



## **DEPARTAMENTO DE POSGRADOS**

**“Método de algoritmo genético (GA-VSS) aplicado en la  
selección de variables significativas del proceso de  
industrialización para un modelo predictivo de la calidad del  
atún en conserva”**

**MAGÍSTER EN GESTIÓN DE LA CALIDAD Y  
SEGURIDAD ALIMENTARIA**

**Autor:**

**Nancy Janneth Contreras Parra**

**Director:**

**Dr. Piercósimo Tripaldi Cappelletti**

**Cuenca, Ecuador**

**2015**

## Dedicatoria

Cuando inicié este camino, jamás hubiese imaginado en dónde terminaría. Solamente supe que era el correcto, supe con certeza que era el destino al que quería llegar, pero lo cierto es que la mejor parte de todo no fue el final, sino el viaje en sí mismo...

Dedico mi trabajo y esfuerzo a mi **Dios Jesucristo**, en quien están escondidos todos los tesoros de la sabiduría y del conocimiento (Col 2:3). En su infinita bondad y misericordia me ha llevado por el buen camino, me ha pedido esforzarme y ser valiente, no temeré ni desmayaré, porque Él está conmigo, dondequiera que vaya (Josué 1:9).

En mi corazón y en mi mente, siempre mis hijos **Milena y Ezequiel**, bendecidos días nos esperan gracias a cada pasito que dieron a mi lado, a cada minuto de compañía del que les privé, por cada comida de prisa, por cada maleta y cada viaje que hicimos juntos. Dedicado para ustedes todo desafío superado, porque por ustedes tiene sentido mi vida y para ustedes vivo y muero.

**César**, no imagino mi vida sin ti a mi lado. Tu apoyo incondicional y tu insistencia me han llevado a culminar esta etapa. Sin duda cada logro es aún más grande si estamos juntos.

Agradezco al Señor haberte puesto en mi camino, por habernos dado la sabiduría para afrontar los problemas, pero sobre todo le agradezco por ser fuente infinita de amor, por ese amor que todo lo sufre, todo lo cree, todo lo espera, todo lo soporta. Por ese amor que nos tenemos y que nunca dejará de ser (1Cor13 7:8).

## Agradecimientos

La mejor manera de agradecer el apoyo, es que el fruto del esfuerzo valga la pena.

Mi agradecimiento y respeto al **Doctor Piero Tripaldi**, perenne investigador, por sus conocimientos y su guía desde la concepción misma del tema de esta tesis. Gracias a su paciencia y perspicacia para trazar un camino inteligible hasta alcanzar la meta planteada.

De manera especial agradezco a los ingenieros **Jelisava y Andrés Cuka Auad**, Directora General y Director de Operaciones de Marbelize S.A. por la confianza depositada al entregar la información y el espacio necesarios para desarrollar esta investigación.

A mi familia, mi madre **Betty**, mi padre **Jaime, Daniel** mi hermano, **abuelita Vicky**. Todos estuvieron presentes de una u otra manera animándome a cada paso, antes de iniciar cada viaje. Fueron testigos del esfuerzo invertido y fueron también una motivación y mi bendición.

Gracias a la **familia Azuero Parra**, por ser un segundo hogar para mí desde siempre.

A la **familia Monge Parra** por darme la mano cuando más lo necesité.

## **Resumen**

El presente estudio plantea la aplicación del Método de algoritmo genético (GA-VSS) para la selección de variables significativas del proceso de industrialización del atún, con el fin de construir un modelo predictivo de la calidad. Basado en el análisis cuasi-experimental de la producción, la función objetivo es el cumplimiento del peso escurrido de conservas en envase flexible. Tres niveles de calidad fueron establecidos para categorizar los casos de estudio, realizándose un pre-tratamiento y depuración de los datos. La selección derivó en cuatro variables significativas que se clasificaron mediante Algoritmo de árbol de decisiones CART para su aplicación en el proceso.


## **Palabras clave:**

Algoritmo genético, GA-VSS, KNN, CART, peso escurrido, optimización.

## ABSTRACT

This study presents the application of the genetic algorithm (GA-VSS) method for the selection of significant variables of tuna industrialization process, in order to build a predictive model of quality. Based on the quasi-experimental analysis of production, the objective function is compliance with drained weight of canned flexible packaging. Three quality levels were established to categorize the case studies, performing a pre-treatment and data cleaning. The selection resulted in four significant variables that were classified using CART decision tree algorithm for its application in the process.

**Keywords:** Genetic Algorithm, GA-VSS, KNN, CART, drained weight, optimization



Translated by,  
Lic. Lourdes Crespo

**Índice de contenido**

<b>Dedicatoria</b> .....	<b>ii</b>
<b>Agradecimientos</b> .....	<b>iii</b>
<b>Resumen</b> .....	<b>iv</b>
Palabras clave: .....	iv
<b>Abstract</b> .....	<b>v</b>
Keywords: .....	v
<b>Índice de contenido</b> .....	<b>vi</b>
<b>Índice de Figuras</b> .....	<b>viii</b>
<b>Índice de Tablas</b> .....	<b>ix</b>
<b>Índice de Anexos</b> .....	<b>x</b>
<b>INTRODUCCIÓN</b> .....	<b>1</b>
Objetivo general:.....	2
Objetivos específicos: .....	2
<b>CAPÍTULO 1: MATERIALES Y MÉTODOS</b> .....	<b>3</b>
1.1. Selección de los datos .....	4
1.2. Pre procesamiento de los datos.....	6
1.3. Depuración de los datos .....	7
1.4. Escalado de los datos (scaling) .....	8
1.5. Hardware y Software.....	8
1.6. Algoritmos genéticos .....	9
1.7. Operadores del Algoritmo Genético:.....	11
1.7.1. Selección.....	11
1.7.2. Reproducción .....	11
1.7.3. Mutación.....	11
1.7.4. Reinserción .....	12

1.8.	Clasificación mediante algoritmo KNN.....	12
1.9.	Método de Clasificación por árboles de decisión (CART) .....	12
<b>CAPÍTULO 2: RESULTADOS.....</b>		<b>14</b>
2.1.	Selección de Variables mediante Algoritmo Genético GA-VSS .....	14
2.2.	Predicción y depuración de los resultados .....	15
2.3.	Validación del modelo .....	15
2.3.1.	Validación cruzada (cross-validation) .....	16
2.3.2.	Validación externa (test set).....	17
<b>CAPÍTULO 3: DISCUSIÓN.....</b>		<b>20</b>
<b>CAPITULO 4: CONCLUSIONES .....</b>		<b>27</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>		<b>29</b>
<b>Bibliografía .....</b>		<b>29</b>
<b>Referencias de consulta.....</b>		<b>30</b>
<b>ANEXOS .....</b>		<b>31</b>

## Índice de Figuras

Figura 1 Procesamiento de información de tiempos de proceso.....	6
Figura 2 Procesamiento de Información ingresada desde registros físicos .....	7
Figura 3 Procedimiento del Algoritmo Genético (Todeschini, Consonni, & Pavan, 2012) ....	10
Figura 4 Distribución de frecuencias en la selección de variables significativas.....	15
Figura 5 Modelo escalado de predicción de la calidad resultante en conservas de atún .....	18
Figura 6 Modelo de árbol de decisión con valores sin normalizar.....	19
Figura 7 Modelo CART. Identificación de escenarios para los productos 1, 3 y 7Kg.....	26



## Índice de Tablas

Tabla 1 Variables de producción que conforman la base de datos.....	5
Tabla 2 Clasificación y criterio de desviación al peso escurrido especificado. ....	5
Tabla 3 Variables seleccionadas por Algoritmo Genético .....	14
Tabla 4 Matriz de confusión del modelo CART .....	16
Tabla 5 Cros validación del Modelo CART utilizado en la clasificación .....	16
Tabla 6 Matriz de confusión del modelo CART .....	17
Tabla 7 Validación externa del modelo de clasificación CART .....	17
Tabla 8 Valores de los umbrales de decisión revertidos del escalado o normalización. ....	19

## **Índice de Anexos**

Anexo 1 Diagrama de flujo del proceso de producción de conservas de atún.....	31
Anexo 2 Productos que conforman el 64% de las exportaciones no petroleras. Enero-abril 2014. Fuente: Dirección de Inteligencia Inversiones, PROECUADOR. ....	32
Anexo 3 Evolución de las exportaciones no petroleras de Ecuador. Sector Pesca. Enero-abril 2014. Fuente: Dirección de Inteligencia Inversiones, PROECUADOR .....	32
Anexo 4 Metodología de evaluación del Peso escurrido según Codex Stan 70-1981.....	33
Anexo 5 Conservas de atún en envase flexible.....	33
Anexo 6 Plantilla de levantamiento de información para base de datos. Fuente: Elaboración propia. ....	34
Anexo 7 Vista Panorámica Planta de Procesamiento de Atún Marbelize S.A. ....	35

Contreras Parra Nancy Janneth.

Trabajo de Graduación

Dr. Tripaldi Cappelletti Piercósimo

Diciembre, 2015

**“Método de algoritmo genético (GA-VSS) aplicado en la selección de variables significativas del proceso de industrialización para un modelo predictivo de la calidad del atún en conserva”**

**INTRODUCCIÓN**

Según cifras de la FAO, Ecuador captura aproximadamente el 4 por ciento del total mundial de atún. Las actividades de pesca se orientan principalmente a la captura de tres especies conocidas como: atún aleta amarilla o Yellowfin (*Thunnus albacares*), atún ojo grande o Bigeye (*Thunnus obesus*) y atún barrilete o Skipjack (*Katsuwonus pelamis*). En la actualidad, el país cuenta con una flota de 90 barcos de bandera nacional y 2 embarcaciones extranjeras asociadas, equipados para conservar alrededor de 46.000 toneladas mediante sistema de refrigeración combinada con salmuera.

El perfil del país, indica que la pesquería industrial del atún inició en 1952 en el puerto de Manta y desde entonces ha venido evolucionando de manera tal que, según informes de Pro Ecuador, en el año 2013 representó el 9.69% de las exportaciones no petroleras, superado únicamente por el banano. Aportó alrededor de \$347.985 FOB USD a la balanza comercial, con un total de 60 mil toneladas de producto en exportaciones. Ecuador es el tercer país Latinoamericano con las mayores exportaciones de productos de la pesca, lo antecede Chile seguido de Perú.

Pro Ecuador indica, que de los principales importadores mundiales del sector pesca, el principal comprador de producto ecuatoriano es España con una participación de mercado del 14.88%, seguido de Estados Unidos de América con un 14.60% y en tercer puesto se ubica Holanda con un 9.65%. Los mercados que más han crecido incluyen Suecia con 8.84%, República de Corea con 6.39% y Tailandia con 6.34%.

Marbelize S.A., ubicada en Jaramijó – Manabí y que ocupa una importante posición entre las empresas exportadoras ecuatorianas del sector pesca, busca continuamente satisfacer las rigurosas exigencias de los mercados internacionales. En este contexto el rol empresarial se direcciona a la búsqueda de estrategias que permita optimizar los procesos en términos de la percepción de la calidad para el consumidor, generando ventajas competitivas en su sector industrial y consolidando las relaciones comerciales internacionales.

La presente investigación se encamina a determinar las mejores prácticas aplicables para el proceso tecnológico y reconocer las características de la materia prima que permitan predecir y maximizar aquellos factores que serán determinantes en la calidad resultante de la fabricación de conservas de atún, expresada como la capacidad de adsorción en la proteína de los lomos que favorezca retención de los líquidos de gobierno, es decir, maximizando el rendimiento pretendido en tanto se mantengan las especificaciones pactadas con los clientes para los productos elaborados.

La identificación de las variables significativas asociadas a la calidad resultante del atún en conserva, permitirán concebir un modelo de predicción estableciendo umbrales de control en etapas críticas del proceso y en las características específicas de la materia prima, infiriendo los resultados esperados acorde a los estándares de producción.

**Objetivo general:**

Aplicar el Método de Algoritmo Genético para seleccionar variables significativas del proceso de industrialización para construir un modelo predictivo de la calidad de las proteínas presentes en los lomos de atún y su capacidad de retención del líquido de cobertura.

**Objetivos específicos:**

- Determinar los factores característicos aportados por el manejo post captura del atún, que benefician la preservación de la calidad de las proteínas del pescado y sus propiedades para retener el líquido de gobierno utilizado en las conservas de atún.
- Determinar los factores críticos a controlar en el proceso tecnológico, beneficiando la preservación de la calidad de las proteínas del pescado y sus propiedades para retener el líquido de gobierno utilizado en las conservas de atún.
- Desarrollar un modelo matemático de predicción del comportamiento de las proteínas del pescado ante el proceso industrial aplicado en la empresa Marbelize S.A., que permita estandarizar las características de calidad y predecir el factor de adsorción de los lomos utilizados en la preparación de las recetas.

## **CAPÍTULO 1: MATERIALES Y MÉTODOS**

El proceso de elaboración de conservas de atún se enfrenta a un cambio sustancial, debido al requerimiento de exclusión de proteínas vegetales de las recetas, utilizado ampliamente en el sector atunero como coadyuvante para beneficiar la retención del líquido de cobertura. Este requerimiento especial de los mercados europeos se debe a la constante evolución en la búsqueda de alimentos libres de alérgenos, demandando la implementación y cumplimiento de nuevos requisitos, así como la necesidad de adaptar los procesos tecnológicos y productivos industriales, de manera que permita conservar los estándares de calidad y por lo tanto la satisfacción de clientes y consumidores con regímenes de dieta específicos.

Este cambio ha derivado en una problemática interna a nivel local, considerando la diversidad del manejo post captura y posterior procesamiento. La incertidumbre sobre el efecto de estos factores en la capacidad de la proteína del lomo para adsorber el líquido de cobertura, pone también en conflicto el rendimiento y por ende la rentabilidad económica de la industria privada.

El eje de la problemática se centra, por lo tanto, en identificar la combinación de condiciones del manejo y procesamiento que aseguren el cumplimiento del peso escurrido que se evalúa en las conservas y consecuentemente la satisfacción de los requisitos establecidos para el producto final

Inicialmente, la compañía ha implementado medidas específicas de control en la dosificación, pesaje de atún e ingredientes, manteniendo cantidades estándar en producto y líquidos de gobierno. Sin embargo, la determinación en laboratorio del peso escurrido de la conserva, a través de la metodología del Codex Stan 70, varía significativamente entre diferentes producciones e incluso entre muestras del mismo lote.

La propuesta desarrollada, se direcciona en evaluar el conjunto de condiciones, especificaciones y características del proceso productivo mediante la aplicación del Método de Algoritmo Genético.

De esta metodología se seleccionarán las variables significativas para la función objetivo y se construirá un modelo predictivo respecto a la calidad resultante de las proteínas presentes en los lomos de atún, clasificándolas o agrupándolas según su capacidad de retención del líquido de cobertura, factor expresado como el peso escurrido o peso drenado analizado en el producto final con una metodología estándar.

Se pretende por lo tanto, determinar los factores característicos aportados por el manejo post captura del atún, que benefician la preservación de la calidad de las proteínas del pescado y sus propiedades para retener o adsorber el líquido de gobierno utilizado en las conservas de atún.

Paralelamente se identificarán factores críticos a controlar en el proceso tecnológico, mediante un modelo de predicción del comportamiento de las proteínas del pescado ante el proceso industrial aplicado en la empresa Marbelize S.A., estandarizando las características resultantes de calidad e infiriendo el factor de adsorción de los lomos.

### **1.1. Selección de los datos**

El estudio se llevó a cabo con una metodología cuasi experimental, basada en el análisis de información histórica del proceso tecnológico de la elaboración de conservas de atún en envase flexible (pouch) correspondiente al año 2014. La construcción de la base de datos incluyó especificaciones de producto terminado, características de la materia prima, tiempos de fabricación y traslados, dosificaciones, rendimientos, características físico químicas en producto terminado y evaluación sensorial o “grading”.

Para el objeto de estudio se dispuso de información de aproximadamente 700 lotes fabricados, identificando a lo largo del proceso de fabricación desde el origen de la materia prima, descarga, almacenamiento, descongelamiento, cocción, nebulizado, limpieza, empackado y esterilización. Esta información ha sido principalmente obtenida de archivo del software de control disponible en la compañía, mismo que almacena información respecto a los distintos tiempos, temperaturas y pesos durante las etapas de procesamiento dentro de un periodo determinado.

Se ha obtenido información adicional por trazabilidad de los lotes a partir de la codificación del producto, relacionándola con información del origen de la materia prima utilizada y las especificaciones determinadas para cada producto según formulaciones y recetas.

Los resultados de la evaluación de producto terminado, tanto de las características físico químicas, organolépticas (grading) y de control de la calidad se obtuvo a partir de registros físicos (archivos de la empresa) de la calificación de los atributos de una muestra individual por cada lote secuencial.

Las variables mencionadas en la Tabla No. 1 fueron correlacionadas con el resultado obtenido en la función objetivo denominada “Peso escurrido” o “peso drenado”. Este resultado se obtiene del análisis de producto terminado con la mencionada metodología del Codex Stan 70-1981 norma para el atún y el bonito en conserva.

Variables Ingresadas	
▪ Especie	▪ Grated especificación (requerido)
▪ Talla de pescado	▪ Peso escurrido especificación (requerido)
▪ Tiempo de almacenamiento en cámara	▪ Líquido de cobertura usado
▪ Tiempo en descongelación por especie y talla	▪ Densidad (kg de producto/volumen del envase)
▪ Variación de temperatura al descongelar	▪ pH en producto terminado
▪ Horno usado para cocción	▪ %sal en producto terminado
▪ Tiempo cocción	▪ Histamina en producto terminado
▪ Pérdida de peso tras la cocción (merma)	▪ Grated o migas producto terminado
▪ Temperatura de cocción	▪ Peso neto producto terminado
▪ Variación de temperatura al cocinar	▪ Color evaluado en grading
▪ Tiempo en nebulizado	▪ Textura evaluada en grading
▪ Número de autoclave utilizado	▪ Sabor evaluada en grading
▪ Tiempo total esterilización	▪ Adsorción esperada por kg de producto
▪ Tiempo total operaciones unitarias	▪ Diferencial del peso neto a la especificación
▪ Diferencial del tiempo de esterilización al estándar	▪ Diferencial de cantidad de migas a la especificación
▪ Fill o dosificación de pescado	▪ Tiempo total fabricación
▪ Peso neto especificación (requerido)	▪ Tiempos muertos (espera)

Tabla 1 Variables de producción que conforman la base de datos.

Conforme los resultados obtenidos por cada muestra y, en función del grado de cumplimiento de pesos escurridos analizados bajo esta metodología, se estableció una calificación enmarcada en tres categorías: buena (clase 1), regular (clase 2) y mala (clase 3).

Los criterios de calificación se detallan a continuación:

	CRITERIO POR DESVIACIÓN	ESP. CLIENTE	RESULTADO MÍNIMO 1KG	ESP. CLIENTE	RESULTADO MÍNIMO 3KG	ESP. CLIENTE	RESULTADO MÍNIMO 7KG
<b>CLASE 1</b>	< o =1%	950	941	2925	2896	6750	6683
<b>CLASE 2</b>	1,1% a 4,9%	950	926	2925	2852	6750	6581
<b>CLASE 3</b>	= o >5%	950	903	2925	2779	6750	6413

Tabla 2 Clasificación y criterio de desviación al peso escurrido especificado.

La categorización de los resultados en función del cumplimiento del peso escurrido, se realizó calificando cada muestra en una escala de 1 a 3 siendo 1 el mejor resultado, con un margen de cumplimiento igual o superior al 99% de la especificación del cliente. En la categoría 2 se encasillaron casos con niveles de cumplimiento entre 95,1% y 98,9% y en la categoría 3 aquellos que se encuentren igual o por debajo del 95% de cumplimiento a la especificación.

En la Tabla 2 se puede observar el impacto de la desviación en el resultado final del peso escurrido, siendo éste cada vez más notorio en tanto mayor sea el tamaño de la presentación del producto. Aun considerando que los requisitos de cliente son de cumplimiento obligatorio, con esta clasificación se pretende determinar el resultado máximo que se puede obtener en las condiciones actuales de procesamiento, conforme se seleccionen y establezcan límites de trabajo para cada variable relevante.

## 1.2. Pre procesamiento de los datos

El pre-procesamiento de datos es un tema importante tanto para el almacenamiento de datos como para la minería de datos, debido a que los datos del mundo real suelen ser incompletos, ruidosos, e inconsistentes. Incluye la limpieza, integración, transformación y la reducción de datos (Han, Jiawei; Kamber, Micheline, 2006).

La compañía cuenta con un software para la captura instantánea de información a lo largo de la producción, misma que almacena la información y permite importar la información consolidada. Esta información fue migrada a Excel y posteriormente fue filtrada y revisada para detectar inconsistencias del registro en línea. La información general se resumió mediante tablas dinámicas y se utilizaron funciones matriciales (índice) para importar información de los casos analizados. En las etapas de descongelación y cocción, se promediaron los tiempos de cada proceso en el día, agrupándolos por barco, especie y talla.

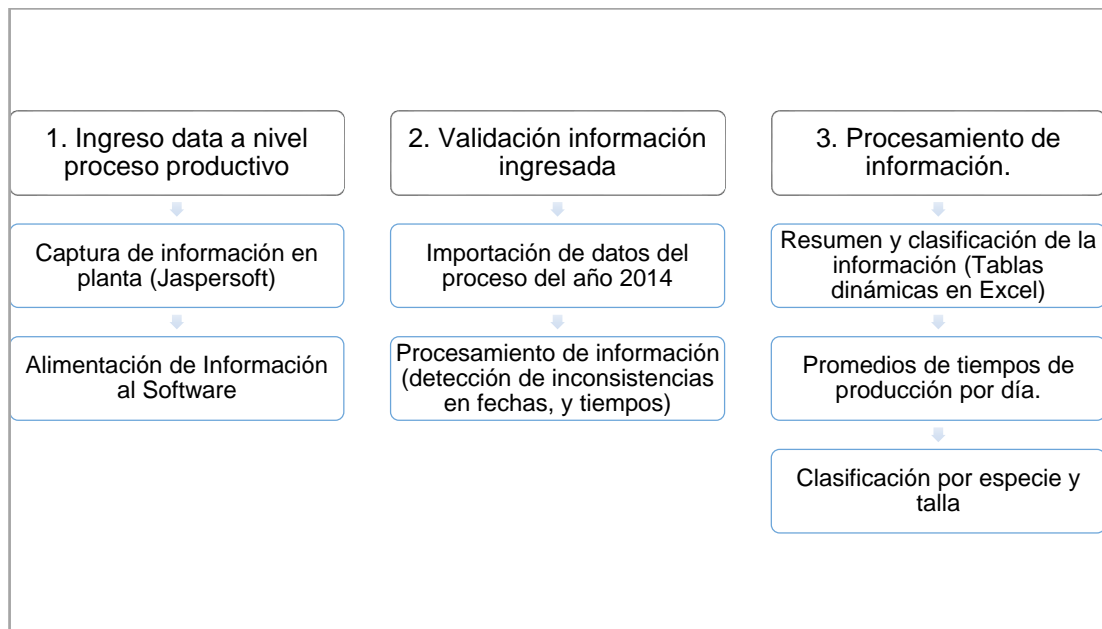


Figura 1 Procesamiento de información de tiempos de proceso

Los resultados de las evaluaciones de producto terminado y evaluación sensorial “grading” se digitaron desde registros físicos, y la información requerida fue importada posteriormente a la base de estudio. Para reducir la probabilidad de errores durante la etapa de levantamiento de información de la trazabilidad se utilizaron herramientas de Excel para realizar una validación de los datos, tanto en extensión de texto como en combinación de caracteres.

Se utilizó como referencia las especificaciones utilizadas para la interpretación de los lotes de fabricación, listados y codificación de barcos y lotes de materia prima, así como el listado general de especificaciones de productos (incluyendo pesos netos, pesos escurridos, porcentajes de migas, líquido de cobertura, y presentación).

Las variables discontinuas o cualitativas se colocaron de manera independiente para ser calificadas como una particularidad de presencia o ausencia de esta condición.



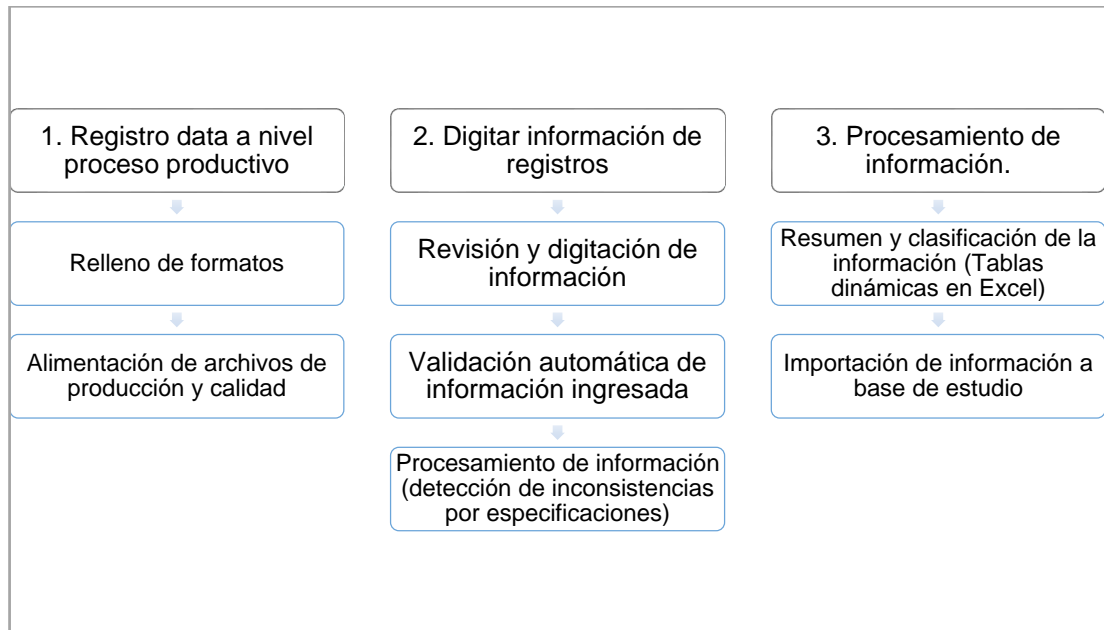


Figura 2 Procesamiento de Información ingresada desde registros físicos

### 1.3. Depuración de los datos

A través de este proceso, se obtiene una representación reducida del conjunto de datos, que sin embargo, produce el mismo (o muy similar) resultado posterior al análisis.

Es de saber que hay un sinnúmero de estrategias para la reducción de datos. Éstas incluyen la agregación de datos, la selección subconjunto, la reducción de dimensionalidad y la reducción de cantidad. (Han, Jiawei; Kamber, Micheline, 2006)

Se fundamenta en la detección de la discrepancia de la información, utilizando como punto de partida, la noción previa respecto a las propiedades de los datos y conocido como “meta-data” (Han, Jiawei; Kamber, Micheline, 2006).

En lo que refiere a la base de datos de estudio, se separaron aquellos casos que no contaron con información completa de las variables ingresadas o cuyo contenido no superó la validación mencionada en el punto anterior. Para ello se aplicó la herramienta de filtrado y se seleccionaron los casos cuyas fórmulas dieron como resultado error de cálculo.

De esta manera se evitan inconvenientes en la etapa de análisis de los datos debido a datos incompletos (falta valores de atributos o ciertos atributos de interés), ruidosos (que contiene errores, o valores atípicos), e inconsistentes (por ejemplo, contiene discrepancias en los códigos de recetas) (Han, Jiawei; Kamber, Micheline, 2006).

#### 1.4. Escalado de los datos (scaling)

Para ayudar a evitar la dependencia de la elección de las unidades de medida, los datos deben ser normalizados. La estandarización de las mediciones intenta dar todas las variables de un peso igual. Esto es particularmente útil cuando se administra sin conocimiento previo de los datos. La normalización implica escalar todos los valores para un atributo dado para que caigan dentro de un pequeño rango especificado, como -1,0 a 1,0, o 0,0 a 0,1. (Han, Jiawei; Kamber, Micheline, 2006).

El valor normalizado es definido como:

$$MM(X_{ij}) = \frac{X_{ij} - X_{\min}}{X_{\max} - X_{\min}}$$

(Ismail Bin & Dauda, 2013)

Donde ( $x_{ij}$ ) es el valor normalizado,  $x_{ij}$  es el valor original,  $x_{\min}$  y  $x_{\max}$  son los valores más bajos y más altos de la serie, respectivamente. (Ismail Bin & Dauda, 2013)

#### 1.5. Hardware y Software

A través de la herramienta de captura y recuperación de la información Microsoft SQL Server Management Studio se obtuvo información del proceso productivo, misma que se almacena en el Software Microsoft SQL SERVER 2000.

Para el análisis de datos se utilizó un ordenador del Laboratorio de Química de la Universidad del Azuay, con un sistema operativo Windows 8 con procesador Core i7. Se utilizó para el análisis y consolidación de la información el programa Microsoft Office Excel 2010.

Para el cálculo de GAs se utilizó el Software MATLAB (versión 10, Mathworks, Inc.). MATLAB dispone de numerosas herramientas llamadas "toolbox", que le añaden funcionalidades especializadas. Es una colección de módulos para cálculos de clasificación (reconocimiento supervisado de patrones) de modelos multivariantes: Árboles de Clasificación (CART), K-Nearest Neighbors (kNN), Algoritmos genéticos GA-VSS, entre otros. (Ballabio & Consonni, 2013).

Estos cálculos se realizaron en el entorno MATLAB, con el complemento o Toolbox de Algoritmos genéticos creado por el Profesor Roberto Todeschini y proporcionado bajo convenio entre el Departamento de Ciencias Ambientales y Ciencias de la Tierra y de la Tierra de la Universidad de Milán–Bicocca y la Universidad del Azuay.

## 1.6. Algoritmos genéticos

Los Algoritmos Genéticos (GAs) son un método basado en variables, presentado por primera vez en 1961 por Bledsoe y matemáticamente formalizada por Holland en 1975, que se inspira en la teoría de Darwin de la evolución (Cassotti & Grisoni, 2012).

Fue posible aplicarlos con tiempos razonables de análisis apenas en la década de 1990, cuando las computadoras se volvieron mucho más rápidas. (Noorizadeh & Farmany, 2011)

Los algoritmos genéticos (GAs) son un método evolutivo ampliamente utilizado para problemas de optimización complejos en varios campos como la robótica, la química y la QSAR (Relaciones cuantitativas estructura-actividad, por sus siglas en inglés). (Todeschini, Consonni, & Pavan, 2012)

En comparación con los procedimientos de búsqueda y optimización tradicionales, GAs es robusto, global y generalmente más sencillo de aplicar a situaciones en las que se dispone de poco o ningún conocimiento a priori sobre el proceso a controlar. No se requiere información derivada o una estimación inicial formal de la función objetivo y debido a la naturaleza estocástica del mecanismo de búsqueda, es capaz de buscar en todo el espacio de la solución con una mayor probabilidad de encontrar el óptimo global (función objetivo). (Noorizadeh & Farmany, 2011)

Dado que los sistemas complejos se describen por varias variables, una meta importante en el análisis del sistema es la extracción de la información pertinente, junto con la exclusión de información redundante y ruidosa. La selección de variables se lleva a cabo por GAs considerando poblaciones de modelos generados a través de un proceso de reproducción y optimizado de acuerdo a una función de objetivo definido en relación con la calidad del modelo. El procedimiento se ilustra en la Figura 1. (Todeschini, Consonni, & Pavan, 2012)

De acuerdo con la terminología de GAs, cada gen corresponde a una variable y una secuencia de genes, es decir, un cromosoma a un modelo. La población de cromosomas es inicializado al azar y la presencia / ausencia de cada variable está codificada por un dígito binario. Los cromosomas se evalúan según su calidad, de acuerdo con una función de aptitud predefinida y son ordenados en consecuencia.

Consiste en la evolución de una población de modelos, es decir, un conjunto de modelos clasificados de acuerdo con alguna función objetivo, sobre la base de los procesos de cruce y mutación, que se repiten alternativamente hasta que se encuentra una condición de parada (por ejemplo, un número máximo definido por el usuario de iteraciones) o el proceso se termina arbitrariamente. (Cassotti & Grisoni, 2012)

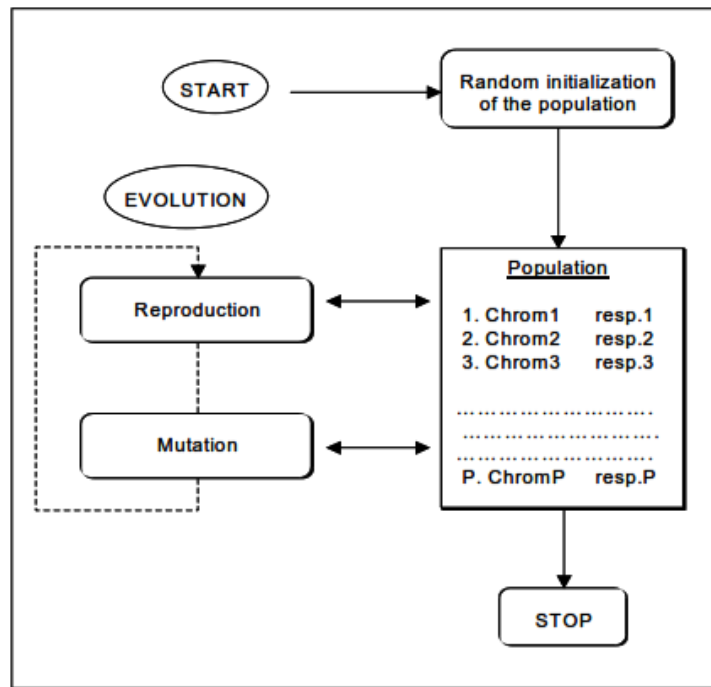


Figura 3 Procedimiento del Algoritmo Genético (Todeschini, Consonni, & Pavan, 2012)

El método de GA-VSS (Variable subset selection) no proporciona un modelo único, sino una población de modelos aceptables; esta característica permite la evaluación de las relaciones variables con respuesta desde diferentes puntos de vista.

Por otra parte, cuando se aplica esta selección del subconjunto de variables a un gran número de variables, la estrategia genética se puede extender a más de una población, cada uno basado en diferentes subconjuntos de variables, que evoluciona de forma independiente entre sí.

En este caso, después de un número de iteraciones, estas poblaciones pueden combinarse de acuerdo a diferentes criterios, para la obtención de una nueva población con diferentes capacidades evolutivas. (Todeschini, Consonni, & Pavan, 2012)

El GA realiza su optimización por variación y selección a través de la evaluación de la función de aptitud. Esta función de aptitud se utiliza para evaluar subconjuntos descriptores alternativos que finalmente serán ordenados de acuerdo con el rendimiento predictivo del modelo relacionándolo con la validación cruzada. (Noorizadeh & Farmany, 2011)

## **1.7. Operadores del Algoritmo Genético:**

### **1.7.1. Selección**

La selección determina qué individuos son elegidos para la recombinación y cuántos descendientes produce cada individuo. La función de aptitud es una función que mide la idoneidad de un individuo. El primer paso es el cálculo del fitness del modelo (aptitud), mismo que se realizó mediante la combinación de los distintos objetivos (multi-objective ranking). (Villalba Montero, 2012)

Un cierto número de condiciones experimentales (cromosomas) se seleccionan al azar al graficar los valores de las variables individuales. (Leardi, 1994)

### **1.7.2. Reproducción**

La recombinación de variables binarias, se denomina "cross-over" (cruce). Esto es debido principalmente a razones históricas, por el equivalente en el área de los Algoritmos Evolutivos. La reproducción engendra nuevos individuos combinando la información contenido en los padres (acoplamiento de la población-padres). (Villalba Montero, 2012)

De la población real una nueva población se obtiene mediante la realización de un cierto número de "apareamientos" entre los cromosomas, elegidos al azar pero de una manera tal que los mejores individuos (es decir, las que producen las mejores respuestas experimentales) tienen una mayor probabilidad de ser seleccionado.

Durante el "cross-over" un intercambio de genes se lleva a cabo de modo que dos nuevos cromosomas se crean, cada uno con una mezcla de las condiciones experimentales que caracterizan los "padres". Estas crías luego pueden sustituir algunos elementos de la población a la que los dos padres pertenecían. (Leardi, 1994)

### **1.7.3. Mutación**

Después de la recombinación, cada descendiente sufre una mutación. Las variables de los descendientes son mutadas por pequeñas perturbaciones (tamaño del paso de la mutación). (Villalba Montero, 2012)

Los componentes de la población pueden sufrir perturbaciones aleatorias de sus genes. Al igual que en las mutaciones naturales, tienen un más bien baja probabilidad de ocurrencia y generalmente resultan en una condición "patológica", pero a veces puede suceder que se obtiene una mejora. Además, las mutaciones pueden ayudar a escapar de una situación de "impase". (Leardi, 1994)

#### **1.7.4. Reinserción**

Después de producir descendientes, éstos deben ser insertados en la población. Esto es especialmente importante si se producen menos descendientes que el tamaño original de la población. (Villalba Montero, 2012)

El individuo mutado entonces puede entrar en la población mediante la sustitución de otro cromosoma. La reproducción y mutación continúan hasta que se alcanza una condición de parada (por ejemplo, solicitando una respuesta, un cierto número de generaciones o de un cierto tiempo de elaboración). (Leardi, 1994)

Estos pasos básicos caracterizan cada GA, pero cada uno de los varios algoritmos conocidos los realiza de diferentes maneras. Para el análisis realizado en este estudio se establecieron 100 iteraciones de cálculo para el Algoritmo.

#### **1.8. Clasificación mediante algoritmo KNN**

Para la clasificación de los atributos, se aplicó en principio el Método del Algoritmo KNN (k-Nearest Neighbor, por su nombre en inglés) cuya función de proximidad puede decidir la clasificación de un elemento atendiendo a la clasificación del modelo o de la mayoría de los k elementos más cercanos.

Es un método no paramétrico, cuya distancia de la versión ponderada de kNN es una variación que pesa la contribución de cada vecino por su semejanza con el elemento de prueba.

Para ello, se evaluó el método de determinación de distancias (Distancia Euclidiana), obteniendo la matriz de distancias y determinando el valor de k para la clasificación del entorno, y finalmente evaluando las clases a las que pertenecen los elementos considerados.

El modelo ha sido aplicado al algoritmo KNN para pre clasificar y depurar los datos mejorando la identificación y el ajuste al modelo.

#### **1.9. Método de Clasificación por árboles de decisión (CART)**

Los árboles de decisión (Decision Trees - DTS) son ampliamente utilizados en la minería de datos a efectos de clasificación. Dada la colección de datos heterogéneos y un conjunto de atributos que describen los datos, los árboles de decisión utilizan los valores de estos atributos para dividir el conjunto de datos en subconjuntos más homogéneos y más pequeños. Estas técnicas de modelado se denominan comúnmente como algoritmos de decisión basados en árboles, o simplemente, algoritmos de árboles de decisión. El objetivo de los DT es minimizar el tamaño del árbol y aumentar al máximo la precisión de la clasificación. (Berry, 2006)

El algoritmo de clasificación y regresión por árboles (CART) utiliza un algoritmo de inducción árbol básico que asigna un atributo a un nodo del árbol basado en la cantidad de información que se obtiene a partir de ese nodo para seleccionar atributos óptimos. A diferencia de otros algoritmos de clasificación mediante árboles de decisión, CART sólo permite la construcción de árboles binarios. Dado un conjunto de datos, el método aplica una partición recursiva binaria basada en los atributos que se pueden expresar en forma de respuestas "sí" o "no". En cada nodo, todas las divisiones posibles se comparan y se selecciona el atributo con el más alto grado de homogeneidad. (Berry, 2006)

## CAPÍTULO 2: RESULTADOS

### 2.1. Selección de Variables mediante Algoritmo Genético GA-VSS

La selección del subconjunto de variables significativas en la calidad resultante del atún en conserva se ha obtenido al aplicar como método el algoritmo genético a un universo de 700 casos con un total de 55 variables cada uno. Una vez depurados los casos incompletos y habiendo escalado los datos se procedió a clasificar los casos acorde a los criterios mencionados en la Tabla No. 2.

De esta manera se etiquetaron los casos a analizar estableciendo una escala en función de los resultados de peso escurrido evaluado en el producto terminado. Consecuentemente la función objetivo priorizará los casos en los que la diferencia entre el peso escurrido establecido en la especificación del producto y el determinado por análisis de producto terminado sea mínima (menor al 1%, clase 1). Es importante recalcar que la especificación de peso escurrido variará tanto en función del tamaño de la presentación (1, 3 y 7 kilos) como en la negociación particular para cada cliente en términos comerciales.

El algoritmo se desarrolló en el entorno Matlab, determinándose un total de 100 iteraciones de reproducción y mutación, hasta obtener los resultados de la búsqueda y selección de las variables significativas, tomando como función objetivo la clase asignada en la etapa de pre procesamiento de los datos (ytrain).

Como resultado de esta selección de variables se han identificado cuatro elementos principales. En la Tabla 3 se observan las variables seleccionadas al aplicar el método del Algoritmo Genético, entre las cuales se encuentran: (1) T Grated (Porcentaje de migas o trozos en la evaluación de producto terminado). (2) M DGrated (Diferencia entre el porcentaje de migas encontrado en el producto terminado y el establecido en la especificación de producto. (3) Sabor4 (Sabor Normal evaluado en producto terminado) (4) S WEsc (Especificación de peso escurrido determinado para el producto). La presencia de estas cuatro variables significativas, detalladas en la Tabla No. 3, presentó un NER (No error rate) de 0,66.

	(T) GRATED	(M) DGRATED	SABOR4	(S) WESC
<b>VAR SEL</b>	38	53	48	30
<b>NO.</b>	1	2	3	4
<b>NER</b>	0,48	0,49	0,51	<b>0,66</b>
<b>FREC</b>	59	49	36	33

Tabla 3 Variables seleccionadas por Algoritmo Genético



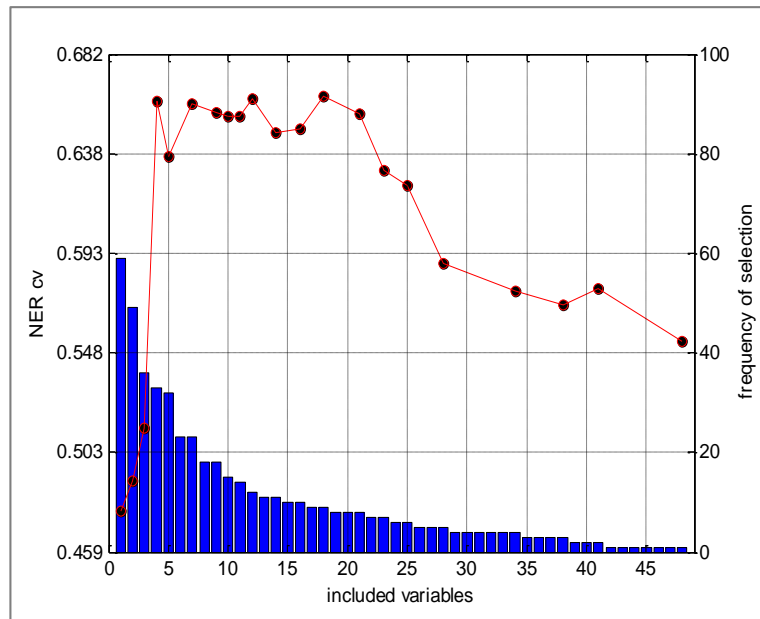


Figura 4 Distribución de frecuencias en la selección de variables significativas

## 2.2. Predicción y depuración de los resultados

Una vez seleccionadas las variables se busca construir el modelo de clasificación, para lo cual se ha utilizado la herramienta de Matlab y en función de predecir las clases para cada caso. Para lograrlo se aplicó el método KNN, evaluando los primeros resultados del modelo de clasificación.

En la búsqueda de un modelo más robusto y preciso que permita calcular las clases según el criterio aplicado en los datos iniciales, se ha realizado un afinamiento de los datos utilizando como punto de partida la diferencia entre la clasificación asignada por el modelo y la clasificación inicial calculada. De esta manera han sido seleccionados y excluidos aquellos casos cuya clasificación del modelo comparada con la clasificación inicial tenga distancia mayor a 1. (Clases uno que se calificaron como clase tres o viceversa). Estos casos fueron separados a manera de una primera depuración, evitando que generen ruido y desestabilicen el modelo evitando se pueda reconocer el patrón de clasificación aplicable.

## 2.3. Validación del modelo

La validación es un aspecto crucial de cualquier modelo. La precisión de los modelos propuestos se ilustra en el uso de las técnicas de evaluaciones como la validación cruzada, procedimiento que se desarrolla al dejar un grupo fuera (Leaving Group Out - Cross Validation) y la validación a través de un grupo externo de prueba (test set). (Noorizadeh & Farmany, 2011)

**2.3.1. Validación cruzada (cross-validation)**

La validación cruzada es una técnica utilizada para explorar la fiabilidad de los modelos estadísticos. Sobre la base de esta técnica, un número de conjuntos de datos modificados se crean mediante la supresión en cada caso uno o un pequeño grupo de los objetos (dejar algunos fuera). Para cada conjunto de datos, un modelo de entrada-salida es desarrollado, basado en la técnica de modelado utilizada. Cada modelo se evalúa, mediante la medición de su exactitud en la predicción de las respuestas de los datos restantes, es decir, de aquellos datos del grupo que no han sido utilizados en el desarrollo del modelo. (Noorizadeh & Farmany, 2011)

En particular, el procedimiento de LGO se utilizó en este estudio. A continuación, el modelo se construyó sobre la base de este conjunto de prueba (training set). La significación estadística (aptitud) en la búsqueda del modelo proyectado fue juzgada en función de los resultados de la matriz de confusión (certeza en la clasificación) y la capacidad predictiva se evaluó por el coeficiente de la validación cruzada.

Una vez realizada la depuración de los datos analizados clases mediante la clasificación KNN, el modelo de clasificación CART fue capaz de predecir las clases del training set con un coeficiente de exactitud (no error rate) igual a 0.70. En cuanto a la clasificación fueron correctamente clasificados 55 casos dentro de la Clase 1, 236 casos en la clase 2 y 96 casos en la clase 3. El modelo presentó una especificidad de 0.98 0.85 y 0.95 para las clases 1, 2 y 3 respectivamente. El nivel de sensibilidad del modelo se calculó en 0.98 0.91 y 0.79 para las clases 1, 2 y 3. La precisión en la clasificación se determinó en 0.9 para las clases 1 y 2 y de 0.85 para la clase 3.

'REAL/PREDICTED'	'clase 1'	'clase 2'	'clase 3'
'CLASE 1'	55	1	0
'CLASE 2'	6	236	17
'CLASE 3'	0	25	96

*Tabla 4 Matriz de confusión del modelo CART*

NER CV VENE5		0,7		
'REAL/PREDICTED'	Especificidad	Sensibilidad	Precisión	
'CLASE 1'	0,98	0,98	0,9	
'CLASE 2'	0,85	0,91	0,9	
'CLASE 3'	0,95	0,79	0,85	

*Tabla 5 Cros validación del Modelo CART utilizado en la clasificación*

**2.3.2. Validación externa (test set)**

La validación del modelo con datos externos (es decir, datos que no se utilizan en el desarrollo de modelos) es el mejor método de validación. La capacidad predictiva de un modelo CART con los descriptores seleccionados se exploró dividiendo el conjunto de datos completo en la plataforma Matlab. El poder predictivo del modelo desarrollado en el conjunto de entrenamiento seleccionado (training set) se calculó sobre la clasificación asignada por el modelo.

Bajo estas condiciones, el modelo presentó una especificidad de 0.88 0.60 y 0.90 para las clases 1, 2 y 3 respectivamente. El nivel de sensibilidad del modelo se calculó en 0.88 0.70 y 0.45 para las clases 1, 2 y 3. La precisión en la clasificación se determinó en 0.54 para la clase 1 y de 0.85 y 0.66 para las clases 2 y 3. El modelo de clasificación CART fue capaz de predecir las clases del training set con un coeficiente de exactitud (no error rate) igual a 0.68.

'real/predicted'	'clase 1'	'clase 2'	'clase 3'
'clase 1'	30	4	0
'clase 2'	24	96	18
'clase 3'	2	41	35

*Tabla 6 Matriz de confusión del modelo CART*

**NER CV vene5 0,68**

'real/predicted'	Especificidad	Sensibilidad	Precisión
'clase 1'	0,88	0,88	0,54
'clase 2'	0,60	0,70	0,68
'clase 3'	0,90	0,45	0,66

*Tabla 7 Validación externa del modelo de clasificación CART*

En la figura 3 se observa el árbol de decisiones generado por el algoritmo de clasificación CART, el cual representa de manera gráfica los umbrales a considerar en cada variable y la interacción entre las mismas hasta llegar a la función objetivo.

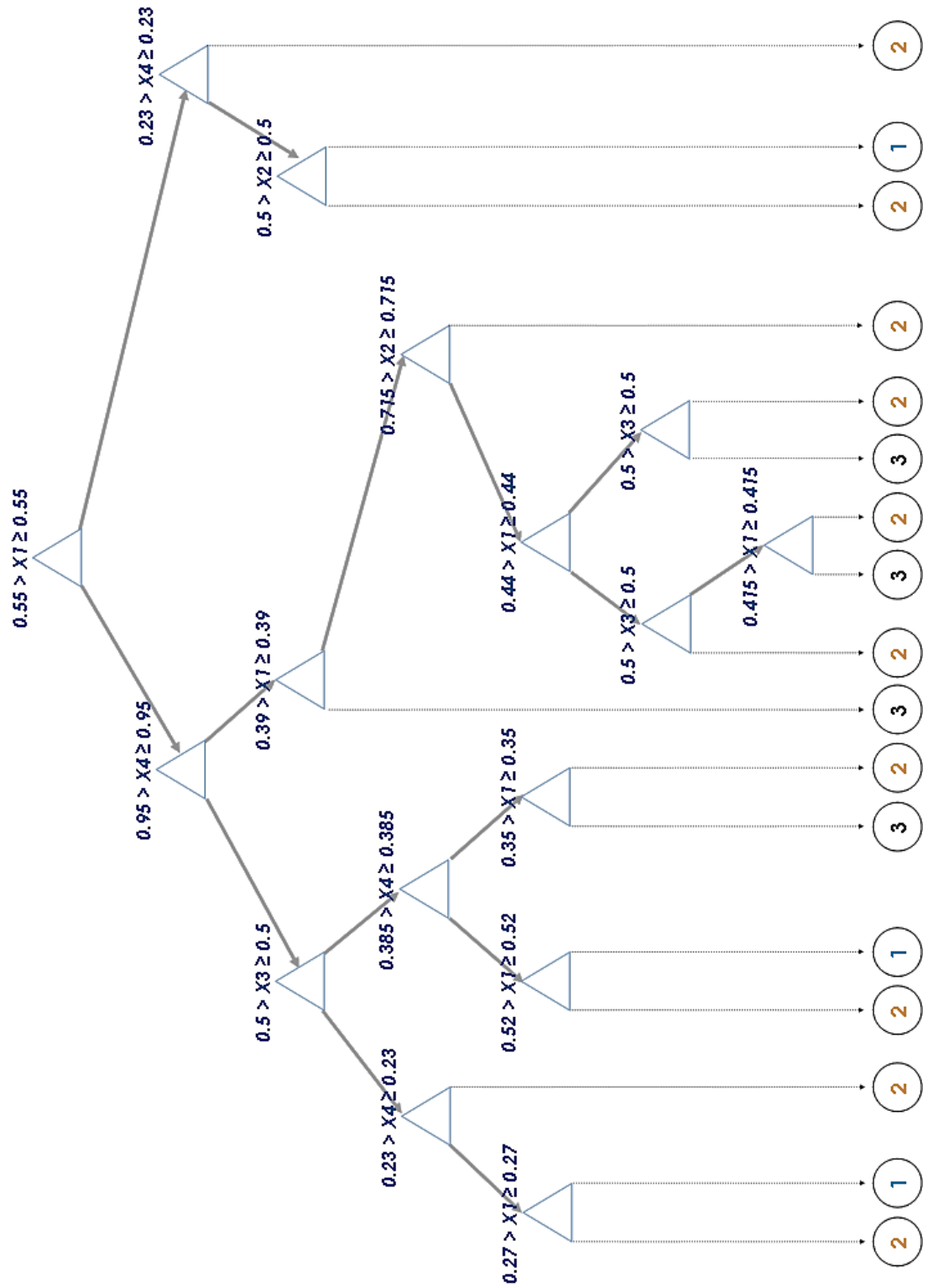


Figura 5 Modelo escalado de predicción de la calidad resultante en conservas de atún  
 Fuente: Elaboración propia en base al modelo obtenido de la plataforma Matlab.

Esta operación se ha realizado para los umbrales determinados para las variables en cada nivel de decisión del modelo CART. La Tabla 6 resume los resultados de la reversión del escalado de los valores, en función de facilitar la interpretación de los resultados obtenidos.

	Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5	Nivel 6	Nivel 7
<b>(T) GRATED X1</b>	30,9%		23,3%		17,7%		24,5%
					29,4%		
					21,5%		
					25,7%		
<b>(M) DGRATED X2</b>			-29%	-2,8%			
<b>SABOR4 X3</b>			0,5			0,5	
<b>(S) WESC X4</b>		6502,75		1934,35			
		1934,35		2917,83			

Tabla 8 Valores de los umbrales de decisión revertidos del escalado o normalización.

Los umbrales una vez normalizados han sido sustituidos en el modelo del árbol de decisión e identificados con las unidades de medida originales, es decir antes de la normalización o escalado.

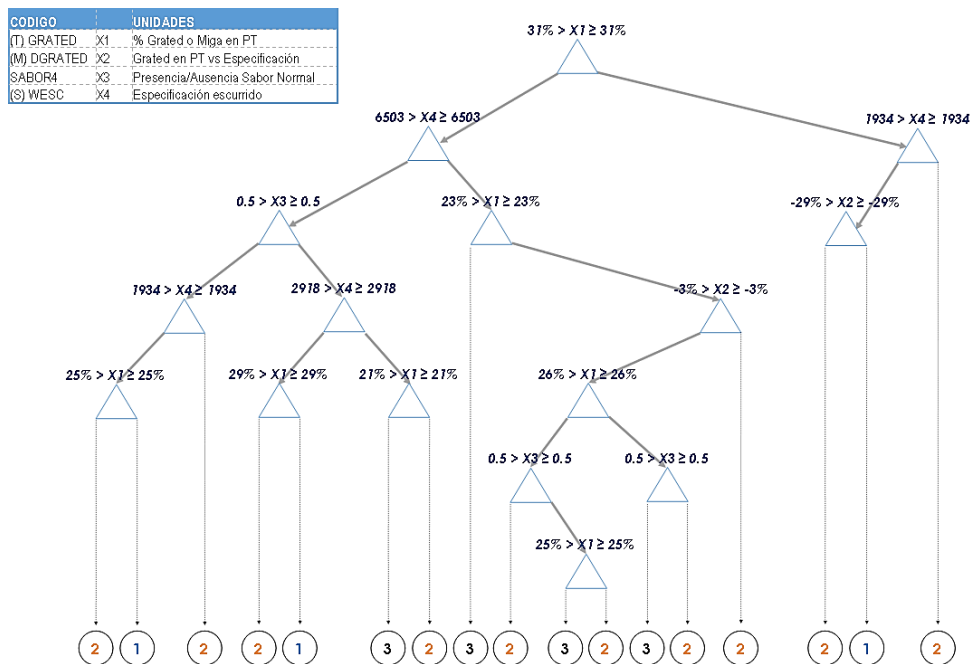


Figura 6 Modelo de árbol de decisión con valores sin normalizar.

### CAPÍTULO 3: DISCUSIÓN

El método de Algoritmo genético ha sido aplicado a una base de datos compuesta por 700 casos reales del histórico de producción de conservas de atún en envase flexible. Este universo de muestras ha sido caracterizado por 55 variables de proceso entre las cuales se incluyeron características propias de la materia prima así como de cada etapa del proceso productivo, especificaciones y finalmente los resultados de la evaluación del producto terminado denominado “grading”.

En el ámbito cotidiano de la producción de alimentos, y en este caso en particular, de la producción de conservas de atún, nos enfrentamos diariamente con un sinnúmero de variables que inciden directa e indirectamente, sea en mayor o menor grado, en la calidad resultante de los alimentos procesados. La exploración de herramientas de optimización como los algoritmos genéticos presenta una alternativa efectiva y práctica que nos permite vislumbrar aquellas características intrínsecas del proceso y que pueden tener una mayor significancia a la hora de establecer y validar nuestros parámetros y especificaciones de calidad de los productos sin dejar de lado los requisitos legales y del cliente.

A nivel operativo, se busca maximizar la eficiencia de los procesos y mejorar la rentabilidad económica generando una mayor competitividad en los mercados. Es por ello que el sector industrial debe encaminarse en implementar soluciones basadas en el análisis estadístico de datos para la toma de decisiones.

Es común encontrar, como se observó durante la etapa de análisis, que a nivel de proceso productivo los resultados no presenten una tendencia necesariamente lineal que describa o permita predecir de manera teórica los resultados en términos de calidad. Esta condición se presenta principalmente en procesos cuyo diseño se ha realizado bajo condiciones diferentes a las adaptadas en la planta o en procesos que nacieron de la experticia de sus actores y que han debido ir reformándose y ajustándose a lo largo del tiempo. Esta situación, conocida como el «know how», un método de transferencia interna de tecnología, ha representado durante décadas el “secreto industrial” y la ventaja competitiva en compañías que le han apostado a evaluar el comportamiento de sus productos en el transcurso tiempo.

En este sentido aquellas características o variables del proceso interactúan y se adaptan a la manera particular de procesar y transformar las materias primas que tiene cada empresa, siendo importante, por lo tanto que cada proceso productivo se analice de manera individual en función del contexto propio de la producción. Por esta razón cabe destacar que, aun siendo la metodología ampliamente aplicable a cualquier industria o método de procesamiento, los resultados podrían variar entre una compañía a otra e incluso modificarse en intervalos significativos de tiempo, si la empresa llegase a reemplazar o modificar una de las variables utilizadas para el estudio.

Los resultados obtenidos con este nivel de análisis, permiten a una compañía identificar y seleccionar las variables relevantes que son propias de su proceso y que deben ser controlados con miras a implementar medidas correctivas al proceso de fabricación, asegurando de esta manera la estabilidad de los resultados en términos de eficiencia y calidad.

Pese a que la recolección de la información puede ser un proceso difícil de estructurar en principio, una vez se determinan las fuentes de información relevante, éstas pueden ser seleccionadas y enfocadas a medir y optimizar los procesos productivos de manera independiente (en cada sub-proceso) o en la forma como se ha desarrollado en el presente estudio, partiendo del conjunto integral de variables.

Una de las ventajas de esta metodología cuasi experimental aplicada es que, en función del modelo obtenido, permite direccionar esfuerzos hacia el estudio de una etapa del proceso en particular, aprovechando de mejor manera los recursos utilizados en el análisis (materiales, humanos y de tiempo). En consecuencia, se evitaría el levantamiento de información irrelevante que no aportará a la consecución del objetivo planteado, hablando en términos de efectividad y de aseguramiento de la calidad. Paralelamente, es importante considerar que cierta información del proceso podría no ser susceptible de modificarse por restricciones de inocuidad o de infraestructura disponible, derivando también en un dispendio innecesario de esfuerzos en la recolección de datos.

Otra ventaja identificada es la aplicación y la facilidad de interpretación a nivel operativo del diagrama de árbol de decisión CART. Es una herramienta útil, altamente comprensible y de aplicación directa, sin necesidad de disponer de las herramientas adicionales de software utilizadas para la construcción del modelo. Sin embargo, en medida que los procesos se optimicen, encontraremos que el levantamiento de una restricción del proceso dará origen a una nueva por lo que la mejora continua se extiende también a lo largo del tiempo en función de la madurez de los procesos. Esto hace que de una u otra manera la investigación realizada se conciba como una herramienta para la optimización a mediano y largo plazo.

Los resultados obtenidos posterior a la aplicación del método de Algoritmo Genético GA nos direccionan a observar el comportamiento de las cuatro variables seleccionadas y mencionadas en la Tabla No.3, cada una con un grado de inferencia diferente:

Con el contenido de Grated (X1), nos encontramos analizando el resultado de la evaluación posterior a la fabricación, es decir en el producto terminado. El porcentaje de “grated” evaluado como la cantidad de trozos y migas menores a 1,2 cm, se evalúa en laboratorio interno, conforme la metodología recomendada por el *Codex stan 70 1981*.

Según se observa en la Tabla No. 6 esta variable (X1) aparece en cuatro de los siete niveles de decisión del Algoritmo de clasificación CART, lo cual nos lleva a profundizar sobre la importancia de este factor en los resultados finales del peso escurrido.

Esta distinción se le puede atribuir a la “superficie de contacto” disponible entre el producto en proceso (lomos limpios y trozos) con el líquido de cobertura, independientemente cuál sea este. Debido a la viscosidad del líquido de cobertura se podría suponer que la fabricación de productos en aceite tendría un mejor comportamiento en la evaluación de peso escurrido, sin embargo esta variable no está presente dentro del modelo obtenido. Esto respaldaría el argumento de que a mayor superficie de contacto entre la superficie del lomo con el líquido de cobertura podremos asegurar una mejor adsorción, independientemente de que el producto se envase con aceite o agua; sin embargo también estará en función del tamaño de los trozos puesto que, entre más desmigado se encuentre el producto, aumentará la permeabilidad del líquido de cobertura durante la evaluación de la masa escurrida en el producto terminado.

Con esta consideración respecto a la variable X1, se presentan paralelamente restricciones de especificación de las conservas, puesto que el consumidor final prefiere una apariencia del producto sólida y la apreciación de lomos enteros. Por lo tanto esta condición podría aplicarse con la observancia de la percepción esperada del cliente o manejarse mediante negociación de las especificaciones de drenado, siempre y cuando se cumpla mínimo con el 75% del volumen total de la conserva, según lo determina la normativa aplicable al producto en el Ecuador, Norma NTE INEN 184 1990 Rev. 7. Así también se observa una restricción en cuanto al exceso de miga en el producto, la cual se analiza más adelante con el umbral para la variable X2.

La cantidad de trozos presentes en el producto final están relacionados, además de la dosificación por receta, a los procesos de moldeo y manipulación posteriores al empaque que dan lugar a un resquebrajamiento de los lomos y por lo tanto a la generación de trozos, adicionales a los incorporados durante la fabricación. Así también la proporción de grated puede variar acorde al método de mezcla y dosificación funda a funda por lo que los procesos deben estructurarse también en función de estandarizar en medida de lo posible estos factores que inciden en la variación del contenido de miga o grated en el producto final.

El modelo CART se estructuró en dos grupos en el nivel de 1 de decisión, para aquellos productos que presentaron una evaluación mayor a 31% en migas y aquellos que se encontraron por debajo de este rango. En ambos casos el siguiente nivel de decisión (Nivel 2) está marcado por la variable WEsc (X4) que representa la especificación de escurrido determinada para cada producto o receta en particular. Esta variable define el contenido efectivo de masa escurrida en el producto, siendo además un aspecto importante a nivel comercial como base para la negociación del valor del producto en el mercado y convirtiéndose en una ventaja comparativa frente a la competencia.

Desde el punto de vista de la competitividad, para una planta exportadora el reto del mercado internacional no solamente radica en el cumplimiento de pedidos, capacidad productiva o satisfacción de requisitos.



En un mundo globalizado como lo es especialmente el mercado europeo, es imprescindible conocer a los clientes y conocer también las preferencias y tendencias de consumo del nicho al que se dirige. Actualmente las compañías tienen el reto de satisfacer a clientes que además de exigir calidad, se preocupan por el consumo de productos saludables, siendo un segmento de consumidores que se preocupan por evaluar a profundidad la composición de sus alimentos y los efectos que éstos puedan tener en su dieta y su organismo tanto a mediano como a largo plazo.

En busca de una tendencia más “naturalista”, las diferentes marcas comerciales buscan mostrar a sus consumidores la reducción o sustitución de aditivos, conservantes y coadyuvantes de proceso la cual viene marcada en su declaración de ingredientes. Por esta razón, la adición de proteínas ajenas a la naturaleza del producto y utilizadas comúnmente para la favorecer la adsorción y retención de líquidos, no son una alternativa a considerar para incrementar o estabilizar resultados en rendimiento y rentabilidad.

El modelo obtenido nos presenta una condición de decisión para aquellos productos que se evaluaron con un porcentaje de grated igual o superior al 31%; para aquellos productos cuya especificación de peso escurrido es superior al umbral definido (1934g), el resultado siempre será Clase 2. Es decir que existirá una diferencia entre 1% y 5% en comparación al escurrido deseado. En este caso se aplicaría para presentaciones de 3 y 7 kilos y representa por lo tanto que en estos productos el porcentaje de migas por encima del umbral de X1 (31%) dará como resultado siempre una clase dos, viéndose el efecto negativo del exceso de miga en estas presentaciones.

Por otra parte, para productos con especificación de peso escurrido inferior a 1934g (presentación de 1Kg y 1.2Kg) existe una opción para obtener una clase 1 la cual representa, además de un contenido de miga superior al 31%, que la especificación del cliente hubiese sido superior a 60%, es decir que en estas condiciones el peso escurrido esperado se cumpliría si se tratase de un producto fabricado habitualmente con el 100% miga, de lo contrario el resultado también será Clase 2.

Considerando aquellos productos que se evaluaron con un porcentaje de miga por debajo de 30,9% se puede derivar en dos escenarios, en función de la especificación de escurrido requerida. Para los productos con escurrido superior a 6502 (usualmente producto 7Kg), se observa a priori que el mejor resultado a obtener será Clase 2. Lo que se determina con ello es que, en las condiciones de producción actuales, esta presentación en particular debería negociarse con un umbral escurrido por debajo de este límite (actualmente establecido en 6820 y 6750 que representan el 97 y 96% del peso total, respectivamente). De esta manera, si adicionamos a la alta especificación de escurrido por alcanzar, un porcentaje de migas inferior al 23.3% el resultado que se obtenga será Clase 3.

La Clase 2 para el producto en presentación 7 Kilos, con porcentaje de migas menor a 30,98%, estaría por lo tanto dentro del rango menor a 6683g y mayor a 6502g, bajo las siguientes condiciones:

**(a)** Que el porcentaje de migas sea igual o superior a la especificación del cliente (X2), sin sobrepasar el umbral del Nivel 1 para la variable X1 (Grated).

**(b)** Si el porcentaje de miga o grated es inferior a la especificación solicitada por el cliente hasta en 2.8 puntos, se vuelve a analizar el intervalo de grated e ingresa una nueva variable: X3 (Sabor 4 - Normal).

DGrated (**X2**) es una variable obtenida a partir del valor resultante de la evaluación de grading del producto terminado vs la especificación inicial solicitada por el cliente. Trata de una variable que integra aspectos tanto de operación (a nivel de receta) así como aspectos inherentes a los requisitos específicos del cliente en cuanto a su percepción de calidad del producto.

La variable Sabor 4 (**X3**) se interpreta como la presencia del sabor característico de la conserva, relacionada también con un proceso adecuado que ha partido de una materia prima en condiciones aceptables o normales. Esto representaría que cualquier factor que altere el sabor normal de la conserva y pudiera derivar en sabor: salado, oxidado o quemado figurará como ausencia del sabor normal, denotado con valor 0.

Indirectamente la presencia o ausencia de esta variable relaciona el modelo con la calidad de la proteína del músculo (lomo del pescado) al ser influida por el contenido de sal (sabor salado), la sobre exposición en tiempo y temperatura durante el proceso productivo (sabor oxidado); y un tratamiento térmico excesivo (sabor quemado).

En el primer caso (sabor salado) se estaría trabajando con un tejido con un nivel de saturación de sustancias salinas elevado (esto es superior al 2% de NaCl). Dicho factor podría intervenir negativamente en la adsorción de líquido debido a que el espacio disponible una vez ha llegado a la estabilidad con el líquido de cobertura, sería menor que en aquellos productos con niveles de sal dentro de parámetros.

Al hablar de un sabor oxidado estaríamos relacionando el tiempo determinado que ha tomado la producción de un lote de materia prima que haya permitido la oxidación de los ácidos grasos presentes en el músculo debido a temperaturas superiores a 21,1°C en un punto del proceso y por un periodo mayor a 12 horas o por un periodo mayor a 24 horas acumulativas en las que el proceso no haya sobrepasado los 21,1°C<sup>1</sup>. (FDA, 2011)

El sabor quemado estaría relacionado a desviaciones asociadas al proceso de esterilización, mismo que si bien está normalizado y regulado en función del tamaño y características del envase (espesor, contenido, etc.) y cumple con los tiempos de seguridad; éste puede extenderse al combinar presentaciones de diferentes tamaños en autoclaves, obligando a aplicar el tiempo del producto de mayor tamaño y derivando en aparición de quemaduras en productos de presentaciones más pequeñas.

---

<sup>1</sup> Tiempos de manejo recomendados para el control del riesgo de formación de escombrotóxina (histamina) en producto en proceso. *Fish and Fishery Products Hazards and Controls Guidance. Fourth Edition – APRIL 2011*

Continuando con el punto (b) si el porcentaje de migas es inferior a 25.7% aun cuando tenga sabor diferente al normal puede ser clase 2. Si en su lugar en el punto (b) presenta sabor normal y el rango del valor de las migas está entre 24,5 hasta 25.7% éste resultado será Clase 2. Si por el contrario el porcentaje de migas fuese menor a 24.5% la clase resultante sería 3.

En caso de que el punto (b) trate de producto con porcentaje de migas entre 25.7 y 30.98% la Clase 2 está sujeta a condiciones normales de sabor (X3) caso contrario el resultado a obtener será Clase 3.

En resumen, las condiciones para que un producto 7 kilos alcance una Clase 2, serían:

- Peso Escurrido (W):  $6683g \leq W \leq 6502g$
- Igualar o superar el Grated requerido por el Cliente

En caso de que el contenido de Grated sea menor a la especificación del cliente:

- Sabor diferente al normal requiere porcentaje de migas inferior a 25.7%
- Sabor normal y porcentaje de migas no menor a 24.5% y máximo 25.7%
- Sabor normal y porcentaje de migas entre 25.7% y 30.98%

Analizando los casos de producto con especificación de peso escurrido menor a 6502 (a considerar productos 1Kg, 1.2Kg y 3Kg) la condición siguiente para discriminar resultados sería:

- a. En ausencia de sabor normal:
  - Producto con escurrido menor a 1934 (1Kg y 1.2Kg) pueden ser Clase 1 si el contenido de grated es mayor a 17.7%, de lo contrario serán Clase 2.
  - Producto con escurrido mayor a 1934 (3Kg) serán Clase 2
- b. Si se presentan condiciones normales tenemos:
  - Si la especificación de escurrido es superior a 2917.8g deberá contener una cantidad igual a 29.4% de migas para ser Clase 2, de lo contrario el resultado es Clase 3.
  - Si la especificación de escurrido se estableciera por debajo de 2917.8g el producto con porcentaje de migas superior a 21.5% será Clase 1, por debajo de este valor será Clase 2.

Para una mejor observación de los casos y condiciones mencionados anteriormente, en la Figura 5 se han identificado los productos que se enmarcarían dentro de las diferentes clases establecidas en el modelo de árbol de decisión CART.

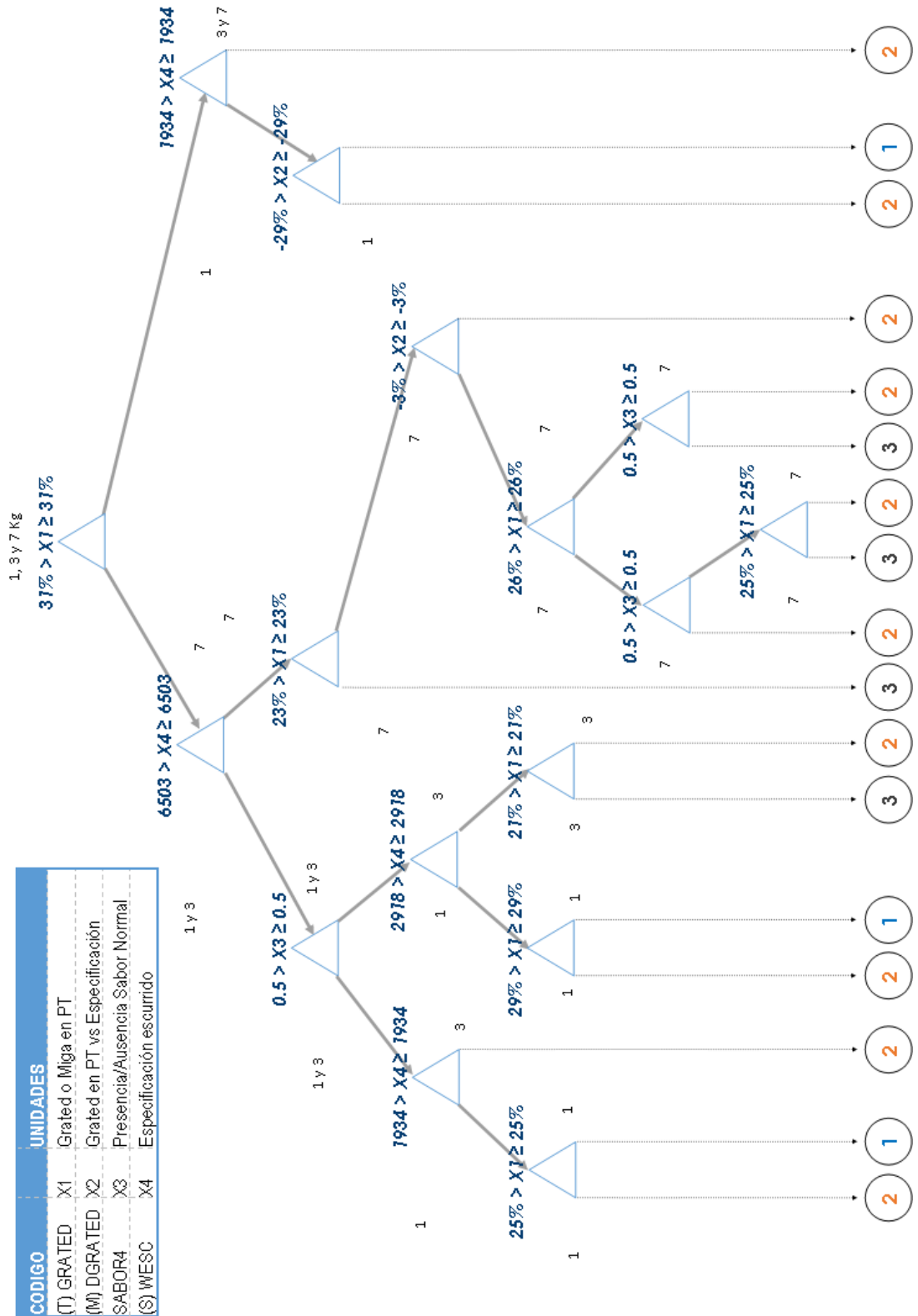


Figura 7 Modelo CART. Identificación de escenarios para los productos 1, 3 y 7Kg.

## CAPITULO 4: CONCLUSIONES

La metodología de Algoritmos genéticos representa una poderosa herramienta de optimización de los procesos productivos y propone una alternativa efectiva y práctica de análisis de información e interrelación de las operaciones con los valores resultantes en términos de calidad de producto y cumplimiento de especificaciones de clientes. Se considera ampliamente aplicable a cualquier industria o método de procesamiento, más se recalca que los resultados pueden variar entre una compañía a otra e incluso modificarse en intervalos significativos de tiempo, en función de la modificación de las variables utilizadas para el presente estudio, por lo que el análisis debe ser personalizado para cada caso.

Mediante la selección aplicada con la metodología de Algoritmos Genéticos, se descartó la relevancia de las condiciones del manejo post captura del atún en la preservación de la calidad de las proteínas del pescado y sus propiedades para retener el líquido de gobierno utilizado en las conservas de atún. De esta manera se observa que los controles aplicados actualmente como criterios de recepción del producto, aseguran un comportamiento estable de la materia prima frente al proceso tecnológico aplicado.

La técnica cuasi experimental aplicada, permitirá en función del modelo obtenido, direccionar esfuerzos hacia el estudio de etapas del proceso en particular, evitando el levantamiento de información irrelevante para los fines investigativos. Así también cada etapa del proceso podría convertirse en un universo de datos con una función objetivo distinta y ser susceptible de optimizarse individualmente, conforme especificaciones más detalladas para cada una.

El modelo CART utilizado, proporciona facilidad de interpretación a nivel operativo, mediante su estructura de diagrama de árbol de decisión. Es útil, altamente comprensible y de aplicación directa, eliminando a corto plazo la necesidad de disponer de las herramientas adicionales de software como las utilizadas en la construcción del modelo.

El levantamiento de una restricción del proceso, en este caso el cumplimiento de peso escurrido en producto, dará origen a una nueva restricción por lo que la mejora continua se extenderá a lo largo del tiempo en función de la madurez del proceso y demandando por lo tanto una herramienta para la optimización a mediano y largo plazo, la inherente incorporación del método de Algoritmos Genéticos al proceso de análisis de información de la producción para la toma de decisiones.

Los resultados obtenidos posterior a la aplicación del método de Algoritmo Genético GA nos llevaron a analizar el contexto de las cuatro variables seleccionadas: (1) T Grated (Porcentaje de migas o trozos en la evaluación de producto terminado); (2) M DGrated (Diferencia entre el porcentaje de migas encontrado en el producto terminado y el establecido en la especificación de producto); (3) Sabor4 (Sabor Normal evaluado en producto terminado) y (4) S WEsc (Especificación de peso escurrido determinado para el producto).

Así también se evaluó y justificó la presencia de las variables seleccionadas por el modelo y su consecuente injerencia en el resultado de la función objetivo (Peso escurrido en producto terminado).

El modelo obtenido permitió no solamente determinar las condiciones de proceso y las clases resultantes en función de los umbrales establecidos para cada variable, sino que también se observa una potencial herramienta para la negociación de requisitos y especificaciones con los clientes, manejando a priori la certeza del cumplimiento y permitiendo potencializar las fortalezas y minimizar las debilidades del proceso de producción.

En función de los resultados se recomienda la gestión de los recursos necesarios para implementar y disponer en la compañía del software y aplicativos necesarios que permitan desarrollar la metodología y paralelamente ampliar la investigación a los diferentes niveles de la organización y de manera especial a las etapas relacionadas a los resultados del presente estudio. Así también será necesario estructurar el levantamiento de información y determinar intervalos de medición de resultados y de actualización del modelo.

## REFERENCIAS BIBLIOGRÁFICAS

### Bibliografía

- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. . *Analytical Methods*, 5, 3790-3798.
- Berry, M. W. (2006). *Lecture Notes in Data Mining*. River Edge, NJ, USA: World Scientific.
- Cassotti, M., & Grisoni, F. (2012). Variable selection methods: an introduction. Milano-Bicocca.
- De Bruyne, S. (2010). *Data and Classifier Models for Accessible Supervised Classification Problem Solving*. Bruxelles: Process.
- FDA. (2011). *Fish and Fishery Products Hazards and Controls Guidance*. Food and Drug Administration. , Department of health and Human Services.
- Han, Jiawei; Kamber, Micheline. (2006). *Data Mining : Concepts and Techniques (2nd Edition)*. Burlington, MA, USA: Elsevier Science & Technology.
- Ismail Bin, M., & Dauda, U. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. Johor Bahru: Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia.
- Leardi, R. (1994). Application of Genetic Algorithms to Feature Selection Under Full Validation. Conditions and to Outlier Detection. *Journal of Chemometrics*, 65-79.
- Moreno, M., Quintales, L., García Peñalvo, F., & Polo Martín , J. (2001). *Aplicación de técnicas de minería de datos en la construcción y validación de modelos predictivos y asociativos a partir de especificaciones de requisitos de software*. Universidad de Salamanca. Departamento de Informática y Automática.
- Noorizadeh, H., & Farmany, A. (2011). *Application of GA-PLS and GA-KPLS calculations for the prediction of the retention indices of essential oils*. Ilam, Iran: Department of Chemistry, Ilam Branch, Islamic Azad University,.
- Todeschini, R., Consonni, V., & Pavan, M. (2012). *Partial Ranking Models by Genetic Algorithm Variable Subset Selection (GA-VSS) Approach for Environmental Priority Settings*. Bicocca, P.za della Scienza: Milano Chemometrics and QSAR Research Group – Dept. of Environmental Sciences, University of Milano.
- Villalba Montero, J. (2012). *Modelo de optimización por simulación para la gestión del transporte en muelles de la terminal de contenedores del Puerto de Algeciras*. Sevilla: Escuela Técnica Superior de Ingenieros de la Universidad de Sevilla. .

### Referencias de consulta

Cassotti, M. Grisoni, F. "Variable selection methods: an introduction". Milano Chemometrics and QSAR Research Group - Dept. of Environmental Sciences, University of Milano-Bicocca, P.za della Scienza 1 – 20126 Milano (Italy)

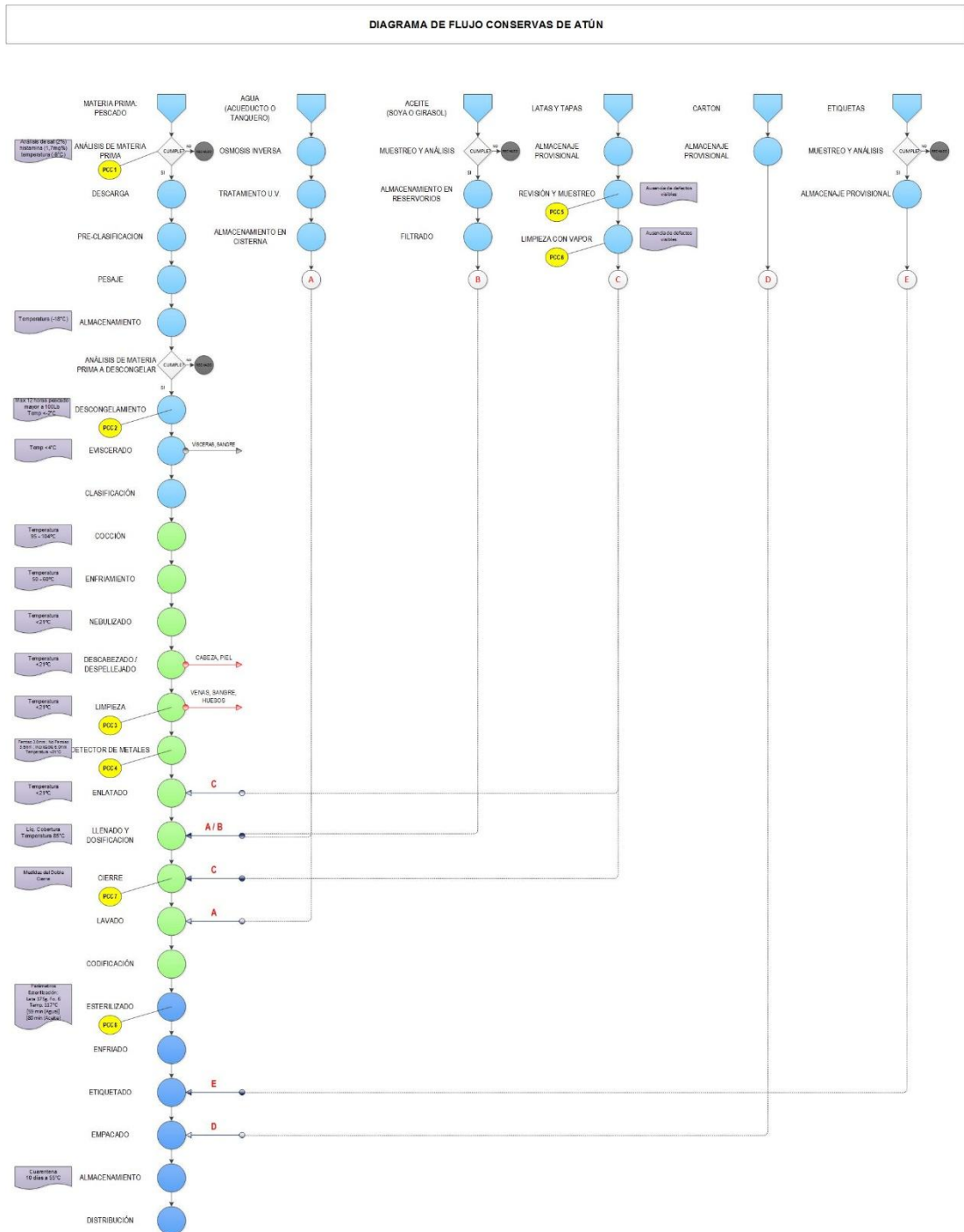
Dirección de Inteligencia Comercial e Inversiones, PROECUADOR. "Evolución de las exportaciones ecuatorianas Enero-Abril 2012-2014". Junio 2014.

Leardi, R. Lupiañez, Amparo. "Genetic algorithms to feature selection in PLS regression: how and when to use them". 1998. Chemometrics and intelligent laboratory systems 41. Pág. 195-207.

Pavan, M. Consonni, V. Todeschini, R. "Partial Ranking Models by Genetic Algorithm Variable Subset Selection. (GAVSS) Approach for Environmental Priority Settings. 2005. MATCH Commun. Math. Comput. Chem. No. 54. Pág 583-609.



ANEXOS



Anexo 1 Diagrama de flujo del proceso de producción de conservas de atún.  
Fuente: Elaboración propia.

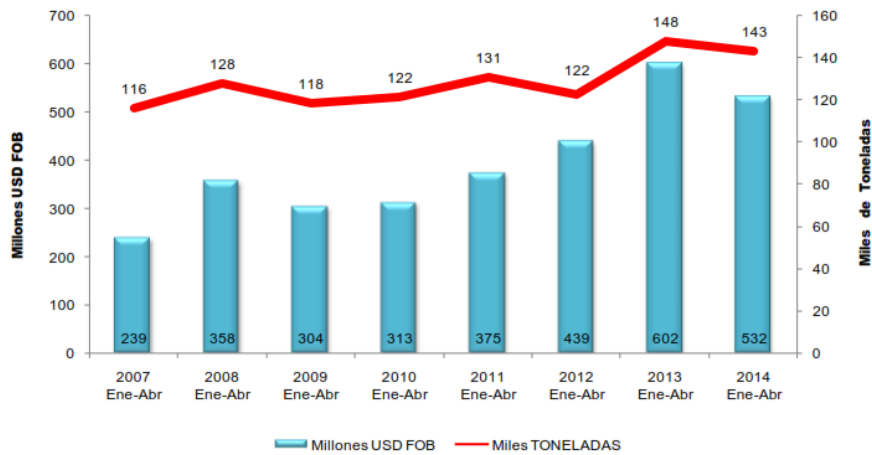
## PRINCIPALES PRODUCTOS NO PETROLEROS EXPORTADOS POR ECUADOR

Subpartida	Descripción	Unidades	Ene-Abr			2012 - 2013	Variación	2013 - 2014	Part. 2013
			2012	2013	2014				
0803.90.11.00*	BANANAS FRESCAS TIPO «CAVENDISH VALERY»	Fob Miles USD	741,995	843,880	856,154	13.73%	1.45%	23.49%	
		Toneladas	1,843,887	1,961,623	1,992,330	6.39%	1.57%	64.96%	
1604.14.10.00	ATUNES EN CONSERVA	Fob Miles USD	213,069	347,985	305,703	63.32%	-12.19%	9.69%	
		Toneladas	42,274	60,412	56,842	42.91%	-5.91%	2.00%	
0603.11.00.00	ROSAS FRESCAS CORTADAS	Fob Miles USD	219,469	231,619	219,670	5.54%	-5.16%	6.45%	
		Toneladas	32,760	37,803	43,584	15.40%	15.29%	1.25%	
0306.17.99.00*	LOS DEMÁS CAMARONES, LANGOSTINOS Y DEMÁS DECAPODOS CONGELADOS NO CONTEMPLADOS EN OTRA PARTE	Fob Miles USD	256,653	213,134	438,253	-16.96%	105.62%	5.93%	
		Toneladas	41,623	30,181	46,856	-27.49%	55.25%	1.00%	
0306.16.00.00*	CAMARONES, LANGOSTINOS Y DEMÁS DECAPODOS DE AGUA FRIA CONGELADOS	Fob Miles USD	133,433	175,140	227,846	31.26%	30.08%	4.88%	
		Toneladas	21,100	26,103	23,753	23.71%	-9.00%	0.86%	
7108.12.00.00	LAS DEMÁS FORMAS DE ORO EN BRUTO PARA USO NO MONETARIO	Fob Miles USD	76,641	160,456	120,387	109.36%	-24.97%	4.47%	
		Toneladas	2	5	4	132.50%	-23.89%	0.00%	
1801.00.19.00	CACAO EN GRANO CRUDO, LOS DEMÁS EXCEPTO PARA SIEMBRA	Fob Miles USD	126,693	137,776	160,576	8.75%	16.55%	3.84%	
		Toneladas	55,354	63,348	56,113	14.44%	-11.42%	2.10%	
1511.10.00.00	ACEITE DE PALMA EN BRUTO	Fob Miles USD	40,897	69,673	63,428	70%	-9%	1.94%	
		Toneladas	36,940	69,066	67,692	87%	-1.99%	2.29%	
1604.20.00.00	LAS DEMÁS PREPARACIONES Y CONSERVAS DE PESCADO	Fob Miles USD	77,480	66,473	42,722	-14.21%	-35.73%	1.85%	
		Toneladas	14,471	12,586	9,012	-13.63%	-28.39%	0.42%	
2101.11.00.00	EXTRACTOS, ESENCIAS Y CONCENTRADOS DE CAFÉ	Fob Miles USD	49,759	61,162	49,089	22.92%	-19.74%	1.70%	
		Toneladas	6,353	7,773	7,247	22.34%	-6.76%	0.26%	
DEMÁS PRODUCTOS		Fob Miles USD	1,230,847	1,285,064	1,575,753	4.40%	22.62%	35.77%	
		Toneladas	740,654	751,057	777,490	1.40%	3.52%	24.87%	
<b>TOTALES</b>		Fob Miles USD	3,166,935	3,592,362	4,059,981	13.43%	13.01%	100.00%	
		Toneladas	2,835,418	3,019,957	3,080,923	6.51%	2.02%	100.00%	

\* La subpartida 0803.90.11.00 proviene de la subpartidas 0803.90.11.00 en la nomenclatura de la IV Enmienda.  
 \* La subpartida 0306.16.00.00 y 0306.17.99.00 provienen de las subpartidas 0306.13.91.00 y 0306.13.99.00 en la nomenclatura de la IV Enmienda. Valores del 2012 estimados por participación en el 2014.

Anexo 2 Productos que conforman el 64% de las exportaciones no petroleras. Enero-abril 2014.  
 Fuente: Dirección de Inteligencia Inversiones, PROECUADOR.

## EVOLUCIÓN EXPORTACIONES PESCA



Variación % Ene - Abr	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	TCPA 2007-2014
En USD	49%	-15%	3%	20%	17%	37%	-12%	12%
En TON	10%	-8%	3%	8%	-6%	21%	-3%	3%

Anexo 3 Evolución de las exportaciones no petroleras de Ecuador. Sector Pesca. Enero-abril 2014.  
 Fuente: Dirección de Inteligencia Inversiones, PROECUADOR

#### 7.4 Determinación del peso escurrido

El peso escurrido de todas las unidades de muestra se determinará mediante el procedimiento siguiente:

- i) mantener el envase a una temperatura de 20 °C a 30 °C durante un mínimo de 12 horas previamente al examen;
- ii) abrir el envase y verter el contenido distribuyéndolo en un tamiz circular previamente pesado que tenga una malla de alambre con aperturas cuadradas de 2,8 mm x 2,8 mm;
- iii) inclinar el tamiz con un ángulo de 17° a 20° aproximadamente y dejar escurrir el pescado durante dos minutos a partir del momento en que el producto se haya vertido en el tamiz;
- iv) pesar el tamiz con el pescado escurrido;
- v) determinar el peso del pescado escurrido se obtiene restando el peso del tamiz del peso del tamiz con el producto escurrido.

*Anexo 4 Metodología de evaluación del Peso escurrido según Codex Stan 70-1981*



*Anexo 5 Conservas de atún en envase flexible*

CASO	
No.	↓
BARCO	↓
FAO	↓
(O) ESPECIE	↓
(O) TALLA	↓
(O) TALAMC	↓
(O) TDESC (min)	↓
(O) DFTM	↓
(C) COCINA	↓
(C) FCOINADA	↓
(C) TTCOCCI	↓
(C) MERMA	↓
(C) DTRISALE	↓
(N) TTNEBU (min)	↓
(A) AUTOCPLY	↓
(A) HESTERIL	↓
(A) TTTEST (min)	↓
(G) TTOPER	↓
(G) TTPROD	↓
(F) FBILL	↓
(F) PER	↓
(S) WNETO	↓
(S) WGRATED	↓
(S) WESC	↓
(S) LIACOB	↓
(T) pH	↓
(T) %SAL	↓
(T) HST	↓
(T) GRATED	↓
(T) PNETO	↓
(T) PESCCZ	↓
(T) PESCC3	↓
(T) COLOR	↓
(T) TEXTURA	↓
(T) SABOR	↓
(T) CALIF	↓
(M) DPNETO	↓
(M) ESPACIO	↓
(M) DESCZ	↓
(M) DESC3	↓
(M) DPESC	↓
(M) DFILL	↓
(M) FABS	↓
(M) DENIS	↓
(M) DGRATED	↓
(M) DTTEST	↓
(M) TRASLADOS	↓

Anexo 6 Plantilla de levantamiento de información para base de datos. Fuente: Elaboración propia.





*Anexo 7 Vista Panorámica Planta de Procesamiento de Atún Marbelize S.A.*