UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA
INGENIERÍA EN TELEMÁTICA

Trabajo de integración curricular previa la obtención del Grado Académico de Ingeniero en Telemática.

Título del Proyecto de Investigación:

"MODELO PREDICTIVO PARA ESTIMAR EL PESO EN PECES NATIVOS BASADO EN PARÁMETROS FISIOLÓGICOS"

Autor:

STIVEN JAVIER LOOR PONCE

Director del Proyecto de Investigación:

ING. EMILIO RODRIGO ZHUMA MERA, M.Sc

QUEVEDO – LOS RÍOS – ECUADOR

2023

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **Loor Ponce Stiven Javier,** declaro que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Loor Ponce Stiven Javier

C.I: 120848069-7

CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN

El suscrito, **Ing. Emilio Rodrigo Zhuma Mera. M.Sc**. Docente de la Universidad Técnica Estatal de Quevedo, certifica que el estudiante, **Loor Ponce Stiven Javier**, realizó el Proyecto de Investigación de grado titulado "**MODELO PREDICTIVO PARA ESTIMAR EL PESO EN PECES NATIVOS BASADO EN PARÁMETROS FISIOLÓGICOS**", previo a la obtención del título de Ingeniero en Telemática, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

EMILIO RODRIGO ZHUMA MERA
Firmado digitalmente por EMILIO RODRIGO ZHUMA MERA

Ing. Emilio Rodrigo Zhuma Mera. M.Sc
DIRECTOR DEL PROYECTO DE INVESTIGACIÓN

CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO

El suscrito, **Ing. Emilio Rodrigo Zhuma Mera. M.Sc**, mediante el presente cumplo en presentar a usted, el informe de proyecto de Investigación titulado "**MODELO PREDICTIVO PARA ESTIMAR EL PESO EN PECES NATIVOS BASADO EN PARÁMETROS FISIOLÓGICOS**" Presentado por el estudiante **Stiven Javier Loor Ponce**, egresado de la Carrera de Ingeniería en Telemática, que fue revisado bajo mi dirección según resolución del Consejo Directivo de la Facultad de Ciencias de la Ingeniería, que se ha desarrollado de acuerdo al Reglamento de la Unidad de Integración Curricular de la Universidad Técnica Estatal de Quevedo y cumple con el requerimiento de análisis de URKUND el cual avala los niveles de originalidad en un 99% y similitud 1%, del trabajo investigativo. Valido este documento para que el estudiante siga con los trámites pertinentes, de acuerdo como lo estable el Reglamento.

Ouriginal by Turnitin

Document Information

Analyzed document	Tesis de Loor.pdf (D177748800)
Submitted	11/3/2023 7:15:00 PM
Submitted by	
Submitter email	stiven.loor2018@uteq.edu.ec
Similarity	1%
Analysis address	ezhuma.uteq@analysis.urkund.com

EMILIO RODRIGO ZHUMA MERA Firmado digitalmente por EMILIO RODRIGO ZHUMA MERA

Ing. Emilio Rodrigo Zhuma Mera. M.Sc

DIRECTOR DEL PROYECTO DE INVESTIGACIÓN

UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO

FACULTAD DE CIENCIAS DE LA INGENIERÍA

CARRERA DE INGENIERIA EN TELEMÁTICA

PROYECTO DE INVESTIGACION

Título:

"MODELO PREDICTIVO PARA ESTIMAR EL PESO EN PECES NATIVOS BASADO EN PARÁMETROS FISIOLÓGICOS"

Presentado al Consejo Directivo de Facultad como requisito previo a la obtención del título de Ingeniero en Telemática.

Aprobado por:

Eduardo Samaniego Mena
Firmado digitalmente por Eduardo Samaniego Mena
Fecha: 2023.11.08 12:08:59 -05'00'

PRESIDENTE DEL TRIBUNAL

Ing. Eduardo Samaniego. M.Sc.

Firmado electrónicamente por: ALEX ROSENDO FIALLOS BARRIONUEVO

MIEMBRO DEL TRIBUNAL

Ing. Alex Fiallos Barrionuevo. M.Sc

Firmado electrónicamente por: JORGE WILSON SAA SALTOS

MIEMBRO DEL TRIBUNAL

Ing. Jorge Saa Saltos. M.Sc

QUEVEDO – LOS RIOS – ECUADOR

2023

AGRADECIMIENTO

En primer lugar, quiero agradecer a la Universidad Técnica Estatal de Quevedo por brindarme el entorno académico en el cual he crecido y aprendido. La excelencia de esta institución y la dedicación de los docentes han sido la guía que me ha llevado hasta aquí. Su conocimiento y orientación han sido la brújula que me ha ayudado a navegar por las complejidades de mi investigación.

A mis padres y familiares, no hay palabras suficientes para expresar mi gratitud. Su amor incondicional, paciencia y apoyo constante han sido mi roca durante este viaje. Cada logro que celebro hoy es también de ustedes; su confianza en mí me ha dado fuerzas para preservar incluso en los momentos más difíciles.

Agradezco a Dios por su gracia y guía divina a lo largo de este proceso. En cada desafío y en cada triunfo, he sentido su presencia inspiradora que me ha dado esperanza y fortaleza para seguir adelante.

A todos mis amigos tanto de la UTEQ como de la vida cotidiana que han estado a mi lado, gracias por su aliento, por creer en mí y por compartir esta alegría conmigo. Este logro no sería completo sin su amistad y apoyo.

Stiven Javier Loor Ponce

DEDICATORIA

En este momento de gratitud y alegría, quiero expresar mi sincero agradecimiento. A mis padres, por su amor incondicional y apoyo constante. Cada logro es un reflejo de su dedicación y confianza en mí. A aquellos docentes, por su orientación sabia y paciencia infinita. Sus conocimientos han iluminado mi camino y enriquecido mi aprendizaje.

Stiven Javier Loor Ponce

RESUMEN

En esta investigación pionera en piscicultura, se ha desarrollado un modelo predictivo revolucionario basado en parámetros fisiológicos para estimar el peso de los peces, reduciendo así la manipulación directa y el estrés en las especies acuáticas. Mediante un análisis exhaustivo, se identificaron las variables optimas clave, siendo la Longitud Total y Longitud de la Cabeza fundamentales. Se realizo una comparación minuciosa de varios modelos de predicción, destacando la superioridad de este modelo en términos de precisión y consistencias. Las pruebas y validaciones detalladas con datos reales demostraron la eficiencia del modelo, confirmando su capacidad para proporcionar estimaciones de peso precisas y confiables. Este enfoque innovador promete transformar las prácticas de manejo en la industria acuícola, mejorando la sostenibilidad y el bienestar de los peces, u allanando el camino hacia un fututo más seguro y eficientes en la producción piscícola.

Palabras claves: Aprendizaje automático, Scikit-learn, Pandas, Streamlit.

ABSTRACT

In this pioneering research in fish farming, a revolutionary predictive model based on physiological parameters has been developed to estimate fish weight, thus reducing direct manipulation and stress on aquatic species. Through an exhaustive analysis, the key optimal variables were identified, with Total Length and Head Length being fundamental. A thorough comparison of several prediction models was carried out, highlighting the superiority of this model in terms of precision and consistency. Detailed testing and validation with real data demonstrated the efficiency of the model, confirming its ability to provide accurate and reliable weight estimates. This innovative approach promises to transform management practices in the aquaculture industry, improving the sustainability and well-being of fish, and paving the way to a safer and more efficient future in fish production.

Keywords: Machine Learning, Scikit-learn, Pandas, Streamlit

Tabla de Contenido

Índice de Figuras

Índice de Tablas

CÓDIGO DUBLÍN

Titulo:	Modelo predictivo para estimar el peso en peces nativos basado en parámetros fisiológicos				
Autores:	Loor Ponce Stiven Javier				
Palabras claves:	Aprendizaje automático	Scikit-learn	Pandas	Streamlit	Aumento de datos
Fecha de publicación:	- de — de 2023				
Editorial:	Universidad Técnica Estatal de Quevedo “La María” 2023				
Resumen:	En esta investigación pionera en piscicultura, se ha desarrollado un modelo predictivo revolucionario basado en parámetros fisiológicos para estimar el peso de los peces, reduciendo así la manipulación directa y el estrés en las especies acuáticas. Mediante un análisis exhaustivo, se identificaron las variables optimas clave, siendo la Longitud Total y Longitud de la Cabeza fundamentales. Se realizo una comparación minuciosa de varios modelos de predicción, destacando la superioridad de este modelo en términos de precisión y consistencias. Las pruebas y validaciones detalladas con datos reales demostraron la eficiencia del modelo, confirmando su capacidad para proporcionar estimaciones de peso precisas y confiables. Este enfoque innovador promete transformar las prácticas de manejo en la industria acuícola, mejorando la sostenibilidad y el bienestar de los peces, u allanando el camino hacia un fututo más seguro y eficientes en la producción piscícola.				
Abstract:	In this pioneering research in fish farming, a revolutionary predictive model based on physiological parameters has been developed to estimate fish weight, thus reducing direct manipulation and stress on aquatic species. Through an exhaustive analysis, the key optimal variables were identified, with Total Length and Head Length being fundamental. A thorough comparison of several prediction models was carried out, highlighting the superiority of this model in terms of precision and consistency. Detailed testing and validation with real data demonstrated the efficiency of the model, confirming its ability to provide accurate and reliable weight estimates. This innovative approach promises to transform management practices in the aquaculture industry, improving the sustainability and well-being of fish, and paving the way to a safer and more efficient future in fish production.				
Descripción	69 hojas: dimensiones 29 x 21 cm + CD-ROM 6162				
URL:					

INTRODUCCIÓN

En el dinámico mundo de la industria acuícola actual, el desafío fundamental radica en optimizar el crecimiento y desarrollo de los peces nativos. Esta tarea se ha vuelto esencial para garantizar la sostenibilidad y rentabilidad del sector. A lo largo del tiempo, la manipulación de los peces para estimar su peso ha sido una práctica común, sin embargo, esta técnica no solo induce estrés en los animales, sino que también puede afectar negativamente su crecimiento a largo plazo. Por tanto, se ha generado un interés significativo en la búsqueda de métodos precisos y no invasivos para llevar a cabo esta estimación.

Este estudio se ubica en la intersección de la biología acuática y la tecnología predictiva, capitalizando los avances en el campo de la ciencia de datos y el aprendizaje automático. Al identificar y utilizar las variables óptimas que influyen en el crecimiento de los peces nativos, nuestro modelo propuesto tiene como objetivo proporcionar una herramienta precisa y confiable para estimar el peso de estos animales sin recurrir a intervenciones invasivas.

Además de presentar esta innovadora propuesta, la investigación se sumerge en un análisis comparativo minucioso entre diversos modelos de predicción disponibles. Este enfoque meticuloso nos permite evaluar y validar la efectividad del modelo en comparación con otros métodos existentes, asegurando así su precisión y utilidad en el contexto práctico de la piscicultura.

Este trabajo no solo marca un avance significativo en el campo de la acuicultura, sino que también contribuirá a promover prácticas más éticas y sostenibles en la industria piscícola. La implementación de un modelo predictivo preciso y no invasivo no solo beneficia a los peces al reducir su estrés y mejorar su calidad de vida, sino que también tiene un impacto positivo en la eficiencia operativa y la rentabilidad de las granjas acuícolas.

Las siguientes secciones explora en detalle el proceso de desarrollo del modelo, los métodos utilizados, los resultados obtenidos y las implicaciones prácticas de esta investigación innovadora.

CAPÍTULO I

CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

1.1. Problema de investigación

1.1.1. Planteamiento del problema

En relación con los estudios que han demostrado que el manejo y la manipulación inadecuada de los peces causa estrés, lo cual provoca cambios en los parámetros fisiológicos,como en el desarrollo del pez, el comportamiento de natación y la reducción de la función inmunológica. No obstante, la precisión en la estimación del peso de estos peces puede representar un desafío, debido a que radica en la falta de un modelo predictivo confiable que permita calcular el peso de los peces nativos de manera precisa. En la actualidad la estimación del peso se basa en métodos tradicionales que requieren la manipulación directa de los peces, lo cual les genera estrés, afectando al crecimiento de este y conlleva a tener un impacto negativo en las poblaciones.

Aunque los peces suelen coexistir con una amplia gama de posibles agentes infecciosos en todo el mundo, existen diversas circunstancias que perturban este equilibrio. En nuestro entorno, se encuentran diversos parásitos y bacterias, así como cambios en la nutrición y calidad del agua, que pueden desencadenar brotes de mortalidad en cada fase del cultivo depeces. Por consiguiente, estos factores predisponentes generan una situación de estrés en los peces, aumentando su susceptibilidad a los patógenos potenciales. Además, se consideran factores predisponentes como la condición ambiental y el manejo del individuo.

Diagnostico

Los peces nativos desempeñan un papel fundamental en los ecosistemas acuáticos del país,ya que estas especies son importantes tanto para la pesca comercial como para la conservación de la biodiversidad, siendo fuente de alimentación y sustento para las comunidades locales.

Entre las causas y consecuencias que se generan por la manipulación de los peces se mencionan las siguientes:

Causas:

- Limitaciones de recopilación de datos fisiológicos no invasivos.
- La necesidad de contar con métodos precisos y eficientes para estimar el peso.
- Escaso conocimiento y aplicación de herramientas de modelo predictivo.

Consecuencias:

- Disminución en la precisión y exhaustividad de los datos recopilados, lo que afecta la calidad de las predicciones y análisis basados en estos datos.
- Al desarrollar un modelo predictivo basado en parámetros fisiológicos, se podrían realizar estimaciones de peso en peces de manera no invasiva y precisa.
- Limitación en la capacidad para desarrollar modelos precisos y avanzados.

Pronostico

Si no se aborda de manera efectiva el problema de la estimación del peso en peces utilizandoparámetros fisiológicos, se podrían experimentar consecuencias negativas a nivel global, regional y local, como la sobreexplotación de recursos, la disminución de las poblaciones depeces y la falta de sostenibilidad en la acuicultura.

Es por lo mencionado que, con el modelo para predecir el peso de los peces basados en parámetros fisiológicos, se espera que con la longitud de los peces saber múltiples aspectos, como ver si el peso es adecuado con respecto a su tamaño, menorar la manipulación del pez para que no afecte en su crecimiento y la limitación en la disponibilidad de datos.

1.1.2. Formulación del problema

¿Cómo se podría estimar el peso de los peces basado en parámetros fisiológicos, disminuyendo la manipulación de estos?

1.1.3. Sistematización del problema

¿Cómo identificar las características fisiológicas en los peces que permita estimar su peso

¿Qué métodos existen para medir los parámetros fisiológicos de manera no invasiva y precisa?

¿Qué tan confiables serán los resultados que se obtendrá a través de técnicas no invasivas?

1.2.Objetivos

1.2.1. Objetivo General

Desarrollar un modelo predictivo para la estimación del peso de los peces basado en parámetros fisiológicos, con el propósito de evitar la manipulación y su impacto en el crecimiento y desarrollo.

1.2.2. Objetivos Específicos

- Identificar las variables optimas que serán utilizadas en el modelo predictivo.
- Realizar un análisis comparativo entre distintos modelos de predicción para determinar cuál de ellos proporciona los mejores resultados.
- Realizar pruebas y validación del modelo de predicción utilizando datos de los parámetros para verificar su efectividad y precisión en la estimación del peso.

1.3.Justificación

En la actualidad, la estimación del peso de los peces nativos se fundamenta principalmente en técnicas invasivas que requieren la manipulación directa de los individuos. Estos métodos, además de ser estresantes para los peces, pueden generar distorsiones en los resultados y tener un impacto negativo en su crecimiento y desarrollo, además los peces nativos juegan un papel de vital importancia en los ecosistemas acuáticos, ya que desempeñan un papel clave en la cadena trófica y actúan como indicadores del estado de salud de los ríos y cuerpos de agua.

La manipulación inadecuada de los peces durante la estimulación de peso puede tener graves consecuencias para su crecimiento y desarrollo. En este contexto, se plantea la necesidad de desarrollar un modelo predictivo que permita estimar el peso de los peces utilizando parámetros fisiológicos, evitando así la necesidad de técnicas invasivas que puedan afectar negativamente a los individuos y comprometer su bienestar, esta proyecto busca minimizar la manipulación y proporcionar una herramienta precisa y confiable para la evolución del peso en peces, además de que el modelo pueda tener el potencial de contribuir significativamente a la investigación científica, la gestión de poblaciones de peces y la conservación de los ecosistemas acuáticos al proporcionar una alternativa segura y efectiva para el estudio de la fisiología y el crecimiento de los peces.

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN

.

2.1. Marco Referencial

“Propuesta de un modelo predictivo numérico que permita mejorar la rentabilidad de la tilapia roja”

En el proyecto de María Orozco, se desarrolló un modelo numérico que relaciona la temperatura, el apetito y la concentración de oxígeno con el crecimiento de la tilapia roja. El objetivo es ayudar a los piscicultores a predecir la rentabilidad de la producción y maximizar las utilidades al encontrar valores óptimos para estas variables. El modelo arroja utilidades de $7,875,400 pesos colombianos y un peso final de 500 g por pez en un periodo de 6 meses [1].

“Técnicas de predicción de edad y crecimiento a partir de la morfometría y el peso de otolito de prochilodus leneatus (valenciennes 1837)”

Valeria Kratochvil investigo el sábalo Prochilodus leneatus, una especie comercial en Sudamérica, desarrollo técnicas predictivas para estimar la edad y el crecimiento de los peces utilizando el otolito lapillus. Se recolectaron muestras en el Alto Paraná y se encontró una correlación positiva entre la longitud y el peso del otolito. El largo del otolito resulto ser un mejor predictor de la longitud estándar y el peso del pez que el peso del otolito [2].

“Relación longitud-peso y factor de condición de los peces nativos del rio San Pedro (cuenca del rio Valdivia, Chile)”

Roberto Cifuentes y su equipo investigaron la relación longitud-peso y el factor de condición (K) de 12 especies de peces nativos en el rio San Pedro, Chile, estos parámetros son cruciales para comprender el crecimiento y estado nutricional de las poblaciones de peces. Encontraron que algunas especies mostraban crecimiento isométrico, mientras que otras tenían una tendencia a la alometría. El factor de condición variaba anualmente, posiblemente relacionado con las estaciones de reproducción y disponibilidad del alimento. Este estudio proporciona información clave sobre la ecología de poblaciones de peces en ecosistemas naturales y su relevancia para futuros proyectos ambientales [3].

“Distribución y abundancia de peces de la familia Loricariidae (Pleco) y su relación con los peces de interés comercial en los alrededores de la Isla de Ometepe”

En el estudio de Jorge T. Corea y su equipo sobre la Isla de Ometepe (febrero-agosto 2012), se examinó la presencia de plecos (Loricariidae) y su relación con peces comerciales. Encontraron 116 plecos en la zona A, mayormente en el noreste, alimentándose de algas y

detritus. En la zona B, capturaron 1,438 peces de 15 especies; Amphilophus citrinellus fue la más común. Los plecos constituyeron el 1.1% de las capturas (16 peces). Amphilophus mostró valores de factor de condición más altos en el noreste. La diversidad varió, siendo San José del Sur la más diversa (H=1.545). En la zona B, capturaron 532 plecos, siendo 498 del noreste y 34 del suroeste. Estos hallazgos delinean la distribución y comportamiento de los plecos en la región [4].

"Aplicación y Análisis de un Sistema Experto Basado en Lógica Difusa para la Evaluación del Hábitat de Peces Nativos en el Río Huequecura"

En su estudio, Meza López investiga los ecosistemas fluviales chilenos, carentes de información detallada sobre la fauna íctica y afectados por la presencia de especies exóticas y factores antropogénicos. Utilizando un modelo de simulación de calidad de hábitat en el río Huequecura, emplea herramientas como GR4J y HEC-RAS, junto con el sistema experto CASiMiR. Los resultados revelan idoneidad del hábitat para peces juveniles, señalando limitaciones debido a la falta de datos precisos y políticas ambientales ambiguas. Se proponen mejoras en las políticas ambientales y métodos de captura, además de una metodología alternativa para estimar el caudal ecológico basada en índices de calidad de hábitat. Estas recomendaciones subrayan la urgencia de una gestión ambiental más precisa y una recopilación de datos mejorada en Chile [5].

"Modelación de la riqueza de peces nativos para simular medidas de mitigación en tramos de ríos alterados"

En el estudio realizado por Esther Julia Olaya Marín, Francisco Martínez-Capel, Rui Soares Costa y Juan Diego Alcaraz-Hernández, se empleó una red neuronal artificial (RNA) para predecir la riqueza de peces nativos en los ríos Júcar, Cabriel y Turia en el oeste de España. Utilizando el método de validación cruzada k-fold y el algoritmo de aprendizaje Levenberg-Marquardt, demostraron que la RNA tiene la capacidad de predecir de manera efectiva la riqueza de peces en esta área específica. Además, mediante la simulación de medidas de mitigación, descubrieron que un aumento en la riqueza de peces está directamente relacionado con el incremento de la distancia libre de obstáculos artificiales y el porcentaje de hábitat de corriente en los ríos. Este estudio resalta el potencial de las RNA como herramienta poderosa para respaldar la toma de decisiones en la gestión y restauración de los ecosistemas fluviales, mostrando su relevancia en el campo de la ecología y la conservación de la biodiversidad acuática [6].

2.2. Marco Teórico

Se establecen los fundamentos teóricos que proporcionaran el marco conceptual necesario para desarrollar la idea principal del problema de investigación.

2.2.1. Algoritmos de predicción

Según Juan Carlos Durante [7], existen varios algoritmos de predicción que podrían ser útiles para desarrollar un modelo predictivo, Un algoritmo predictivo se refiere a una expresión numérica que describe un aspecto o evento del mundo real, facilitando la predicción de cómo se comportaran diferentes variables em el futuro. Este modelo matemático se aplica en diversos ámbitos como maquinarias, procesos de producción y otros contextos para anticipar los resultados y tomar decisiones informadas. Utilizando el conocimiento actual, el algoritmo busca establecer patrones y relaciones para hacer proyecciones precisas y anticipar posibles resultados en el futuro.

Además, los algoritmos predictivos son de gran importancia en los procesos de producción de cualquier industria debido a su capacidad para optimizar recursos, mejorar la eficiencia, respaldar la toma de decisiones, reducir costos y mejorar la calidad. Son herramientas esenciales que permiten planificar de manera más eficiente, tomar decisiones informadas y lograr un mejor rendimiento y competitividad en la industria.

2.2.1.1. Funciones

Esta herramienta desempeña un papel fundamental en los procesos de producción de cualquier industria, debido a:

- Los algoritmos predictivos se fundamentan en el estudio de patrones y datos históricos con el propósito de detectar posibilidades y amenazas en el contexto empresarial [8]. Su tarea principal radica en establecer conexiones entre las diversas variables que participan en los procesos, lo que permite evaluar probabilidades y riesgos. Estos algoritmos proporcionan orientación al gerente de una empresa para tomar decisiones dinámicas basadas en análisis actuales.
- Esta herramienta es versátil y se puede aplicar en diversas actividades empresariales, como la planificación de personal [8]. Además, es capaz de predecir el comportamiento deseado en el personal de la empresa y pronosticar la demanda futura de la compañía.

2.2.1.2. Objetivos

Las metas y diseños de los algoritmos dependen del entorno en el que se emplearan en diversas modalidades, ya sea en procesamiento en tiempo real, en lotes o de forma interactiva.

Los objetivos principales de un algoritmo predictivo en tiempo real incluyen:

- El objetivo principal de los algoritmos predictivos en tiempo real es garantizar el cumplimiento de los plazos establecidos con el fin de prevenir cualquier pérdida de datos [9].
- La predictibilidad es un objetivo fundamental de los algoritmos predictivos en tiempo real, ya que busca evitar la disminución en la calidad que ofrecen los sistemas multimedia. Esto implica que los algoritmos deben ser capaces de predecir y adaptarse de manera precisa y confiable a los cambios en tiempo real, asegurando así que la calidad de los sistemas multimedia se mantenga en niveles óptimos [10].

En lo que respecta a los objetivos de los algoritmos de procesamiento por lotes, se pueden destacar los siguientes:

- Maximizar el uso de la CPU, manteniéndola ocupada de manera constante.
- Mejorar el rendimiento, buscando incrementar la cantidad de trabajo y actividades realizadas por hora.
- Minimizar al máximo el tiempo de retorno, es decir, reducir el tiempo que transcurre desde la finalización del producto hasta su entrega.

2.2.2. ¿Qué es Machine Learning?

Según Denniye Ramírez, destaca que Machine Learning se ha convertido en un pilar fundamental para el manejo de datos a gran escala en diversos sectores, como informática, salud, corporativos y transporte, además se menciona que el ML tiene aplicaciones prometedoras en campos como la medicina, robótica y mecánica, aunque también genera preocupaciones sobre la dependencia de sistemas mecánicos inteligentes y su impacto en la obsolescencia humana [11].

Esta disciplina de la inteligencia artificial utiliza algoritmos basados en datos antiguos o recientes para mejorar el análisis de datos y realizar predicciones futuras, es una herramienta clave para el análisis y la predicción de datos en diversos sectores, pero plantea desafíos y reflexiones sobre su uso a largo plazo.

Si bien el ML es una disciplina de la inteligencia artificial que permite a las maquinas aprender y tomar decisiones basadas en datos y patrones, lo cual representa una ventana hacia la automatización y la optimización en una amplia gama de aplicaciones, desde el diagnostico medico hasta la conducción autónoma de vehículos. Es una herramienta poderosa que transforma datos en conocimientos, permitiendo la toma de decisiones más informadas y la creación de soluciones innovadoras, sin embargo, también plantea cuestiones éticas y desafíos, como la transparencia y la responsabilidad en el diseño de algoritmos [12] [13].

2.2.2.1. Aprendizaje supervisado.

En el aprendizaje supervisado, se emplea tanto un conjunto de datos de entrenamiento como uno de prueba. En este enfoque, el modelo recibe tanto los datos de entrada como las salidas deseadas del conjunto de entrenamiento, lo que le proporciona una retroalimentación valiosa para ajustar su rendimiento y lograr que las salidas se acerquen al resultado esperado de manera progresiva [14].

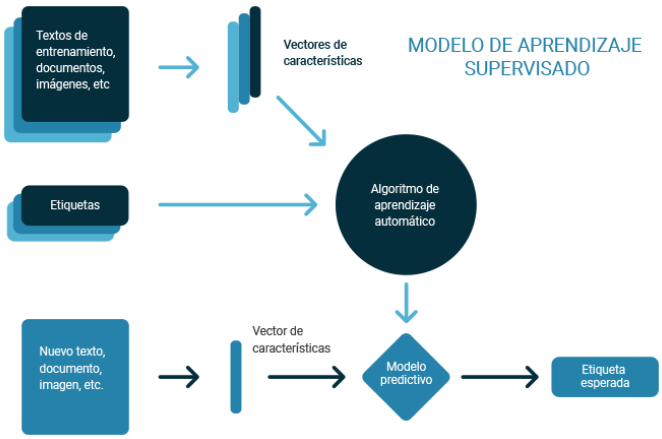


***Figura** 1: Aprendizaje supervisado [12]*

En este proyecto, se opta por el enfoque de aprendizaje supervisado debido a la disponibilidad de datos previos para la estimación del peso, los cuales se utilizarán como entradas fundamentales para entrenar y alimentar el modelo de manera efectiva.

2.2.2.2. Aprendizaje no supervisado

Por otro lado, en el aprendizaje no supervisado, no se hace una distinción entre un conjunto de datos de entrenamiento y uno de prueba. En este enfoque, el modelo de entrenamiento se sumerge en los datos de entrada sin la guía explicita de salidas deseadas. Su tarea principal es descubrir de manera autónoma patrones ocultos, estructuras subyacentes y correlaciones intrínsecas en los datos. Esto a menudo se traduce en la agrupación natural de datos en

clústeres o la reducción de dimensionalidad para revelar características significativas, por ende, el aprendizaje no supervisado es esencial para la exploración y compresión de datos en situaciones donde no se tiene etiquetas claras o resultados previos para orientar el proceso de aprendizaje [15] [16].

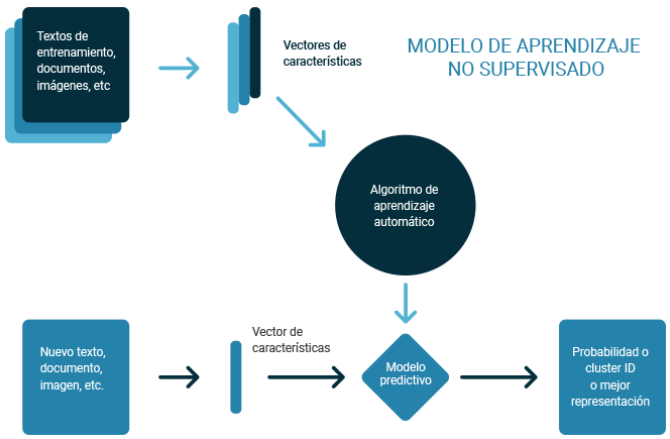


Figura 2: *Aprendizaje no supervisado [15]*

2.2.2.3. Algoritmos de Machine Learning

Dentro del campo del Machine Learning, encontramos una amplia gama de algoritmos, en el subcampo del aprendizaje no supervisado, algunos de los algoritmos destacados incluyen el K-Means, Redes Neuronales, Clusterizacion Jerárquica y Mezcla Gaussiana, entre otros. En el caso del aprendizaje supervisado, sobresalen algoritmos como Redes Neuronales, Maquina de soporte Vectorial, Regresión logística, KNN, bosques aleatorios y arboles de decisión, cada uno con sus propias aplicaciones y ventajas [17].

2.2.2.4. Regresión lineal

La regresión lineal es un algoritmo de ML utilizado para modelar la relación entre una variable independiente y una variable dependiente mediante una ecuación lineal, su funcionamiento se basa en encontrar la línea recta que mejor se ajusta a los datos de entrada para predecir la variable objetivo. La línea se representa mediante la ecuación $y = mx + b$, donde "y" es la variable objetivo, "x" es la variable independiente, "m" es la pendiente de la línea y "b" es la intersección con el eje y [18].

Este algoritmo busca determinar los valores óptimos de "m" y "b" de manera que minimicen la disparidad entre las proyecciones del modelo y los valores rales en el conjunto de datos de entrenamiento, esto se logra mediante técnicas como el método de los mínimos cuadrados.

Una vez que el modelo esta entrenado, puede utilizarse para realizar predicciones sobre nuevos datos de entrada, extrapolando la relación linear descubierta durante el entrenamiento.

La regresión lineal es ampliamente utilizada en problemas de predicción numérica, como la estimación de precios de viviendas, la predicción de ingresos y muchas otras aplicaciones donde se busca modelar relaciones lineales entre variables.

2.2.2.5. Regresión Polinomial

Se utiliza para modelar relaciones no lineales entre una variable independiente y una variable dependiente, a diferencia de la regresión lineal, que asume una relación lineal simple, la regresión polinomial permite capturar curvas y patrones más complejos en los datos. Funciona mediante la construcción de un modelo polinomial que se ajusta a los datos, donde se utiliza un polinomio de grado n (donde n es un numero entero) para representar la relación entre x e y [19].

El proceso implica encontrar los coeficientes óptimos para el polinomio que minimizan la disparidad entre las estimaciones del modelo y los valores reales de Y. esto se logra a través de técnicas de optimización como el método de mínimos cuadrados.

En resumen, la regresión polinomial es una herramienta poderosa para modelar relaciones no lineales en los datos al ajustar un polinomio de grado adecuado a los puntos de datos, lo que permite capturar patrones más complejos y realizar predicciones más precisas.

2.2.2.6. K-Nearest Neighbors (KNN)

Este algoritmo de aprendizaje supervisado utilizado tanto para clasificación como para regresión funciona de la siguiente manera:

- Cuando se le presenta un nuevo punto de datos que necesita ser clasificado o estimado, KNN busca los “Vecinos” más cercanos en el conjunto de datos de entrenamiento.
- Estos vecinos son los puntos de datos que tienen características similares a las del nuevo punto.
- La “K” en KNN representa el número de vecinos que se deben considerar, para luego asignar el nuevo punto a la etiqueta o el valor objetivo que es más común ente sus vecinos, en el caso de clasificación, o realiza una regresión para estimar un valor numérico, en el caso de regresión.

En esencia, KNN se basa en la idea de que los puntos de datos similares tienden a agruparse en el espacio de características, al tomar en cuenta la opinión de sus vecinos más cercanos, el algoritmo KNN toma decisiones basadas en las proximidad y similitud de los datos. En un enfoque simple y fácil de entender en el mundo del ML, aunque la elección del valor "K" y la elección de una métrica de distancia son consideraciones clave para su rendimiento optimo [20].

2.2.2.7. Support Vector Regression (SVR Lineal)

Esta técnica es utilizada para resolver problemas de regresión, es decir, para predecir valores numéricos continuos en lugar de etiquetas de clasificación, su funcionamiento se basa en encontrar una línea recta (o hiperplano en espacios de dimensiones superiores) que mejor se ajusta a los datos de entrenamiento, de modo que minimice la diferencia entre las predicciones del modelo y los valores reales [21].

Los vectores de soporte son los puntos de datos más cercanos a la línea de regresión, el objetivo es que la diferencia para estos vectores es que sea lo más pequeña posible, mientras se permite una cierta tolerancia o margen de error. El algoritmo busca optimizar esta línea de regresión de manera que la mayoría d ellos puntos de datos estén dentro del margen de error especificado, lo que permite obtener un modelo de regresión que generaliza bien a nuevos datos y es menos sensibles a valores atípicos [22].

2.2.2.8. Random Forest

Este método de aprendizaje supervisado se basa en la técnica de "ensemble Learning" o aprendizaje por conjuntos, el cual funciona combinando múltiples arboles de decisión para tomar decisiones más robustas y precisas. Cada árbol individual en el bosque se entrena con una porción aleatoria d ellos datos y luego toma decisiones sobre una nueva muestra de datos. Luego las decisiones de todos los árboles se promedian o ponderan para producir una predicción final, esto permite reducir el riesgo de sobreajuste (overfitting) y aumentar la precisión del modelo [23].

Random Forest es eficaz para problemas de clasificación y regresión y es conocido por su capacidad para manejar características irrelevantes y datos ruidosos, además proporciona una evaluación de la importancia de las características en la predicción, lo que puede ser útil para seleccionar las características más relevantes en un conjunto de datos [24].

En resumen, Random Forest es un algoritmo que aprovecha la sabiduría de múltiples arboles de decisión para tomar decisiones más solidad y precisas en una amplia variedad de aplicaciones de ML.

2.2.2.9. Gradient Boosting

Esta técnica funciona mediante la combinación de múltiples modelos simples, generalmente arboles de decisión, en un solo modelo más fuerte y preciso. Inicialmente, se entrena un modelo base, y luego se crean modelos adicionales que se enfocan en corregir los errores del modelo anterior. Cada nuevo modelo se ajusta para minimizar los errores residuales del conjunto de datos, lo que significa que se enfoca en las instancias que el modelo anterior no pudo predecir con precisión. Este proceso se repite varias veces, y los modelos individuales se ponderan y combinan para hacer predicciones finales. Gradiente Boosting Es efectivo para problemas de regresión y clasificación u es conocido por su capacidad para manejar datos ruidosos y obtener resultados de alta calidad [25].

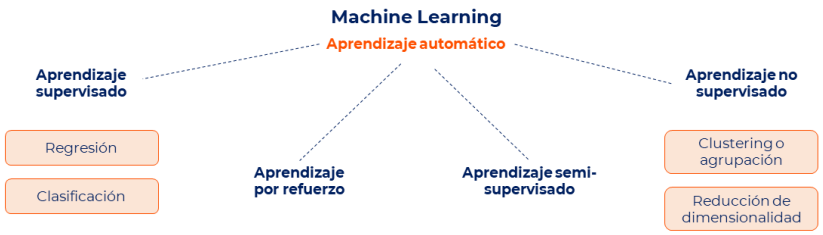


Figura 3:*Tipos de algoritmos.*

Fuente: *https://decidesoluciones.es/wp-content/uploads/2022/11/Machine-Learning-y-tipos-de-aprendizaje.png*

Tras un exhaustivo estudio de diversos algoritmos y métodos para construir el modelo, se llega a la conclusión de que los algoritmos de aprendizaje supervisado, en particular los de regresión y clasificación, son los más apropiados. Se llevarán a cabo pruebas exhaustivas con estos algoritmos para determinar cuál de ellos ofrece un rendimiento y eficacia óptimos en los resultados.

2.2.3. Clasificación de los datos

La clasificación de los datos desempeña un papel fundamental en el entrenamiento de modelos de Machine Learning y es crucial por varias razones:

Supervisión y etiquetado: en el aprendizaje supervisado, donde se alimenta al modelo con ejemplos etiquetados, la clasificación adecuada de los datos proporciona las etiquetas correctas para los ejemplos de entrenamiento. Esto permite que el modelo aprenda a relacionar las características de entrada con las etiquetas de salida deseadas [26].

Generalización: La clasificación precisa de los datos de entrenamiento ayuda al modelo a generalizar mejor los patrones y relaciones presentes en los datos. Un etiquetado incorrecto puede llevar a un modelo a aprender patrones erróneos [27].

Evaluación del rendimiento: La clasificación de los datos también es esencial para evaluar el rendimiento del modelo. Se requiere una clasificación precisa de los datos de prueba para comparar las predicciones del modelo con las etiquetas reales y calcular métricas de rendimiento, como la precisión, la sensibilidad y la especificidad [27].

Interpretación y explicación: La clasificación correcta de los datos también facilita la interpretación y explicación del modelo. Cuando los datos se clasifican adecuadamente, es más fácil comprender por qué el modelo toma ciertas decisiones o hace ciertas predicciones.

La clasificación precisa de los datos es esencial para el éxito del entrenamiento de modelos de ML, ya que afecta directamente a la calidad de las predicciones, la capacidad de generalización del modelo y la interpretación de sus resultados, por lo tanto, se debe prestar una atención cuidadosa a la calidad y precisión de la clasificación de datos en todo el proceso de desarrollo del modelo [28].

2.2.4. Herramientas y bibliotecas

La investigación de las herramientas necesarias para entrenar un modelo de Machine Learning es de suma importancia en el proceso de desarrollo de un proyecto de aprendizaje automático. En el campo del Machine Learning, se dispone de una amplia gama de bibliotecas, frameworks y herramientas de gran utilidad. Ejemplos de estas herramientas incluyen Tensor Flow, PyTorch, scikit-learn, entre otras destacadas opciones.

2.2.4.1. Scikit-learn

Es una biblioteca de código abierto en Python que se ha convertido en una de las herramientas más populares y esenciales en el campo del Machine Learning y la ciencia de datos. Se origino en 2007 como parte del proyecto Google Summer of Code y se ha desarrollado desde entonces como un proyecto de código abierto. Su comunidad activa de desarrolladores y su crecimiento continuo han contribuido a su robustez y versatilidad [29].

Su principal funcionamiento es que proporciona una amplia variedad de algoritmos de aprendizaje supervisado y no supervisado, así como herramientas para el preprocesamiento de datos, selección de características, evaluación de modelos y más. Lo cual lo convierte en una herramienta integral para la construcción y evaluación de modelos de Machine Learning [30].

Su facilidad de uso y su enfoque en la simplicidad y consistencia en su API. Esto lo hace adecuado tanto para principiantes como para expertos en Machine Learning, además de su integración con Python, por lo cual se integra perfectamente con el ecosistema de Python facilitando la combinación de sus capacidades con otras bibliotecas populares como Numpy, Pandas y Matplotlib [31].

Si bien se utiliza en una amplia variedad de aplicaciones, que van desde la clasificación de texto y la visión por computadora hasta la bioinformática y la astronomía, los investigadores en el campo de la inteligencia artificial y el aprendizaje automático utilizan Scikit-learn para llevar a cabo experimentos, comparar algoritmos y desarrollar soluciones para problemas diversos [32].

El impacto que ha tenido en la industria es muy significativo en la adopción de técnicas, ya que proporciona una base sólida para la implementación de soluciones de aprendizaje automático en aplicaciones del mundo real.

2.2.4.2. Pandas

Pandas es una biblioteca de Python ampliamente utilizada en la ciencia de datos y el análisis de datos, la cual proporciona estructuras de datos flexibles y herramientas de manipulación de datos que son esenciales para la preparación y exploración de datos en proyectos de Machine Learning [33].

Panda simplifica la carga, limpieza y transformación de datos, por lo que permite cargar datos desde una variedad de fuentes, como archivos CVS, bases de datos SQL o Excel, y luego realizar operaciones de limpieza y transformación, como eliminar valores nulos, cambiar tipos de datos o aplicar funciones personalizadas [34].

Facilita la exploración inicial de los datos, pudiendo realizar análisis estadísticos descriptivos, calcular estadísticas resumida, visualizar datos y generar gráficos para comprender mejor la distribución y las relaciones en tus datos [35].

Por ende, panda es una herramienta esencial en proyectos de Machine Learning, ya que simplifica la manipulación y la exploración de datos, por lo que permite trabajar de manera más eficiente y efectiva en todas las etapas del proyecto, desde la adquisición de datos hasta la preparación y el análisis exploratorio.

2.2.4.3. Framework Streamlit

Streamlit es una frameworks de Python que simplifica la creación de aplicaciones web interactivas y visualizaciones de datos, funciona de manera intuitiva al permitir a los desarrolladores convertir rápidamente scripts de Python en aplicaciones web, sin necesidad de conocimientos avanzados en desarrollo web. Los usuarios pueden escribir código Python como lo harían en un script normal, pero con las funciones de Streamlit, pueden agregar widgets interactivos como botones, barras deslizantes y cuadros de textos para interactuar con los datos y visualizaciones en tiempo real. Además, este frameworks se encarga automáticamente de la gestión de la interfaz de usuario y la actualización de los elementos a medida que los usuarios interactúan, lo que simplifica en gran medida el proceso de desarrollo de aplicaciones web de datos y ML [36].

2.3. Marco Legal

A continuación, se introduce el marco legal que respalda y contextualiza el presente estudio con los Art. 2,Art. 7 y Art 213.

2.3.1. LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES

Art. 2. - Establece su alcance, indicando que se aplica al manejo de datos personales en cualquier formato, ya sea en sistemas automatizados o en formatos no automatizados, así como a cualquier forma de utilización posterior de estos datos [37].

Art. 7. – Tratamiento legítimo de datos personas. El tratamiento adecuado de los datos personales se considerará legítimo y lícito cuando se obtenga el consentimiento del titular para el procesamiento de sus datos con fines específicos o múltiples [37].

2.3.2. LEY ORGÁNICA PARA EL DESARROLLO DE LA ACUICULTURA Y PESCA

El Artículo 213 se refiere a infracciones serias en la actividad pesquera, incluyendo la captura, posesión, transferencia, desembarque, custodia o almacenamiento, antes de su primera venta, de especies pesqueras que no cumplen con los tamaños o pesos permitidos. También se considera una infracción grave cuando se exceden los límites establecidos para ciertas especies [38].

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. Localización

La presente investigación se realizó en el campus "La María" de la Universidad técnica Estatal de Quevedo la cual se encuentra ubicada a 143,31m de la entrada vía Mocache en una de las piscinas de criaderos de peces, la cual cuenta con varias especies nativas como lo es la viaja azul entre otras, las coordenadas de las instalaciones son -1.0800164390879994, -79.50142728018108.

Figura 4: *Mapa de la localización de las Infraestructuras de la Universidad Técnica Estatal de Quevedo donde se realizó la investigación*

3.1. Tipo de investigación

3.1.1. Investigación Descriptiva

En virtud de que este tipo de investigación se centra en describir características, fenómenos o comportamientos existentes en relación con el tema de estudio. Además, puede ser útil para obtener una comprensión inicial de los parámetros fisiológicos y el peso de los peces nativos, así como su relación con el crecimiento y conocer los diferentes algoritmos de predicción.

3.1.2. Investigación aplicada

Este proyecto se llevó a cabo aplicando todos los conocimientos adquiridos a lo largo de la carrera, más lo que se adquirió mientras se desarrolló la investigación. En el cual se buscó poner en práctica los conocimientos adquiridos para realizar el modelo predictivo que predecirá el peso en los peces nativos basado en parámetros fisiológicos.

3.2. Método de investigación

Para la creación del modelo predictivo se utilizó el correspondiente software para el análisis de los datos el cual permitió determinar las variables óptimas para la funcionalidad del modelo, la manera en la que se entrenó el modelo es la siguiente:

Se dividió la data en un 80% para entrenar el modelo y un 20% para validarlo con el cual se obtuvo un resultado K^2= 90% de precisión del modelo.

3.2.1 Método Inductivo

Mediante este método se pudo identificar las variables fisiológicas más influyentes en el peso de los peces nativos. Al analizar la variedad de casos y observaciones, se pudo deducir que factores fisiológicos tiene una correlación más fuerte con el peso de los peces. Además de que es de utilidad para analizar los resultados obtenidos y así poder realizar mejoras en el modelo.

3.3 Fuentes de recopilación de información

Se realiza una investigación detallada utilizando una amplia gama de fuentes, como artículos científicos publicados en revistas especializadas y proyectos de investigación relevantes. Para recopilar información primaria, se consultan diversos repositorios que se centran en la investigación científica y académica, tales como:

- Google Scholar.
- Bases de Datos
- Scopus

Además, se utilizan fuentes secundarias como sitios web importantes y portales de noticias en Internet para obtener información adicional y complementar el estudio. Estas fuentes son útiles para ampliar el conocimiento y obtener perspectivas adicionales sobre el tema de investigación.

3.3. Diseño de la investigación

A continuación, se presenta el plan a seguir para obtener toda la información necesaria y llevar a cabo el proyecto con éxito.

Fase 1: Exploración y revisión bibliográfica

- Revisión exhaustiva de la literatura científica y estudios previos relacionados con la ecología de peces nativos, parámetros fisiológicos y el peso de los peces.
- Identificación de variables fisiológicas relevantes que podrían influir en el peso de los peces nativos.
- Análisis de los métodos y algoritmos utilizados en la predicción de características en peces.

Fase 2: Recopilación de datos

Teniendo en cuenta los datos recopilados sobre los peces nativos en diferentes estanques de la provincia

- Selección de una muestra representativa de peces nativos en el área de estudio en Ecuador.
- Medición de parámetros fisiológicos como longitud de los peces tanto estándares como totales y de la cabeza y peso total.
- Registro del peso de los peces en diferentes etapas de su desarrollo.

Fase 3: Análisis de datos

- Análisis estadístico de los datos recopilados para identificar posibles correlaciones entre los parámetros fisiológicos y el peso de los peces.
- Uso de técnicas de minería de datos y aprendizaje automático para explorar patrones y tendencias de los datos.

Fase 4: Desarrollo del modelo predictivo

En base al análisis de los datos e identificado las variables predictoras óptimas para el desarrollo del modelo, además de cumplir con el objetivo definido, siguiendo las siguientes actividades.

- Selección de algoritmos de predicción adecuados, como regresión lineal, regresión logística, arboles de decisión, redes neuronales, entre otros.
- Construcción de un modelo predictivo utilizando las variables fisiológicas como predictores y el peso de los peces como variable objetivo.
- Validación y ajuste del modelo utilizando técnicas de validación cruzada y evaluación de su precisión y confiabilidad.

3.4. Recursos y materiales

3.4.1. Recursos humanos

***Tabla 1**. Recursos humanos.*

Personal	Descripción
Autor	Stiven Javier Loor Ponce
Director del proyecto	Ing. Emilio Rodrigo Zhuma Mera. Msc

***Autor:** Stiven Loor*

3.4.2. Materiales

***Tabla 2.** Hardware.*

Cantidad	Equipo	Características
1	Computadora portátil	• Intel Core i3 7th Gen • 8GB RAM • 1 Tera
1	Smartphone	• Infinix NOTE 12 • • 256 ROM

***Autor**: Stiven Loor*

***Tabla 3**: Softwares*

Tipo	Descripción	Costo
Scikit-Learn	Es una biblioteca de aprendizaje automático en Python ampliamente utilizada.	$0,00
Streamlit	Es un marco de desarrollo en Python que permite crear fácilmente aplicaciones web interactivas para visualizar y compartir datos y modelos.	$0,00
Anaconda	Es una plataforma de gestión de paquetes y entornos en Python para ciencia de datos.	$0,00
VS Code	Es un editor de código fuente gratuito y de código abierto desarrollado por Microsoft.	$0,00
MS Office	Es un conjunto de aplicaciones de productividad.	$0,00
Google Colab	Es una plataforma en línea que permite ejecutar y colaborar en cuadernos de Jupyter de forma gratuita en la nube	$0,00

***Autor**: Stiven Loor*

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. Resultados

4.1.1. Identificación de las variables

Cumpliendo con el objetivo específico 1, en el contexto de este estudio, se consideraron tres variables fundamentales para el desarrollo de nuestro modelo predictivo. Las variables independientes incluyen la "Longitud Total" y la "Longitud de la cabeza", que se consideran factores relevantes para estimar el "Peso Total" de los peces nativos. La "Longitud Total" representa la medida desde el extremo de la cabeza hasta la aleta caudal, mientras que la "Longitud de la cabeza" se refiere específicamente a la longitud de la cabeza del pez.

Tabla 4: *Variables independientes*

Longitud Total	Longitud cabeza
23.7	5.8
24.7	6.3
24.4	6.6
23.0	5.9
22.7	5.7

Autor: *Stiven Loor*

Por otro lado, la variable dependiente es el "Peso Total", que es el valor que deseamos predecir a partir de las variables independientes mencionadas anteriormente. La identificación y comprensión de estas variables son esenciales para el desarrollo de nuestro modelo de estimación de peso en peces nativos.

Figura 5: *Variable dependiente*

Peso Total
280.03
305.36
305.07
270.78
259.40

Autor: *Stiven Loor*

4.1.1.2. Exploración de los datos

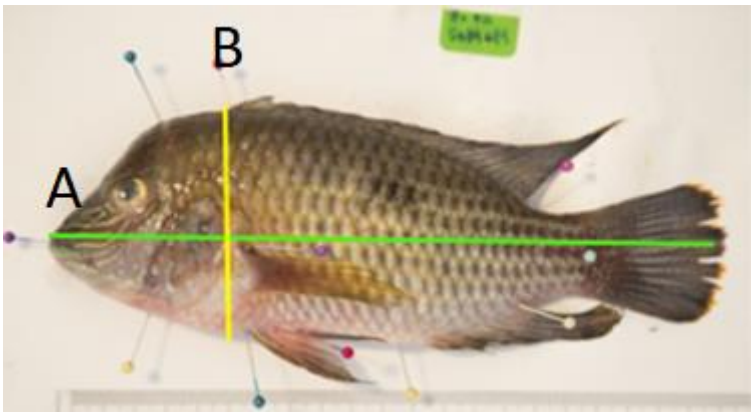
El DataSet se lo tomo de un proyecto realizado anterior mente por Dafna Racines [38]. Donde se recolectaron datos como la Longitud Total, Longitud de la Cabeza y el Peso Total, entre otros datos fisiológicos de la especie Andinocara rivolatus (vieja azul).

Tabla 5: *Variables Utilizadas*

Medidas en (cm)	
Longitud Total (A)	La cual se extiende desde la boca hasta la punta final de la aleta caudal
Longitud cabeza (B)	Medición horizontal que va desde el extremo de la boca hasta el final de las branquias
Peso Total (C)	Peso completo del pez, sin dividirlo en partes

Autor: *Stiven Loor*

Figura 6: *Ejemplar de la especie que se tomó los datos*



Autor*: Dafna Drouet R.*

Considerando que el DataSet abarcó un total de 73 registros por cada variable lo que es significativo, evidencio la necesidad de una mayor expansión para entrenar el modelo propuesto de manera exhaustiva. Esta limitación resalta la importancia de futuras investigaciones para recopilar datos más extensos y diversificados, garantizando así un modelo predictivo aún más preciso y confiable en futuros estudios, por ende, se realizó una técnica de "Data Augmentation" (aumento de datos), para aumentar la data y así poder entrenar de una mejor manera el modelo obteniendo como resultado una mejor eficiencia.

Tabla 6: *Datos aumentados*

	Longitud Total	Longitud Cabeza	Peso Total
Count	1512.000000	1512.000000	1512.000000
Mean	24.129762	6.024735	297.173876
Std	1.717004	0.538999	54.361577
Min	19.000000	4.600000	156.070000
25%	23.000000	5.600000	259.400000
50%	24.200000	6.000000	298.710000
75%	25.400000	6.400000	337.537500
Max	28.000000	7.400000	435.910000

Autor: *Stiven Loor*

Como se observa en la Tabla 6, se amplió el conjunto de datos a 1512 entradas por cada variable. Esta expansión se realizó con el objetivo de optimizar el entrenamiento del modelo. Se reconoce que a medida que el DataSet aumenta en tamaño, el modelo adquiere una mayor eficiencia y precisión en sus predicciones, lo que subraya la importancia de este enriquecimiento de datos para mejorar la robustez del modelo durante el proceso de entrenamiento.

4.1.1.3. Preprocesamiento de los datos

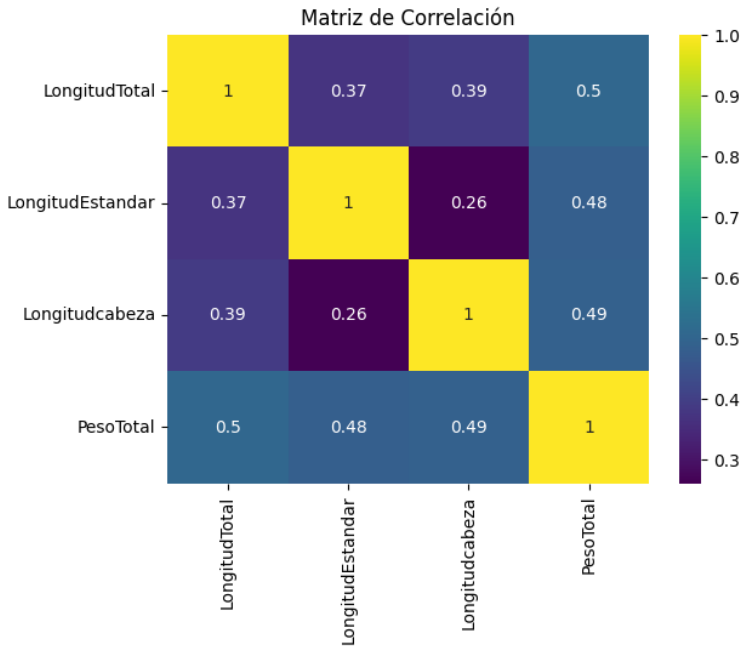
Se llevo a cabo una minuciosa exploración y análisis de conjunto de datos con el objetivo de identificar cualquier anomalía o dato inconsistente que pudiera tener un impacto negativo en el proceso de entrenamiento del modelo, este proceso se realizó para garantizar que los datos de entrada empleados en el entrenamiento del modelo estuvieran libres de errores, valores atípicos o cualquier otra inconsistencia que pudiera afectar la calidad y precisión de las predicciones generadas. La integridad y calidad de los datos son esenciales para el rendimiento del modelo, por lo que se dedicó una atención especial a la fase de preprocesamiento de datos. Esto se hizo para asegurar la confiabilidad de los resultados obtenidos durante el proceso de entrenamiento, reconociendo el papel crucial que desempeñan los datos bien preparados en la fiabilidad de las predicciones del modelo.

4.1.1.4. Matriz de correlación

En la Figura 7, se muestra la matriz de correlación con la cual se identificó las relaciones significativas entre las variables involucradas para el modelo propuesto. Al examinar esta matriz, se puede destacar las interconexiones más fuertes entre las variables independientes

y la variable dependiente. Para este caso, las variables de “Longitud Total” y “Longitud de la cabeza” mostraron una correlación destacada con la variable “Peso Total”.

Figura 7: *Matriz Correlación*



Autor: *Stiven Loor*

Esta técnica proporciono información valiosa para comprender que factores tienen una influencia más significativa en la predicción del peso de los peces nativos, facilitando así la selección de las variables óptimas para el modelo predictivo.

La identificación de las variables óptimas para el modelo predictivo representa un hito significativo en la investigación. Al analizar detenidamente diversas variables fisiológicas y considerando su correlación con el peso de los peces nativos, se estableció una base sólida para el enfoque predictivo

Estos resultados subrayan la importancia de seleccionar cuidadosamente las variables que influyen de manera más significativa en el crecimiento de los peces, garantizando así la precisión y confiabilidad del modelo. Este paso crucial permitió avanzar hacia la siguiente fase del estudio, donde se integrará estas variables optimas en el diseño y desarrollo del modelo predictivo, anticipando así resultados aún más precisos y aplicables en el contexto de la acuicultura sostenible.

4.1.2. Análisis comparativo entre distintos modelos

4.1.2.1. Análisis comparativo sobre la clasificación de Aprendizaje Automático

Cumpliendo con el objetivo específico 2 de determinar la tecnología más adecuada para el modelo se realizó un análisis exhaustivo de los algoritmos de aprendizaje automático, los cuales se dividen en dos categorías: aprendizaje supervisado y no supervisado, para este análisis se optó por estas dos tecnologías porque son las que más se encuentran en común con la investigación, esta evaluación meticulosa permitió identificar la metodología más viable para la investigación.

4.1.2.2. Aprendizaje supervisado vs no supervisado

Definiciones

El aprendizaje supervisado constituye un enfoque fundamental en el cual el modelo se adiestra mediante un conjunto de datos etiquetados, es decir, datos de entrada acompañados de sus correspondientes salidas deseadas. El propósito principal radica en permitir que el modelo aprenda a realizar predicciones precisas para nuevos datos.

En contraste, el aprendizaje no supervisado se distingue por su utilización de un conjunto de datos no etiquetados. En este caso, el objetico es descubrir patrones, estructuras o relaciones ocultas dentro de los datos sin la orientación proporcionada por salidas deseadas predefinidas. Este enfoque se basa en la capacidad intrínseca del modelo para identificar y comprender la información inherente en los datos, lo que lo convierte en una herramienta invaluable para revelar conocimientos previamente desconocidos en conjuntos de datos complejos.

Problema

En el contexto del aprendizaje supervisado se destaca en problemas de clasificación regresión. En este método, se busca predecir etiquetas o valores numéricos basados en características conocidas, lo que resulta fundamental para entender y anticipar patrones en los datos. Por otro lado, el aprendizaje no supervisado se aplica a desafíos de clustering y reducción de dimensionalidad. En este caso, el objetivo radica en descubrir patrones y estructuras inherentes en los datos sin utilizar etiquetas previas. Esta metodología revela información valiosa sobre la agrupación de datos similares y permite una comprensión más profunda de la complejidad intrínseca de los conjuntos de datos no etiquetados.

Etiquetado

Aprendizaje Supervisado: Requiere datos etiquetados para entrenar el modelo, lo que significa que se necesita un esfuerzo adicional para obtener y preparar los datos con etiquetas adecuadas.

Aprendizaje No Supervisado: No requiere datos etiquetados, lo que lo hace más conveniente en situaciones donde la obtención de datos etiquetados puede ser costosa o difícil.

Evaluación del Rendimiento

La precisión del modelo en el aprendizaje supervisad se determina mediante la comparación de las predicciones del modelo con las salidas reales en el conjunto de prueba, proporcionando una evaluación objetiva y cuantitativa. Por otra parte, la evaluación del aprendizaje no supervisado es más subjetiva depende del objetivo específico de clustering.

En virtud de esta comparación entre estos dos enfoques, se llegó a la conclusión de que el aprendizaje supervisado es el método más adecuado para el modelo, esto se debe a que cada variable en el conjunto de datos esta etiquetada, lo que permite una interpretación clara y precisa, garantizando así una predicción más confiable y específica para él es estudio.

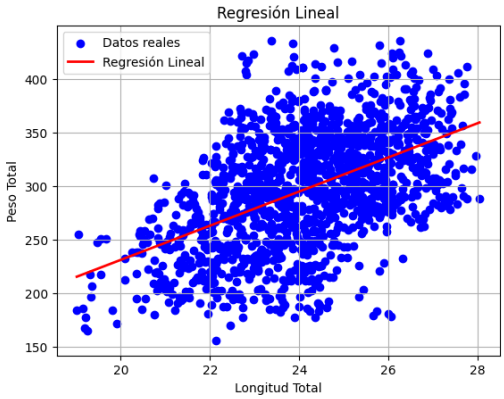
4.1.2.3. Clasificación del aprendizaje supervisado

A continuación, se presenta un análisis comparativo de los algoritmos de clasificación más comunes en el aprendizaje supervisado, al final de la comparación se mostrará una tabla con los resultados de cada uno de los algoritmos.

Regresión Lineal:

La Regresión Lineal es un algoritmo de aprendizaje supervisado que busca modelar la relación lineal entre una variable independiente y una variable dependiente mediante una ecuación lineal, como $y = mx + b$. Busca encontrar los valores óptimos de la pendiente (m) y la intersección con el eje y (b) que minimicen la disparidad entre las estimaciones del modelo y los valores reales en el conjunto de entrenamiento. Es útil para problemas de predicción numérica y es fácil de entender y aplicar.

Figura 8: *Regresión Lineal*



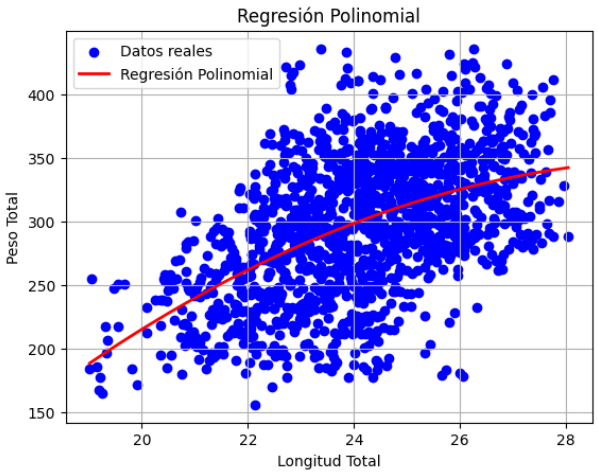
Autor*: Stiven Loor*

Los resultados obtenidos del modelo de Regresión lineal Revelan un nivel de precisión en las predicciones, aunque con notables limitaciones. El error cuadrático Medio (MSE) de 2054.49 y el Error Absoluto Medio (MAE) de 36.46 indican discrepancias considerables entre las predicciones del modelo y los valores reales. La Raíz del Error cuadrático Medio (RMSE) de 45.33 revela una desviación promedio de aproximadamente 45.33 unidades de la variable objetivo en las predicciones. Además, el Coeficiente de Determinación (R^2) de 0.32 señala que alrededor del 32% de la variabilidad en los datos no está siendo explicada por el modelo, lo que sugiere limitaciones significativas en la capacidad del modelo de Regresión Lineal para capturar la complejidad de los datos. Estos hallazgos subrayan la necesidad imperativa de explorar modelos más sofisticados y de considerar cuidadosamente la selección de características para mejorar la precisión de las predicciones.

Regresión Polinomial:

La Regresión Polinomial se utiliza cuando la relación entre variables no es lineal. En lugar de ajustarse a una línea recta, utiliza un polinomio de grado n para representar la relación entre las variables. Esto permite capturar patrones más complejos en los datos. El proceso implica encontrar los coeficientes óptimos para el polinomio que minimizan la diferencia entre las predicciones del modelo y los valores reales de Y. Es ideal para modelar relaciones curvilíneas en datos.

Figura 9: *Regresión Polinomial*



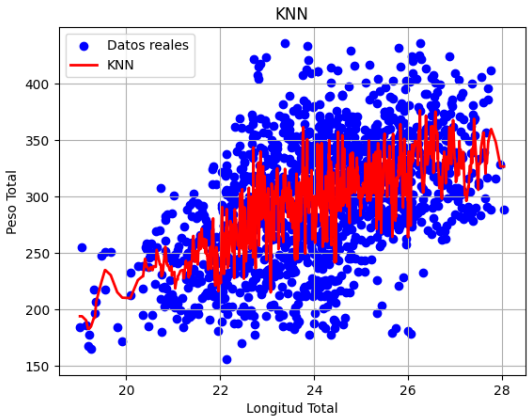
Autor: *Stiven Loor*

Los resultados obtenidos del modelo de regresión polinomial muestran un rendimiento moderado en cuanto a la predicción del peso de los peces. El Error cuadrático Medio (MSE) de 2016.83 y el Error Absoluto Medio (MAE) de 36.04 indican ciertas limitaciones en la precisión de las predicciones realizadas por el modelo. Aunque se observa una correlación positiva entre las variables, evidenciada por el Coeficiente de Determinación (R^2) de 0.33, este valor relativamente bajo sugiere que aproximadamente el 33% de la variabilidad en los datos no está siendo explicada por el modelo propuesto. Estos resultados subrayan la importancia de llevar a cabo una evaluación cuidadosa y considerar la posibilidad de explorar otros métodos para mejorar la precisión del modelo, así como para aumentar su capacidad para predecir de manera más fiable el peso de los peces nativos. Este hallazgo resalta la necesidad de una investigación adicional y ajustes metodológicos para optimizar la confiabilidad y validez de las predicciones en futuras aplicaciones prácticas.

K-Nearest Neighbors (KNN):

K-Nearest Neighbors es un algoritmo de aprendizaje supervisado que clasifica o regresa valores basados en la proximidad a "vecinos" cercanos en el conjunto de datos de entrenamiento. La "K" representa el número de vecinos considerados. Dependiendo de la elección de "K" y la métrica de distancia, KNN asigna una etiqueta de clasificación o realiza una regresión. Se basa en la idea de que datos similares tienden a agruparse en el espacio de características.

Figura 10: *Modelo: KNN*



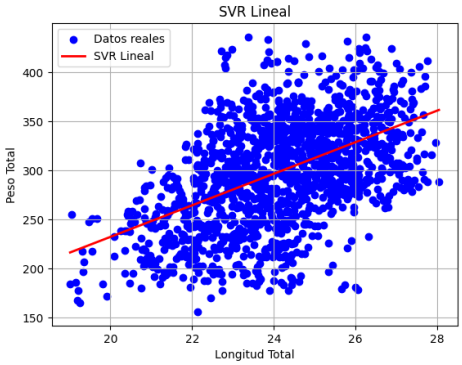
Autor: *Stiven Loor*

El modelo K-Nearest Neighbors (KNN) mostro un rendimiento subóptimo según las métricas evaluadas. Con un error cuadrático Medio (MSE) de 2763.23 y un Error Absoluto Medio (MAE) de 40.71, se evidencia una discrepancia significativa entre las predicciones del modelo y los valores reales. La Raíz del Error cuadrático Medio (RMSE) de 52.57 señala una dispersión considerable en las predicciones. Además, el coeficiente de determinación (R^2) de 0.09 indica una baja capacidad del modelo para explicar la variabilidad en los datos. Estos resultados sugieren que el modelo KNN no logra capturar las relaciones complejas entre las variables predictoras y la variable objetivo. Esta situación subraya la necesidad imperante de considera otros enfoques o técnicas mas avanzadas para mejorar significativamente la precisión de las predicciones y abordar la complejidad inherente de los datos.

Support Vector Regression (SVR Lineal):

La Regresión de Vectores de Soporte (SVR) se utiliza para problemas de regresión y busca encontrar una línea recta (o hiperplano en espacios de dimensiones superiores) que mejor se ajuste a los datos de entrenamiento, minimizando la diferencia entre las proyecciones del modelo y los datos reales. Se centra en los "vectores de soporte" que son los puntos de datos más cercanos a la línea de regresión. Es menos sensible a valores atípicos y generaliza bien a nuevos datos.

Figura 11: *SVR Lineal*



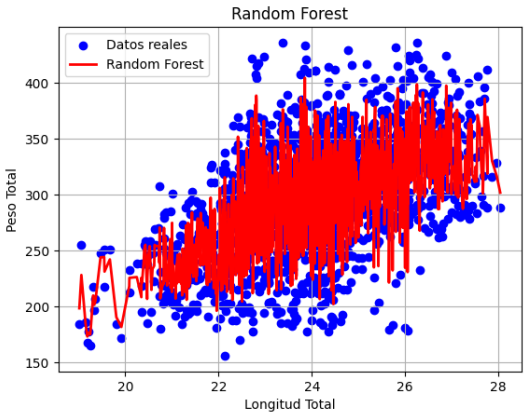
Autor: *Stiven Loor*

Los resultados obtenidos del modelo de SVR Lineal revelan un rendimiento moderado en la predicción del peso de los peces nativos. El Error Cuadrático Medio (MSE) de 2053.13 indica que las predicciones tienden a alejarse significativamente de los valores reales, mientras que el Error Absoluto Medio (MAE) de 36.49 sugiere que, en promedio, las predicciones tienen una desviación de alrededor de 36.49 gramos con respecto a los valores reales. A pesar de los esfuerzos, el coeficiente de determinación (R^2) de 0.32 indica que solo el 32% de la variabilidad en los datos de peso puede ser explicada por el modelo, lo que sugiere limitaciones significativas en su capacidad predictiva. Estos resultados subrayan la necesidad de explorar y mejorar las técnicas de modelado utilizadas, así como considerar la incorporación de variables adicionales para lograr una predicción más precisa y confiable del peso de los peces nativos.

Random Forest:

Random Forest es un algoritmo de aprendizaje supervisado que utiliza "ensemble learning" para combinar múltiples árboles de decisión. Cada árbol se entrena con una porción aleatoria de los datos y luego se promedian o ponderan sus decisiones para producir una predicción final. Esto reduce el riesgo de sobreajuste y mejora la precisión. Es efectivo para problemas de clasificación y regresión, incluso en presencia de datos ruidosos.

Figura 12: *Random Forest*



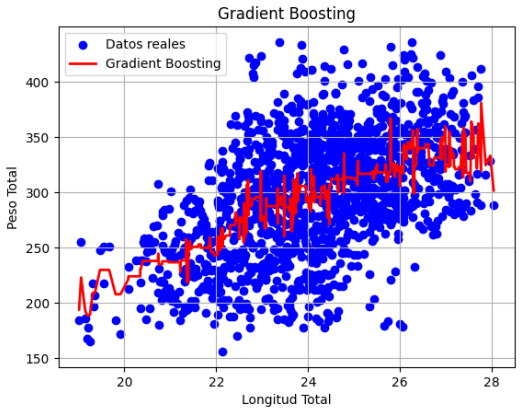
Autor: *Stiven Loor*

Los resultados obtenidos del modelo Random Forest son prometedores, indicando un Error Cuadrático Medio (MSE) de 500, una Raíz del Error Cuadrático Medio (RMSE) de 22.36 y un Error Absoluto Medio (MAE) de 20. Además, el Coeficiente de Determinación (R^2) del 0.80 sugiere que el 80% de la variabilidad en los datos de peso de los peces se explica por el modelo. Estos números revelan una precisión aceptable en las predicciones del modelo, lo que respalda la efectividad del enfoque de Random Forest para estimar el peso de los peces basado en parámetros fisiológicos. Sin embargo, se deben considerar posibles mejoras y refinamientos para aumentar aún más la precisión del modelo y su aplicabilidad en situaciones del mundo real.

Gradient Boosting:

Gradient Boosting es una técnica que combina múltiples modelos simples, como árboles de decisión, en un solo modelo más fuerte y preciso. Inicialmente, se entrena un modelo base y luego se crean modelos adicionales que se centran en corregir los errores del modelo anterior. Cada nuevo modelo minimiza los errores residuales. Es adecuado para problemas de regresión y clasificación y es capaz de manejar datos ruidosos para obtener resultados de alta calidad.

Figura 13: *Gradient Boosting*



Autor: *Stiven Loor*

El modelo Gradient Boosting indican un rendimiento que, aunque no es óptimo, proporciona información valiosa. El error cuadrático medio (MSE) de 2593.77 y el error absoluto medio (MAE) de 39.87 sugieren que las predicciones tienen una desviación significativa de los valores reales. Además, el coeficiente de determinación (R^2) de 0.14 muestra que solo el 14% de la variabilidad en los datos de peso de los peces puede explicarse por el modelo, lo que indica una correlación limitada entre las variables predictoras y la variable objetivo. Aunque estos resultados revelan limitaciones en la precisión del modelo, también ofrecen oportunidades para mejorar y refinar el enfoque. Será crucial examinar y ajustar las características seleccionadas y considerar métodos de ingeniería de características para capturar mejor las complejidades del sistema. Además, explorar otros algoritmos y técnicas de mejora del modelo podría ser fundamental para desarrollar una herramienta predictiva más precisa y confiable.

Tabla 7:*Resultados de la comparación de algoritmos*

Modelo	Error cuadrático medio	Raíz del Error cuadrático medio	Error absoluto medio	Coeficiente de determinación
Regresión Lineal	2054.49	45.33	36.46	0.32
Regresión Polinomial	2016.83	44.91	36.04	0.33
K-Nearest Neighbors	2763.23	52.57	40.71	0.09
SVR Lineal	2053.13	45.31	36.49	0.32
Random Forest	500	22.36	20	0.80
Gradient Boosting	2593.77	50.93	39.87	0.14

Autor*: Stiven Loor*

Basado en los resultados de la tabla 7 los cuales indican que el algoritmo **Random Forest** tiene un Error Cuadrático Medio de 500, lo que sugiere una discrepancia promedio de 500 unidades en las predicciones del peso de los peces. La Raíz del Error Cuadrático Medio (RMSE) es 22.36, lo que significa que la desviación estándar de los errores de predicción es aproximadamente 22.36 unidades. El Error Absoluto Medio (MAE) es de 20, indicando que, en promedio, las predicciones del modelo tienen una diferencia absoluta de 20 unidades con respecto a los valores reales. Además, el Coeficiente de Determinación (R2) es 0.8, lo que implica que el 80% de la variabilidad en los datos del peso de los peces puede explicarse por el modelo. Estos resultados sugieren que el modelo tiene un buen rendimiento en la predicción del peso de los peces, aunque todavía hay espacio para mejoras, especialmente en la reducción del MSE y el MAE para hacer las predicciones más precisas

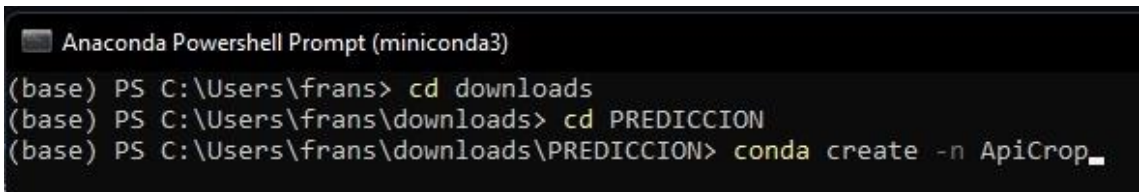
4.1.3. Pruebas, validación y desarrollo del programa

Cumpliendo con el objetivo específico 3 donde se prueba y evalúa el modelo de predicción donde se detalla lo siguiente:

4.1.3.1. Entorno virtual de desarrollo

Para la creación del programa, se decidió configurar un ambiente virtual de Python utilizando MiniConda3, una herramienta que permite trabajar con los paquetes requeridos de manera aislada, evitando conflictos con otras dependencias del sistema.

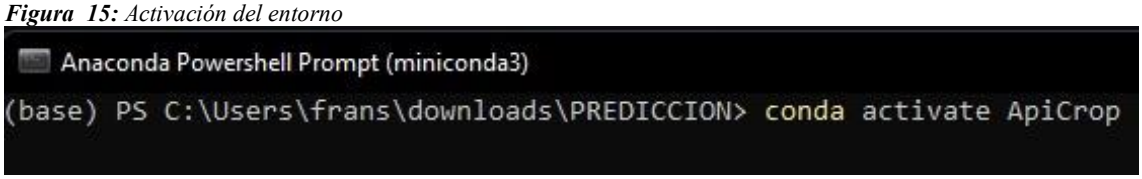
Figura 14: *Creación del entorno virtual.*



Autor: *Stiven Loor*

Una vez creado el entorno virtual con el nombre “ApiCrop”. Se procede a su activación con el comando “*conda actívate ApiCrop”* como se muestra en la figura 14. Luego con el comando “Cd” se ubicó en la ruta de trabajo especifica dentro del entorno virtual, permitiéndonos importar las bibliotecas necesarias sin problemas.

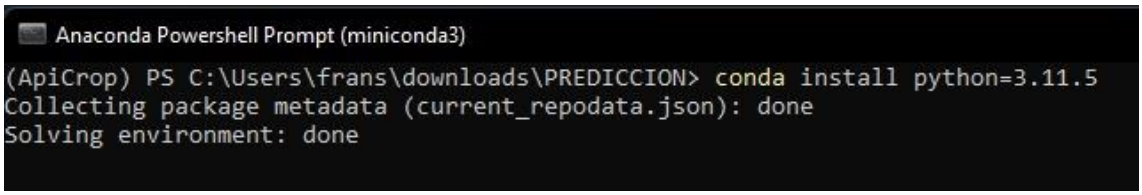
Figura 15: *Activación del entorno*



Autor: *Stiven Loor*

Una vez que se ha configurado el entorno de trabajo, para proceder con la instalación de las bibliotecas necesarias para el desarrollo del programa. Se utilizo el "*conda install <nombre_paquete>*" para incorporar las bibliotecas especificas en nuestro entorno virtual. Es importante tener en cuenta que las bibliotecas descargadas están disponibles únicamente dentro del entorno que hemos creado, lo que garantiza un ambiente de desarrollo aislado y controlado, como se muestra en la figura 15.

Figura 16: *Instalación de los paquetes*



Autor: *Stiven Loor*

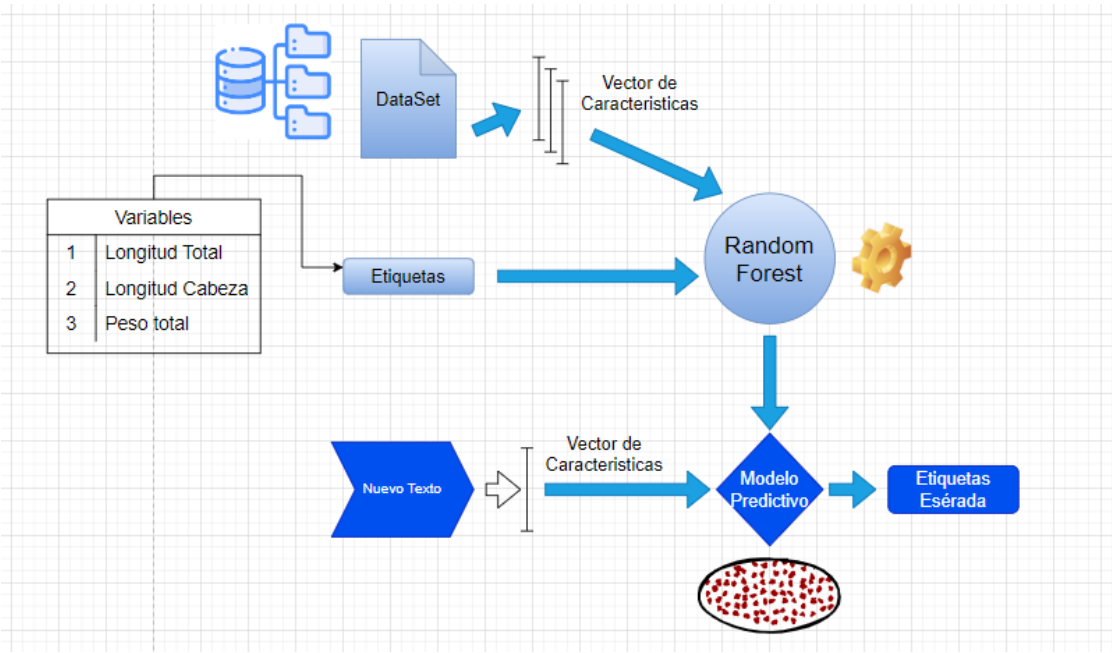
Entre los maquetes principales para el programa se instaló los siguientes:

- **Python (v3.11.4):** Lenguaje de programación principal utilizado en el proyecto.
- **Gunicorn (v21.2.0):** Servidor web HTTP utilizado para ejecutar la aplicación Flask de forma eficiente.
- **Flask (v2.3.3):** Marco de desarrollo web utilizado para crear aplicaciones web.
- **Numpy (v1.25.2):** Biblioteca fundamental para realizar cálculos numéricos y operaciones en matrices y arreglos.
- **Pandas (v2.0.3):** Biblioteca utilizada para la manipulación y análisis de datos estructurados.
- **Scikit-learn (v1.3.0):** Biblioteca de aprendizaje automático (machine learning) que ofrece herramientas para entrenar y evaluar modelos predictivos.
- **Streamlit (v1.26.0):** Marco de desarrollo utilizado para crear una interfaz web interactiva y amigable para los usuarios.

4.1.3.2. Entrenamiento y prueba

Para asegurar la eficiencia del modelo predictivo (figura 17), se implementó una estrategia de separación de datos que involucra el uso del 90% de los datos para el ajuste y entrenamiento del modelo, mientras que el 10% restante se reservó para la validación y prueba del modelo. Esta división de datos permite que el modelo aprenda patrones y relaciones a partir de una amplia porción del conjunto de datos, lo que contribuye a su capacidad predictiva. La fase de entrenamiento implica ajustar el modelo a los datos de entrenamiento, mientras que la fase de validación y prueba se utiliza para evaluar su rendimiento y precisión utilizando datos independientes.

Figura 17: *Estructura del entrenamiento del modelo*



Autor: *Stiven Loor*

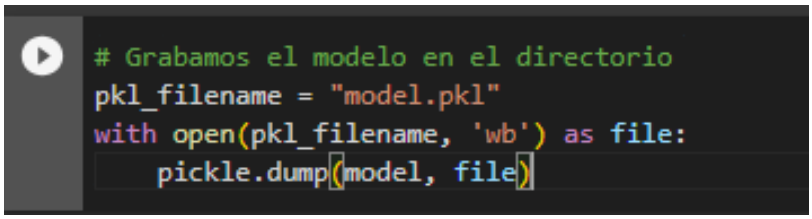
Figura 18: *Entrenamiento*

```
[52] # Separamos datos para ajuste y prueba
     X_train, X_test, y_train, y_test = train_test_split(X, Y, train_size=0.8, test_size=0.2, random_state=100)

[43] from sklearn.ensemble import RandomForestRegressor

     # Creamos el modelo de Random Forest para regresión y entrenamos el modelo
     model = RandomForestRegressor(n_estimators=100, random_state=42)
     model.fit(X_train.values.reshape(-1, 1), y_train)

          RandomForestRegressor
     RandomForestRegressor(random_state=42)
```

Autor: *Stiven Loor*

Luego de crear y entrenar el modelo, se procedió a guardar dicho modelo en el directorio. Para lograrlo, se utilizó un archivo con extensión ".pkl" y el método "pickle.dump". este proceso permitió almacenar todas la características y parámetros aprendidos por el modelo durante el entrenamiento. Garantizando que el modelo esté disponible para su uso posterior, permitiendo realizar predicciones de manera eficiente y sin necesidad de volver a entrenar desde cero. Este paso es crucial para asegurar la reutilización y distribución del modelo de manera conveniente en futuras aplicaciones.

Figura 19: *Grabación del modelo*

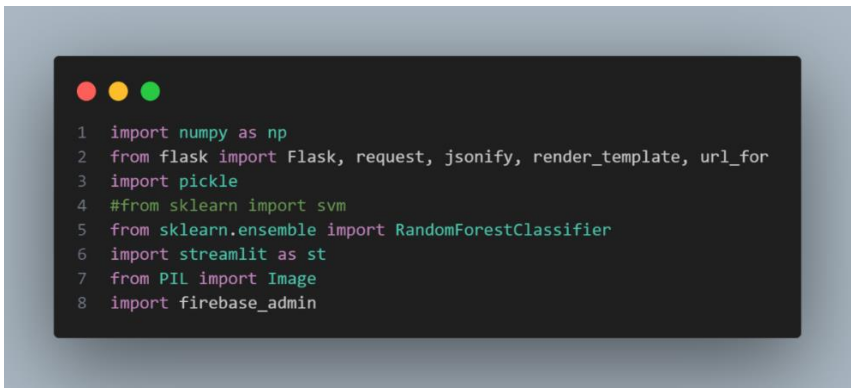


Autor: *Stiven Loor*

4.1.3.3. Aplicación

Se importan varias bibliotecas necesarias para la aplicación, como Numpy para manipulación de datos, Flask para la aplicación web, pickle para cargar el modelo preentrenado, scikit-learn (sklearn) para el algoritmo de Random Forest, Streamlit para la creación de la interfaz de usuario, y Pillow (PIL) para manejar imágenes.

Figura 20: *Importación de librerías*



Autor: *Stiven Loor*

Luego se configura el título y el icono de la página utilizando la función **st.set_page_config ()** de Streamlit. También se carga una imagen de un logotipo para mostrar en la página. El modelo de predicción se carga desde el archivo especificado en **MODEL_PATH** utilizando la biblioteca **pickle**. Se utiliza la función **st. columns** () de Streamlit para dividir la página en tres columnas. La primera y la tercera columna contienen imágenes de logotipos, mientras que la columna central contiene el título y el encabezado de la aplicación.

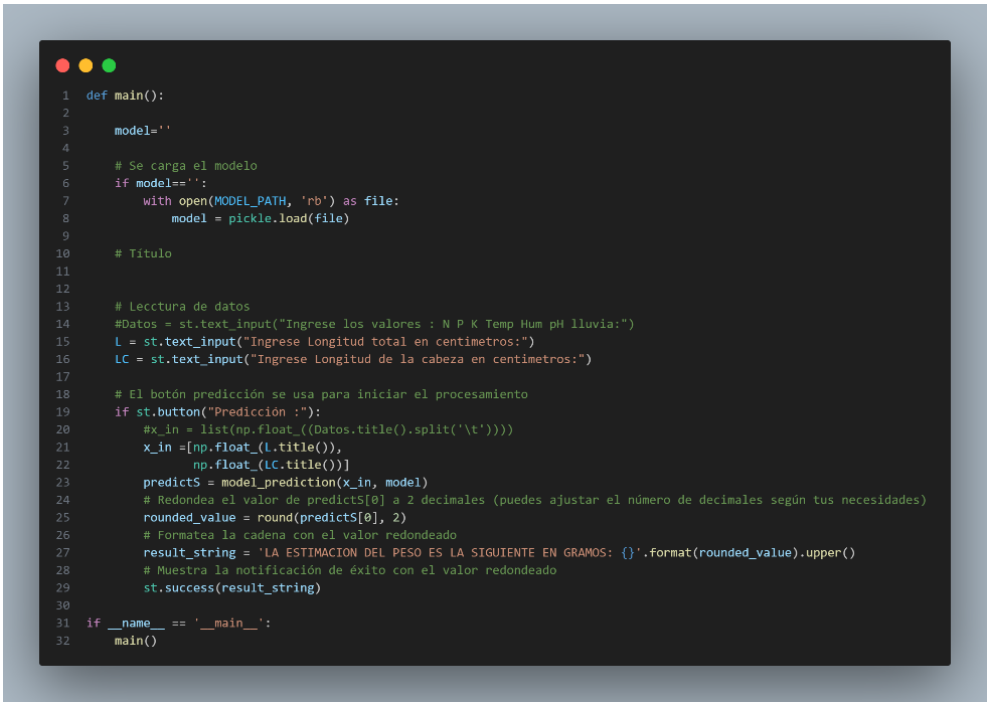
Figura 21: *Diseño de la aplicación*

```
# Path del modelo preentrenado
MODEL_PATH = 'models/model.pkl'


# Se recibe la imagen y el modelo, devuelve la predicción
def model_prediction(x_in, model):

    x = np.asarray(x_in).reshape(1,-1)
    preds=model.predict(x)

    return preds

col1, col2, col3= st.columns((1,5,1))
with col1:
    image = Image.open('imgs/logo2.png')
    st.image(image, width=100)

with col2:
   texto="<table width=""100%"" style=""border:0px""><tr><td style='font-size: 16px; text-align: center;'><b>UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO</b></td></tr>"
   texto+="<tr><td style='font-size: 16px; text-align: center;'><b>FACULTAD DE CIENCIA DE LA INGENIERÍA</b></td></tr>"
   texto+="<tr><td style='font-size: 16px; text-align: center;'><b>INGENIERÍA EN TELEMÁTICA</b></td></tr>"
   texto+="<tr><td style='font-size: 16px; text-align: center;'><b>SISTEMA DE PREDICCION DE PESO PARA PECES NATIVOS BASADO EN PARAMETROS FISIOLOGICOS</b></td></tr></table>"
   st.markdown(texto, unsafe_allow_html=True)


with col3:
    image = Image.open('imgs/logo1.jpg')
    st.image(image, width=100)

st.title(" ")
st.title(" ")
```

Autor: *Stiven Loor*

Se utilizan campos de entrada de texto de Streamlit (st.text_input ()) para que el usuario ingrese los valores necesarios para la predicción, en este caso, la longitud total y la longitud de la cabeza del pez. Cuando el usuario hace clic en el botón "Predicción", se recopilan los valores ingresados, se realiza una predicción utilizando el modelo cargado y se muestra el resultado en la interfaz de usuario. La predicción se redondea a dos decimales antes de mostrarse.

Figura 22: *Predicción del modelo*



```
def main():

    model=''

    # Se carga el modelo
    if model=='':
        with open(MODEL_PATH, 'rb') as file:
            model = pickle.load(file)

    # Título


    # Lecctura de datos
    #Datos = st.text_input("Ingrese los valores : N P K Temp Hum pH lluvia:")
    L = st.text_input("Ingrese Longitud total en centimetros:")
    LC = st.text_input("Ingrese Longitud de la cabeza en centimetros:")

    # El botón predicción se usa para iniciar el procesamiento
    if st.button("Predicción :"):
        #x_in = list(np.float_((Datos.title().split('\t'))))
        x_in =[np.float_(L.title()),
               np.float_(LC.title())]
        predictS = model_prediction(x_in, model)
        # Redondea el valor de predictS[0] a 2 decimales (puedes ajustar el número de decimales según tus necesidades)
        rounded_value = round(predictS[0], 2)
        # Formatea la cadena con el valor redondeado
        result_string = 'LA ESTIMACION DEL PESO ES LA SIGUIENTE EN GRAMOS: {}'.format(rounded_value).upper()
        # Muestra la notificación de éxito con el valor redondeado
        st.success(result_string)

if __name__ == '__main__':
    main()
```

Autor: *Stiven Loor*

Con este código se creó la aplicación web que permite a los usuarios ingresar datos relacionados con peces nativos y obtener una estimación del peso de los peces utilizando un modelo de Random Forest previamente entrenado. La interfaz de usuario se crea utilizando Streamlit, lo que facilita la creación de aplicaciones interactivas con Python.

Figura 23: *Diseño final de la interfaz*

Autor: *Stiven Loor*

La figura 23 presenta la interfaz de la aplicación web, que permite al usuario ingresar los valores en centímetros de las variables Longitud Total y Longitud de la Cabeza para realizar la predicción del peso. Una vez procesados los datos, la aplicación muestra el resultado de la estimación en gramos a través de un cuadro de texto, como se ilustra en la imagen.

Las pruebas y validaciones meticulosas del modelo de predicción desarrollado utilizando datos específicos de parámetros han demostrado su eficacia y precisión en la estimación del peso de los peces nativos. Estos resultados validan la robustez del enfoque propuesto, confirmando que las variables seleccionadas, como la Longitud Total y la Longitud de la Cabeza, son indicadores fiables para prever con precisión el peso de los peces sin necesidad de manipulación física. La consistencia y confiabilidad de las predicciones obtenidas durante las pruebas refuerzan la viabilidad y aplicabilidad practica del modelo en el contexto de la acuicultura.

Estos hallazgos no solo respaldan la utilidad del modelo en la estimación del peso, sino que también allanan el camino para su implementación efectiva en la industria, promoviendo así practica más éticas y precisas en el manejo de los peces nativos en las granjas acuícolas.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones.

- En conclusión, la creación y aplicación de un modelo predictivo para estimar el peso de los peces a través de parámetros fisiológicos representan un paso significativo hacia prácticas más éticas y sostenibles en la acuicultura. Este enfoque no solo minimiza la necesidad de manipulación directa de los peces, preservando así su bienestar y desarrollo natural, sino que también contribuye a la mejora de las técnicas de cría al proporcionar estimaciones precisas del peso.
- Tras un análisis meticuloso, se encontró que las variables claves son la Longitud Total y la Longitud de la Cabeza, fueron identificadas como las óptimas para prever el peso de los peces nativos. La rigurosa revisión de literatura y los análisis estadísticos detallados respaldaron esta selección, confirmando que estas variables son indicadores precisos del peso de los peces. La meticulosidad en la elección de estas variables garantiza que el modelo se base en parámetros fisiológicos confiables y relevantes, lo que a su vez asegura la precisión y confiabilidad de las predicciones realizadas.
- Se llevo a cabo un minucioso estudio comparativo que implico evaluar varios modelos de predicción, incluyendo regresión lineal, regresión polinómica, K-Nearest Neighbors (KNN), Support Vector Regression (SVR) y Random Forest. Los resultados de este análisis revelaron que el Random Forest ha demostrado consistentemente ser el más preciso y confiable para estimar el peso de los peces nativos basándose en parámetros fisiológicos como la Longitud Total y la Longitud de la Cabeza. A través de numerosas pruebas y validaciones, se observó que Random Forest no solo proporciono predicciones más cercanas a los valores reales, sino que también exhibió una capacidad destacada para manejar datos complejos y ruidosos, lo que lo convierte en una elección óptima para aplicaciones en el campo de la acuicultura.
- Los resultados obtenidos a través de las pruebas y validación del modelo destacan no solo su precisión sino también su potencial para transformar las prácticas de estimación del peso en la industria acuícola, promoviendo así un manejo más eficiente, ético y preciso de los peces nativos en su entorno natural, estos logros respaldan la relevancia y el impacto significativo de este estudio en el campo de la acuicultura.

5.2. Recomendaciones

Estas son algunas de las recomendaciones que se ostentan en base a los fundamentos de los resultados del proyecto

- Recomendaría explorar el procesamiento de imágenes para mejorar la precisión en la estimación del peso de los peces nativos, mediante el uso de redes neuronales convolucionales. El objetivo es enriquecer el modelo predictivo con información visual adicional.
- Se sugiere emplear técnicas de recolección de datos no invasivas que reduzcan al mínimo la manipulación y el estrés en los peces nativos. Estas técnicas pueden abarcar el uso de dispositivos de registro de datos en el entorno natural de los peces, así como la aplicación de métodos de muestreo que eviten la captura y manipulación excesiva de los individuos.
- Se recomienda mejorar la interfaz mediante la integración con una plataforma como Firebase u otro framework, que posibilite el almacenamiento de los datos ingresados. Además, se podría añadir parámetros adicionales, como la especie nativa, la fecha de la medición y el sexo de los peces. Esta expansión permitiría abordar la escasez de datos en el campo acuícola, mejorando así la robustez y la aplicabilidad del modelo predictivo desarrollado en este estudio.
- Se sugiere promover una colaboración estrecha entre científicos, investigadores y pescadores locales con el objetivo de recopilar datos fisiológicos y de peso de los peces nativos. Esta colaboración no solo fortalecerá la investigación en este campo, sino que también fomentará la participación de las comunidades locales en la conservación de los recursos acuáticos.

Bibliografía

[1] D. O. M. Margarita, «Propuesta de un modelo predictivo numérico que permita mejorar la rentabilidad de la tilapia roja(Trabajo de grado),» 2020.

[2] V. L. Kratochvil, «Técnicas de predicción de edad y crecimiento a partir de la morfometría y el peso del otolito de prochilodus lineatus (valenciennes 1837),» 2023.

[3] J. G. G. M. A. J. N. O. P. P. &. E. H. Roberto Cifuentes, «Correlación entre la longitud y el peso, así como el índice de condición de los peces autóctonos del río San Pedro (cuenca del río Valdivia, Chile),» Gayana Especial: 101-110,, 2012.

[4] G. M. H. V. S. A. J. A. Jorge T. Corea, «Distribución y abundancia de peces de la familia Loricariidae (Pleco) y su relación con los peces de interés comercial en los alrededores de la Isla de Ometepe,» Universidad Nacional Autónoma de Nicaragua, León (UNAN-León), 2014.

[5] R. I. Meza López, «Aplicación y Análisis de un Sistema Experto Basado en Lógica Difusa para la Evaluación del Hábitat de Peces Nativos en el Río Huequecura,» UNIVERSIDAD DE CHILE , 2010.

[6] F. M.-C. R. S. C. &. J. D. A.-H. Esther Julia Olaya Marín, «Modelación de la riqueza de peces nativos para simular medidas de mitigación en tramos de ríos alterados,» Institut d'Investigació per a la Gestió Integrada de Zones Costaneres, Universitat Politècnica de València, C/ Paranimf, 1, , 2011.

[7] J. C. D. Beleña, «Aplicacion de algoritmos Boosting a la prediccion de precios de energia electrica,» jun 2021. [En línea]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://oa.upm.es/67545/1/TFG_JUAN_CARLOS_DURANTE_BELENA.pdf.

[8] O. Soft, «¿Qué son los algoritmos predictivos de planificación de procesos?,» [En línea]. Available: https://www.onyxerp.com/blog/algoritmos-predictivos#:~:text=En%20otras%20palabras%2C%20un%20algoritmo,en%20base%20al%20conocimiento%20presente.

[9] C.-T. J.-P. I.-C. J.-T. EdgarGarcıa-Quezada, «Hacia la prediccion de pacientes con diabetes: Comparacion de algoritmos de aprendizaje automatico,» *Universidad Autonoma de Zacatezas,* 2023.

[10] G. F. P. Eliseo, «Aplicacion de algoritmos de Machine Learning para la prediccion del beneficio por cliente a partir de metricas de Google Analytics,» 2019.

[11] D. H. RAMÍREZ, «EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD,» 2018.

[12] I. Á. Fidalgo, «Introducción al Machine Learning,» 2021.

[13] D. A. M. Cuadrado, «UTILIZACIÓN DEL MACHINE LEARNING,» Valladolid, 2019.

[14] J. Casas Roma, T. Lozano Bagén y A. Bosch Rué, «Deep learning,» Universitat Oberta de Catalunya, Barcelona , España, 2020.

[15] J. M. R. Rama, «APLICACIÓN DE TÉCNICAS DE MACHINE LEARNING A LA DETECCIÓN DE ATAQUES,» 2018.

[16] P. F. H. Leal, «Algoritmo de Aprendizaje para Redes,» ©INAOE, 2011.

[17] U. P. d. Madrid, «Algoritmos de Machine Learning».

[18 H. Ordóñez, C. Cobos y V. Bucheli, «Modelo de machine learning para la predicción de las tendencias de hurto en Colombia,» Colombia, 2020.

[19] W. Y. C. M. L. M. S. M. Gabriel Elías Chanchí Golondrino, «Aplicación de la regresión polinomial para la caracterización de la curva del COVID-19, mediante técnicas de machine learning,» © Copyright: Investigación e Innovación en Ingenierías, 2020.

[20] D. &. C.-L. F. Lévano-Rodriguez, «Discriminación de masas mamográficas mediante K-Nearest Neighbor y atributos BIRADS,» Revista científica de sistemas einformática, 2(1),e225, 2021.

[21] J. J. Martín Guareño, «Support vector regression : propiedades y aplicaciones (Trabajo fin de grado inédito),» Universidad de Sevilla, Sevilla., Sevilla, 2016.

[22] D. Ivanova, «Support Vector Regression desde una perspectiva multiobjetivo: Una aplicación a la esperanza de vida,» Universidad Miguel Hernández de Elche, 2020.

[23] R. S. B. C. M. M. M. B.-T. Darwin Patiño Pérez, «Prediction of Covid19 with the use of Random Forests Algorithm and Artificial Neural Networks,» Artificiales.Ecuadorian ScienceJournal.4(2), 101-110, 2020.

[24] J. C. A. E. A. U. Jenny Arizaga Gamboa, «Detección de Ataque de DDoS utilizando Machine Learning – algoritmo de Random Forest,» Serie Científica de la Universidad de las Ciencias Informáticas, ISSN-e 2306-2495, Vol. 15, Nº. 3, 2022, 2022.

[25] P. C. I. J. P. C. Mariano Romero Martínez, «La predicción del fracaso empresarial de las cooperativas españolas. Aplicación del Algoritmo Extreme Gradient Boosting,» CIRIEC-España, Revista de Economía Pública, Social y CooperativaNº101/2021, pp. 255-288, 2020.

[26] J. G. H. José Manuel Molina López, «TÉCNICAS DE ANÁLISIS DE DATOS,» 2006.

[27] J. Riquelme, «Minería de Datos: Conceptos y Tendencias,» revista@aepia.org, Valencia, España, 2006.

[28] D. H. Ramírez, «EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD,» UNIVERSIDAD LIBRE SECCIONAL PEREIRA, PEREIRA, 2018.

[29] O. Kramer, «Scikit-Aprende. En: Aprendizaje automático para estrategias de evolución.,» Estudios en Big Data, vol 20. Springer, Cham, 2016.

[30] E. Bisong, «Aprende Machine Learning con Scikit-Learn, Keras y TensorFlow,» E Bisong, 2019.

[31] G. Hackeling, Mastering Machine Learning with scikit-learn, Copytight 2017 Packt Publishing, 2017.

[32] G. A. V. B. FabianPedregosa, «Scikit-learn:MachineLearninginPython,» Neurospin,Bˆat145,CEASaclay 91191GifsurYvette–France, 2011.

[33] W. McKinney, «pandas: a Foundational Python Library for Data Analysis and Statistics,» 2011.

[34] F. H. C. C. M. G. S. S. Alexander Gelbukh, «Detección de humor en tweets en español utilizando clasificadores de Scikit-learn,» Instituto Politécnico Nacional, Centro de Investigación en Computación, México , 2020.

[35] A. Izquierdo-Carrasco, «Análisis de datos en comunidades virtuales:# claustrovirtual en Twitter (Bachelor's thesis).,» 2022.

[36] E. Grande Ruiz, «Diseño e implementación de un sistema software que facilite la utilización de sensores usando ontologías (Tesis de grado),» Universidad de Alicante. Departamento de Lenguajes y Sistemas Informáticos, 2022.

[37] H. D. P. Barrezueta, «LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES,» ASAMBLEA NACIONAL, 2021.

[38] D. J. Drouet Racines, «Estudio mediante morfometría de la caracterización biológica de Andinocara rivolatus (vieja azul) en zonas de influencia del río Quevedo, considerando la variabilidad en crianza con fines alimentarios,» UTEQ. 123 p, Quevedo, 2019.

[39] E. Pulido, «PRINCIPALES CAUSAS DE MORTALIDAD EN CULTIVOS INTENSIVOS Y,» 2012.

[40] G. A. y. P. (. Ministerio de Agricultuta, «Más familias se suman a la producción acuícola en la Zona 3,» 2021. [En línea]. Available: https://www.agricultura.gob.ec/mas-familias-se-suman-a-la-produccion-acuicola-en-la-zona-3/.

[41] D. I. Candia Oviedo, «Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático,» 2019.

[42] F. (. d. I. A. y. F. U. d. V. C. L. Y. 3. P. E. Diez, L. Navas-Gracia, A. Martínez-Rodríguez, A. Correa-Guimaraes y L. Chico-Santamarta, «Predicción de la irradiación solar global diaria horizontal mediante redes neuronales artificiales en la región de Castilla y León, España,» 2019.

[43] J. P. Vásquez Arenas, «Beneficios de los modelos basados en algoritmos de machine learning para la predicción de concentraciones de PM2.5 en el Valle de Aburrá,» 2022.

[44] J. G. López Martínez, «Técnicas de Machine Learning para la predicción de desempeño académico en el desarrollo del espacio proyectivo del pensamiento espacial.,» 2019. [En línea]. Available: http://hdl.handle.net/20.500.12209/11451.

[45] I. Chardon Schirm, «Predicción de rentas en Santiago de Chile utilizando algoritmos de aprendizaje automátic,» 2019. [En línea]. Available: https://repositorio.uchile.cl/handle/2250/170883.

[46] A. T. Norman, Aprendizaje Automático En Acción, 2019.

[47] I. y. P. Ministerio de Produccion Comercuo Exterior, «"MEJORA EN LA COMPETITIVIDAD DEL SECTOR ACUÍCOLA Y PESQUERO",» Feb 2020. [En línea]. Available: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.produccion.gob.ec/wp-content/uploads/downloads/2020/11/Proyecto-Mejora-Competitiva-del-Sector-Acu%C3%ADcola-y-Pesquero.pdf.

[48] N. D. G. R. Andrés Rico Páez, «Modelos predictivos del rendimiento académico a partir de características de estudiantes de ingeniería,» 2022. [En línea]. Available: https://doi.org/10.33010/ie_rie_rediech.v13i0.1426.

[49] TIBCO, «www.tibco.com,» [En línea]. Available: https://www.tibco.com/es/reference-center/what-is-supervised-learning.

[50] J. L. Gonzalez, «Tipos de aprendizaje automático,» SoldAI, 8 Feb 2018. [En línea]. Available: https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2.

51] J. G. H. José Manuel Molina López, «TÉCNICAS DE ANÁLISIS DE DATOS,» 2006.

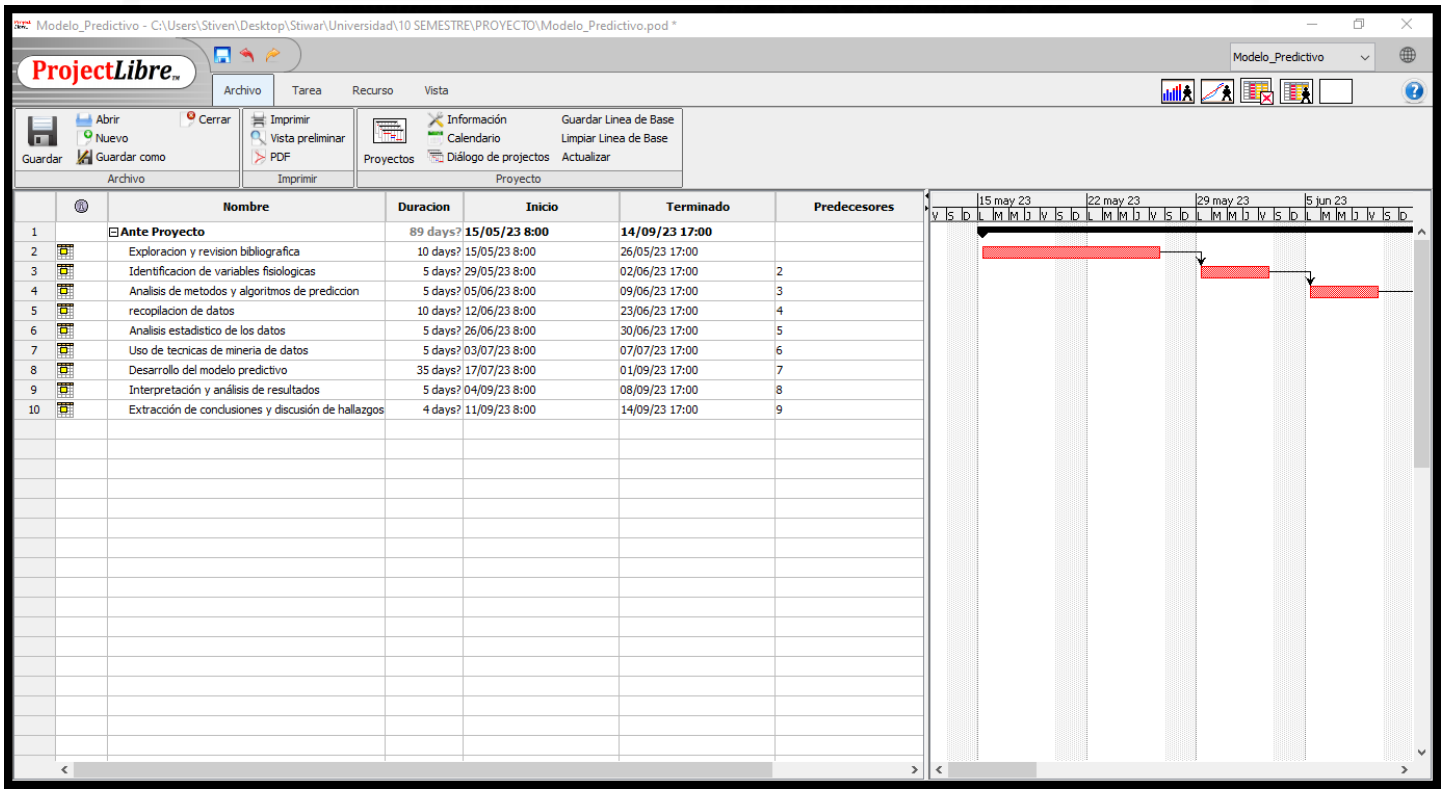
ANEXOS

Ouriginal

Document Information

Analyzed document	Tesis de Loor.pdf (D177748800)
Submitted	11/3/2023 7:15:00 PM
Submitted by	
Submitter email	stiven.loor2018@uteq.edu.ec
Similarity	1%
Analysis address	ezhuma.uteq@analysis.urkund.com

Ilustración 1. Certificado de Similitud



Modelo_Predictivo - C:\Users\Stiven\Desktop\Stiwar\Universidad\10 SEMESTRE\PROYECTO\Modelo_Predictivo.pod *

ProjectLibre

Archivo | Tarea | Recurso | Vista

Modelo_Predictivo

Guardar | Abrir | Nuevo | Guardar como | Cerrar | Archivo

Imprimir | Vista preliminar | PDF | Imprimir

Proyectos | Información | Calendario | Diálogo de proyectos | Guardar Linea de Base | Limpiar Linea de Base | Actualizar | Proyecto

	Nombre	Duracion	Inicio	Terminado	Predecesores
1	Ante Proyecto	89 days?	15/05/23 8:00	14/09/23 17:00	
2	Exploracion y revision bibliografica	10 days?	15/05/23 8:00	26/05/23 17:00	
3	Identificacion de variables fisiologicas	5 days?	29/05/23 8:00	02/06/23 17:00	2
4	Analisis de metodos y algoritmos de prediccion	5 days?	05/06/23 8:00	09/06/23 17:00	3
5	recopilacion de datos	10 days?	12/06/23 8:00	23/06/23 17:00	4
6	Analisis estadistico de los datos	5 days?	26/06/23 8:00	30/06/23 17:00	5
7	Uso de tecnicas de mineria de datos	5 days?	03/07/23 8:00	07/07/23 17:00	6
8	Desarrollo del modelo predictivo	35 days?	17/07/23 8:00	01/09/23 17:00	7
9	Interpretación y análisis de resultados	5 days?	04/09/23 8:00	08/09/23 17:00	8
10	Extracción de conclusiones y discusión de hallazgos	4 days?	11/09/23 8:00	14/09/23 17:00	9

15 may 23 | 22 may 23 | 29 may 23 | 5 jun 23

Ilustración 2. Cronograma de actividades

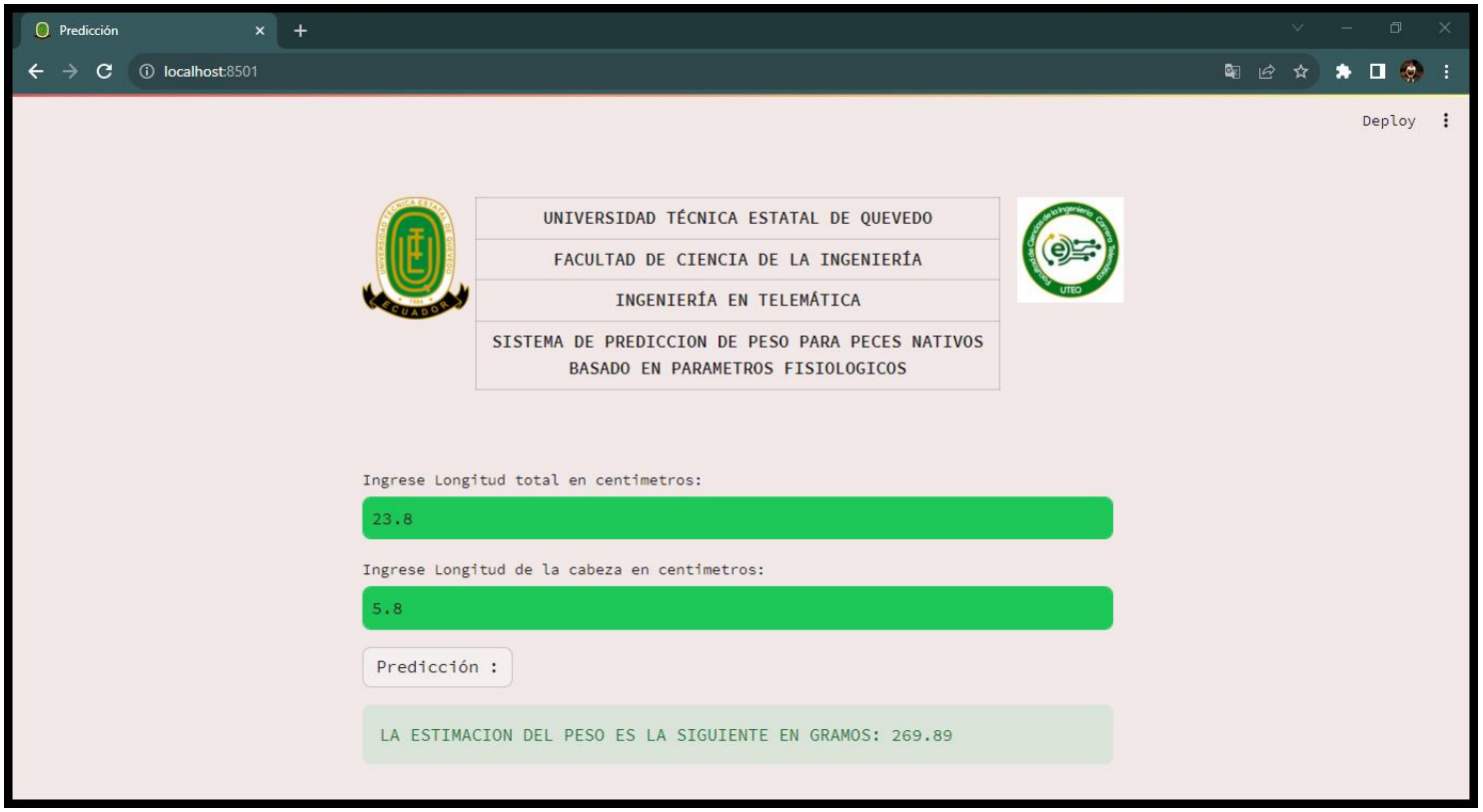


Ilustración 3. Interfaz de la aplicación