

Benchmarking Machine Learning Models for Predicting the Average Weight of Oncorhynchus Mykiss in High Andean Fish Farming

Patricia Uceda, Dr.¹ ; Manuel Torrel, Mg.¹ ; Santos Chavez¹ ; Reyles Ruiz¹ ; Samir Sanchez¹ 

¹ Universidad Privada del Norte, Perú, patricia.uced@upn.edu.pe, manuel.torrel@upn.pe, N00219167@upn.pe, N00339788@upn.pe, N00316769@upn.pe

Abstract– *The present research compares Machine Learning models to predict growth (average final weight) and support profitability decisions in a high Andean fish farm. Linear and Random Forest Regression were evaluated using a set of 500 weekly records, integrating historical data provided by the fish farm and synthetic data generated within validated physicochemical ranges. The results concluded with a high performance in the growth prediction ($R^2 > 0.98$) and adequate performance in profitability ($R^2 \approx 0.87$). In addition, the profitability classifier achieved 0.93 accuracy, suggesting operational utility as an early warning.*

Keywords– *aquaculture, rainbow trout, prediction, machine learning, profitability.*

Evaluación Comparativa de Modelos de Machine Learning para la Predicción del Peso Promedio de *Oncorhynchus Mykiss* en Piscicultura Altoandina

Patricia Uceda, Dr.¹; Manuel Torrel, Mg.¹; Santos Chavez¹; Reyles Ruiz¹; Samir Sanchez¹

¹ Universidad Privada del Norte, Perú, patricia.uced@upn.edu.pe, manuel.torrel@upn.pe, N00219167@upn.pe, N00339788@upn.pe, N00316769@upn.pe

Resumen– La presente investigación compara modelos de Machine Learning para predecir el crecimiento (peso final promedio) y apoyar decisiones de rentabilidad en una piscigranja altoandina. Se evaluaron Regresión Lineal y Random Forest utilizando un conjunto de 500 registros semanales, integrando datos históricos proporcionados por la piscigranja y datos sintéticos generados dentro de rangos fisicoquímicos validados. Los resultados concluyeron con un alto desempeño en la predicción de crecimiento ($R^2 > 0.98$) y desempeño adecuado en rentabilidad ($R^2 \approx 0.87$). Además, el clasificador de rentabilidad alcanzó 0.93 de precisión, lo que sugiere utilidad operativa como alerta temprana.

Palabras clave– acuicultura, trucha arcoíris, predicción, machine learning, rentabilidad.

I. INTRODUCCIÓN

La producción mundial de acuicultura superó los 130 millones de toneladas en 2022, representando más del 50% del pescado destinado al consumo humano, según reportes recientes [1] [2]. En sistemas intensivos y semi intensivos de cultivo de truchas, la alimentación representa entre el 50% y 70% de los costos operativos, por lo que decisiones ineficientes generan pérdidas económicas y riesgos ambientales.

En piscigranjas altoandinas del Perú, el monitoreo del crecimiento depende todavía en gran medida de observación manual y muestreos periódicos. Sin embargo, la literatura indica que la alimentación automatizada basada en señales de comportamiento y sensores es crítica para evitar excesos o déficit en la alimentación [3], [4], [5]. La sobrealimentación no sólo incrementa costos, sino que deteriora la calidad del agua, afectando oxígeno disuelto y estabilidad productiva.

Diversos estudios han demostrado que modelos predictivos de consumo y crecimiento basados en oxígeno disuelto, temperatura y peso promedio presentan alta correlación con el desempeño productivo [6]. En países con alto nivel de automatización acuícola como Noruega, el uso de inteligencia artificial y digital twins está transformando la toma de decisiones operativas [1] [2].

En Cajamarca, la piscigranja “El Valle”, ubicada en La Encañada, enfrenta variaciones constantes de temperatura y pH, además de recursos tecnológicos limitados, lo que complica el manejo manual y afecta su eficiencia. Por ello, se

vuelve necesario implementar un modelo automatizado que ayude a monitorear parámetros en tiempo real, predecir el crecimiento de las truchas y optimizar el uso de recursos, mejorando así la productividad y sostenibilidad del cultivo.

A nivel de antecedentes, en los últimos años, la predicción de crecimiento en acuicultura ha evolucionado hacia modelos supervisados y enfoques ensemble, donde se reportan mejoras frente a modelos lineales tradicionales [7]. En paralelo, se han propuesto sistemas de estimación de biomasa mediante visión computacional; por ejemplo, Zhang et al. plantearon enfoques multimodales que combinaron visión y datos de sensores, logrando mejoras en la estimación de biomasa y mayor robustez en escenarios reales [5].

En cuanto a automatización, la literatura describe experiencias en países con alta tecnificación, como Noruega, donde ya existen sistemas que regulan la alimentación según señales de comportamiento, reduciendo desperdicio y mejorando eficiencia [8]. Sin embargo, en muchas piscigranjas de América Latina todavía prevalecen métodos manuales, lo que dificulta el control de parámetros críticos del agua y puede derivar en estrés, mortalidad y pérdidas económicas. En el contexto peruano, se han reportado experiencias de Machine Learning para predecir crecimiento de truchas en regiones como Junín con buenos resultados, aunque su adopción en piscigranjas altoandinas sigue siendo limitada [9].

A nivel internacional, también se han desarrollado modelos para predecir tamaño y biomasa de trucha arcoíris con impacto directo en prácticas de alimentación, reportando reducción de error en planificación productiva y potencial ahorro de alimento y costos operativos [10]. De manera complementaria, revisiones sistémicas resaltan que la detección automática del comportamiento alimenticio mediante visión y aprendizaje profundo ayuda a disminuir desperdicio y mejorar eficiencia [3] - [5].

Otras áreas donde los modelos predictivos van tomando importancia, son los sistemas de recirculación (RAS), los cuales han permitido anticipar mortalidad y optimizar parámetros críticos en acuicultura [11]. Asimismo, la integración de modelos predictivos dentro de arquitecturas de digital twin representa una tendencia emergente para

acuicultura inteligente [12], [13]; pero aún no es masiva en acuíferos peruanos, a diferencia de escenarios de la Unión Europea.

Finalmente, soluciones industriales como Aquabyte han implementado estimación de biomasa y bienestar mediante análisis de imágenes y datos recolectados en piscigranjas, con el objetivo de optimizar alimentación y reducir costos [14].

En investigaciones relacionadas con monitoreo visual, se mostraron métricas cuantitativas para detección y segmentación (por ejemplo YOLOv5 y U-Net) con desempeños consistentes en tareas de seguimiento y conteo [15]. Tal es así que, estos antecedentes justifican la necesidad de propuestas reproducibles y adaptables a contextos altoandinos, donde la variabilidad ambiental y la limitada infraestructura tecnológica siguen siendo barreras para la adopción.

Ante esta situación, se propone desarrollar un modelo automatizado que permita monitorear en tiempo real la calidad del agua y predecir el crecimiento de las truchas mediante técnicas de Machine Learning, con el fin de optimizar la alimentación, mejorar la toma de decisiones y fortalecer la productividad y sostenibilidad de la piscigranja.

Resumiendo, la presente investigación busca contribuir con:

1. Un análisis aplicado en contexto altoandino peruano con variables operativas reales.
2. Comparación reproducible de modelos supervisados (Regresión Lineal vs Random Forest).
3. Validación temporal para evitar sobreestimación por fuga de información.
4. Interpretabilidad mediante análisis de importancia de variables.

A continuación, se describen los objetivos que se persiguen:

Objetivo general:

Diseñar y evaluar un modelo predictivo (en Google Colab) que estime el crecimiento y apoye decisiones de rentabilidad en una piscigranja altoandina, usando variables operativas accesibles.

Objetivos específicos:

1. Comparar Regresión Lineal y Random Forest para predecir el peso final promedio (crecimiento).
2. Comparar ambos modelos para estimar la rentabilidad del lote.
3. Clasificar si un lote será rentable o no, usando Random Forest Classifier.

4. Identificar qué variables influyen más en crecimiento y rentabilidad (importancia de variables) y analizar coherencia con la correlación entre variables.

II. METODOLOGÍA

A. Diseño

El presente estudio adopta un diseño cuantitativo, aplicado y comparativo, orientado a evaluar el desempeño predictivo de modelos supervisados para estimar el peso promedio de *Oncorhynchus mykiss*.

El enfoque es no experimental y longitudinal retrospectivo, dado que se analizan registros históricos semanales sin manipulación de variables independientes.

El componente comparativo se basa en la evaluación de dos algoritmos de regresión (Regresión Lineal y Random Forest), contrastando su desempeño mediante métricas estadísticas bajo un protocolo de validación temporal.

B. Contexto del estudio y conjunto de datos

El estudio se desarrolló en la piscigranja “El Valle”, ubicada en el distrito de La Encañada, Cajamarca (Perú), enfocándose en el cultivo de trucha arcoíris (*Oncorhynchus mykiss*) en etapa de crecimiento. Este entorno productivo presenta condiciones ambientales propias de zonas altoandinas, caracterizadas por variaciones de temperatura y disponibilidad limitada de infraestructura tecnológica, lo que hace relevante la aplicación de herramientas predictivas para mejorar la toma de decisiones operativas dentro del sistema productivo.

Se construyó un dataset compuesto por 500 registros semanales, integrando datos históricos reales proporcionados por la piscigranja y datos sintéticos generados para complementar la información disponible. La incorporación de datos sintéticos tuvo como propósito ampliar la variabilidad del conjunto de datos y permitir una mejor exploración del comportamiento de los modelos predictivos en distintos escenarios productivos posibles, práctica que ha sido reportada como válida cuando existe disponibilidad limitada de datos reales [17], [18].

Del total de registros considerados, aproximadamente 320 correspondieron a datos históricos reales los cuales fueron utilizados para tener parámetros en la generación de los 500 datos sintéticos. La generación de datos sintéticos se realizó con el objetivo de representar escenarios plausibles dentro de rangos fisicoquímicos conocidos y evitar la presencia de valores fuera de condiciones productivas realistas.

La generación de datos sintéticos se realizó siguiendo principios de simulación estadística basada en distribuciones normales, técnica empleada en procesos de aumento de datos (*data augmentation*) y validación experimental en modelos predictivos [19], [20]. Este procedimiento permitió preservar las características estadísticas esenciales del conjunto original y evitar la introducción de patrones artificiales inconsistentes.

Los datos sintéticos fueron generados utilizando rangos fisicoquímicos validados en el cultivo de trucha arcoíris, considerando parámetros críticos reportados en la literatura científica especializada. En particular, se consideraron los siguientes rangos productivos: temperatura del agua entre 8 y 14 °C, pH entre 6.8 y 8.0, y oxígeno disuelto entre 5 y 9 mg/L. Asimismo, se aplicaron perturbaciones controladas sobre variables productivas, tales como la cantidad de alimento suministrado y el peso promedio semanal, con el fin de preservar las relaciones estadísticas observadas en los datos históricos y mantener la coherencia biológica del sistema productivo [21], [22].

Las variables consideradas en el estudio incluyeron temperatura del agua (°C), pH, oxígeno disuelto (mg/L), cantidad de alimento suministrado (g) y peso promedio semanal (g), siendo esta última definida como la variable objetivo del modelo predictivo. Estas variables fueron seleccionadas por su disponibilidad operativa en el contexto productivo y por su relevancia documentada en estudios relacionados con el crecimiento de peces en sistemas acuícolas.

Finalmente, se reconoce que la incorporación de datos sintéticos representa una estrategia metodológica válida cuando existe limitada disponibilidad de datos reales, siempre que se mantenga coherencia estadística con el conjunto original [17], [18]. No obstante, los resultados obtenidos deben interpretarse considerando esta característica del dataset, recomendándose que futuros estudios incorporen conjuntos de datos completamente reales provenientes de múltiples ciclos productivos para mejorar la robustez y capacidad de generalización del modelo predictivo.

C. Proceso metodológico (CRISP-DM)

El flujo experimental se organizó con CRISP-DM para asegurar trazabilidad del trabajo: comprensión del problema productivo, comprensión y preparación de datos, modelado, evaluación e implementación.

El desarrollo y la ejecución se realizaron en Google Colab para garantizar reproducibilidad del pipeline.

D. Preparación y validación de datos

El preprocesamiento incluyó eliminación de registros duplicados y corrección de inconsistencias, tratamiento de

valores faltantes mediante interpolación lineal, detección de valores atípicos utilizando el rango intercuartílico (IQR) y estandarización de variables numéricas.

Se generaron variables derivadas como variación semanal del peso, promedio móvil de temperatura e índice alimento/peso (Feed Ratio).

E. Modelado predictivo

Se evaluaron dos algoritmos de aprendizaje supervisado para regresión: (1) Regresión Lineal con regularización Ridge ($\alpha = 1.0$), seleccionada por su interpretabilidad; (2) Random Forest Regressor, elegido por su capacidad para capturar relaciones no lineales y reducir sensibilidad al ruido.

El dataset se dividió utilizando particiones temporales (80% entrenamiento, 20% prueba), evitando fuga de información.

Los hiperparámetros fueron optimizados mediante Grid Search, utilizando como criterio la minimización del RMSE y la maximización del R^2 .

Para evaluar estabilidad y generalización, se aplicó validación cruzada k-fold ($k = 5$), reportando media y desviación estándar.

F. Evaluación del desempeño e interpretabilidad

El desempeño se evaluó mediante MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) y R^2 (Coeficiente de determinación).

Se compararon los resultados estimados por el modelo con los valores reales observados, analizando las diferencias para evaluar la precisión de las predicciones.

Para facilitar la adopción en contexto productivo, se aplicaron técnicas de interpretabilidad basadas en importancia de variables, alineadas con enfoques contemporáneos de Explainable Artificial Intelligence (XAI) [16].

G. Implementación

La implementación final fue desarrollada en Google Colab utilizando Python y Scikit-Learn.

El flujo experimental permite cargar nuevos registros en formato CSV, aplicar automáticamente preprocesamiento, ejecutar el modelo entrenado y generar predicciones del peso esperado junto con visualizaciones comparativas.

De esta manera, el modelo puede utilizarse y volver a ejecutarse sin necesidad de contar con equipos especializados.

III. RESULTADOS

Para responder al primer objetivo específico, orientado a evaluar el desempeño de modelos supervisados para estimar el peso final promedio por lote, se entrenaron y compararon dos algoritmos de regresión: Regresión Lineal y Random Forest. El desempeño fue evaluado mediante el error cuadrático medio (MSE) y el coeficiente de determinación (R^2) sobre el conjunto de prueba.

TABLA I
DESEMPEÑO PREDICTIVO PARA EL MODELO DE CRECIMIENTO

Modelo	MSE	R^2
Regresión Lineal	99.99	0.993
Random Forest	171.82	0.988

Los resultados de la Tabla I muestran que ambos modelos explican más del 98% de la variabilidad del peso final, lo que evidencia una alta capacidad predictiva.

No obstante, la Regresión Lineal obtuvo un menor MSE y un R^2 ligeramente superior, lo que indica que, en el conjunto de datos analizado, el crecimiento responde predominantemente a relaciones lineales, particularmente asociadas al alimento total suministrado.

Las Fig. 1 y 2 muestran la comparación entre valores reales y esperados.

```
# Rentabilidad final
data["rentabilidad_lote"] = data["ingresos"] - data["costo_operativo"]

# Clasificación = 1 si el lote fue rentable
data["es_rentable"] = (data["rentabilidad_lote"] > 0).astype(int)

# Vista previa
display(data.head())
display(data.describe())
```

lote_id	fecha_inicio	fecha_cosecha	temperatura_prom	oxigeno_prom	ph_prom	peces_semrados	peso_inicial_prom	alimento_tot
0	1	2023-01-01	2023-06-21	8.554890	9.730577	6.975087	782	13.207201
1	2	2023-01-08	2023-05-22	8.806224	6.866547	8.442790	4178	24.190852
2	3	2023-01-15	2023-07-25	8.145775	6.353723	6.861954	1986	47.669028
3	4	2023-01-22	2023-07-21	8.755544	8.219998	8.208770	4582	36.742906
4	5	2023-01-29	2023-06-18	13.464054	7.043671	7.484556	4683	37.146798

lote_id	fecha_inicio	fecha_cosecha	temperatura_prom	oxigeno_prom	ph_prom	peces_semrados	peso_inicial_prom
count	500.000000	500	500	500.000000	500.000000	500.000000	500.000000
mean	250.500000	2027-10-13 12:00:00	2028-03-24 19:26:24	11.977933	7.518639	7.488818	2640.404000
min	1.000000	2023-01-01 00:00:00	2023-05-22 00:00:00	8.037056	5.031929	6.509880	507.000000
25%	125.750000	2025-05-23 06:00:00	2025-11-06 18:00:00	9.851779	6.163188	6.975088	1583.750000
50%	250.500000	2027-10-13 12:00:00	2028-03-20 12:00:00	11.971152	7.611520	7.513377	2582.000000
75%	375.250000	2030-03-04 18:00:00	2030-08-12 00:00:00	14.037532	8.808362	7.981710	3638.000000
max	500.000000	2032-07-25 00:00:00	2033-01-13 00:00:00	15.997741	9.898671	8.488827	4999.000000
std	144.481833	NaN	NaN	2.354085	1.486041	0.581893	1238.245577

Fig. 1 Comparación Real vs Esperada - Regresión Lineal (Crecimiento)

En ambas figuras se observa una fuerte alineación respecto a la línea ideal, con baja dispersión de puntos. Esto confirma que el modelo captura adecuadamente la dinámica del crecimiento bajo las condiciones evaluadas.

X (features):

	dias_cultivo	temperatura_prom	oxigeno_prom	ph_prom	peces_semrados	peso_inicial_prom	alimento_total_kg	mortalidad_total	costo_operativo
0	171	8.554890	9.730577	6.975087	782	13.207201	1393.954382	13.383044	8182.750484
1	134	8.806224	6.866547	8.442790	4178	24.190852	1436.207173	4.816581	7300.725365
2	191	8.145775	6.353723	6.861954	1986	47.669028	950.259362	10.530582	4095.844556
3	180	8.755544	8.219998	8.208770	4582	36.742906	664.257507	14.895140	10533.324046
4	140	13.464054	7.043671	7.484556	4683	37.146798	1223.813812	6.876969	14624.961796

Y - Crecimiento:

	peso_final_prom
0	456.509086
1	490.506567
2	329.739464
3	251.440045
4	435.845890

Fig. 2 Comparación Real vs Esperado – Random Forest (Crecimiento)

Para complementar el análisis del Objetivo 1, se revisó la importancia de variables del modelo Random Forest. Como se observa en la Fig. 3, el factor alimento_total_kg es el que más aporta a la predicción del crecimiento. En términos simples, con los datos usados, la cantidad de alimento explica gran parte del peso final esperado.

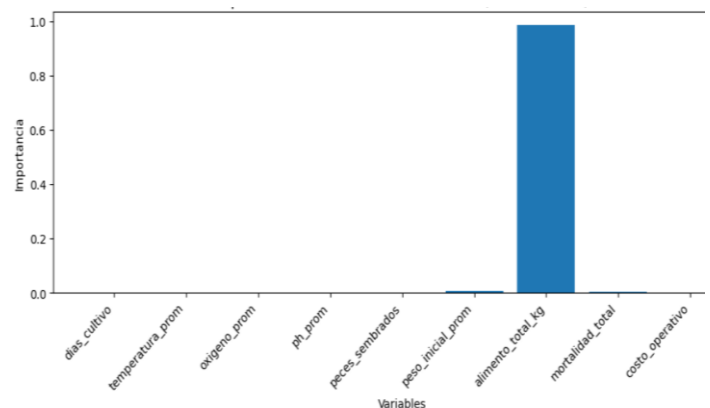


Fig. 3 Importancia de variables - CRECIMIENTO (Random Forest)

Para responder al Objetivo 2 (predicción de rentabilidad), se modeló la rentabilidad del lote como variable continua. Al depender de factores productivos y económicos, se esperaba mayor complejidad en la predicción.

TABLA II
DESEMPEÑO PREDICTIVO PARA EL MODELO DE RENTABILIDAD

Modelo	MSE	R^2
Regresión Lineal	883,419.72	0.866
Random Forest	864,637.64	0.869

En rentabilidad, los valores de R^2 bajan a ~0.87, lo cual es normal porque la rentabilidad combina producción, costos y precio de venta. Random Forest tuvo una leve mejora frente a Regresión Lineal, lo que sugiere que el modelo de ensamble se adapta mejor cuando las relaciones no son completamente lineales.

Las Fig. 4 y 5 presentan la comparación real versus la esperada para ambos modelos.

Y - Rentabilidad:		Y - Clasificación:	
rentabilidad_lote		es_rentable	
0	-4466.807865	0	0
1	19181.056945	1	1
2	2341.808716	2	1
3	887.284007	3	1
4	4396.411559	4	1

dtype: float64

dtype: int64

Fig. 4 Comparación Real vs Esperado – Regresión Lineal (Rentabilidad)

```

✓ División 80/20 realizada correctamente.

♦ REGRESIÓN - CRECIMIENTO
Linear Regression R2: 0.9927550766433969
Random Forest R2: 0.987551334468675

♦ REGRESIÓN - RENTABILIDAD
Linear Regression R2: 0.8655860567027639
Random Forest R2: 0.8685329245058767

♦ CLASIFICACIÓN - ¿ES RENTABLE?
Accuracy: 0.91
Matriz de Confusión:
[[50  4]
 [ 5 41]]

```

Fig. 5 Comparación Real vs Esperado – Random Forest (Rentabilidad)

Se observa mayor dispersión que en el modelo de crecimiento, especialmente en valores elevados de rentabilidad, lo que sugiere variabilidad económica no completamente capturada por modelos lineales simples.

Para el Objetivo 2, la Fig. 6 sugiere que la rentabilidad se ve más influenciada por peces_semrados, alimento_total_kg y costo_operativo.

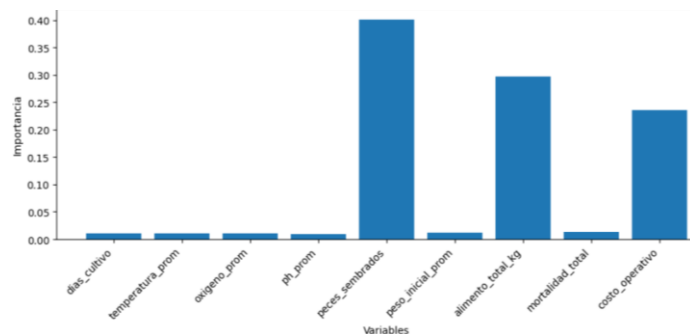


Fig. 6 Importancia de variables - RENTABILIDAD (Random Forest)

El tercer objetivo fue clasificar si un lote era rentable o no. El modelo Random Forest Classifier alcanzó una precisión global (accuracy) de 0.93.

El modelo logró identificar correctamente 52 lotes no rentables y 41 rentables. Los errores fueron limitados (2 falsos positivos y 5 falsos negativos), lo que demuestra buena

capacidad de discriminación y utilidad operativa para toma de decisiones, tal como se observa en la Tabla III y Fig. 7.

TABLA III
MATRIZ DE CONFUSIÓN DEL MODELO DE CLASIFICACIÓN

	Predicho No	Predicho Sí
Real No	52	2
Real Sí	5	41

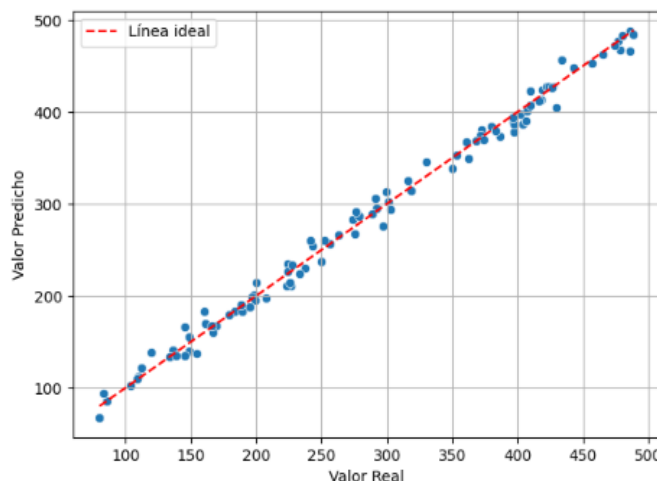


Fig. 7 Matriz de confusión visual del modelo de clasificación

Posteriormente, se evaluó un escenario simulado para validar aplicabilidad práctica. En la Fig. 8 se observa que el sistema proyectó un crecimiento promedio de 304.95g y una rentabilidad estimada de S/. 2,512.26, clasificando el lote como rentable.

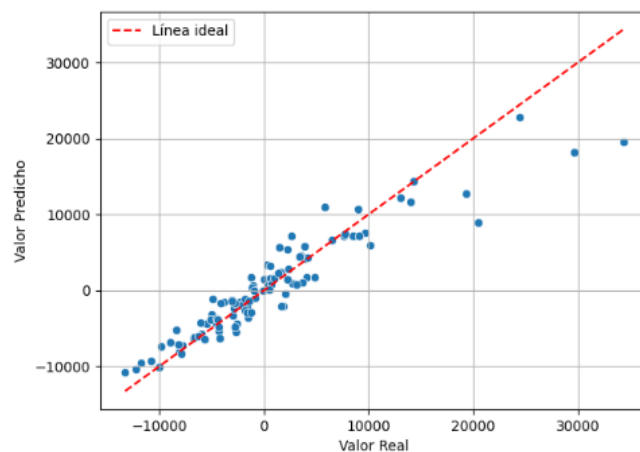


Fig. 8 Resultado de simulación para nuevo lote evaluado

Estos resultados evidencian que el sistema no sólo posee capacidad predictiva, sino también potencial aplicación como herramienta de soporte a decisiones en piscicultura.

Finalmente, la Fig. 9 muestra una correlación positiva muy alta entre alimento_total_kg y peso_final_prom ($\rho \approx 0.99$), lo que explica el elevado desempeño de los modelos lineales. Asimismo, se observa fuerte correlación entre producción_kg e ingresos ($\rho > 0.95$) y entre ingresos y rentabilidad_lote, lo que confirma la coherencia estructural del sistema productivo modelado.

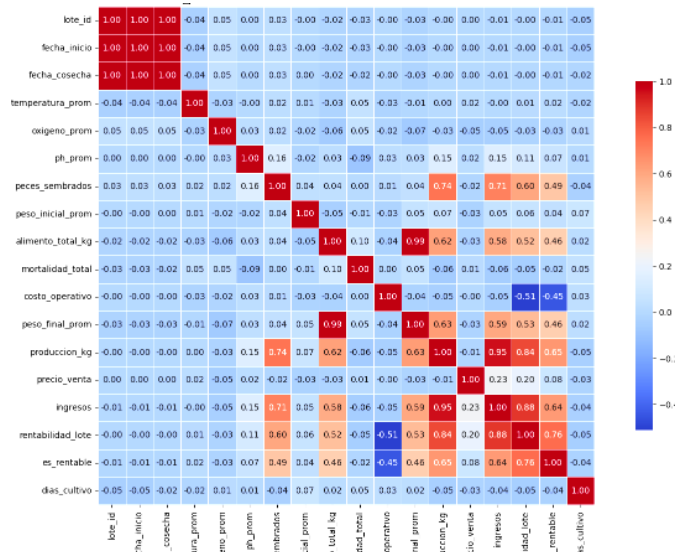


Fig. 9 Matriz de correlación entre variables del lote

IV. DISCUSIÓN

Los resultados obtenidos permiten analizar el desempeño de los modelos en función de la naturaleza del fenómeno modelado.

En el caso del crecimiento, la Regresión Lineal alcanzó un $R^2=0.993$, superando ligeramente a Random Forest ($R^2=0.988$). Este comportamiento se explica por la fuerte correlación observada entre alimento_total_kg y peso_final_prom (Fig. 9), lo cual sugiere que, bajo las condiciones del dataset analizado, la dinámica de crecimiento responde predominantemente a relaciones lineales.

Este hallazgo es consistente con estudios que muestran que el consumo alimenticio es el principal determinante del crecimiento en sistemas intensivos [6], y que, cuando las variables ambientales se mantienen dentro de rangos controlados, los modelos lineales pueden ofrecer alto desempeño predictivo [7].

Por el contrario, la predicción de rentabilidad presentó menor R^2 (aprox. 0.87), lo cual es coherente con la naturaleza multivariable y económica del fenómeno estudiado.

Random Forest tuvo un mejor resultado por poco, lo que sugiere que este modelo funciona mejor cuando las variables

económicas se relacionan de forma más compleja. Esto está en línea con investigaciones actuales que promueven el uso de herramientas digitales e inteligencia artificial en la acuicultura. [8], [12], [13].

La clasificación binaria alcanzó una precisión del 93%, lo que demuestra alta capacidad de discriminación. La baja tasa de falsos positivos (2 casos) es especialmente relevante desde el punto de vista operativo, ya que reduce el riesgo de considerar rentable un lote que no lo es.

Finalmente, se debe considerar que el dataset incluyó datos resumidos generados dentro de rangos validados. Esto puede hacer que las relaciones entre variables sean más ordenadas y que las métricas salgan altas. Por eso, los hallazgos deben interpretarse como una validación del enfoque en condiciones controladas. El siguiente paso es validar con campañas reales adicionales y mayor variabilidad ambiental.

A pesar de ello, el estudio demuestra que es técnicamente posible implementar modelos predictivos reproducibles en piscigranjas ubicadas en zonas altoandinas, incluso cuando se cuenta con recursos tecnológicos limitados.

V. CONCLUSIONES Y RECOMENDACIONES

A continuación, se presentan las conclusiones obtenidas:

Se logró desarrollar y comparar modelos supervisados para la predicción del crecimiento de *Oncorhynchus mykiss*, alcanzando un R^2 superior a 0.98, confirmando que, el crecimiento está fuertemente asociado al alimento suministrado

La rentabilidad fue más difícil de predecir ($R^2 \approx 0.87$), porque depende de varias variables económicas y productivas. Random Forest mostró una leve mejora (menor MSE y R^2 ligeramente mayor), lo que indica mejor ajuste cuando las relaciones no son totalmente lineales.

Se obtuvo un accuracy de 0.93 y pocos errores, por lo que el clasificador puede funcionar como una alerta rápida de viabilidad del lote.

Respecto a crecimiento, el alimento_total_kg fue el factor principal. En rentabilidad, destacaron peces_semrados, alimento_total_kg y costo_operativo. Finalmente, como se observó en la Fig. 9, se refuerza la coherencia entre variables productivas y económicas.

Y a nivel de recomendaciones, es posible que se requiera a futuro:

Validar el modelo con datasets completamente reales provenientes de múltiples campañas productivas.

Integrar sensores IoT para captura automática de temperatura, pH y oxígeno disuelto; para ello, los productores requieren iniciar proyectos de implementación de equipos complementarios.

Incorporar variables adicionales como tasa de conversión alimenticia (FCR) y densidad por estanque, para que el modelo sea más robusto.

AGRADECIMIENTO

Agradecemos al equipo de la piscigranja “El Valle” por brindarnos la oportunidad de realizar investigación aplicada en sus instalaciones.

REFERENCIAS

- [1] A.-Q. Wu *et al.*, «Deep Learning for Sustainable Aquaculture: Opportunities and Challenges», *Sustainability*, vol. 17, n.º 11, p. 5084, ene. 2025, doi: 10.3390/su17115084.
- [2] T. Aung, R. Abdul Razak, y A. Rahiman Bin Md Nor, «Artificial intelligence methods used in various aquaculture applications: A systematic literature review», *J. World Aquac. Soc.*, vol. 56, n.º 1, p. e13107, 2025, doi: 10.1111/jwas.13107.
- [3] D. Li, Z. Wang, S. Wu, Z. Miao, L. Du, y Y. Duan, «Automatic recognition methods of fish feeding behavior in aquaculture: A review», *Aquaculture*, vol. 528, p. 735508, nov. 2020, doi: 10.1016/j.aquaculture.2020.735508.
- [4] D. An, J. Hao, Y. Wei, Y. Wang, y X. Yu, «Application of computer vision in fish intelligent feeding system—A review», *Aquac. Res.*, vol. 52, n.º 2, pp. 423-437, 2021, doi: 10.1111/are.14907.
- [5] L. Zhang, B. Li, X. Sun, Q. Hong, y Q. Duan, «Intelligent fish feeding based on machine vision: A review», *Biosyst. Eng.*, vol. 231, pp. 133-164, jul. 2023, doi: 10.1016/j.biosystemseng.2023.05.010.
- [6] C. Lan, Y. Xinting, S. Chuanheng, W. Yizhong, X. Daming, y Z. Chao, «Feed intake prediction model for group fish using the MEA-BP neural network in intensive aquaculture», vol. 7, n.º 2, pp. 261-271, 2020, doi: <https://doi.org/10.1016/j.inpa.2019.09.001>.
- [7] N. M. Khiem, Y. Takahashi, T. Masumura, G. Kotake, H. Yasuma, y N. Kimura, «A machine learning ensemble approach for predicting growth of abalone reared in land-based aquaculture in Hokkaido, Japan», *Aquac. Eng.*, vol. 103, p. 102372, nov. 2023, doi: 10.1016/j.aquaeng.2023.102372.
- [8] S. Budaev, G. M. Cusimano, y I. Rønnestad, «FishMet: A Digital Twin Framework for Appetite, Feeding Decisions and Growth in Salmonid Fish», *Aquac. Fish Fish.*, vol. 5, n.º 2, p. e70064, 2025, doi: 10.1002/aff.7.0064.
- [9] «Estrategias para la producción de truchas en el centro piscícola “El Ingenio” mediante la dinámica de sistemas». [En línea]. Disponible en: <https://repositorio.uncp.edu.pe/items/1dc68e4d-a67f-4eb2-9d42-0bd35be2b80b>
- [10] «Predicting farmed rainbow trout weight ... | Open Research Europe». [En línea]. Disponible en: <https://open-research-europe.ec.europa.eu/articles/2-63>
- [11] V. Costa, E. Rocha, y C. Marques, «Robust mortality prediction on a recirculating aquaculture system», *Rev. Sci. Instrum.*, vol. 95, n.º 2, feb. 2024, doi: 10.1063/5.0196248.
- [12] M. Føre *et al.*, «Digital Twins in intensive aquaculture — Challenges, opportunities and future prospects», *Comput. Electron. Agric.*, vol. 218, p. 108676, mar. 2024, doi: 10.1016/j.compag.2024.108676.
- [13] J. Chen *et al.*, «The Application of Digital Twin Technology in the Development of Intelligent Aquaculture: Status and Opportunities», *Fishes*, vol. 10, n.º 8, p. 363, ago. 2025, doi: 10.3390/fishes10080363.
- [14] S. Noticias, «Aquabyte lanza inédito sistema de monitoreo basado en inteligencia de datos para trucha». [En línea]. Disponible en: <https://www.salmonexpert.cl/aquabyte-trucha-trucha-arcoiris/aquabyte-lanza-inedito-sistema-de-monitoreo-basado-en-inteligencia-de-datos-para-trucha/1785248>
- [15] W. F. Manrique Chalco, «Detección y segmentación de peces utilizando técnicas de deep learning», jun. 2023. [En línea]. Disponible en: <http://repositorio.ulasalle.edu.pe/handle/20.500.12953/176>
- [16] «A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME - Salih - 2025 - Advanced Intelligent Systems - Wiley Online Library». [En línea]. Disponible en: <https://advanced.onlinelibrary.wiley.com/doi/10.1002/aisy.202400304?>
- [17] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- [18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. <https://www.deeplearningbook.org>
- [19] García, S., Luengo, J., & Herrera, F. (2016). Data Preprocessing in Data Mining. Springer. <https://doi.org/10.1007/978-3-319-10247-4>
- [20] Provost, F., & Fawcett, T. (2013). Data Science for Business. O'Reilly Media. <https://www.oreilly.com/library/view/data-science-for/9781449374273/>
- [21] Prado, L., et al. (2025). Machine learning for genomic prediction of growth traits in aquaculture: a case study of the Australasian snapper (*Chrysophrys auratus*). *BMC Bioinformatics*. <https://doi.org/10.1186/s12859-025-05821-3>
- [22] García-Moreno, C. A., et al. (2021). Environmental variables influencing fish growth in aquaculture systems. *Aquaculture Reports*. <https://doi.org/10.1016/j.aqrep.2021.100737>