

**Escuela Superior Politécnica del Litoral**

**Facultad de Ingeniería en Electricidad y Computación**

Modelo para pronosticar la demanda de alimentos balanceados en el sector  
acuícola

TECH-330

**Proyecto Integrador**

Previo la obtención del Título de:

**Ingeniera en Ciencias de la Computación**

Presentado por:

Noelia Alejandra Intriago Sánchez

Daniela Cithalli Landeta Púa

Guayaquil – Ecuador

Año: 2023 – 2024

## Dedicatoria

---

Dedico este proyecto a mis padres, Humberto y Penélope, y a mi hermano Daniel, quienes siempre supieron apoyarme a su manera durante toda mi vida. A mi peludo Nando, quien me cambió la vida cuando llegó y por quien me he esforzado para que tenga la vida que se merece.

A mi mejor amiga Nataly, quien me enseñó el significado de la lealtad y ha sido mi apoyo incondicional en cada paso de mi vida desde el colegio.

A mi amistad Daniela, quien terminó siendo mi dúo dinámico en mi aventura en Espol. A mis amistades del colegio y trabajo, por enseñarme que salir de la rutina es necesario para no desistir.

A todos los que me acompañaron durante este viaje y a los que se bajaron antes de tiempo, gracias.

**Noelia Alejandra Intriago Sánchez.**

## Dedicatoria

---

Con profundo cariño y gratitud, dedico este proyecto a: Mis padres, Gina y Gilbert, por su amor incondicional y su énfasis en una educación de calidad. Alaska, mi fiel compañera canina, fuente de alegría diaria. Mis hermanos, Gilbert y Michelle, mi inspiración constante. Totola y Mimí, mis segundas madres, por su amor y sabiduría. Ximena, mi mejor amiga desde la infancia, por su apoyo y lealtad inquebrantables. Ema, por su ayuda invaluable en los momentos más desafiantes. Noelia, mi primera amiga de la carrera, compañera de estudio excepcional y verdadera amiga. A mis amiTAWS y a todos mis compañeros de la universidad, por su amistad y compañerismo.

A todos, gracias por ser parte esencial de este viaje.

**Daniela Cithalli Landeta Púa.**

## Agradecimientos

---

Queremos expresar nuestra más sincera gratitud a todos aquellos que han contribuido al éxito de este proyecto:

A Eduardo Cruz, nuestro tutor, por su invaluable orientación y apoyo constante; a Ángel Catagua, por brindarnos la oportunidad de embarcarnos en este proyecto; a Juan Cadena, por sus consejos y recomendaciones técnicas; y a Jonathan Gorotiza, por su asistencia y conocimientos especializados.

Su ayuda y guía han sido fundamentales en cada paso de este camino.

## Declaración Expresa

---

Nosotras Noelia Alejandra Intriago Sánchez y Daniela Cithalli Landeta Púa, acordamos y reconocemos que:

La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores. La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por nosotras durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que nos corresponda de los beneficios económicos que la ESPOL reciba por la explotación de nuestra innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique las autoras que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 1 de Febrero del 2024.

NOELIA ALEJANDRA  
INTRIAGO SANCHEZ

Firmado digitalmente por NOELIA  
ALEJANDRA INTRIAGO SANCHEZ  
Fecha: 2024.02.07 14:52:42 -05'00'



---

Noelia Alejandra Intriago Sánchez

---

Daniela Cithalli Landeta Púa

## Evaluadores



Firmado digitalmente por:  
LUIS EDUARDO  
MENDOZA MORALES

---

**Luis Eduardo Mendoza Morales, Ph.D.**

Profesor de Materia

EDUARDO  
SEGUNDO  
CRUZ RAMIREZ

Firmado digitalmente  
por EDUARDO  
SEGUNDO CRUZ  
RAMIREZ  
Fecha: 2024.02.07  
15:02:11 -05'00'

---

**Eduardo Segundo Cruz Ramírez, MsC.**

Tutor de proyecto

## Resumen

ExpoCam, empresa en la industria de alimentos balanceados para acuicultura, enfrenta desafíos de sobreproducción y escasez debido a predicciones inexactas de demanda. El proyecto se centró en desarrollar un modelo de aprendizaje automático para mejorar la predicción de demanda y optimizar la producción. Utilizando datos históricos de ventas, exportaciones, precio del camarón y materias primas, se implementó un modelo de redes neuronales LSTM (Long-Short Term Memory, por sus siglas en inglés). Este enfoque permitió capturar patrones temporales complejos en los datos. Se creó también un dashboard interactivo para visualizar predicciones de ventas futuras por cliente y producto. El modelo de series de tiempo logró predecir con éxito las ventas del mes siguiente, mejorando la gestión de la cadena de suministro y el inventario en ExpoCam, y reduciendo notablemente los problemas de sobreproducción y escasez. Estos resultados evidencian la efectividad del aprendizaje automático en la optimización de la planificación y gestión de la producción. Como conclusión, el proyecto demuestra que la integración de tecnologías avanzadas, como el aprendizaje automático y series de tiempo, en los procesos de producción puede conducir a operaciones más eficientes y rentables, resaltando su valor en la industria de acuicultura.

**Palabras Clave:** Alimentos balanceados, Aprendizaje Automático, Series de Tiempo

## ***Abstract***

*ExpoCam, a company in the aquaculture feed industry, faces challenges of overproduction and shortages due to inaccurate demand predictions. The project focused on developing a machine learning model to improve demand prediction and optimize production. Using historical data on sales, exports, shrimp price and raw materials, a Long-Short Term Memory (LSTM) model was implemented. This approach allowed capturing complex temporal patterns in the data. An interactive dashboard was also created to visualize future sales predictions by customer and product. The time series model successfully predicted next month's sales, improving supply chain and inventory management at ExpoCam, and significantly reducing overproduction and shortage problems. These results demonstrate the effectiveness of machine learning in optimizing production planning and management. In conclusion, the project demonstrates that the integration of advanced technologies, such as machine learning and time series, into production processes can lead to more efficient and profitable operations, highlighting their value in the aquaculture industry.*

***Keywords:*** *Balanced Feeds, Machine Learning, Time Series*

## ÍNDICE GENERAL

Resumen .....	I
<i>Abstract</i> .....	II
Índice general .....	III
Abreviaturas.....	V
Índice de figuras .....	VI
Índice de tablas .....	VI
Capítulo 1 .....	1
1. Introducción .....	2
1.1 Descripción del problema .....	2
1.2 Justificación del Problema .....	3
1.3 Objetivos.....	4
1.3.1 Objetivo general .....	4
1.3.2 Objetivos específicos .....	4
1.4 Marco teórico.....	5
1.4.1 Trabajos relacionados .....	5
1.4.2 Bases teóricas .....	7
Capítulo 2 .....	9
2. Metodología .....	10
2.1 Análisis de la solución .....	10
2.1.1 Requerimientos funcionales.....	11
2.1.2 Requerimientos no funcionales.....	11
2.2 Alcance del proyecto.....	11
2.3 Riesgos y beneficios de la solución .....	12
2.3.1 Riesgos .....	12
2.3.2 Beneficios .....	13
2.4 Prototipado.....	13

2.5	Diseño de la solución .....	15
2.5.1	Vista de Escenarios – Historias de usuario .....	15
2.5.2	Vista lógica .....	16
2.5.3	Vista de desarrollo .....	17
2.5.4	Vista de proceso.....	18
2.5.5	Vista física .....	18
2.6	Plan de desarrollo.....	19
Capítulo 3	.....	21
3. Resultados y análisis .....		22
3.1	Desarrollo.....	22
3.1.1	Modelo de aprendizaje automático .....	22
3.1.2	Sistema .....	28
3.2	Resultados .....	28
3.2.1	Modelo de aprendizaje automático .....	28
3.2.2	Sistema .....	30
3.3	Análisis de costos.....	33
Capítulo 4	.....	34
4. Conclusiones y recomendaciones.....		35
4.1	Conclusiones .....	35
4.2	Recomendaciones.....	36
Referencias .....		37
Apéndices .....		39
Apéndice A .....		39
Apéndice B.....		41
Apéndice C.....		42

## **ABREVIATURAS**

LSTM Long-Short Term Memory

DNN Deep Neural Network

RSME Root Mean Square Error

ARIMA AutoRegressive Integrated Moving Average

MAE Mean Absolute Error

PIB Producto Interno Bruto

MSE Mean Squared Error

## ÍNDICE DE FIGURAS

Figura 1.....	10
Figura 2.....	14
Figura 3.....	14
Figura 4.....	15
Figura 5.....	16
Figura 6.....	18
Figura 7.....	18
Figura 8.....	20
Figura 9.....	23
Figura 10.....	23
Figura 11.....	24
Figura 12.....	24
Figura 13.....	25
Figura 14.....	26
Figura 15.....	29
Figura 16.....	30
Figura 17.....	31
Figura 18.....	32
Figura 19.....	32

## ÍNDICE DE TABLAS

Tabla 1.....	16
Tabla 2.....	33

# Capítulo 1

# **1. Introducción**

## **1.1 Descripción del problema**

En los últimos años, el sector acuícola en Ecuador ha experimentado un auge en la producción y exportación de camarones, con un notable aumento del 65% en las libras exportadas de camarón [1], consolidándose como uno de los principales actores en el mercado global. Sin embargo, la implementación de soluciones tecnológicas como herramientas de predicción y análisis de datos no es sencilla, puesto que requiere una comprensión profunda del sector, así como una inversión significativa en tecnología y formación para la obtención de resultados más coherentes, tal como señala Antonio José Boada en su estudio sobre la proyección de ventas por catálogo mediante pronósticos [2].

Otro de los problemas de la producción es referente a la determinación de los niveles apropiados de producción. Uno de estos desafíos es la sobreproducción, que implica el uso de recursos y la generación de contaminantes como principales consecuencias. Estos problemas emergen desde el inicio de la cadena alimenticia debido a una gestión inadecuada y una inapropiada disposición de almacenamiento [3]. La sobreproducción también implica un incremento en el consumo de recursos naturales, como el agua, la tierra y la energía y, al mismo tiempo, se traduce en un aumento de las emisiones de gases de efecto invernadero. Por último, la acumulación excesiva de materia prima indica también que la inversión de capital y presupuesto de la empresa se realizó de manera subóptima, ya que la demanda del mercado no resultó suficiente para cubrir los costos de producción.

Por otro lado, la producción también se ve afectada por la escasez, la cual se origina como consecuencia de eventos como la pandemia, conflictos bélicos y disputas internacionales, ya que limitan la disponibilidad de insumos necesarios para la producción de bienes, lo que a su vez lleva a una reducción de la cosecha con el fin de optimizar la gestión de recursos dentro de los límites presupuestarios [4]. Esta situación, a su vez, conlleva a un aumento en los precios de los

productos para el consumidor, quienes se ven afectados por el encarecimiento de los productos y buscan productos que se ajusten a su disponibilidad económica.

Por último, la falta de información precisa acerca de la evolución de la demanda del mercado, en lo que respecta a sus intereses y necesidades, tiene un impacto negativo en la planificación de una campaña de marketing efectiva destinada a atraer clientes de la competencia y fortalecer relaciones con los actuales [5]. La elaboración de estrategias de marketing requiere una meticulosa planificación, ya que deben ajustarse lo más posible a la situación actual del mercado. Esto es esencial para evitar la promoción de productos que se agoten demasiado pronto o, por otro lado, para prevenir una inversión excesiva en una campaña que no logre alcanzar al público deseado.

## **1.2 Justificación del Problema**

La toma de decisiones en la producción de alimentos debe ser precisa para evitar gastos innecesarios debido a la sobreproducción y el almacenamiento de materia prima, así como la pérdida de clientes y, por consiguiente, de ingresos debido a la insatisfacción de la demanda del mercado. Este proceso es fundamental para gestionar eficazmente el presupuesto y el capital de la empresa, construir confianza con clientes potenciales y realizar inversiones en mejoras. Sin embargo, esta problemática conlleva desafíos que requieren un enfoque que simplifique el proceso de toma de decisiones.

En el Simposio Argentino sobre Tecnología y Sociedad (STS) de 2017 [6], se resaltó la importancia de las técnicas de predicción en el proceso de toma de decisiones. Estas técnicas, respaldadas por el análisis de indicadores clave y la utilización de datos históricos, han demostrado ser fundamentales para lograr predicciones más confiables y precisas. Combinando estos datos con la implementación de herramientas de Business Intelligence (BI) y Machine Learning (ML), se pueden revelar tendencias en el mercado y ajustar la producción en consecuencia.

Estas técnicas poseen un gran potencial en lo que respecta a la optimización de la producción y la retención de clientes existentes, así como la atracción de nuevos clientes en un mercado competitivo. Para alcanzar este objetivo, es esencial desarrollar herramientas tecnológicas que mejoren la capacidad de reconocer patrones en el comportamiento de los clientes, especialmente en relación con las temporadas de exportación. Al identificar estos patrones, se puede tomar decisiones más acertadas que conducen a una gestión más eficiente de los recursos y el capital disponibles.

La propuesta de solución basada en la previsión de producción de balanceado de camarón según su demanda en el mercado puede mejorar el manejo y gestión de recursos de estas empresas, y un mejor control del capital invertido en la producción del balanceado, y mejorar sus relaciones con los clientes que vean sus necesidades cubiertas.

### **1.3 Objetivos**

#### **1.3.1 *Objetivo general***

Desarrollar un modelo de aprendizaje automático, utilizando datos históricos de la producción de alimentos de balanceado, junto con tendencias del mercado con el fin de predecir de manera precisa la demanda en el sector acuícola y contribuir a una planificación y producción más eficiente.

#### **1.3.2 *Objetivos específicos***

1. Realizar un análisis exploratorio de los datos históricos relacionados con la demanda de alimentos balanceados en el sector acuícola e integrar factores externos relevantes, como tendencias del mercado, para comprender mejor las influencias en la demanda.
2. Construir un modelo de aprendizaje automático, utilizando datos históricos y un conjunto de datos de prueba para predecir la demanda futura en ExpoCam S.A.

3. Desarrollar un dashboard centrado en el usuario que visualice las predicciones del modelo de demanda de alimentos, permitiendo a los stakeholders tomar decisiones informadas y adaptarse a las tendencias del mercado en tiempo real.

## 1.4 Marco teórico

### 1.4.1 Trabajos relacionados

La predicción de la demanda es una herramienta esencial en la industria, especialmente en sectores tan especializados como el acuícola. En este sector, la demanda de alimentos balanceados es fundamental para garantizar la estabilidad y rentabilidad de las operaciones. La precisión en el pronóstico de la demanda puede tener un impacto directo en la eficiencia de la cadena de suministro, la gestión de inventarios y la satisfacción del cliente. A continuación, se presentan algunos trabajos relacionados con esta temática desde diferentes perspectivas:

En el estudio realizado por Yan, Chen Zheng, Zhang, & Xin, se empleó un modelo de redes neuronales LSTM (Long-Short Term Memory, por sus siglas en inglés) para predecir la demanda de bicicletas compartidas en New York y Jersey City [7]. Tras analizar los datos, se identificaron variaciones en el alquiler por hora entre ambas ciudades y una escasa relación en los viajes entre ellas, lo que presentó desafíos como la escasez de datos. Para su modelo, incorporaron variables como la hora, estación de inicio y fin del viaje, y tres indicadores meteorológicos. Al comparar los resultados, el modelo LSTM demostró ser más preciso que el DNN (Deep Neural Network), evidenciado por una raíz del error cuadrático medio (RSME, por sus siglas en inglés) más bajo en los datos de prueba.

Oscar William Sánchez Córdova [8], realizó un estudio meticuloso sobre la demanda del alimento referencia 65, evaluando diversos modelos para ajustar su curva de demanda.

Utilizando métricas como RMSE y MAE (Mean Absolute Error o *Error Absoluto Medio* en español), Sánchez Córdova analizó tanto modelos ARIMA (AutoRegressive Integrated Moving Average) como regresiones estáticas y dinámicas. Aunque el modelo ARIMA (0,1,1) mostró un

buen ajuste según el MAE, no fue el más preciso en los datos de prueba. Sorprendentemente, la regresión estática que incorpora variables como el producto interno bruto (PIB), precio y exportaciones emergió como el modelo óptimo, con un MAE de 1044.86. Esta investigación destaca la importancia de la precisión en la selección de modelos, y cómo diferentes enfoques pueden variar en su eficacia dependiendo de los datos con los que se trabajen. La ecuación (1) representa la expresión final del modelo seleccionado.

$$DDDDDDDDDDDD = 0.00656(PPPPPP) - 0.0583(EEEEEEEEEEDDEEEEEEDDDEE) - 0.271(PPEEDDEEEEEE) - 10670 \quad (1)$$

La era del Big Data ha traído consigo una revolución en la forma en que se abordan los problemas de pronóstico. Herramientas y plataformas como Spark permiten la implementación de algoritmos escalables que pueden procesar grandes conjuntos de datos con eficiencia. Sin embargo, dada la magnitud de las muestras y características, a veces la potencia computacional no es suficiente, haciendo necesario la reducción de dimensiones. En un estudio reciente, se empleó el Análisis de Componentes Principales (PCA) para la extracción de características [9]. Este estudio evaluó el rendimiento de un sistema de pronóstico en el SOK Market, que vende 21 grupos diferentes de artículos. Se observó que la estrategia de integración (S2) proporcionó una mejora del 38.99% sobre la primera estrategia (S1). Además, con la incorporación de un modelo de Aprendizaje Profundo (DL), se logró una mejora adicional en la precisión del pronóstico de la demanda. Específicamente, el modelo que combinó la estrategia de integración novedosa con el enfoque de aprendizaje profundo resultó ser el más eficiente, logrando un error de predicción medio del 24.7%. Estos avances subrayan la importancia de la adaptabilidad y la innovación en el sector minorista, y cómo diferentes enfoques pueden variar en su eficacia según los datos con los que se trabajen [9].

Estos trabajos, junto con otros en el campo, subrayan la importancia de una investigación continua y adaptativa en el pronóstico de la demanda en el sector acuícola. La elección adecuada de modelos y técnicas puede tener un impacto significativo en la sostenibilidad y rentabilidad de

las operaciones acuícolas. Además, es esencial reconocer que el sector acuícola no opera en un vacío. Factores externos, como los cambios climáticos, las políticas gubernamentales y las tendencias del mercado global, pueden influir significativamente en la demanda de alimentos balanceados. Estos factores, a menudo impredecibles, añaden otra capa de complejidad al desafío del pronóstico.

## **1.4.2 Bases teóricas**

### **1.4.2.1 Modelos de Aprendizaje Automático**

El aprendizaje automático, a diferencia de los enfoques estadísticos tradicionales, pueden aprender patrones complejos a partir de grandes conjuntos de datos. En el sector acuícola, modelos como las redes neuronales artificiales [10] y las máquinas de soporte vectorial [11] han demostrado ser eficaces en la predicción de la demanda. Además, los árboles de decisión y los bosques aleatorios [12] se han utilizado para clasificar y predecir tendencias basadas en factores múltiples. La ventaja de estos modelos es su capacidad para manejar datos no lineales y la posibilidad de mejorar su precisión a medida que se disponga de más datos.

### **1.4.2.2 Series de tiempo**

Las series temporales son una secuencia de observaciones registradas en puntos sucesivos en el tiempo o en intervalos de tiempo espaciados de manera uniforme. En el ámbito de la predicción, las series temporales son esenciales para analizar tendencias, estacionalidades y ciclos en los datos [13]. En el sector acuícola, las series temporales pueden ayudar a identificar patrones en la demanda de alimentos balanceados, permitiendo a las empresas anticipar necesidades futuras. Los modelos ARIMA [14] y los modelos de descomposición estacional [15] son ejemplos de técnicas que se han utilizado ampliamente para analizar y predecir series temporales. Con la llegada de técnicas más avanzadas, como las redes neuronales LSTM [16], la precisión y eficiencia de las predicciones basadas en series temporales han mejorado significativamente.

### **1.4.2.3 Predicción de la demanda**

La predicción de la demanda es un componente esencial en la planificación y gestión de cualquier negocio, y su precisión puede tener un impacto directo en la eficiencia operativa y la rentabilidad. Los métodos tradicionales, como los modelos ARIMA, han sido complementados o incluso superados por técnicas de aprendizaje automático, como las máquinas de soporte vectorial y las redes neuronales [17]. Además, con la integración de factores externos, como las condiciones climáticas y las tendencias del mercado, las predicciones se han vuelto más robustas y precisas [18].

### **1.4.2.4 Integración de Tecnologías en la Predicción**

La integración de diversas tecnologías en la predicción ha revolucionado la forma en que las empresas abordan la planificación y la toma de decisiones. La combinación de tecnologías de almacenamiento de datos, como Hadoop, con herramientas de procesamiento, como Spark, ha permitido a las empresas procesar y analizar datos a una escala sin precedentes [19]. Además, la integración de la inteligencia artificial y el aprendizaje automático ha mejorado la precisión y la eficiencia de las predicciones. Las soluciones basadas en la nube, como AWS y Azure, ofrecen plataformas escalables que pueden adaptarse a las necesidades cambiantes de las empresas, permitiendo una implementación rápida de modelos predictivos y garantizando la disponibilidad de datos en tiempo real [20]

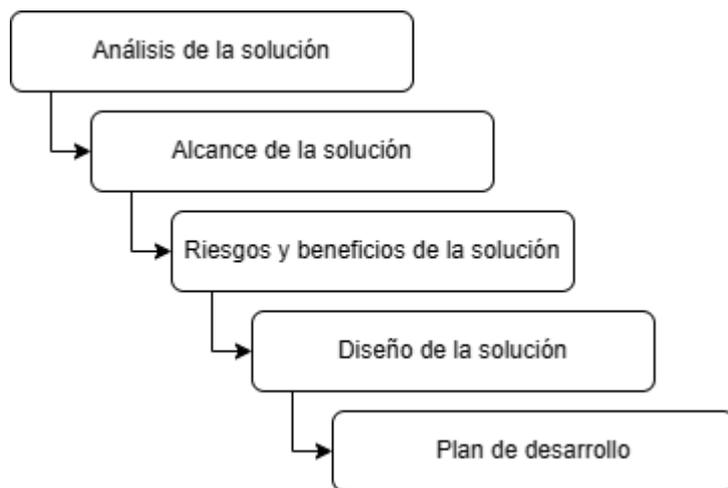
## **Capítulo 2**

## 2. Metodología

El proyecto se centró en la implementación de un modelo de aprendizaje automático para predecir la demanda de producción de balanceado de camarón mediante el análisis de datos históricos y factores externos. El objetivo principal consistió en generar pronósticos precisos y, además, presentar estos resultados de manera visual a través de gráficos claros y comprensibles. La meta fundamental del modelo fue proporcionar una herramienta que respalde la toma de decisiones gerenciales al permitir una gestión más eficiente de los recursos disponibles. El proceso de desarrollo de esta solución se detalla en la Figura 1.

### Figura 1

*Pasos a seguir en la metodología*



#### 2.1 Análisis de la solución

El primer paso de la metodología consistió en levantar los requerimientos, para comprender las necesidades del cliente y establecer un punto de partida para el desarrollo posterior de la solución. Para lograr un mejor entendimiento de la problemática, se llevaron a cabo varias reuniones con usuarios clave de la empresa del cliente, donde se recopiló información detallada sobre los desafíos en la producción de balanceado y sus impactos en diversas áreas. Durante estas reuniones, se identificaron aspectos fundamentales del problema y expectativas de la solución que luego se usaron para definir los requerimientos. Finalmente, se

procedió con el levantamiento de los requerimientos funcionales y no funcionales del proyecto, los cuales son detallados a continuación:

### **2.1.1 *Requerimientos funcionales***

1. El sistema debe mostrar las predicciones de demanda de manera clara y comprensible mediante gráficos de visualización de tendencias.
2. El sistema debe identificar tendencias atípicas en la demanda del mercado y proporcionar información contextual para entenderlas.
3. El sistema debe estimar la cantidad de producto a vender en base a registros pasados.
4. El sistema debe mostrar un registro histórico de transacciones de ventas y exportaciones anteriores; de tal manera que pueda evaluarse la precisión del pronóstico.
5. El sistema debe presentar información detallada y resultados en tiempo real para poder tomar decisiones con mayor eficacia.

### **2.1.2 *Requerimientos no funcionales***

1. La interfaz de usuario del sistema debe ser fácil de usar e intuitiva, de manera que sea accesible para cualquier usuario del sistema.
2. El sistema debe estar disponible de manera continua para garantizar una respuesta oportuna al momento de tomar decisiones en cuanto al ajuste de producción.
3. El sistema debe ser capaz de manejar factores externos al momento de generar un pronóstico.
4. El tiempo de respuesta del sistema a los requerimientos de generación de pronósticos no pueden exceder de 1 minuto.

## **2.2 Alcance del proyecto**

Una vez establecidos los requerimientos de este proyecto, se procede con la definición del alcance. La solución se enfocó en el desarrollo e implementación de un modelo de

aprendizaje automático capaz de realizar pronósticos de demanda de balanceado de camarón, fundamental para ajustar la producción de esta empresa del sector acuícola, siendo que el alcance se detalla de la siguiente manera:

- El modelo debe ser capaz de considerar factores externos a la producción de balanceado para generar pronósticos más acercados a la realidad.
- El modelo debe identificar variaciones en la tendencia de producción.
- El modelo debe poder mostrar registros anteriores para poder comparar los resultados.

Al mismo tiempo, se identificaron ciertas limitantes respecto al desarrollo de la solución, las cuales fueron definidas de acuerdo con la complejidad de desarrollo y el tiempo disponible para poder abordarlas:

- El modelo no brinda recomendaciones de mejora en cuanto a la producción del balanceado.
- El modelo será implementado para trabajar con datos proporcionados por el cliente; por ende, no se toma en consideración la implementación o distribución para otras empresas del sector acuícola.

## **2.3 Riesgos y beneficios de la solución**

### **2.3.1 Riesgos**

El principal riesgo es la imprecisión en los resultados de los pronósticos, puesto que no se puede asegurar que el modelo siempre acertará en los pronósticos. Esto puede deberse a la complejidad del conjunto de datos con el que se entrene al modelo, la variabilidad en los factores externos y limitaciones propias de los modelos de aprendizaje automático.

La complejidad de los factores externos es un riesgo para considerar, pues estos deberían elegirse tras un análisis minucioso del impacto de cada uno en la producción del balanceado. El no escogerlos de manera adecuada puede afectar negativamente la precisión del modelo.

La exclusividad es un riesgo para nuestro proyecto si se quiere trabajar con datos proporcionados por otra empresa, ya que es una solución dirigida a un cliente en particular. Además, los datos proporcionados podrían estar incompletos o mal estructurados, lo que puede afectar la precisión de los pronósticos del modelo.

### **2.3.2 Beneficios**

La creación de un modelo de aprendizaje automático para generar pronósticos de demanda representa una herramienta valiosa en el sector productivo, puesto que la producción puede ajustarse de mejor manera, reduciendo costos operativos y maximizar la eficiencia de la maquinaria. Por consiguiente, el uso de este modelo puede mejorar la toma de decisiones en base a la posible demanda futura, ya que se pueden elaborar planes estratégicos con mayor agilidad y eficacia.

El trabajar con factores externos le da al modelo una mejor comprensión de la influencia de los mismos respecto a la demanda del balanceado, así como la detección de variaciones en las tendencias. Esto permite tomar acciones de manera temprana y adaptarse a los cambios inesperados del mercado, derivando en la satisfacción del cliente y fortalecimiento de las relaciones comerciales.

## **2.4 Prototipado**

Se realizó un prototipo de alta fidelidad del dashboard para visualizar lo que sería la propuesta de solución para este proyecto. Para el prototipo se diseñó la interfaz gráfica de la aplicación web y los componentes de esta que servirán, resultando en una pantalla para cargar el archivo con el cual el modelo generará el pronóstico como se visualiza en la Figura 2, y otra pantalla para generar el dashboard con los gráficos de los pronósticos generados, tal como se muestra en la Figura 3. El prototipo fue realizado en Figma debido a su rapidez y facilidad de aprendizaje al momento de desarrollar el prototipo.

**Figura 2**

*Prototipo: carga de archivo*



**Figura 3**

*Prototipo: visualización de gráficos*

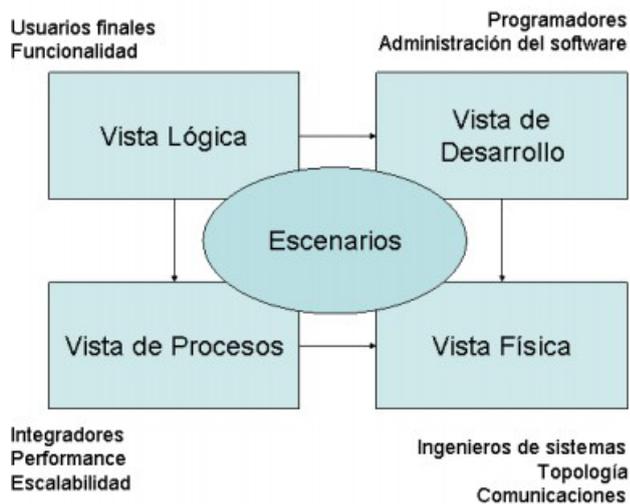


## 2.5 Diseño de la solución

Posterior al levantamiento de requerimientos, la definición del alcance del proyecto, sus riesgos y beneficios, se realizó el diseño de la solución. Para el diseño de la solución se usó el modelo “4+1” vistas de Kruchten [21], con el cual se puede describir el sistema desde diferentes puntos de vista (ver Figura 4). En este modelo se describirá la estructura y funcionalidad del sistema, los componentes, el comportamiento del sistema, las conexiones de los componentes y los escenarios de la solución.

**Figura 4**

*Modelo 4+1 vistas de Kruchten [21]*



### 2.5.1 Vista de Escenarios – Historias de usuario

Después del levantamiento de requerimientos, la definición del alcance del proyecto, sus riesgos y beneficios, se procedió con la elaboración de las historias de usuario, para representar los escenarios o situaciones que podrían enfrentar los usuarios del sistema.

También se definieron los criterios de aceptación de la solución, que sirven de guía sobre lo que se espera lograr de la historia de usuario especificada y se pueden visualizar con más detalle en el apéndice A de este documento. A continuación, se presentan de manera breve las historias de usuario creadas para este proyecto:

**Tabla 1**

*Historias de usuario*

<b>Identificador</b>	<b>Historia de usuario</b>
<b>ABI-001</b>	<i>Como Analista BI <b>necesito acceder a una plataforma que visualice las predicciones de demanda de alimentos balanceados</b> con la finalidad de tomar decisiones informadas y adaptarme a las tendencias del mercado</i>
<b>PRD-001</b>	<i>Como Gerente de Producción <b>necesito saber la cantidad de alimento a producir</b> con la finalidad de evitar sobreproducción o escasez</i>
<b>MKI-001</b>	<i>Como Responsable de Marketing Intelligence <b>necesito saber la demanda del mercado de manera precisa y efectiva</b> con la finalidad de realizar una planificación óptima de estrategias de marketing</i>

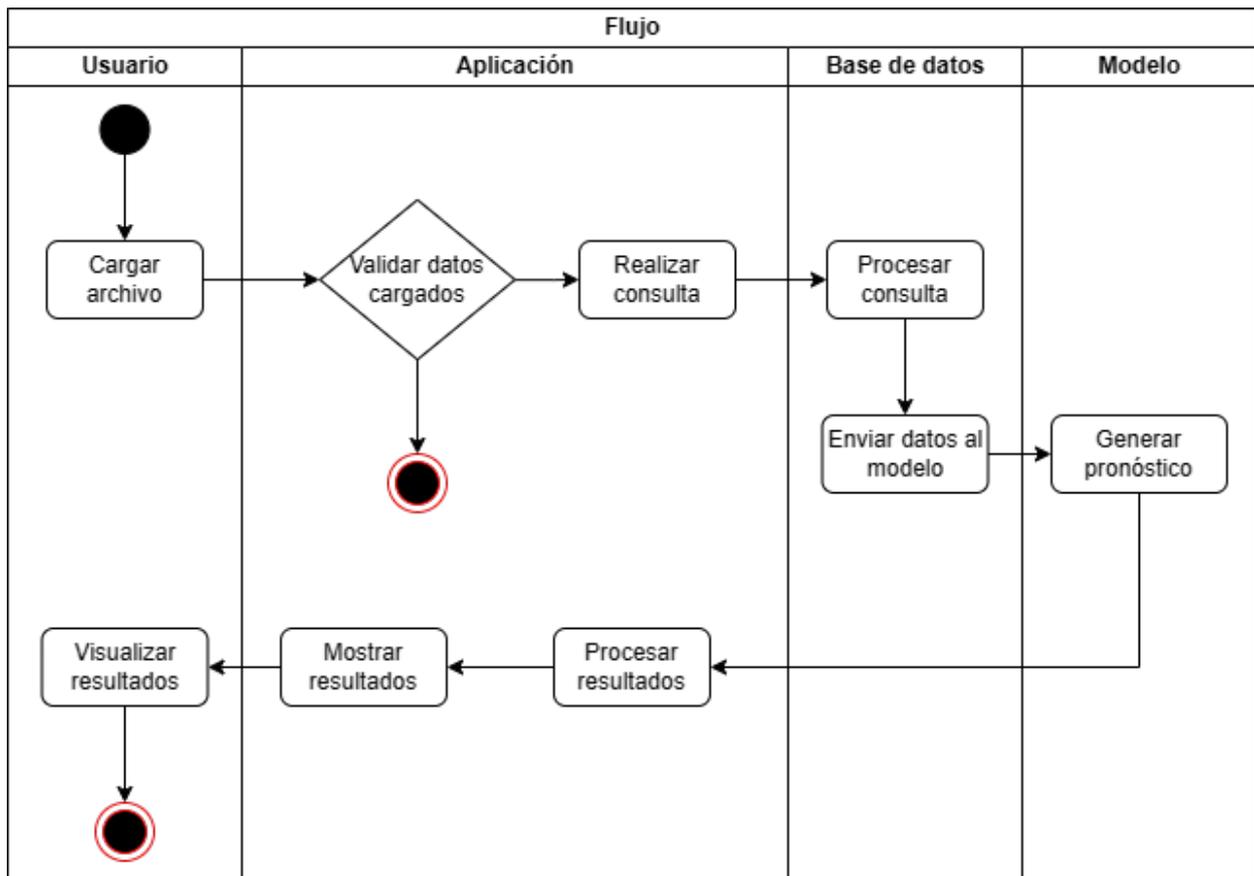
### **2.5.2 Vista lógica**

El diagrama de la Figura 5 muestra el flujo de actividades de la aplicación web, donde se contempla el escenario en el que el usuario carga un archivo de ventas para generar un pronóstico. Cuando el archivo se carga, la aplicación, que actúa como servidor y cliente a la vez, maneja la solicitud. Primero, valida los datos cargados y, de ser válidos, procede a consultar la base de datos; caso contrario, el flujo acaba. De realizarse la consulta, los datos obtenidos se envían al modelo de aprendizaje automático para ser procesados y generar el pronóstico solicitado.

Finalmente, la aplicación procesa los resultados del modelo y los presenta en la interfaz de usuario a través de gráficos intuitivos y detallados que ayudan al usuario a analizar y comprender los resultados del pronóstico de manera clara y efectiva, finalizando con el flujo de actividad de la aplicación.

### **Figura 5**

*Diagrama de actividades de la solución*



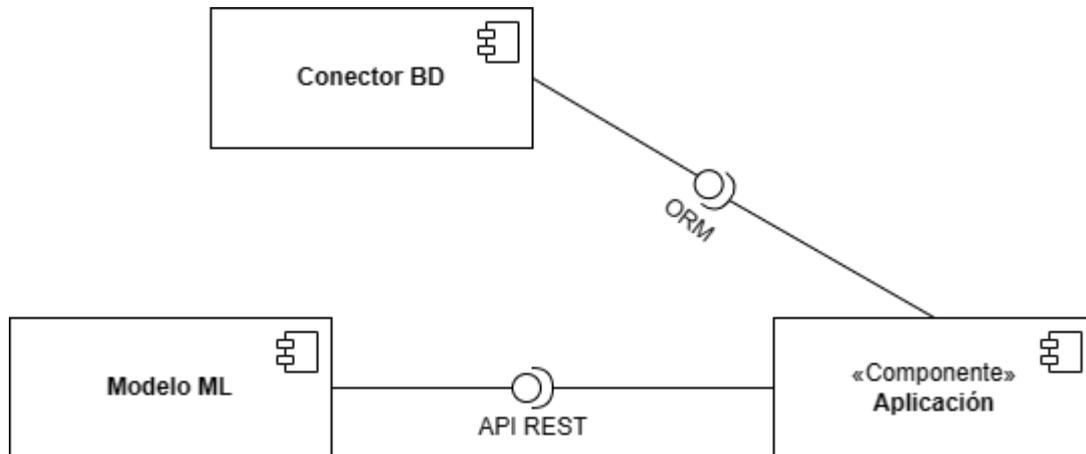
### 2.5.3 Vista de desarrollo

El diagrama de componentes de la aplicación, representado en la Figura 6, muestra las relaciones entre los distintos componentes del sistema. La capa de presentación contendrá al componente front-end de la aplicación, encargado de procesar y mostrar los resultados obtenidos del pronóstico del modelo de aprendizaje automático. Este componente funciona al mismo tiempo como el componente back-end de la aplicación, simplificando la lógica de la aplicación y la interfaz de usuario.

Este componente interactúa tanto con el componente del modelo de aprendizaje automático como con el correspondiente a la base de datos alojada en Azure. Con esta última interactúa mediante un ORM para gestionar las consultas o actualizaciones que se realicen en el proceso de generación de pronósticos. La interacción con el modelo puede ocurrir sin necesidad de una interfaz entre estos componentes.

**Figura 6**

*Diagrama de componentes de la solución*



#### **2.5.4 Vista de proceso**

La vista de proceso se centra en la solución a posibles problemas de concurrencia en el sistema. Sin embargo, como se implementó tecnología encargada de manejar la concurrencia entre la aplicación y la base de datos, esta vista no se desarrolló para su diseño.

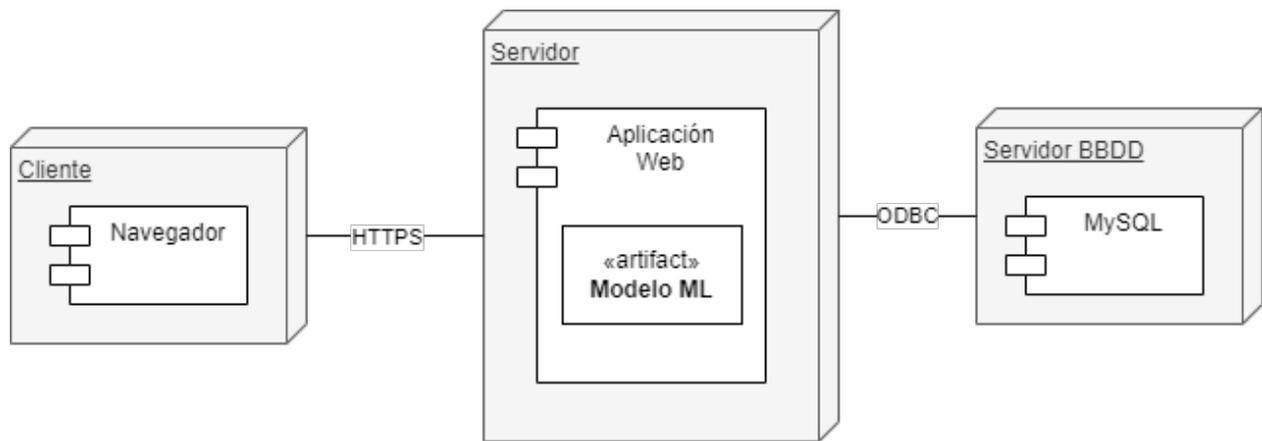
#### **2.5.5 Vista física**

La Figura 7 muestra el despliegue del sistema, donde se puede visualizar la integración de los componentes. El navegador del usuario accede directamente a la aplicación, donde se maneja la lógica de negocio y la interfaz de usuario en un solo entorno. La renderización del dashboard y la gestión de interacciones del usuario se llevan a cabo dentro del mismo ambiente de desarrollo gracias a la tecnología usada para el desarrollo de la aplicación.

El componente del modelo de aprendizaje automático puede ser manejado sin inconvenientes dentro del mismo ambiente que el componente de la aplicación. Para la base de datos, ya no es requerido un componente exclusivamente dedicado al manejo la comunicación HTTPS entre el front-end y back-end, puesto que el componente de la aplicación actúa como punto único de interacción.

**Figura 7**

*Diagrama de despliegue de la solución*



## 2.6 Plan de desarrollo

Para la ejecución del diseño de la solución se consideraron 3 meses de desarrollo, y, usando la metodología Scrum, se realizaron sprints con una duración de dos semanas. La entrega del modelo de aprendizaje automático se realizó al final del tercer sprint, mientras se mantenían reuniones semanales con el cliente donde se presentaban avances y retroalimentación respecto al desempeño del modelo.

De igual manera, la entrega de las historias de usuario referentes a la visualización e interacción con los gráficos del dashboard se realizó al final del tercer sprint, donde se presentó la distribución de la pantalla y su funcionalidad base. Finalmente, las historias de usuario referentes a la visualización de producciones pasadas y generación de alertas de variaciones en tendencias se realizaron en el cuarto y quinto sprint respectivamente.

Considerando las entregas planteadas, los módulos creados para la fase de desarrollo fueron: el modelo de aprendizaje automático, incluidos su entrenamiento y validación, el desarrollo de la aplicación web, el desarrollo de la base de datos del sistema y la integración de los componentes para su posterior despliegue:

- **Modelo de pronósticos de demanda:** Para que el modelo genere un pronóstico, debe trabajar con series de tiempo. Los modelos ARIMA son ideales al trabajar en estadística. Sin embargo, para el enfoque de este proyecto, el modelo se basó en una



## **Capítulo 3**

### 3. Resultados y análisis

#### 3.1 Desarrollo

##### 3.1.1 *Modelo de aprendizaje automático*

###### 3.1.1.1 **Recolección de datos**

Para el desarrollo de la solución, se utilizaron datos proporcionados por el cliente que permiten analizar las tendencias de demanda de alimento para camarón. Los datos recolectados incluyen:

- **Share of Wallet:** cantidad que cada cliente podría comprar mensualmente y el porcentaje destinado a los productos del cliente del proyecto.
- **Transacciones de ventas:** fechas de compra, cliente, grupo de producto adquirido y cantidad en toneladas.
- **Materia prima:** componentes utilizados en la producción del balanceado de camarón, con precios, costos de transporte y cantidad adquirida.
- **Precio del camarón:** variaciones en el precio del camarón en función de su talla.
- **Transacciones de exportaciones:** fechas, volumen en libras y el costo de la exportación del balanceado.

###### 3.1.1.2 **Análisis exploratorio**

Durante el desarrollo de la solución, se realizó un análisis exploratorio de los datos, divididos en categorías claves fundamentales para construir el modelo de aprendizaje automático. A continuación, se presenta detalladamente el análisis exploratorio realizado:

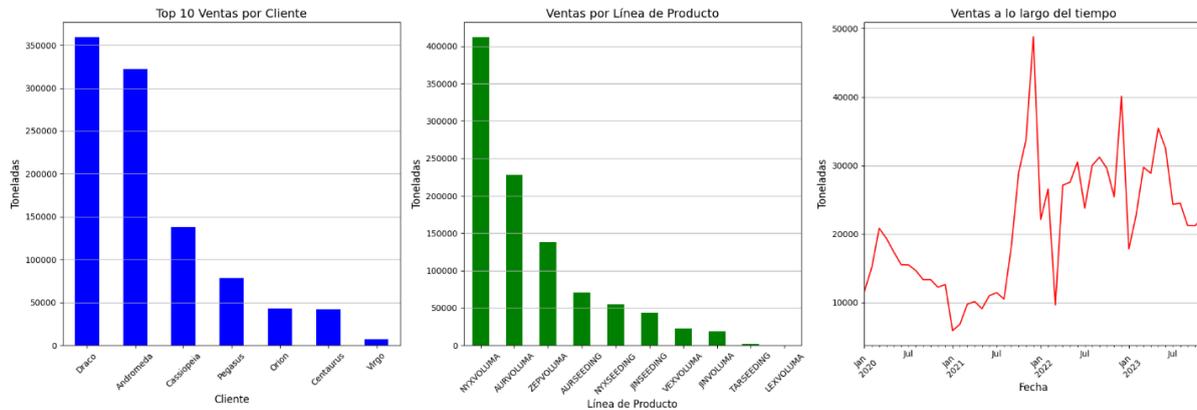
###### 3.1.1.2.1 *Ventas*

La figura 9 muestra una concentración de ventas en pocos clientes y productos, siendo Draco y Andromeda los más destacados y VOLUMA como la línea de producto dominante. Las ventas a lo largo del tiempo exhiben una volatilidad notable con picos pronunciados y una

tendencia general decreciente, sugiriendo patrones estacionales o eventos de mercado específicos que afectan la demanda.

**Figura 9**

*Gráficos de ventas por cliente y producto*

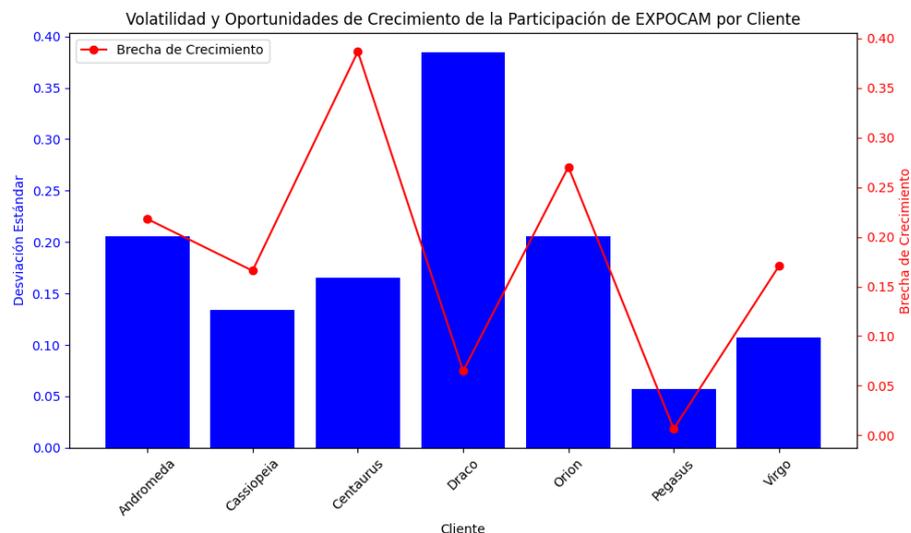


3.1.1.2.2 *Share of Wallet*

La figura 10 presenta la volatilidad y potencial de crecimiento de la participación de EXPOCAM por cliente. Draco muestra la mayor volatilidad en su participación y la mayor brecha de crecimiento, sugiriendo una oportunidad para incrementar su participación de mercado. Centaurus presenta la menor volatilidad, indicando una participación de mercado estable, mientras que Orion y Pegasus tienen oportunidades de crecimiento moderadas.

**Figura 10**

*Tendencia de participación de EXPOCAM por cliente*

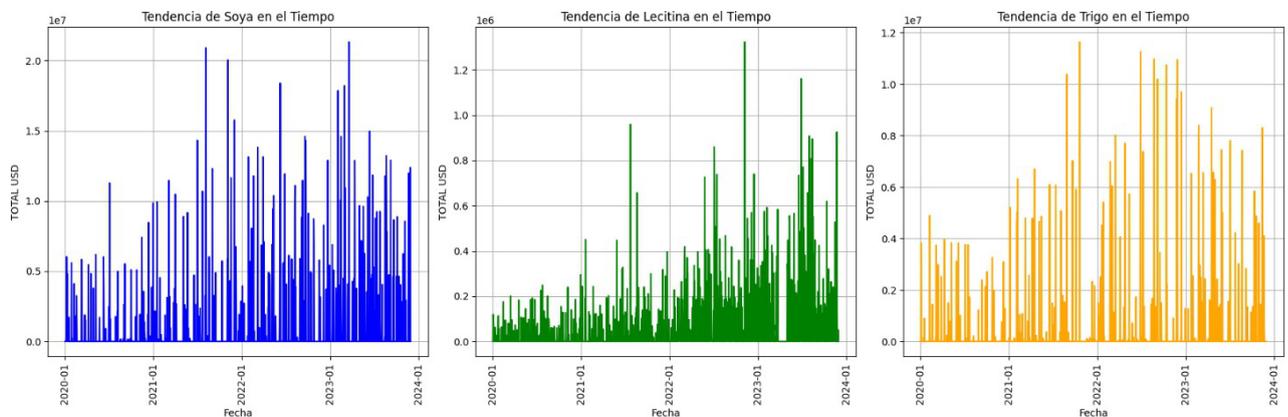


### 3.1.1.2.3 *Materia prima*

La figura 11 muestra las fluctuaciones de precios de soya, lecitina y trigo, ingredientes clave en la producción de alimento para camarón. La soya muestra la mayor volatilidad con picos erráticos, la lecitina presenta una variabilidad moderada, y el trigo tiene la menor volatilidad y picos más bajos. Estos picos sugieren posibles incrementos en los costos que podrían afectar la rentabilidad de los productores.

**Figura 11**

*Tendencia de soya, lecitina y trigo en el transcurso del tiempo*

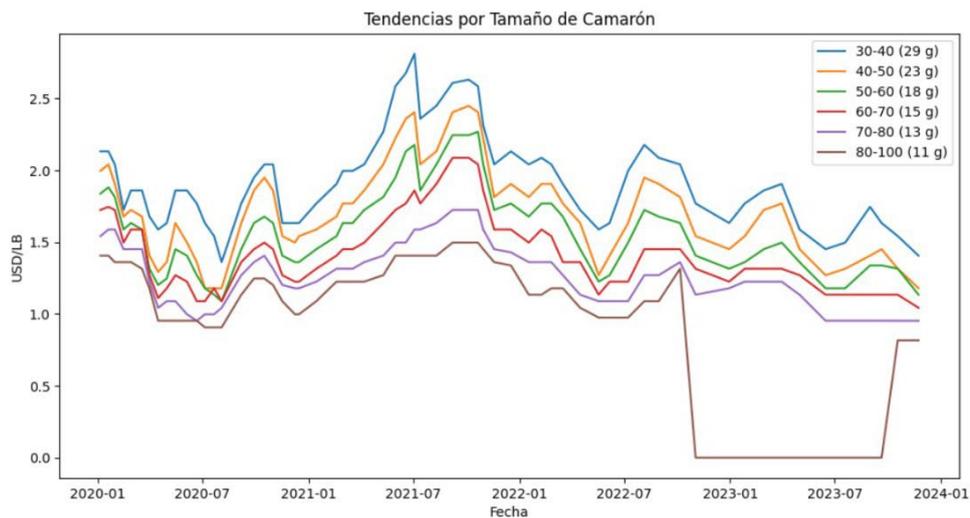


### 3.1.1.2.4 *Precio del camarón*

La Figura 12 exhibe las variaciones en los precios de distintos tamaños de camarón a lo largo de casi cuatro años, con un precio superior para los tamaños mayores. Existe una estacionalidad evidente, con fluctuaciones regulares en los precios, indicando posibles cambios en la demanda o en la producción. Recientemente, una caída pronunciada en los precios de todos los tamaños sugiere una perturbación del mercado, seguido de una tendencia de recuperación, especialmente en los camarones más grandes. Estos patrones indican una correlación entre los tamaños y subrayan la volatilidad del mercado.

**Figura 12**

*Tendencia de precios según el tamaño del camarón*

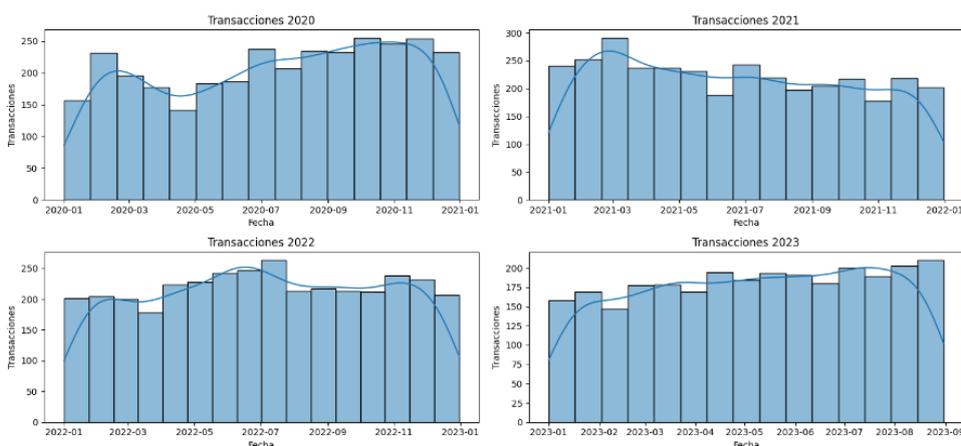


### 3.1.1.2.5 Transacciones de exportaciones

La figura 13 muestra que en 2020 la exportación de balanceado tuvo un declive entre marzo y mayo, hemos de suponer que por la pandemia del COVID-19. En 2021, la exportación de balanceado tuvo un aumento entre enero y marzo, debido a la recuperación de la economía mundial, pero con un ligero declive a partir de mayo. En 2022, la exportación de balanceado se mantuvo estable a lo largo del año. En 2023, la exportación de balanceado ha seguido una tendencia creciente a lo largo del año.

**Figura 13**

### Histogramas de exportaciones de 2020, 2021, 2022 y 2023

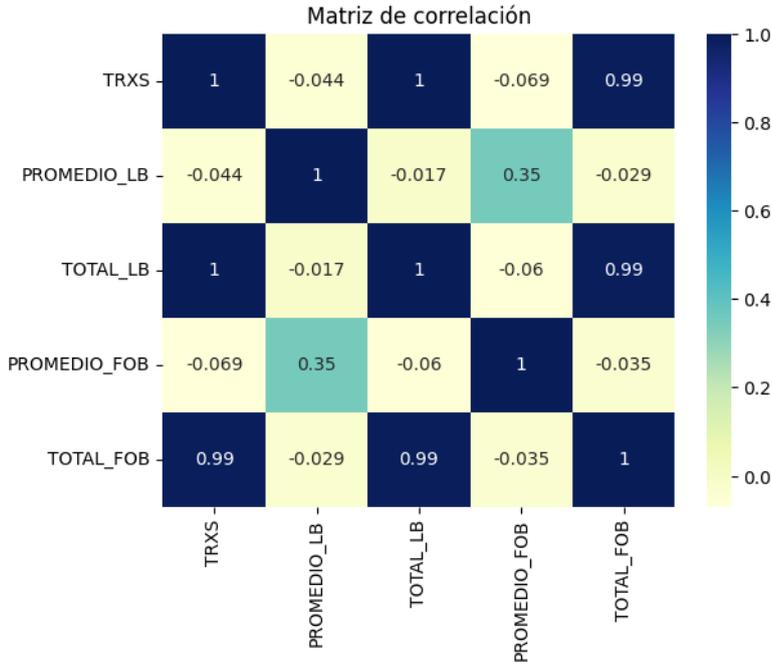


En la figura 14, se observa una correlación positiva (menor a 0.5) entre el promedio de libras y el promedio de U\$S FOB. También, una fuerte correlación positiva entre los totales de libras y los totales de U\$S FOB con el total de transacciones, lo que indica que a mayor cantidad

de transacciones, mayor será el total de libras y el total de U\$\$ FOB. Se observan correlaciones muy cercanas a 0 para los demás casos, por lo que no se puede concluir que exista una correlación entre las variables.

**Figura 14**

*Matriz de correlación entre variables de exportación*



**3.1.1.3 Tratamiento de datos**

El siguiente paso consistió en el tratamiento y transformación de los datos que se utilizaron para entrenar nuestro modelo, incluyendo la anonimización de información confidencial de los clientes, con el objetivo de salvaguardar la privacidad y confidencialidad de los datos durante el análisis y desarrollo del modelo.

El tratamiento de los datos consistió en las siguientes etapas:

- **Reducción de dimensionalidad:** selección de columnas relevantes para el análisis, incluyendo fechas de transacciones, identificadores de clientes y productos, así como cantidad y costos de productos vendidos o exportados

- **Limpieza de datos:** eliminación o transformación de datos nulos mediante imputación de valores o descarte de columnas en aquellas seleccionadas con datos numéricos para asegurar la calidad y predicción del análisis.
- **Transformación de datos:** transposición de columnas para poder realizar la unión entre conjuntos de datos con datos resumidos y calculados de acuerdo con el contexto de los datos.
- **Integración de datos:** combinación de datos en un único conjunto, donde cada fila representa una semana, incluyendo todas las ventas, exportaciones y demás datos relevantes para el pronóstico de demanda

#### 3.1.1.4 Construcción del modelo

Para este proyecto, se ha desarrollado y entrenado una red neuronal LSTM que procesa datos históricos para detectar patrones temporales. La arquitectura del modelo consiste en:

- **Capa LSTM:** Compuesta por tres capas de 350, 400 y 300 neuronas respectivamente, cada una empleando 'return\_sequences' para preservar la secuencia de datos a través de las capas.
- **Capa de Dropout:** Se incorpora un dropout de 0.2 después de cada capa LSTM para prevenir el sobreajuste.
- **Capa de Salida:** Utiliza una capa TimeDistributed con 70 neuronas y activación 'relu', además de regularización L1 y L2 para optimizar el rendimiento y la generalización.

#### 3.1.1.5 Compilación y optimización

El modelo se compila utilizando el optimizador Adam con una tasa de aprendizaje de 0.001, y se emplea la función de pérdida 'mean\_squared\_error' adecuada para problemas de regresión como la predicción de series temporales.

### **3.1.1.6 Entrenamiento del modelo**

Para el entrenamiento, se establecen 500 epochs con un batch size de 32. Se utilizan dos técnicas de callback: EarlyStopping para prevenir el sobreentrenamiento deteniendo el entrenamiento cuando el 'val\_loss' deja de mejorar, y ReduceLRonPlateau para ajustar la tasa de aprendizaje si no se observan mejoras en el 'val\_loss'. Estas técnicas ayudan a mejorar la eficacia del entrenamiento y evitar el sobreajuste.

### **3.1.2 Sistema**

El sistema consistió en una aplicación web que permite al usuario generar predicciones para las siguientes 4 semanas a partir de la fecha seleccionada a partir de datos que se encuentran almacenados en una base de datos alojada en un servidor en la nube. Además, permite la visualización de datos históricos de ventas mediante gráficos que permitan una mejor comprensión al usuario y la carga de archivos para mantener la información actualizada.

## **3.2 Resultados**

### **3.2.1 Modelo de aprendizaje automático**

#### **3.2.1.1 Evaluación del modelo**

La evaluación del desempeño del modelo LSTM se llevó a cabo usando la función de pérdida Mean Squared Error (MSE) y el coeficiente de determinación  $R^2$  score en los conjuntos de entrenamiento, validación y prueba.

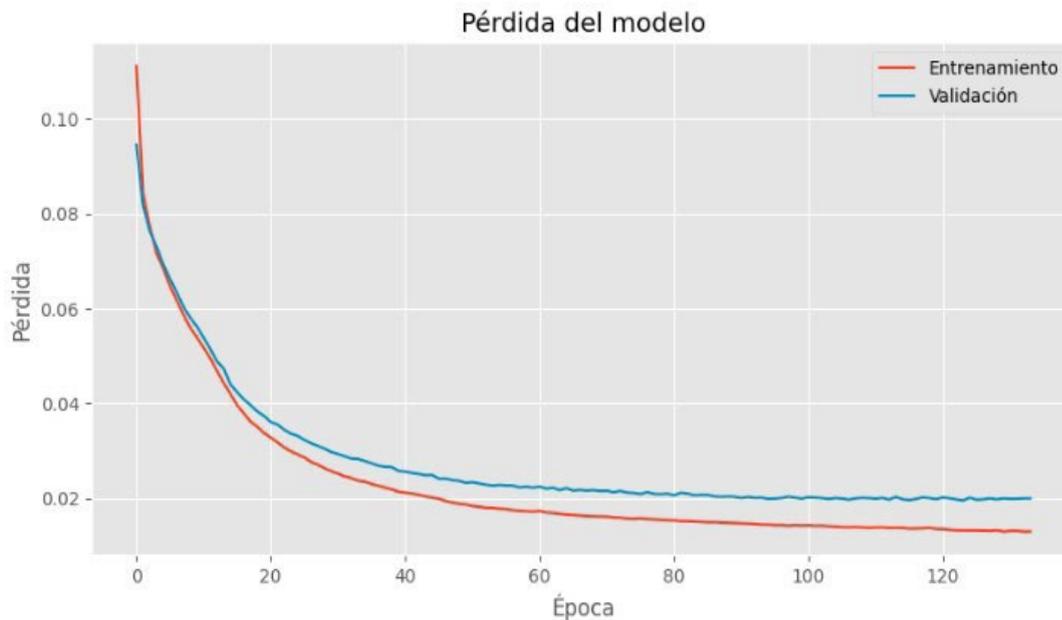
La Figura 15 ilustra cómo la pérdida disminuye rápidamente durante las primeras etapas del entrenamiento, lo que indica una fuerte convergencia inicial del modelo a la solución óptima. Posteriormente, la tasa de disminución se estabiliza, lo que sugiere que el modelo comienza a alcanzar un estado de mínima pérdida.

Por otro lado, la curva de pérdida de validación disminuye de forma similar en las primeras épocas y luego se asienta paralelamente a la curva de entrenamiento. Esto demuestra que el modelo generaliza bien a datos no vistos y que el sobreajuste ha sido controlado

efectivamente, como se pretendía con la implementación de técnicas como Dropout y regularización L1 y L2.

**Figura 15**

*Pérdida del modelo*



Al aplicar el modelo al conjunto de prueba, se obtuvo un MSE de 0.018, indicando un error promedio cuadrático bajo entre las predicciones y los valores reales, reflejando así un rendimiento satisfactorio.

Además, el modelo alcanzó un  $R^2$  score de 0.77 en el conjunto de prueba, lo que sugiere que aproximadamente el 77% de la varianza en los datos de prueba puede ser explicada por el modelo, lo cual es un indicativo de una fuerte capacidad predictiva.

### 3.2.1.2 Análisis de resultados

Los resultados obtenidos sugieren que el modelo LSTM tiene una habilidad notable para predecir series temporales, manteniendo un equilibrio entre aprendizaje y generalización. El MSE relativamente bajo y el  $R^2$  score alto en el conjunto de prueba respaldan su eficiencia. Estas métricas y las curvas de pérdida validan la calidad del modelo LSTM para tareas de predicción.

## 3.2.2 Sistema

### 3.2.2.1 Análisis de pruebas de la implementación

Durante la implementación y desarrollo del dashboard, se realizó una prueba de usabilidad con los analistas BI de ExpoCam para evaluar su funcionalidad y efectividad en condiciones reales. Cada analista interactuó con el tablero de manera autónoma, formulando preguntas pertinentes al proyecto.

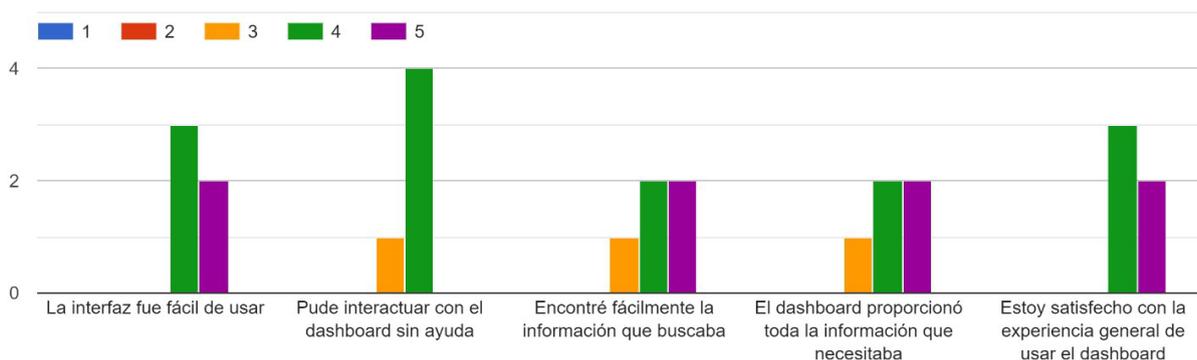
Al finalizar, completaron un cuestionario que incluía tanto la escala de Likert como preguntas abiertas para evaluar el sistema. Los hallazgos de esta prueba mostraron que:

- Todos los analistas BI estuvieron de acuerdo o totalmente de acuerdo en que la interfaz del dashboard ShriMP fue fácil de usar, destacando su accesibilidad.
- 80% pudo interactuar con el dashboard sin ayuda, indicando que la navegación e interacción con el sistema fueron intuitivas.
- 80% confirmó que encontraron fácilmente la información que buscaban en el dashboard., indicando que la disposición de los datos y las funcionalidades de búsqueda fueron efectivas.
- 80% coincidió que el dashboard proporcionó toda la información que necesitaban, cumpliendo con las expectativas en términos de contenido y datos disponibles.
- La satisfacción por parte de todos los analistas evaluados, indicando una recepción positiva general del sistema.

### Figura 16

*Preguntas en escala de Likert*

Indique su nivel de acuerdo con las siguientes afirmaciones, utilizando la siguiente escala:



En las respuestas a las preguntas abiertas, los usuarios sugirieron una mayor personalización de la interfaz y una disposición más flexible de los elementos visuales, además de sugerir mejoras en la legibilidad de ciertos gráficos. Se resaltó la necesidad de una función de búsqueda avanzada y opciones para exportar datos a otros formatos. Los analistas BI expresaron interés en la inclusión de paneles personalizables, notificaciones automatizadas y la creación de informes personalizados.

Además, se mencionó la importancia de proporcionar una mejor documentación y capacitación para maximizar el uso del dashboard ShriMP, así como mantener la plataforma actualizada con las últimas tendencias tecnológicas. Estos resultados son fundamentales para la toma de decisiones y mejora continua del dashboard ShriMP en función de las necesidades reales de los analistas de BI.

Las figuras 18, 19 y 20 ilustran las mejoras realizadas al sistema tras considerar la retroalimentación recibida en reuniones con el cliente. Estas incluyen una pantalla para visualizar los datos de ventas de ExpoCam mediante un dashboard, otra para la realización de predicciones de producción de camarón a 4 semanas vista, y una tercera para la carga de archivos y actualización de datos en la base de datos, respectivamente.

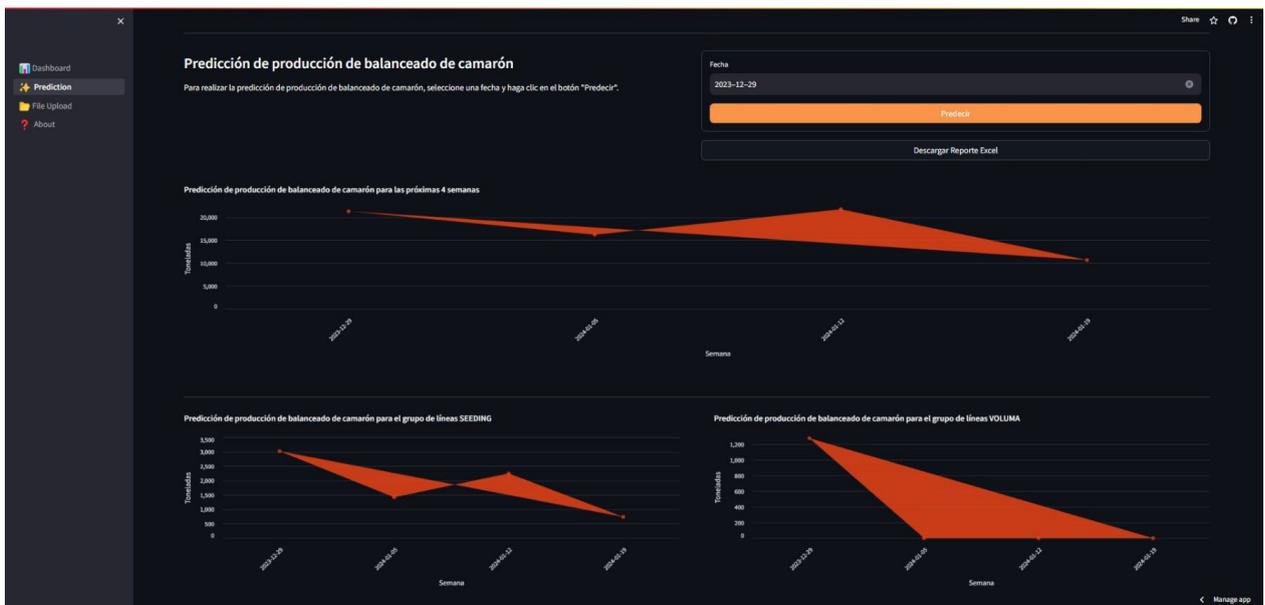
### Figura 17

*Dashboard de ventas del sistema*



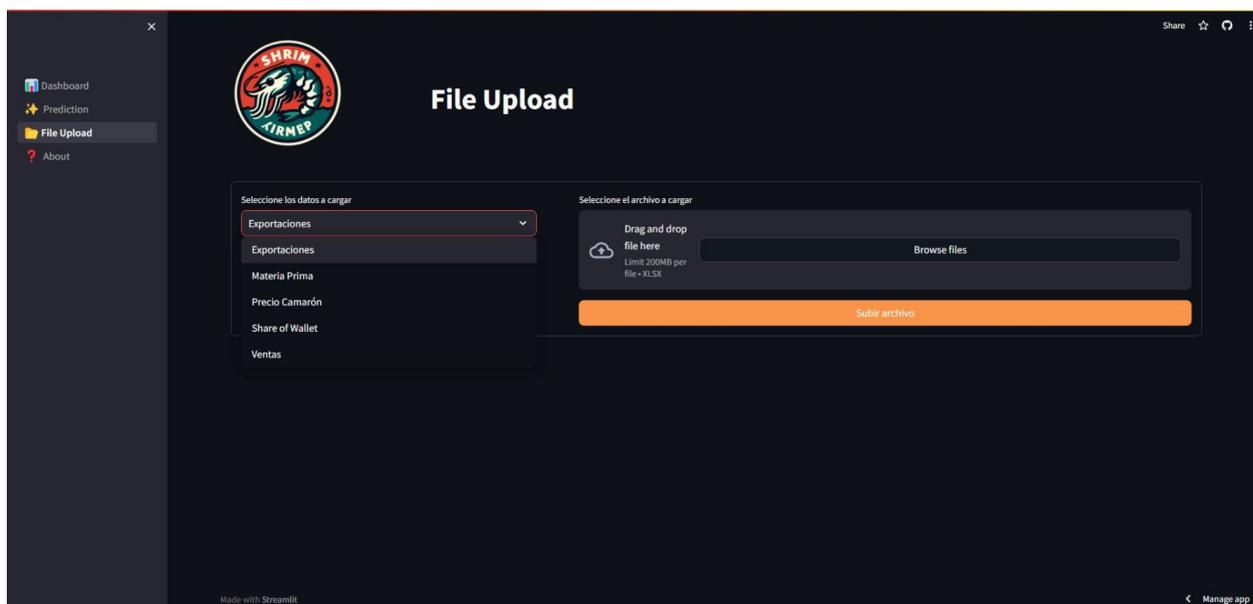
**Figura 18**

*Pantalla de generación de predicción*



**Figura 19**

*Pantalla para carga de archivos*



### 3.3 Análisis de costos

El proyecto aprovechó herramientas de código abierto como MySQL y Streamlit eliminando por completo los gastos asociados a terceros.

Los gastos principales provinieron del desarrollo del modelo de aprendizaje automático y el dashboard, realizado por dos personas durante 10 semanas, con jornadas de 8 horas diarias durante los días laborables y una tarifa de \$10 USD por hora para el trabajo de desarrollo, sumando un costo total del proyecto de \$8,000.00.

**Tabla 2**

*Análisis de costos del proyecto*

Actividad	Cantidad	Costo
Horas de desarrollo por persona	400 * 2	\$8,000.00
Recursos de código abierto	-	\$0.00
<b>Total</b>	-	<b>\$8,000.00</b>

## Capítulo 4

## 4. Conclusiones y recomendaciones

### 4.1 Conclusiones

El proyecto ejecutado se centró en abordar los retos que planteaba ExpoCam, en particular, el problema de la sobreproducción o escasez de alimentos balanceados que resultaba en una experiencia negativa para el cliente. Con la finalización exitosa del proyecto, se logró lo siguiente:

- Se desarrolló y puso en práctica un modelo de aprendizaje automático que ha demostrado ser eficiente en la predicción de la demanda de alimentos balanceados en el sector acuícola. Esto confirmó la hipótesis inicial de que es posible utilizar datos históricos, junto con indicadores del mercado como las exportaciones, los precios del camarón y las materias primas para el balanceado, para prever con precisión las necesidades futuras.
- A través del análisis exploratorio de los datos históricos, se identificaron patrones clave y se evaluó la influencia de los factores del mercado en la demanda de alimentos balanceados. Este análisis proporcionó insights valiosos que han enriquecido el entendimiento de las dinámicas de mercado y han reforzado la fundamentación del modelo predictivo.
- Se implementó un modelo de Aprendizaje Automático basado en redes LSTM, que aprovecha los datos históricos de ventas por cliente y producto, así como la información sobre variables del mercado que inciden en las ventas de los productos balanceados. Este modelo genera predicciones de ventas precisas para las siguientes cuatro semanas, permitiendo la toma de decisiones estratégicas en el negocio.
- Se desarrolló un dashboard para la visualización de datos, que facilita a los stakeholders la interpretación rápida de las tendencias. Esto les permite tomar decisiones informadas y ágiles basándose en las predicciones generadas por el modelo.

## 4.2 Recomendaciones

A continuación, se detallará algunas recomendaciones y consideraciones a tener en cuenta para trabajos futuros:

- Se recomienda ampliar la recolección de datos para incluir información de diversas fuentes como factores económicos globales, tendencias de consumo y datos ambientales. El enriquecimiento del conjunto de datos con estas variables puede mejorar significativamente la capacidad predictiva del modelo y proporcionar una comprensión más holística de los patrones de demanda.
- Se recomienda explorar el uso de técnicas emergentes en aprendizaje automático, como el aprendizaje profundo y los modelos de ensamble. La experimentación con estas metodologías avanzadas podría descubrir patrones más complejos y ofrecer mejoras en la exactitud y la confiabilidad de las predicciones.
- Se aconseja el desarrollo de sistemas que puedan actuar sobre las predicciones del modelo en tiempo real, permitiendo ajustes dinámicos en la producción y la logística. La capacidad de responder con agilidad a las predicciones del modelo puede minimizar el desperdicio, optimizar el uso de recursos y mejorar la sostenibilidad general del sector acuícola.
- Se sugiere invertir esfuerzos en mejorar la interpretación de los modelos de aprendizaje automático empleados. Adoptar técnicas de inteligencia artificial explicativa permitirá a los usuarios finales y a las partes interesadas comprender mejor cómo se generan las predicciones y cuáles son los factores más influyentes, lo que resulta esencial para ganar confianza y facilitar la toma de decisiones basada en estos modelos predictivos.

## REFERENCIAS

- [1] C. Zúñiga, «Estas son las diez mayores exportadoras de camarón de Ecuador, el producto que bate récord,» *El Universo | Noticias de Ecuador y del mundo*, 1 Enero 2023. [En línea]. Available: <https://www.eluniverso.com/noticias/economia/estas-son-las-diez-mayores-exportadoras-de-camaron-de-ecuador-el-producto-que-bate-record-nota/>. [Último acceso: 18 Octubre 2023].
- [2] A. J. Boada, «Sistema de proyección de la demanda. Caso práctico de predicción automatizada en empresas de venta por catálogo,» *Revista Perspectiva Empresarial*, vol. 4, nº 1, 2017.
- [3] M. Mazariegos, «El desperdicio de alimentos y su impacto en el medio ambiente - Sostenibilidad,» 17 Agosto 2016. [En línea]. Available: <https://blogs.iadb.org/sostenibilidad/es/food-loss-waste-negative-impact-environment/>. [Último acceso: 31 Octubre 2023].
- [4] Diario La Hora, «¿Es real el peligro de escasez de alimentos en el mundo y en Ecuador?,» 3 Junio 2022. [En línea]. Available: <https://www.lahora.com.ec/pais/peligro-escasez-alimentos-mundo-ecuador/>. [Último acceso: 31 Octubre 2023].
- [5] C. Ortegón, «Principales retos y desafíos que impiden el crecimiento de tu empresa - inMarketing,» 2 Mayo 2019. [En línea]. Available: <https://blog.inmarketing.co/blog/retos-y-desafios-que-impiden-el-crecimiento-de-tu-empresa>. [Último acceso: 31 Octubre 2023].
- [6] A. Garcete, R. Benítez, D. Pinto Roa y A. Vazquez, «Técnica de pronóstico de la demanda basada en Business Intelligence y Machine Learning,» *Simposio Argentino sobre Tecnología y Sociedad*, pp. 193-202, 2017.
- [7] Y. Pan, R. C. Zheng, J. Zhang y X. Yao, «Predicting bike sharing demand using recurrent neural networks,» *Procedia Computer Science*, vol. 147, pp. 562-566, 2019.
- [8] O. W. Sánchez Córdova y N. L. Cárdenas Escobar, «Construcción de un modelo estadístico para el pronóstico de la demanda de alimento balanceado para camarones en una empresa comercializadora de la ciudad de Guayaquil,» 2020.
- [9] Z. H. Kilimci, A. O. Akyuz, M. Uysal, S. Akyokus, M. O. Uysal, B. Atak Bulbul y M. A. Ekmis, «An Improved Demand Forecasting Model Using Deep Learning Approach and

Proposed Decision Integration Strategy for Supply Chain,» *Complexity*, vol. 2019, pp. 1-15, 2019.

- [10] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*, MIT Press, 2016.
- [11] C. Cortes y V. Vapnik, «Support-vector networks,» *Machine Learning*, vol. 20, nº 3, pp. 273-297, 1995.
- [12] L. Breiman, «Random Forests,» *Machine Learning*, vol. 45, nº 1, pp. 5-32, 2001.
- [13] R. Hyndman y G. Athanasopoulos , *Forecasting: principles and practice*, OTexts, 2018.
- [14] G. Box y G. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1976.
- [15] R. B. Cleveland, W. S. Cleveland y I. Terpenning, «STL: A Seasonal-Trend Decomposition Procedure Based on Loess,» *Journal of Official Statistics*, vol. 6, nº 1, pp. 3-73, 1990.
- [16] F. Chollet, *Deep Learning with Python*, Manning Publications, 2017.
- [17] W. Wang, «An overview of the recent advances in delay-time-based maintenance modelling,» *Reliability Engineering & System Safety*, vol. 106, pp. 165-178, 2012.
- [18] A. A. Syntetos y J. E. Boylan, «On the stock control performance of intermittent demand estimators,» *International Journal of Production Economics*, vol. 103, nº 1, pp. 36-47, 2006.
- [19] M. Zaharia, R. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave y X. Meng, «Apache Spark,» *Communications of the ACM*, vol. 59, nº 11, pp. 56-65, 2016.
- [20] A. Botta, W. de Donato, V. Persico y A. Pescapé, «Integration of Cloud computing and Internet of Things: A survey,» *Future Generation Computer Systems*, vol. 56, pp. 684-700, 2016.
- [21] P. Krutchen, «The 4+1 View Model of architecture,» *IEEE Software*, vol. 12, nº 6, pp. 42-50, 1995.

## APÉNDICES

### Apéndice A

#### *Historias de Usuario*

##### Analista BI

<b>ABI-001</b>	Como Analista BI, necesito <b>acceder a una plataforma</b> que visualice las predicciones de demanda de alimentos balanceados.
Productos	Modelo Machine Learning, Dashboard
Dependencias	
Detalles	El Dashboard presentará: <ul style="list-style-type: none"><li>- Visualización de predicciones</li><li>- Interactividad con las predicciones</li><li>- Alertas sobre tendencias emergentes</li></ul>

##### Gerente de producción

<b>PRD-001</b>	Como Gerente de Producción, necesito <b>acceder a la cantidad aproximada de alimento balanceado a producir</b> el mes entrante con la finalidad de evitar sobreproducción o escasez.
Productos	Dashboard
Dependencias	ABI-001
Detalles	Se necesita observar: <ul style="list-style-type: none"><li>- Estimaciones de producción basadas en predicciones</li><li>- Alertas de variaciones inesperadas en la demanda</li><li>- Historial de producciones pasadas</li></ul>

##### Responsable de Marketing Intelligence

<b>MKI-001</b>	Como Responsable de Marketing Intelligence, necesito <b>conocer la demanda del mercado</b> con la finalidad de realizar una planificación óptima de estrategias de marketing
----------------	--

Productos	Dashboard
Dependencias	ABI-001
Detalles	Se necesita observar: <ul style="list-style-type: none"><li>- Personalización de oferta</li><li>- Visualización de tendencias y pronósticos</li><li>- Alertas sobre cambios repentinos en predicciones</li></ul>

## **Apéndice B**

### *Documentación*

La documentación de ShriMP se encuentra en el siguiente enlace:

<https://drive.google.com/file/d/1PRbKkcsEk-ExZ1U-bTmPW6nTs0ni06-6/view?usp=sharing>

## Apéndice C

### Manual de usuario

Esta guía te ayudará a comprender el uso del dashboard ShriMP

#### 1. Introducción

##### ¿Qué es ShriMP?

ShriMP es un sistema que permite visualizar información de ventas de productos de ExpoCam. Esta información puede ser filtrada por cliente, producto y etapa de producción. ShriMP también proporciona una herramienta para predecir la producción de balanceado de camarón de Expocam.

##### ¿Cómo funciona?

ShriMP se conecta a una base de datos de Expocam para obtener la información de ventas. Esta información es procesada y visualizada en gráficos y tablas.

La predicción de producción de balanceado de camarón se realiza utilizando un modelo LSTM entrenado con datos históricos de producción de balanceado de camarón.

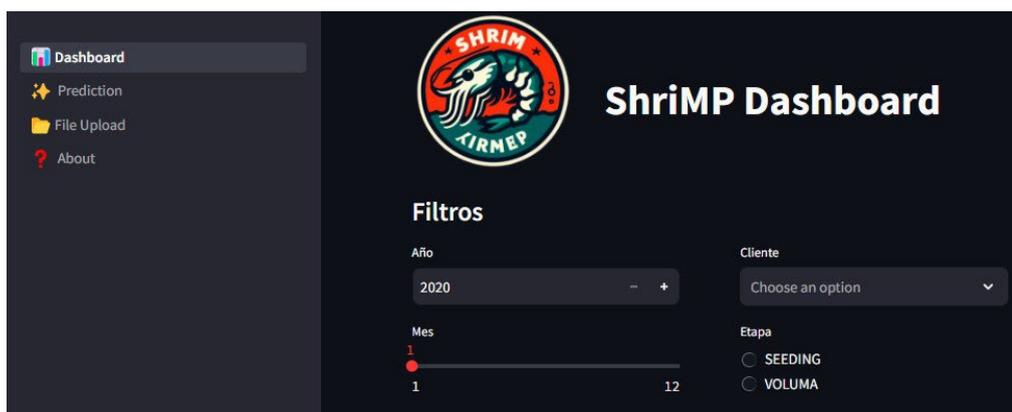
#### 2. Interacción del analista BI

##### Acceso al dashboard

Para comenzar, debes acceder al siguiente enlace: <https://dashboard-shrimp-ml.streamlit.app/>

##### Dashboard

En esta sección podrás visualizar la data histórica de las ventas de alimento balanceado. Además, se puede filtrar según año, mes, cliente y etapa.



Debajo de los filtros, se puede observar el resumen de las ventas.

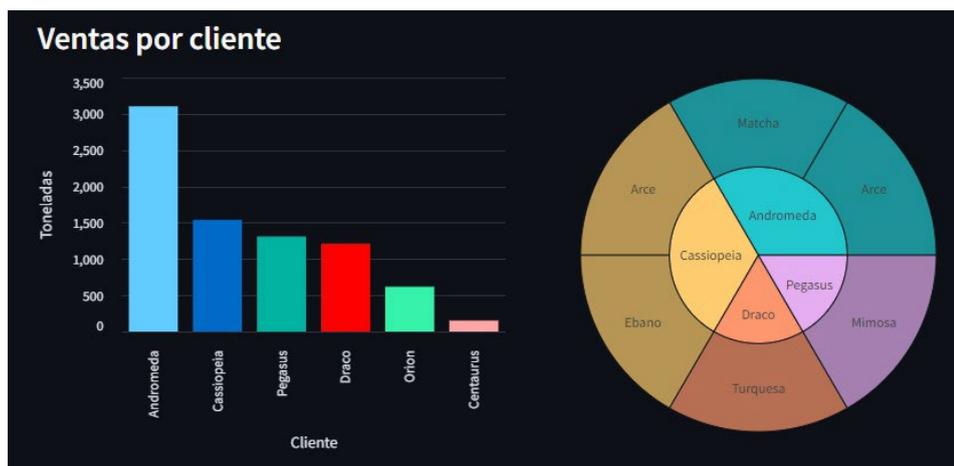
Resumen de ventas			
Total de ventas	Ventas promedio por cliente	Producto más vendido	Cliente que más compra
7930.4 Ton	1321.73 Ton	Arce	Andromeda

Debajo, aparecerá una sección de dos pestañas:

## Tendencia



Detalle de ventas: aparecen visualizaciones de ventas por producto y cliente, y detalle de ventas



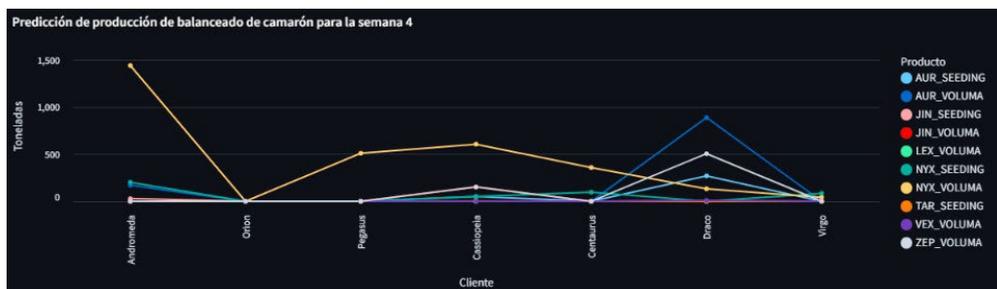
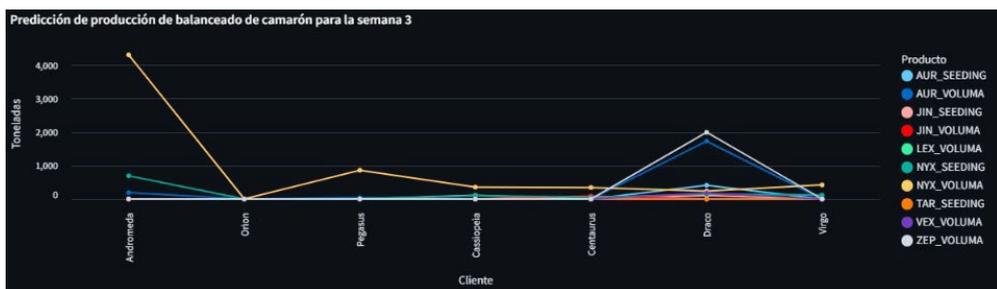
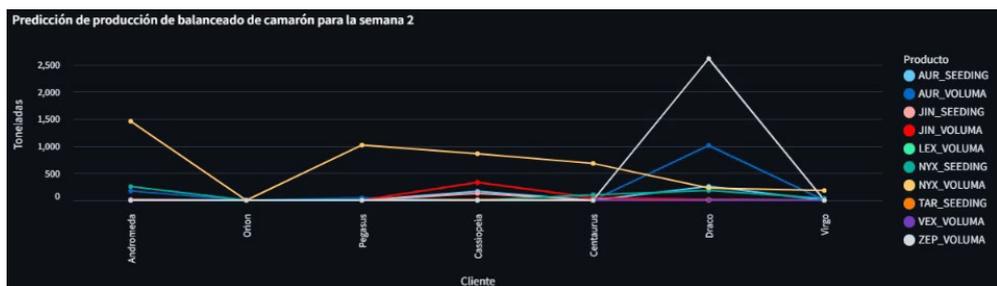
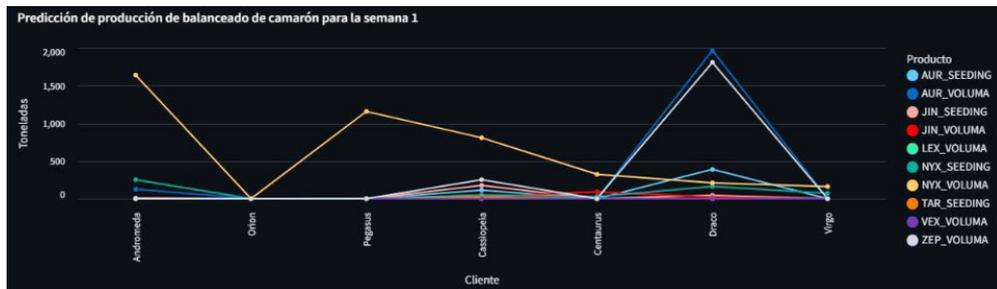
Detalle de ventas					
Fecha	Cliente	Producto	Familia	Etapas	Toneladas
2021-03-01	Andromeda	Coral	AUR	VOLUMA	59.54
2021-03-01	Andromeda	Ebano	AUR	VOLUMA	25.25
2021-03-02	Andromeda	Coral	AUR	VOLUMA	52.49
2021-03-02	Andromeda	Mimosa	AUR	VOLUMA	1.32
2021-03-02	Andromeda	Ebano	AUR	VOLUMA	18.3
2021-03-02	Andromeda	Coral	AUR	VOLUMA	48.11
2021-03-02	Andromeda	Tulipan	NYX	VOLUMA	42.5
2021-03-02	Andromeda	Nina	VEX	VOLUMA	6.38
2021-03-02	Andromeda	Ebano	AUR	VOLUMA	1.62
2021-03-02	Andromeda	Matcha	NYX	VOLUMA	12.62

## Prediction

ShriMP Prediction se conecta a una base de datos de Expocam para obtener la información de ventas. Esta información es procesada y visualizada en gráficos y tablas. La predicción de producción de balanceado de camarón se realiza utilizando un modelo LSTM entrenado con datos históricos de producción de balanceado de camarón. La predicción mostrará la producción de balanceado de camarón para las siguientes 4 semanas a partir de la fecha seleccionada.

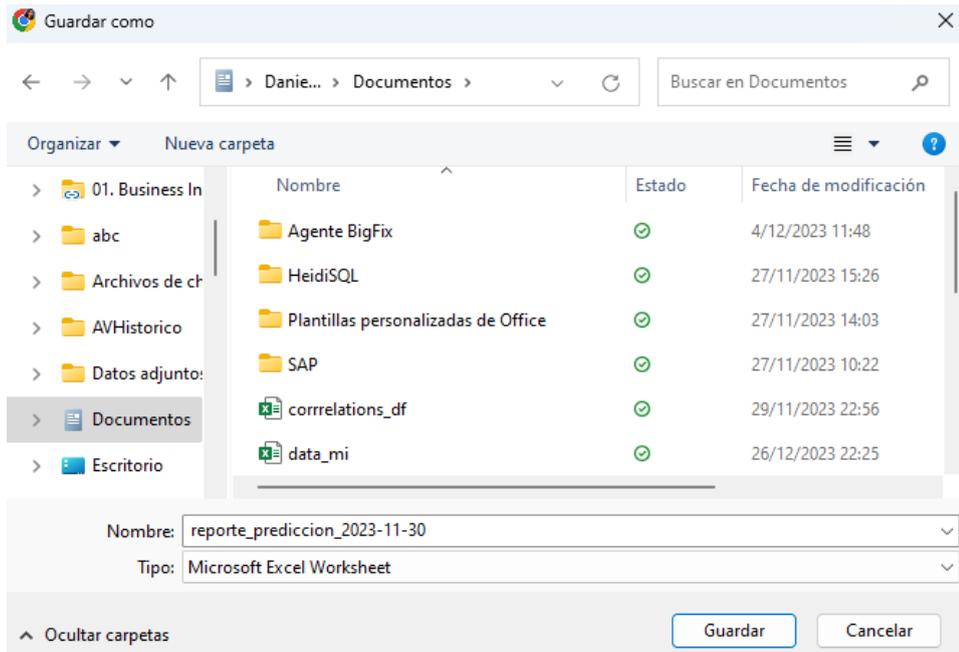
Para realizar la predicción de producción de balanceado de camarón, debe seleccionar una fecha y hacer clic en el botón "Predecir". Tener en cuenta que en el calendario aparecerá hasta la última fecha que tiene la base de datos.

Luego, aparecerán visualizaciones de las predicciones por cliente y familia-etapa para cada semana.



Además, se podrá descargar un archivo de excel con la información detallada.

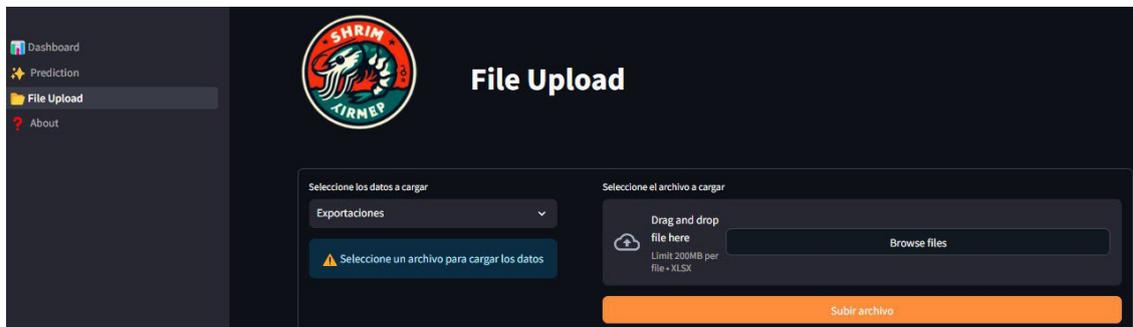




	CLIENTE	AUR_SEEDING	AUR_VOLUMA	JIN_SEEDING	JIN_VOLUMA	LEX_VOLUMA	NYX_SEEDING	NYX_VOLUMA	TAR_SEEDING	VEX_VOLUMA	ZEP_VOLUMA
0	Andromeda	0.130170941	127.6096039	8.523752213	0	0.00489718	252.5313416	1639.784424	0.022351367	0.027281884	0
1	Orion	0.13129279	0.068665668	0.016495151	0	0.003915038	0	0.080041058	0.004346136	0	0.026486274
2	Pegasus	0.079785123	0.127663732	0.020767609	0.009561084	0.029242806	3.88391614	1157.745605	0.047516659	0	0.025329784
3	Cassiopeia	114.3599548	0.149007365	176.815033	25.31446457	0	50.5	808.0866699	22.18638992	0	252.5134125
4	Centaurus	0.009875625	0.044967379	0.065727428	89.79381561	0	28.05999947	322.7352295	0.052896593	0	0
5	Draco	388.6932373	1966.972534	44.74721909	19.19347382	0	163.7607117	209.9332428	0.017414663	0.061728787	1808.929565
6	Virgo	0.04977975	0.030205689	0	0.009285253	0	72.26000214	161.5200043	0.033713147	0	0.005416255

## File Upload

En esta pestaña se cargará la data a la base de datos.



Se debe seleccionar qué archivo se desea subir.

Seleccione los datos a cargar

Exportaciones



Exportaciones

Materia Prima

Precio Camarón

Share of Wallet

Ventas