

## PROGRAMA DIPLOMADO

# ANÁLISIS DE DATOS CON R Y MODELAMIENTO PREDICTIVO CON APRENDIZAJE AUTOMÁTICO PARA LA ACUICULTURA

*4ra versión con machine learning e inteligencia artificial*

**1 ABRIL - 31 JULIO 2023**

### DESCRIPCIÓN DEL DIPLOMADO

El **Diplomado en Análisis de Datos y Modelamiento Predictivo con Aprendizaje Automático para la Acuicultura** brinda a los participantes la oportunidad de adquirir las habilidades y conocimientos necesarios para convertirse en analistas de datos o investigadores júnior en el campo de la acuicultura. Tanto para aquellos recién ingresados en la industria como para profesionales con experiencia, este programa ofrece una formación actualizada en Ciencia de Datos aplicada a la Acuicultura, utilizando el lenguaje de programación R. Esto les permitirá mejorar y consolidar sus habilidades, preparándolos para enfrentar los desafíos y oportunidades del mercado laboral actual.

La cuarta versión del Diplomado en Análisis de Datos y Modelamiento Predictivo para la Acuicultura ofrece un enfoque renovado en la aplicación de la inteligencia artificial y el aprendizaje automático en la industria acuícola. De esta manera, se ha reestructurado el plan de estudios para brindar una mayor comprensión de las herramientas y técnicas avanzadas en estos temas.

Este diplomado se enorgullece de proporcionar un buen equilibrio entre teoría y práctica, lo que permite a los estudiantes aplicar lo que han aprendido en situaciones reales. Durante cada clase, los estudiantes recibirán una introducción teórica a los conceptos, seguida de ejemplos prácticos y casos de estudio. Además, se utilizarán las mejores herramientas de programación en R y metodologías de investigación reproducible para mejorar la efectividad y confiabilidad de los análisis de datos.

Como parte integral del proceso de aprendizaje y evaluación, los estudiantes desarrollarán su propio proyecto de análisis de datos. Este proyecto incluirá un análisis exploratorio de datos

y la elaboración de un modelo estadístico o de modelamiento predictivo adecuado a sus datos. Esta experiencia práctica permitirá a los estudiantes aplicar lo que han aprendido en contextos relacionados con su interés en la acuicultura, ya sea en el ámbito productivo o de investigación.

En resumen, el Diplomado en Análisis de Datos y Modelamiento Predictivo para la Acuicultura brinda a los estudiantes todas las herramientas necesarias para convertirse en analistas de datos o investigadores juniors en la industria acuícola, o para consolidar sus competencias en el uso de herramientas avanzadas de ciencia de datos.

## OBJETIVOS DE APRENDIZAJE

Al finalizar el Diplomado los alumnos serán capaces de:

- Aplicar técnicas avanzadas de análisis de datos, inferencia estadística y modelamiento predictivo con aprendizaje automático (machine learning), para extraer información valiosa de datos relacionados a la acuicultura.
- Seleccionar y utilizar modelos estadísticos apropiados para el análisis de datos de acuicultura, incluyendo modelos lineales, no lineales y multivariados.
- Comunicar y presentar sus resultados de análisis de manera clara y atractiva mediante el uso de reportes dinámicos generados con Rmarkdown.
- Trabajar en un ambiente de investigación reproducible utilizando GitHub para mantener un control riguroso de versiones y documentación de sus proyectos.
- Adoptar herramientas basadas en inteligencia artificial para mejorar la eficiencia y la precisión de sus análisis de datos.

## PERFIL DEL PARTICIPANTE

Profesionales, graduados o investigadores relacionados con la industria de la acuicultura.

## DURACIÓN

Horas sincrónicas: 64 horas

Horas asincrónicas: 56 horas

Total: 120 horas

## DIRECTOR Y PROFESOR DEL PROGRAMA

### **José Gallardo Matus**

Doctor en Ciencias por la Universidad de Chile.

Profesor Adjunto de la Pontificia Universidad Católica de Valparaíso.

## PROFESOR

### **María Angélica Rueda Calderón**

Doctora en Ciencias Agropecuarias por la Universidad Nacional de Córdoba, Argentina.

Investigadora post-doctoral de la Pontificia Universidad Católica de Valparaíso.

## CONTENIDOS

### **UNIDAD 1. ANÁLISIS EXPLORATORIO DE DATOS E INVESTIGACIÓN REPRODUCIBLE CON R.**

#### *Introducción a la unidad*

Se entregan los fundamentos de la investigación reproducible y del análisis exploratorio de datos usando el lenguaje de programación R. Se discuten los aspectos beneficiosos de hacer investigación reproducible para el investigador, así como los principales criterios que determinan que una investigación sea reproducible o no. Respecto del análisis exploratorio de datos se definen y clasifican los distintos tipos de variables aleatorias, y se entregan herramientas que permiten observar y predecir el comportamiento de las variables aleatorias bajo distintos tipos de distribución de probabilidad, lo que contribuye a la correcta toma de decisiones tanto productivas como de investigación.

#### *Resultado de aprendizaje de la unidad*

Al finalizar la unidad, el estudiante será capaz de aplicar los fundamentos de la investigación reproducible y del análisis exploratorio de datos usando el lenguaje de programación R.

#### *Palabras clave*

R, Rstudio, Rmarkdown, Posit, Git, Github, GPT-3, OpenAI, variables aleatorias, distribución de probabilidad, análisis exploratorio de datos, investigación reproducible, control de versiones.

#### *Subtópicos*

Subtópico 1.1.- Introducción a la programación con R, Rstudio y Posit cloud.

Subtópico 1.2.- Análisis exploratorio y manipulación de datos con ggplot y dplyr.

Subtópico 1.3.- Reportes automáticos con Rmarkdown.

Subtópico 1.4.- Investigación reproducible y control de versiones con Git y Github.

Subtópico 1.5.- Herramientas de análisis de datos basadas en inteligencia artificial.

## UNIDAD 2. INFERENCIA ESTADÍSTICA Y PRUEBAS DE HIPÓTESIS CON R.

### *Introducción a la unidad*

Se entregan los fundamentos de la inferencia estadística y de las pruebas de contraste de hipótesis. Respecto del análisis de datos de acuicultura y sobre la base de estudios de caso se entregan herramientas para aplicar diferentes test estadísticos tanto paramétricos como no paramétricos, incluyendo los análisis de supervivencia.

### *Resultado de aprendizaje de la unidad*

Al finalizar la unidad, el estudiante será capaz de aplicar los fundamentos de la inferencia estadística en el análisis de datos de acuicultura mediante el lenguaje de programación R.

### *Palabras clave*

Parámetro, estadístico, correlación, permutación, combinación, inferencia estadística, contraste de hipótesis, análisis de supervivencia.

### *Subtópicos*

Subtópico 2.1.- Pruebas de contraste de hipótesis paramétrica: test de correlación de Pearson y Prueba de comparación de medias T-student.

Subtópico 2.2.- Prueba de F para ANOVA de una vía, dos vías, dos vías con interacción.

Subtópico 2.3.- Evaluación de supuestos de los análisis paramétricos y pruebas *a posteriori*.

Subtópico 2.4.- Pruebas de contraste de hipótesis no paramétrica: test de correlación Spearman, prueba de Man-Whitney, prueba de Wilcoxon, Chi-cuadrado.

Subtópico 2.5.- Análisis de supervivencia Kaplan Meier.

Subtópico 2.6.- Modelo de riesgos proporcionales COX.

## UNIDAD 3. MODELAMIENTO PREDICTIVO CON APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING).

### *Introducción a la unidad*

Se entregan los fundamentos del modelamiento predictivo con aprendizaje automático (machine learning) para resolver tareas concretas de producción o de investigación en acuicultura. En esta unidad, se presentan y ponen en práctica los algoritmos más utilizados de aprendizaje supervisado y no supervisado. Se enseña a seleccionar los mejores algoritmos de predicción de acuerdo al tipo de variable respuesta y de variables predictoras disponibles, para luego aplicar diferentes técnicas que permiten entrenar, testear y comparar la capacidad predictiva de los modelos realizados usando datos reales o simulados de acuicultura.

### *Resultado de aprendizaje de la unidad*

Al finalizar la unidad, el estudiante será capaz de seleccionar, aplicar y validar diferentes modelos predictivos con aprendizaje automático, poniendo en práctica lo aprendido con diferentes estudios de caso aplicados a la producción y a la investigación en acuicultura.

### *Palabras clave*

Aprendizaje automático, algoritmos de regresión y clasificación, entrenamiento y testeo de modelos, algoritmo de regresión lineal simple y múltiple, algoritmo de regresión logística, algoritmos de agrupamiento y de reducción de dimensionalidad (PCA).

### *Subtópicos*

Subtópico 3.1.- Introducción al modelamiento predictivo con machine learning: aprendizaje supervisado y no supervisado.

Subtópico 3.2.- Algoritmos de regresión: Regresión lineal, Regresión logística.

Subtópico 3.3.- Algoritmos de clasificación: Naive Bayes, Random forest, Máquinas de soporte vectorial (del inglés support-vector machines, SVM).

Subtópico 3.4.- Algoritmos de reducción de dimensionalidad: PCA, Análisis de Conglomerados Jerárquicos (UPGMA) y no Jerárquicos (K-Means).

## REQUISITOS DE POSTULACIÓN

- **Título profesional o licenciatura.**

- **Programación con R:** Deseable pero no excluyente. Los alumnos sin experiencia previa en programación con R deben considerar 4 horas de estudio y autoaprendizaje adicional por semana para poder alcanzar una comprensión avanzada de los objetivos de aprendizaje del curso.

- **Inglés:** Los softwares R, Rstudio, Rmarkdown, GitHub, Posit cloud y todos los paquetes de análisis estadístico que se usarán en el curso solo están disponibles en inglés. Alumnos sin competencias de lectura en inglés no deberían tomar el curso.

## RECURSOS DE APRENDIZAJE

- **Software R y Rstudio:** Posit.cloud por 5 mes.

- **Material docente:** Clases, guías de aprendizaje y códigos de programación para el análisis de datos y el modelamiento predictivo con aprendizaje automático.

## EVALUACIÓN

La evaluación se realiza mediante un proyecto personal de análisis de datos o de modelamiento predictivo. El proyecto se desarrolla en dos partes:

Parte 1. Análisis exploratorio de datos. Considera evidencias de aprendizaje de la unidad 1. Pondera un 40% de la nota final.

Parte 2. Modelar y comunicar. Considera evidencias de aprendizaje de la unidad 2 - 3, pondera un 40% de la nota final.

**Nota de aprobación:** La nota mínima de aprobación es un 4,0, en una escala de 1-7, con un 60% de exigencia.