoritmo de propagación hacia atrás. Esto es, las entradas (o estímulos) se propagan por la red desde la primera capa hasta la salida. Esta salida se compara con el objetivo, calculándose el error de cada salida. Posteriormente, el algoritmo de entrenamiento ajusta los pesos y *biases*<sup>2</sup> (o sesgos), de las distintas capas basándose en (aproximadamente) en la contribución relativa de cada neurona a la salida. Se aplica iterativamente el proceso hasta cumplir distintas condiciones de paro como son, alcanzar la meta de error u obtener un mínimo valor de gradiente en la curva de aprendizaje, entre otros. Al final si se seleccionaron adecuadamente los parámetros de la arquitectura y se contó con suficientes épocas (iteraciones) o diversos parámetros, se conseguirá que “aprenda” a convertir la entrada en la salida deseada con un aceptable margen de error. Finalmente suministrando una entrada arbitraria la red neuronal será capaz de convertirla en la salida deseada, en este caso clasificando los elementos. Ver **Fig. 9**

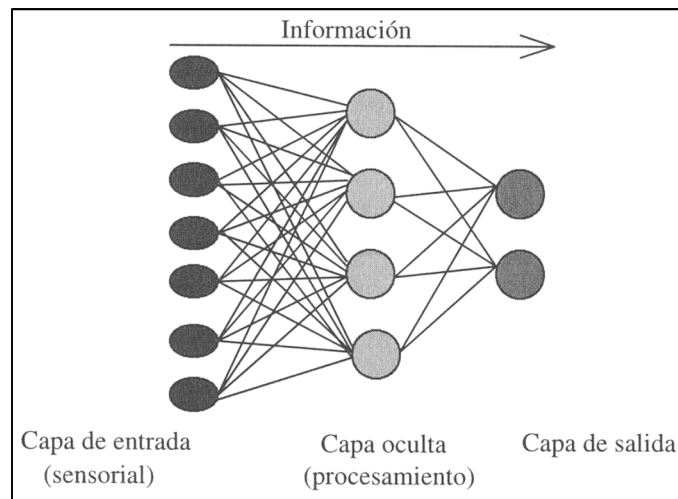


Figura 9: Arquitectura unidireccional con tres capas de neuronas: una capa de entrada, una capa oculta y una de salida.

<sup>2</sup>Biases: Oblicuidad o torcimiento de una cosa hacia un lado, o también definido como tendencia o inclinación.

Una de las desventajas de este tipo de clasificadores es que se requiera un nivel de error tal que no se pueda alcanzar en un tiempo finito, por lo que el algoritmo nunca podrá terminar satisfactoriamente o que se llegue a un mínimo local en la curva de aprendizaje, parando prematuramente el entrenamiento. Por ello se han desarrollado una gran cantidad de técnicas de aprendizaje que resuelven estos problemas.

## **2.2 Antecedentes.**

Como se mencionó anteriormente, la industria de alimentos es muy rigurosa con respecto al control de la calidad, debido a los riesgos sanitarios asociados con el incorrecto manejo de los alimentos [4]. Prueba de esto es la múltiple cantidad de normas (nacionales e internacionales) que deben cumplir los productos para su venta al público o su exportación [4, 7-9].

Por ejemplo, la Norma Oficial Mexicana **NOM-251-SSA1-2009, Prácticas de higiene para el proceso de alimentos, bebidas o suplementos alimenticios**, que tiene como objetivo establecer los requisitos mínimos de buenas prácticas de higiene que deben observarse en el proceso de alimentos, bebidas o suplementos alimenticios y sus materias primas a fin de evitar su contaminación a lo largo de su proceso y es de observancia obligatoria para las personas físicas o morales que se dedican al proceso de alimentos, bebidas o suplementos alimenticios, destinados a los consumidores en territorio nacional, además establece un apéndice.



Mismo que marca las directrices para aplicar o implantar un sistema de Análisis de Peligros y Puntos Críticos de Control (HACCP por sus siglas en inglés (*Hazard Analysis and Critical Control Points*)), el cual considera siete principios. Ver **Fig. 10**.



Figura 10: Siete principios HACCP.

La **ISO 22000 Gestión de la inocuidad alimentaria**, es una norma internacional que incluye los requerimientos necesarios para implantar un sistema de gestión de la Inocuidad Alimentaria. Es utilizada por organizaciones que necesitan evidenciar su aptitud para controlar los riesgos vinculados con la inocuidad alimentaria. El objetivo es garantizar que el alimento tiene las condiciones adecuadas para ser consumido y es aplicable a cualquier tipo de organización que tenga relación con alguna actividad de la cadena alimentaria, independientemente de su tamaño [4, 7-9].

Uno de los rubros que cubren estas normas, es la clasificación de los productos, para su facturación o la determinación de la vida de anaquel entre otros aspectos. Por ejemplo, existen sistemas de clasificación para frutas usando visión por computadora, en los cuales se han aplicado diversos métodos de inteligencia artificial como KNN (K-Nearest-Neighbor) [10, 11], clasificadores bayesianos [11], redes neuronales artificiales [12], entre otros.

Estos sistemas se aplican tanto a frutas [12] como a hortalizas (vegetales) [13], apoyándose en cámaras de diverso tipo y actuadores electromecánicos para la selección, en su mayoría sobre banda transportadora.

Un ejemplo de ello, se puede apreciar en la **Fig. 11**.

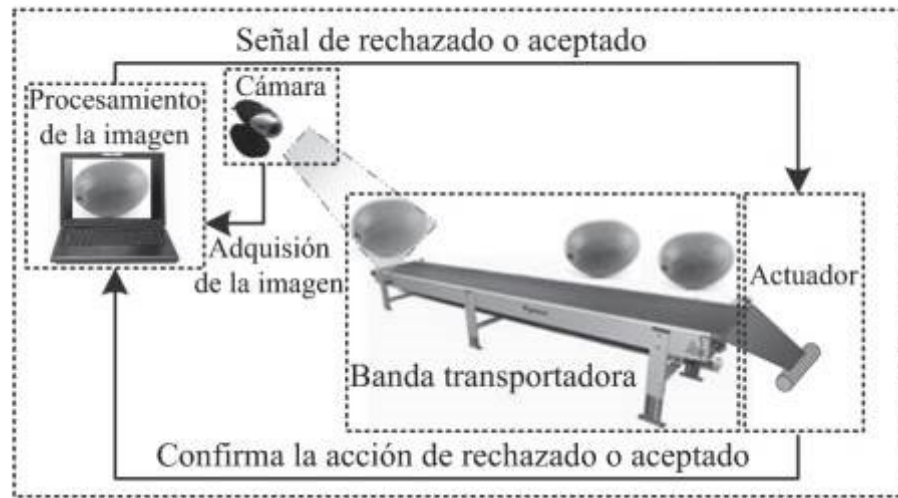


Figura 11: Descripción del proceso de clasificación del mango.

En estos trabajos aplicaron procesamiento digital de imágenes, para segmentar y abstraer características como color o textura, para clasificar los productos sin importar la degradación (y la calidad de los mismos). Ver **Fig. 12**.

Los

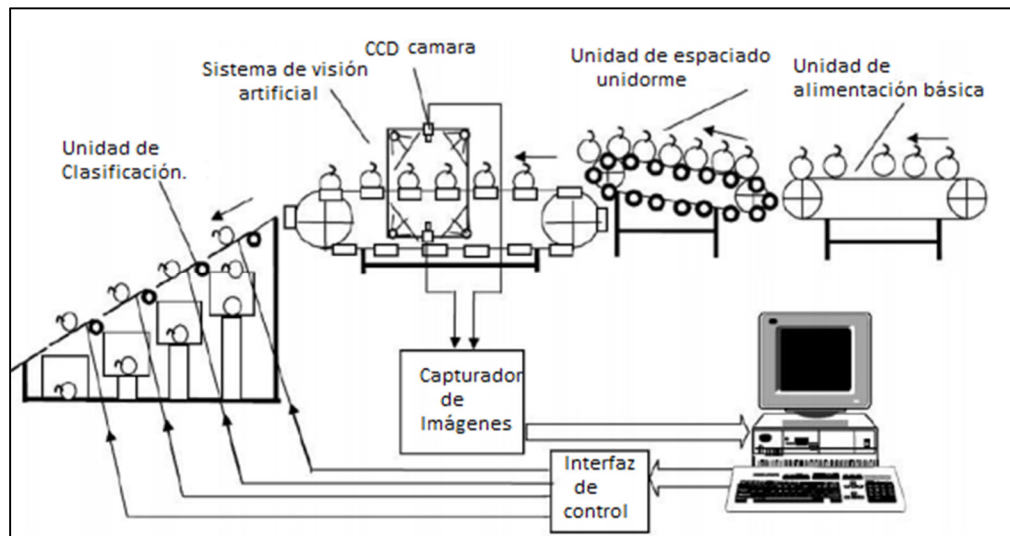


Figura 12: Descripción del proceso de clasificación del mango.

sistemas de visión artificial que se utilizan en estos entornos están dispuestos en una zona de la instalación, donde capturan las imágenes de las frutas que se desplazan

sobre la cinta transportadora. El sistema de visión es capaz de clasificar la fruta según su forma o tamaño, y determinar su grado de maduración a través del color.

Los sistemas de inspección de alimentos frescos en Europa tienen una gran aceptación y difusión. La adopción de estos sistemas ha sido determinada principalmente por dos factores: el envejecimiento de los trabajadores, que tiene como resultado el hecho de que haya menos personas disponibles para este tipo de trabajo de clasificación, y los altos costes laborales en los países desarrollados.

La inmensa mayoría de las compañías relacionadas con la alimentación utilizan sistemas de visión artificial en color, mientras que, en otros mercados, donde tradicionalmente se ha utilizado estos sistemas, como el de la automoción o el de los semiconductores, se usan mayoritariamente sistemas basados en cámaras monocromas. La inspección en la industria alimentaria se lleva a cabo, tanto con cámaras lineales como con cámaras matriciales, dependiendo del tipo de fabricante y del tipo de clasificación.

La complejidad de los productos y los procesos de fabricación de la industria alimentaria, así como el volumen y disparidad de datos que se generan, obligan a avanzar de forma decidida en la aplicación de técnicas de aprendizaje automático ("*Machine Learning*") para extraer información que ayude a la toma de decisiones.

Este tipo de técnicas permite clasificar las ingentes cantidades de información de datos en tiempo real y, por lo tanto, utilizar al máximo el potencial de la industria 4.0, la cual

se refiere a una nueva fase en la revolución industrial que se enfoca en gran medida en la interconectividad, la automatización, el aprendizaje automatizado y los datos en tiempo real, también conocida como IIoT<sup>3</sup> o manufactura inteligente, integra la producción y las operaciones físicas con tecnología digital inteligente, aprendizaje automatizado y *big data*<sup>4</sup> para crear un ecosistema más holístico y mejor conectado para las compañías que se enfocan en la manufactura y la administración de la cadena de suministro. Si bien cada compañía y organización que opera hoy es

---

<sup>3</sup> Internet Industrial de las Cosas, un concepto que se refiere a las conexiones entre personas, datos y máquinas en su relación con la manufactura.

<sup>4</sup> Grandes conjuntos de datos estructurados o no estructurados que pueden recopilarse, almacenarse, organizarse y analizarse para revelar patrones, tendencias, asociaciones y oportunidades.

diferente, todas enfrentan el mismo desafío, la necesidad de conectividad y acceso a información en tiempo real de todos los procesos, socios, productos y personas.[32] Si entrenamos a las máquinas para que al igual que los sistemas neuronales resuelvan problemas complejos como la detección de defectos o la identificación de patrones en tiempo real, las potencialidades del procesamiento digital de imágenes son inmensas.

# **Capítulo 3. Metodología.**

## 3.1 Materiales y Métodos.

### 3.1.1 Metodología

Se seleccionó una metodología por prototipos incrementales, en una visión genérica, el proceso se divide en 4 partes: análisis, diseño, código y prueba. Sin embargo, para la producción del Software, se usa el principio de trabajo en cadena o “Pipeline”, utilizado en muchas otras formas de programación. Con esto se mantiene al cliente en constante contacto con los resultados obtenidos en cada incremento. Es el mismo cliente el que incluye o desecha elementos al final de cada interacción a fin de que el software se adapte mejor a sus necesidades reales. El proceso se repite hasta que se elabore el producto completo. [27].

A continuación, se describe la consecución de los objetivos señalados:

1. Definición del estado del arte sobre los clasificadores, por un lado, identificar los avances más importantes con respecto a los desarrollos tecnológicos similares en los últimos 5 años, así como el marco teórico de los algoritmos de inteligencia artificial para clasificación, enfocándose en las técnicas de clasificación: KNN (K-Nearest-Neighbor) y Redes Neuronales Artificiales, con la intención de contar con los fundamentos teóricos para desarrollar el proyecto.
2. Programar y entrenar algoritmos de clasificación (KNN (K-Nearest-Neighbor), Bayes y Redes Neuronales Artificiales), partiendo de la base de conocimiento, y las herramientas de preprocesamiento, con la intención de verificar su robustez y adaptabilidad al problema.
3. Realizar una prueba piloto, con muestras de concentrados de frutas, que presenten incertidumbre, esto con el objetivo de probar la robustez del sistema desarrollado. En caso de que se requiera realizar una adecuación al sistema o volver a entrenar el algoritmo para aumentar las prestaciones, se deberá de hacer, pero se documentarán los cambios.

4. Se documentarán por completo el desarrollo de los clasificadores para posteriores etapas del proyecto.

### **3.1.2 Materiales**

Para desarrollar el sistema experto, se requirió de una infraestructura de computo de mediano procesamiento, se empleó una cámara web Logitech C170, con una resolución de 640 por 480, y profundidad de color de 24 bits, esto con la finalidad de obtener una óptima calidad en las imágenes y procesamiento de las mismas.

Se ocupó un gabinete (de hechura casera), para la recolección de muestras en ambiente controlado, 130 muestras de productores artesanales de mermelada de fresa, de la región de Irapuato, Gto., los cuales forman parte de la Fundación Guanajuato Produce A.C., estas pruebas fueron medidas con un colorímetro comercial para obtener un patrón de comparación y validar los clasificadores implementados en el sistema. Se etiquetaron con el día de producción, agrupándose en 10 distintos días, se emplearán como el elemento (clase) para la clasificación.

En cuanto a la programación se empleó Matlab y la metodología de software antes mencionada.

## **Capítulo 4. Discusión y análisis de resultados.**



## 4.1. Resultados.

### 4.1.1. Clasificador de *Naïve-Bayes*.

Empleando la plataforma Anaconda (con Python 3) y el IDE Jupyter para codificación junto a las librerías de OpenCV para el pre-procesamiento de imagen, NumPy para el cálculo matricial, Pandas para la generación de las gráficas y *scikit-learn* para los modelos de *machine-learning*, se programó el clasificador de Naïve-Bayes, empleando un modelo gaussiano.

En la **Fig. 13** se observa como el software modela por defecto la distribución de predictores dentro de cada clase usando una distribución gaussiana que tiene 0.1327 de desviación estándar y una media de 39.7736 en este caso **Fig. A**. En la **Fig. B** se muestran las tres características para la clase de 97 días.

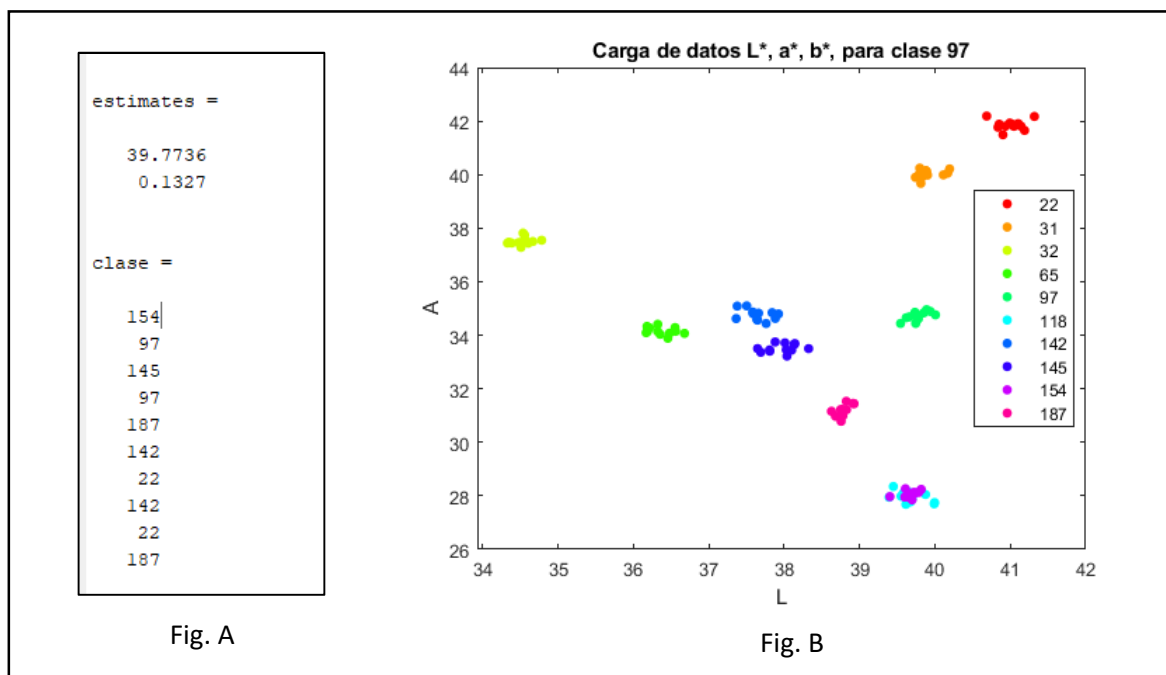


Figura 13: Modelado de datos para clase 97, con media y desviación estándar.

En la **Fig. 14** se muestra un histograma perteneciente a la clase 97.

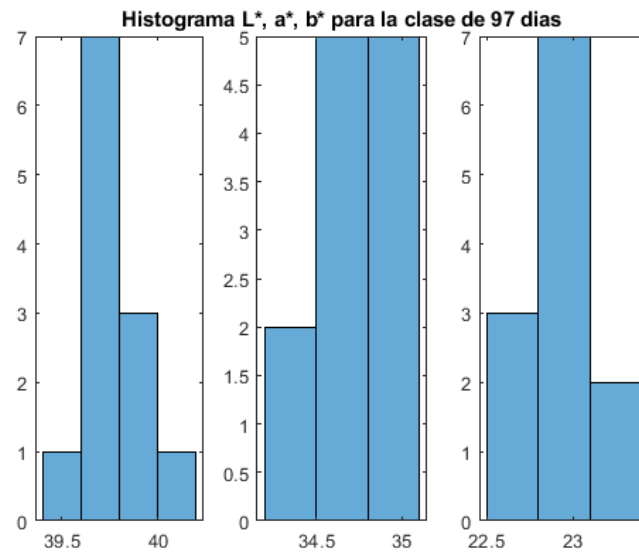


Figura 14: Histograma L, a\*, b\*.

En la **Fig. 15**, se seleccionó la base de datos de 120 muestras de mermeladas y se eligió el grupo pertenecientes a la clase 97, las cuales conforman un total de 12 datos, se puede observar en el grafico que solo 4 de ellas no se encuentran dentro del rango requerido. Con ayuda de estas graficas podemos concluir que la selección del modelo gaussiano para este clasificador es relativamente aceptable, ya que el comportamiento de los datos es normal (gaussiano) en la mayoría de las características.

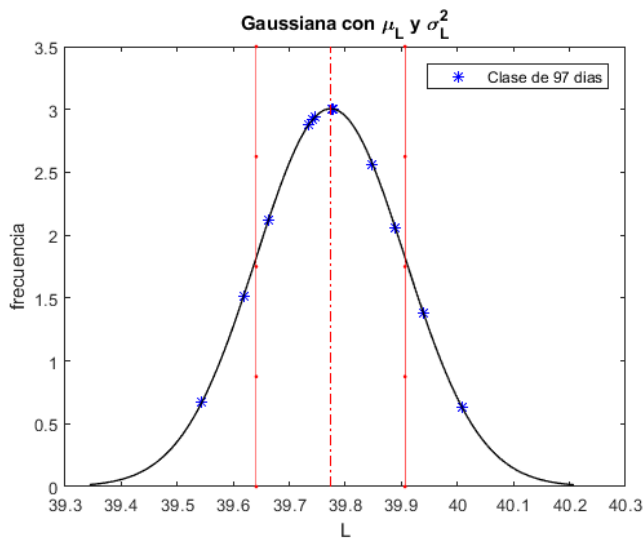


Figura 15: Grafica Gaussiana.

Una matriz de probabilidades posteriores ( Posterior) que indica la probabilidad de que una etiqueta provenga de una clase particular y una matriz de costos de clasificación errónea ( Cost ). Para cada observación en X, la etiqueta de clase predicha corresponde a los costos mínimos de clasificación esperados entre todas las clases. Ver Fig. 16.

```

posteriori =
0      0      0      0      0.0000  0.0000  0.0000  0.0000  1.0000  0.0000
0.0000 0.0000  0      0.0000  1.0000  0.0000  0.0000  0.0000  0      0.0000
0      0      0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  1.0000  0      0.0000
0      0.0000  0      0.0000  1.0000  0.0000  0.0000  0.0000  0.0000  0      0.0000
0      0      0      0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  1.0000
0      0.0000  0.0000  0.0000  0.0000  0      1.0000  0.0000  0      0.0000
1.0000 0.0000  0      0      0      0      0      0      0      0      0
0      0      0      0.0000  0.0000  0.0000  1.0000  0.0000  0      0.0000
1.0000 0.0000  0      0      0      0      0      0      0      0      0
0      0      0      0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  1.0000

costo =
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0000 1.0000
1.0000 1.0000 1.0000 1.0000 0.0000 1.0000 1.0000 1.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 0.0000 1.0000 1.0000 1.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0000 1.0000 1.0000 1.0000
0.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0000 1.0000 1.0000 1.0000
0.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 0.0000

```

Figura 16: Probabilidad de posterior y clasificación errónea.

#### 4.1.2. Programación de KNN (K-Nearest-Neighbor).

Empleando las mismas herramientas que en el caso del clasificador de Naïve-Bayes, se programó el clasificador de K-Vecinos más cercanos, para calcular a que clase

pertenece cada una de las nuevas muestras que aparecen en la **Fig. 17** en rojo, es la parte más simple y sencilla, tan sólo hay que calcular para cada nuevo ejemplo su prototipo más cercano, mirar la **K** local asociada que tenga y clasificar el ejemplo teniendo en cuenta tantos vecinos como indique la **K**.

En la **Fig. 17**, se muestran 3 muestras nuevas a clasificar, diferenciadas con los marcadores \*, +, x, los cuales conformaron el conjunto de prueba, estas fueron asignadas a los diferentes conjuntos de entrenamiento mediante el algoritmo de KNN.

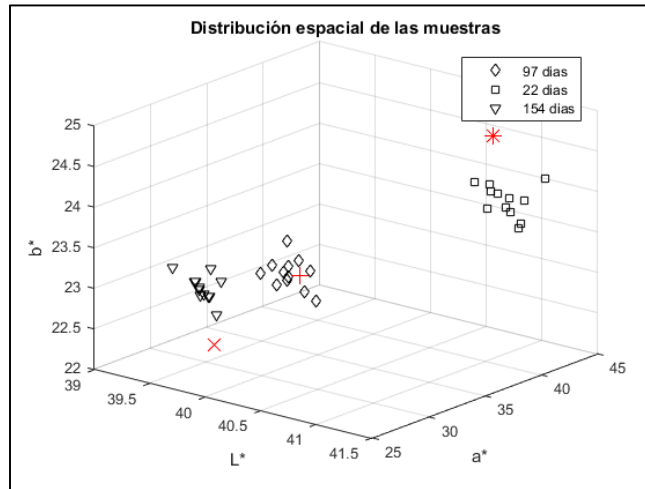


Figura 17: Gráfica con los conjuntos de entrenamiento y prueba.

Aquí se puede notar que en el caso de la clase de 97 días la muestra desconocida está clasificada, mientras que para la clase de 154 días la muestra está bien ubicada, sin embargo, en el caso de 22 días la muestra se encuentra clasificada, pero es evidente que está lejos del conjunto por lo que es fácil pensar que teniendo más datos dicha muestra se clasificará diferente. Asimismo, el número K juega un papel relevante en los resultados, ya que se sabe que debe ser impar (con la finalidad de evitar empates en la clase candidata), por otro lado, el máximo de dicho número debe ser menor o igual a la cantidad de elementos de la clase (de lo contrario se generarán votos en clases que realmente no están cerca pero como se debe cumplir con la cantidad, se tendrán en cuenta, generando falsos positivos), esto dice para el caso que K debe ser impar y menor o igual a 11 (puesto que se tiene 12 elementos en cada clase). Esta característica obliga que los datos deben ser rectangulares, es decir para toda clase debe haber la misma cantidad de elementos.

#### 4.1.3. Programación Red Neuronal (*Pattern Recognition*).

Se programó una red neuronal de reconocimiento de patrones el cual clasifica características únicas que identifican un sujeto de los demás de la misma especie, en este caso se emplearon los mismos elementos que en los otros dos clasificadores. El único cambio fue en el patrón de salida, ya que se debió generar una matriz binaria rectangular de  $m \times n$ , siendo  $m$  la cantidad de clases (en nuestro caso 10) y  $n$  el número de muestras (en nuestro caso 120), con un 1 en donde dicho elemento si pertenece a la clase y un 0 en donde no pertenece.

Se empleó una arquitectura de 3 capas (compuesta por un conjunto de nodos de entrada que componen la capa de entrada, un conjunto de una o más capas ocultas de neuronas y una capa de neuronas de salida, como se mencionó anteriormente a este proceso se le conoce como *back propagation*, el cual consiste en dos pasadas a través de las diferentes capas de la red, una pasada hacia adelante (los pesos sinápticos de la red son fijos y no se modifican) y una pasada hacia atrás (los pesos si se modifican de acuerdo con la regla de corrección del error)), con  $n = 120$  neuronas en la capa de entrada, 60 en la capa oculta y 10 en la capa de salida. La función de transferencia fue competitiva (para producir resultados binarios). Ver **Fig. 18**.

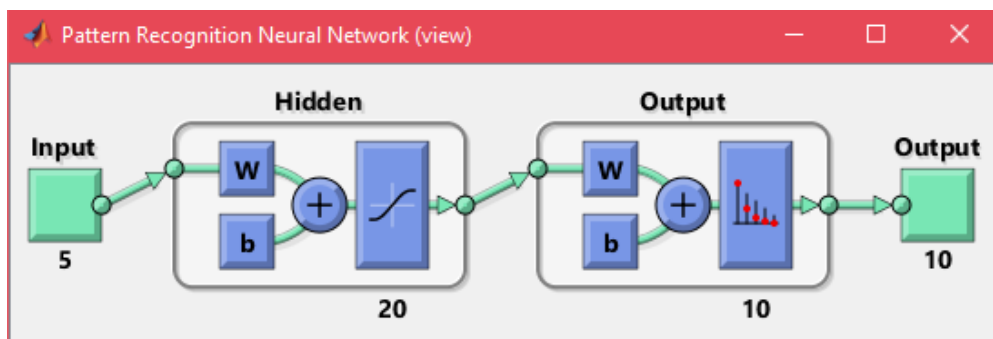


Figura 18: Red Neuronal.

En la **Fig. 19**, se observa cómo mejoró el rendimiento de la red durante el entrenamiento. El rendimiento se mide en términos de error cuadrático medio y se

muestra en una escala logarítmica. Disminuyó rápidamente a medida que se entrenaba la red.

El rendimiento se muestra para cada uno de los conjuntos de entrenamiento, validación y prueba.

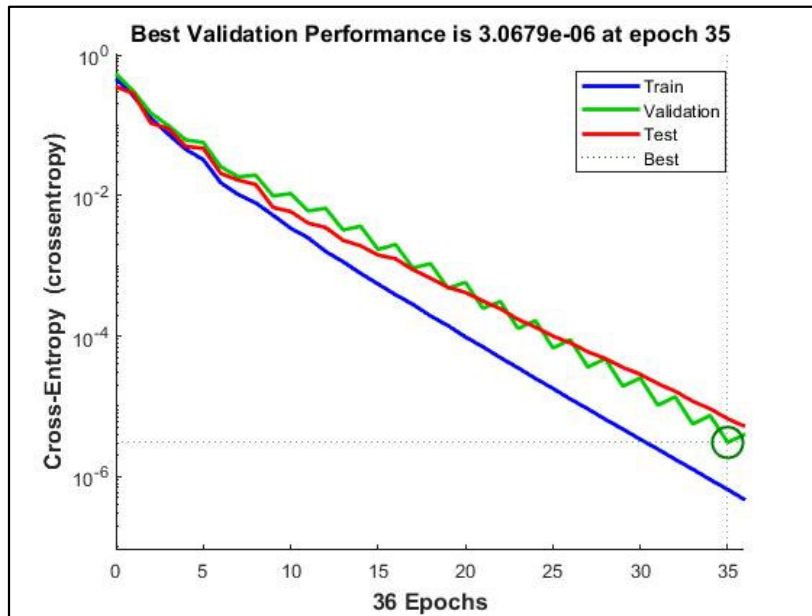


Figura 19: Rendimiento de la red neuronal.

La función de entrenamiento fue de gradiente escalado conjugado, con 1000 época y un gradiente de  $6.507 \times 10^{-7}$ , un gradiente mínimo de  $1 \times 10^{-6}$  y un error de  $2.7 \times 10^{-5}$ . Ver Fig. 20.

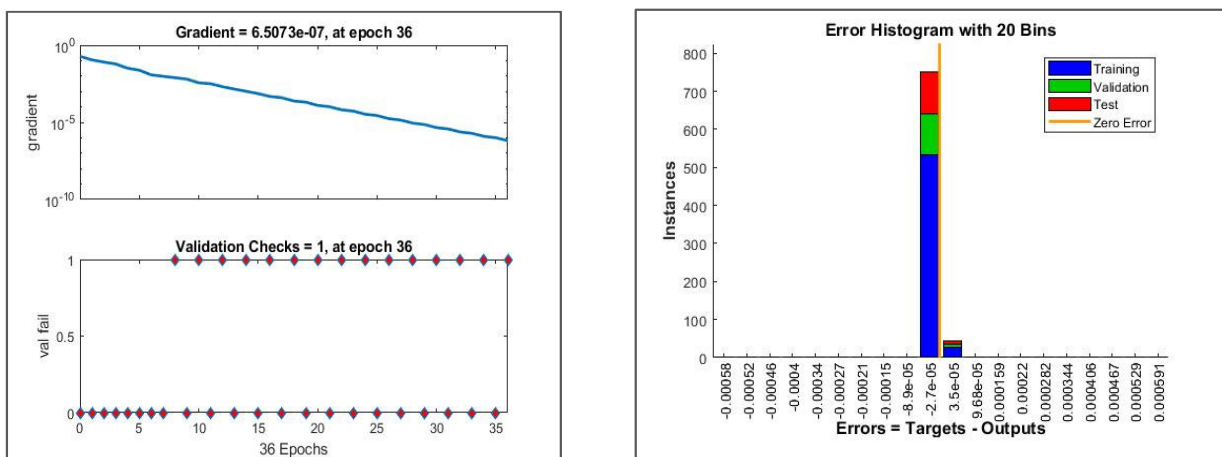


Figura 20: Entrenamiento de red neuronal.

## **4.2. Importancia o trascendencia.**

Los clasificadores que se aplicaron en el desarrollo del software son el principio del sistema experto, cabe mencionar que falta por desarrollarse en su totalidad, ya que faltó el equipo de cómputo para que todo el proceso se hiciera en tiempo real, pero con los resultados obtenidos hasta el momento, se logra observar que es viable el desarrollo del proyecto.

Una vez finalizado el software proporcionará apoyo a los productores locales del estado de Guanajuato, respecto a determinar la vida útil o de anaquel de sus productos de mermelada. Esto conlleva una reducción de los costes de producción importante que se verá reflejada en una disminución del gasto total de la fabricación y al tratarse de un sector que ofrece productos de mejor calidad (conocimiento exacto de la fecha donde el alimento se mantiene inocuo), consigue satisfacer mejor las necesidades del consumidor, lo que repercute en su propio beneficio al venderlos, ya que buenos productos y consumidores satisfechos favorecerán nuevas ventas en el futuro.

Así mismo, el conocimiento exacto de esta variable en los alimentos sirve de aviso para evitar problemas potenciales de daño a la salud de los consumidores.

## **4.3 Conclusiones y recomendaciones.**

La implementación de un sistema que apoye a los agricultores locales a calcular la fecha de anaquel de sus productos de una manera más eficiente y económica, es de suma importancia para subir los niveles de producción de estos, de tal manera que el impacto se ve reflejado en la calidad y cumplimiento de las normativas, satisfaciendo las necesidades de los clientes al comercializar un producto que mantendrá las características esenciales de este.

En este trabajo se han implementado algoritmos de K-NN, Bayes y Red neuronal, el algoritmo de KNN, que consisten en obtener a partir del conjunto de entrenamiento, un valor local de K, que clasifique mejor el vecindario de cada prototipo y clasificar cada nuevo ejemplo utilizando para ello tantos vecinos como indique el valor local de



K de su prototipo más cercano. Sin embargo, este método no ha demostrado tener un mejor rendimiento que el resto de los métodos antes comparados, al menos no en todos los casos.

Otra posibilidad es que el método sólo sea mejor en ciertas bases de datos (más robustas, con una mayor cantidad de características). Con las aplicadas en este proyecto no se logró replicar los resultados esperados.

El algoritmo de Bayes en este caso tuvo una eficiencia del 100%, incluso con menos datos de entrenamiento, aunque es un clasificador óptimo, los algoritmos Naïve Bayes son conocidos por ser pobres estimadores ya que este asume que la presencia o ausencia de una característica particular no está relacionada con la presencia o ausencia de cualquier otra característica, dada la clase variable. Por ello, no se deben tomar muy en serio las probabilidades que se obtienen.

Se recomienda que las investigaciones futuras a este tipo de problemas, se implementen algoritmos de aprendizaje no supervisado (*Deep learning*), con la finalidad de comprobar si se obtiene una mayor eficiencia de estos, así mismo se pueden utilizar de aprendizaje supervisado, pero con base de datos más robustas en ambos casos.

## Datos de la empresa.

Empresa: Instituto Tecnológico Superior de Guanajuato.

### a) Organigrama.

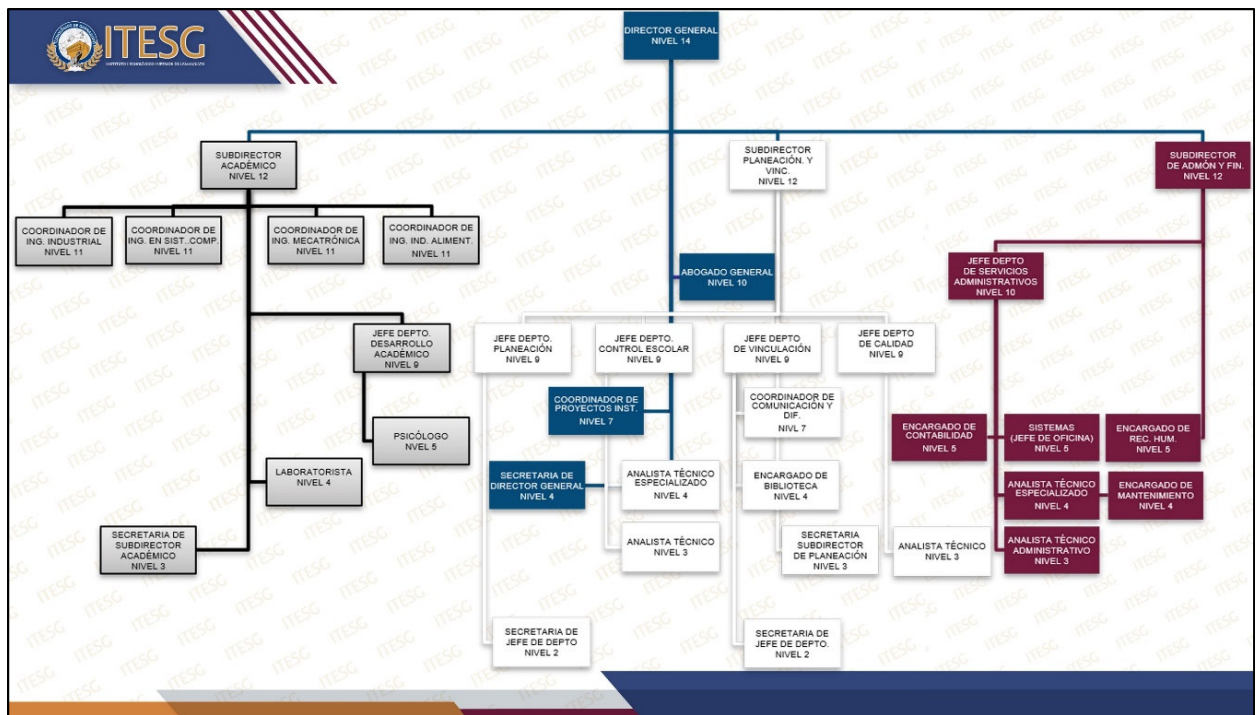


Figura 21: Organigrama del ITESG.

### b) Misión.

El ITESG es una Institución Pública de Educación Superior que forma parte del Tecnológico Nacional de México (TecNM), contribuye en el campo de la ingeniería al desarrollo tecnológico de la región y del país; dedicado a formar profesionistas competitivos con sentido social y humano. Para ello, el Instituto cuenta con certificación en calidad y cuidado ambiental, personal docente calificado, un modelo educativo que atiende las necesidades de los sectores productivo y social; con programas de vinculación que fortalecen el desarrollo académico, actividades complementarias y experiencia profesional que genera en los estudiantes una formación integral de calidad.

c) Política de calidad.

En el Instituto Tecnológico Superior de Guanajuato estamos comprometidos con la calidad y el cuidado ambiental a través de la implementación del Sistema Integral de Gestión de la Calidad y Ambiente (SIGCA), con el cual logramos mejorar continuamente los procesos y satisfacción de las partes interesadas.

Para ello se tiene que:

- Cumplir los objetivos y metas.
- Cumplir el marco normativo y legal aplicable.
- Implementar programas de comunicación y sensibilización del SIGCA.
- Procurar la disponibilidad de los recursos humanos, financieros y materiales.

d) Visión 2023.

Lograr un modelo de formación académica de excelencia y transformación integral humana con enfoque de calidad, innovación, Inclusión y equidad: con proyección nacional e internacional.

e) Objetivos.

1. Mantener y mejorar la eficacia del SIGCA.
2. Asegurar la satisfacción de nuestros estudiantes.
3. Ofrecer con pertinencia y oportunidad la cobertura educativa en el área de influencia del ITESG.
4. Fortalecer la calidad de los servicios educativos.
5. Disponer adecuadamente de los recursos humanos y de infraestructura asignados y asegurar su eficaz aprovechamiento en los procesos del ITESG.
6. Contribuir de manera proactiva a la sustentabilidad ambiental.

# **Anexos.**

## Clasificador Naïve-Bayes.

Adaptación del clasificador Bayes Ingenuo para los datos de 120 muestras de mermeladas, donde clasifica a que clase pertenece cada dato prueba y arroja la eficiencia total de este.

```
%% Clasificador Naïve-Bayes
```

```
clc  
clear all
```

```
%% Carga del conjunto de datos de entrenamiento con 120 valores.
```

```
Da=readtable('Entrenamiento120.csv');  
Datos=table2array(Da);  
x = Datos(:,1:3);  
y = Datos(:,6);
```

```
%% Carga de conjunto de datos de prueba
```

```
Pru=readtable('Prueba.csv');  
Prueba=table2array(Pru);  
x_prueba=Prueba(:,1:3);  
y_prueba=Prueba(:,6);
```

```
clear Da Pru
```

```
%% Definicion del modelo
```

```
ClasificadorNB = fitcnb(x,y);  
Mdl = fitcnb(x,y,...  
    'ClassNames',{'97','187','32','118','31','145','22','65','142','154'});  
    %devuelve un modelo Bayes ingenuo multiclase ( Mdl),  
    %entrenado por los predictores en la tabla x y las etiquetas de clase y
```

```
%%
```

```
%Por defecto, el software modela la distribución de predictores dentro de cada clase usando una distribución gaussiana
```

```
%que tiene alguna desviación estándar y media. Se usa la notación de puntos para mostrar los parámetros de un ajuste
```

```
%Gaussiano en particular, por ejemplo, mostrar el ajuste para la primera característica dentro de 'días 97'.
```

```
pruebaIndex = strcmp(Mdl.ClassNames,'97');  
estimates = Mdl.DistributionParameters{pruebaIndex,1}
```

```
%%
```

```
% Trazar los contornos gaussianos.
```

```

figure
gscatter(x(:,1),x(:,2),y);
h = gca;
xlim = h.XLim;
ylim = h.YLim;
hold on

title('Carga de datos L*, a*, b*, para prueba 97')
xlabel('L')
ylabel('A')
hold off

%%

[clase,posteriori,costo]=predict(ClasificadorNB,x_prueba) % Devuelve un vector de etiquetas
de clase pronosticadas para los datos
%del predictor en la tabla o matriz x_prueba, según el clasificador Bayes ingenuo entrenado,
completo o compacto ClasificadorNB.

%% Medición de la eficiencia del Clasificador Bayes Ingenuo
suma=0;
for indice=1:10
    if clase(indice)==y_prueba(indice)
        suma=suma+1;
    end
end

eficiencia=(suma/10)*100;

disp(strcat('Eficiencia: ',num2str(eficiencia)));
%%
nClase=12;
L1=Datos(1:nClase,1);
a1=Datos(1:nClase,2);
b1=Datos(1:nClase,3);

figure

subplot(1,3,1)
histogram(L1)

subplot(1,3,2)
histogram(a1)
title('Histograma L*, a*, b* para la clase de 97 dias')

subplot(1,3,3)
histogram(b1)

```

```

%%
x2=linspace(min(L1)-0.2,max(L1)+0.2,100);
mu=mean(L1);
sigma=std(L1);
y2=normpdf(x2,mu,sigma);
%%
minimo=0;
maximo=3.5;
n=5;
muVx=[mu,mu,mu,mu,mu];
muVy=linspace(minimo,maximo,5);
s=sigma;
sigma1Vx=[mu+s,mu+s,mu+s,mu+s,mu+s];
sigma1Vy=linspace(minimo,maximo,5);
sigma2Vx=[mu-s,mu-s,mu-s,mu-s,mu-s];
sigma2Vy=linspace(minimo,maximo,5);
y3=normpdf(L1,mu,sigma);
figure
plot(L1,y3,'b*');
% histogram(L1)
hold on
plot(x2,y2,'k-','LineWidth',1)
plot(muVx,muVy,'r-')
plot(sigma1Vx,sigma1Vy,'-r')
plot(sigma2Vx,sigma2Vy,'-r')
xlabel('L')
ylabel('frecuencia')
title('Gaussiana con \mu_L y \sigma_L^2')
legend('Clase de 97 dias')

```

## Clasificador KNN.

Adaptación del clasificador KNN para los mismos datos que se usaron en el clasificador Bayes, donde devuelve un modelo de clasificación de vecino más cercano y arroja la eficiencia total de este.

```
%% Clasificador KNN
clc
clear all
%% Carga del conjunto de datos de entrenamiento.
Da=readtable('Entrenamiento120.csv');
Pru=readtable('Prueba.csv');

Datos=table2array(Da);
x = Datos(:,1:3);
y = Datos(:,6);

Prueba=table2array(Pru);
x_prueba=Prueba(:,1:3);
y_prueba=Prueba(:,6);
clear Da Pru
%%
ClasificadorKNN = fitcknn(x,y); %devuelve un modelo de clasificación de vecino más
cercano k basado en los datos X y la respuesta del predictor Y.
%%
[clase,score,costo]=predict(ClasificadorKNN,x_prueba)%predecir la clasificación para un
clasificador vecino más cercano k .
rloss = resubLoss (ClasificadorKNN)%pérdida de restitución, que, por defecto, es la fracción
de clasificaciones erróneas
%de las predicciones de ClasificadorKNN
%%
suma=0;
for indice=1:10
    if clase(indice)==y_prueba(indice)
        suma=suma+1;
    end
end
eficiencia=(suma/10)*100;

disp(strcat('Eficiencia: ',num2str(eficiencia)));
%%
nClase=12;
ini=1;
fin=ini+nClase-1;
```



```
L1=Datos(ini:fin,1);
a1=Datos(ini:fin,2);
b1=Datos(ini:fin,3);
```

```
ini=73;
fin=ini+nClase-1;
L2=Datos(ini:fin,1);
a2=Datos(ini:fin,2);
b2=Datos(ini:fin,3);
```

```
ini=109;
fin=ini+nClase-1;
L3=Datos(ini:fin,1);
a3=Datos(ini:fin,2);
b3=Datos(ini:fin,3);
```

```
hold on
grid on
plot3(L1,a1,b1,'dk')
% legend('97 dias');
plot3(L2,a2,b2,'sk')
% legend('22 dias');
plot3(L3,a3,b3,'vk')
legend('97 dias','22 dias','154 dias');
plot3(x_prueba(4,1),x_prueba(4,2),x_prueba(4,3),'+r')
plot3(x_prueba(7,1),x_prueba(7,2),x_prueba(7,3),'*r')
plot3(x_prueba(1,1),x_prueba(1,2),x_prueba(1,3),'xr')
xlabel('L*')
ylabel('a*')
zlabel('b*')
```

## Red Neuronal.

Codificación de la red Neuronal para la clasificación de muestras de mermeladas, en relación a una base de datos de 80 pruebas de entrenamiento.

```
clc
clear all
%%
load DatosEntrenamiento80.mat
load ClasesEntrenamiento80.mat

%%
Clases=10;
Muestras=8;
total=Muestras*Clases;
Objetivo=zeros(total,Clases);
for indice=1:Clases
    Objetivo(((indice-1)*8)+1:indice*8,indice)=1;
end
%%
x=Datos';
t=Objetivo';
%%
net = patternnet(20);
net = train(net,x,t);
%%
view(net)
y = net(x);
perf = perform(net,t,y);
classes = vec2ind(y);

%%
Clasificados=net(x);
```

## Referencias.

1. M. L. Carrillo Inungaray, A. Reyes Munguía, “Vida útil de los alimentos”, Revista Iberoamericana de las Ciencias Biológicas y Agropecuarias, Vol. 2, Núm. 3, ISSN 2007-9990, pp 1-2 5.
2. N. Zuluaga Arroyave, “El análisis sensorial de alimentos como herramienta para la caracterización y control de calidad de derivados lácteos”, Tesis de Maestría, Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Medellín, Colombia, 2017.
3. J. Mercado-Flores, M López-Orozco, G. Martínez-Soto y M. R. Abraham-Juárez. “Estimación de la Vida de Anaquel Mediante Pruebas Aceleradas en Fresa Entera en Bolsa de Polietileno y Pulpa de Fresa Congelada”, Investigación y Desarrollo en Ciencia y Tecnología de Alimentos, Vol. 1 No.1, 2016, pp 636-640.
4. T. P. Labuza, “Shelf-Life Dating of Foods”, Food and Nutrition Press. USA, 1982.
5. D. S. Badui, “Química de los alimentos”. Pearson Educación: 5ta.Ed., México, 2013.
6. M. A. Poiana, M., M. F. Munteanu, D. M. Bordean, R. Gligor, E. Alexa. “Assessing the effects of different pectins addition on color quality and antioxidant properties of blackberry jam”. Chemistry Central Journal, Vol. 7 Issue 121, 2013, pp 1-13.
7. Norma Oficial Mexicana NOM-251-SSA1-2009, Prácticas de Higiene para el Proceso de Alimentos, Bebidas o Suplementos Alimenticios. Diario Oficial de la Federación, 01/03/10.
8. Food Safety System Certification (FSSC) 22,000. Framework for effectively managing food safety responsibilities. <http://www.fssc22000.com>.
9. ASTM E2454-05(2011), Standard Guide for Sensory Evaluation Methods to Determine the Sensory Shelf Life of Consumer Products, ASTM International, 2011, [www.astm.org](http://www.astm.org).

- 10.A. Bhargava A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review", *Journal of King Saud University - Computer and Information Sciences*, 2018, pp 1-31.
- 11.C. Montoya Holguin, J. A. Cortés Osorio y J. A. Chaves Osorio, "Sistema automático de reconocimiento de frutas basado en visión por computador", *Ingeniare. Revista chilena de ingeniería*, Vol. 22 No. 4, 2014, pp 504-516.
- 12.Y. Zhang, S. Wang, G. Ji, P. Phillips, "Fruit classification using computer vision and feedforward neural network", *Journal of Food Engineering*, Volume 143, 2014, pp 167-177.
- 13.S. Tippannavar & S. Soma, "A Machine Learning System For Recognition Of Vegetable Plant And Classification Of Abnormality Using Leaf Texture Analysis". *International Journal of Scientific & Engineering Research*. Vol. 8 No. 6, 2017, pp 1558-1563.
- 14.M. G. Corradini, "Shelf Life of Food Products: From Open Labeling to Real-Time Measurements", *Annual Review of Food Science and Technology*, 2018.
- 15.S. Russell, P. Norving, "Artificial Intelligence: A Modern Approach, 3rd Edition", Prentice Hall, USA, 2009.
- 16.S. Lim, K. Lee, O. Byeon, T. Kim, "Efficient Iris Recognition through Improvement of Feature Vector and Classifier", *ETRI Journal*, Vol. 23 Issue 2, 2001, pp 61-70.
- 17.O. Kufandirimbwa, R. Gatora, "Spam Detection Using Artificial Neural Networks (Perceptron Learning Rule)", *Online Journal of Physical and Environmental Science Research*, Vol. 1 Issue 2, 2012, pp 22-29.
- 18.E. N. Malamas, E. G. M. Petrakis, M. Zervakis, L. Petit, J.-D. Legat, "A survey on industrial vision systems, applications and tools", *Image and Vision Computing*, Vol. 21 Issue 2, 2003, pp. 171-188.
- 19.M. Karabatak, M. Cevdet-Ince, "An expert system for detection of breast cancer based on association rules and neural network", *Expert Systems with Applications*, Vol. 36 Issue 2 (Part 2), 2009, pp. 3465-3469.
- 20.R. O. Duda, P. E. Hart, D. G. Stork, "Pattern Classification, 2nd Edition", Wiley, USA 2000.

21. T. M. Mitchell, "Machine Learning", McGraw-Hill Science/Engineering/Math, USA, 1997.
22. B. Martín-Del-Brio, A. Sanz-Molina, "Redes Neuronales Y Sistemas Borrosos, 3ª Edición", Ra-Ma, España, 2006.
23. C. M. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, USA, 1996.
24. J. R. Hilera-González, V. J. Martínez-Hernando, "Redes neuronales artificiales: fundamentos, modelos y aplicaciones", Ra-Ma, España, 1995.
25. Disponible en: <https://www.gob.mx/agenda2030/articulos/objetivo-de-desarrollo-sostenible-2-hambre-cero>, ultimo acceso enero 2019.
26. El UNIVERSAL, Compañía Periodística Nacional S. A. de C. V. Ciencia y Tecnología. Recuperado de: <https://www.eluniversalqueretaro.mx/ciencia-y-tecnologia/05-11-2017/la-dulce-mermelada-de-guanajuato>.
27. Guillermo Pantaleo, Ludmila Rinaldo, "Ingeniería de Software", Alfaomega Ediciones; Edición: 1, 2015.
28. Disponible en: <https://noticias.guanajuato.gob.mx/2018/02/guanajuato-ocupa-tercer-lugar-produccion-fresas-nivel-nacional-mmm/>
29. Carpenter, R. P., Lyon, D. H. & Hasdell, T. A. (2002). Análisis sensorial en el desarrollo y control de la calidad de alimentos. Zaragoza, España: Ed. Acribia.
30. J. R. Hilera; V. J. Martínez; "Redes neuronales artificiales. Fundamentos, modelos y aplicaciones". Ed. Rama, 1995.
31. S. Haykin; "Neural networks. A comprehensive foundation". IEEE Press, 1994.
32. Klaus Schwab; "La cuarta revolución industrial", Penguin Random House Grupo Editorial (19 de junio de 2017).
33. Disponible en: <https://www.kioscodelahistoria.com/single-post/fresa-gto-tercer-lugar>