


Syllabus
Curso Online 

Machine Learning con R y Python

Docente: MSc. Aldo M. Lezama Benavides



1. Docente

I. MSc. Aldo M. Lezama Benavides

MSc en Finanzas Corporativas de la Universidad del Pacífico y MSc© Estadística Aplicada de la UNALM. Economista de la UNMSM. Jefe de Ciencia de Datos en empresa líder de telecomunicaciones, con experiencia bancaria en los principales Bancos del Sistema Financiero, desarrollo de modelos analíticos con consultoras internacionales e Instructor en temas de analytics para ejecutivos.

2. Descripción del curso.

El alumno adquirirá los conocimientos necesarios para afrontar un problema de explotación de los datos utilizando las diferentes técnicas de Machine Learning. Este curso cubre los principales algoritmos de Machine Learning, así como el uso de las herramientas más populares para la Ciencia de Datos R y Python.

3. Dirigido a

El curso está dirigido a estudiantes de Economía, Economistas o personas interesados en formarse en la materia de Machine Learning con el fin de introducir los conocimientos adquiridos y la práctica en la tarea diaria de la empresa en la que trabajan.

4. Requisitos

Conocimientos en cálculo, estadística y programación a nivel básico en R y Python.

5. Contenido del curso

- I. Introducción a Machine Learning.
 - Machine Learning vs Modelos Estadísticos.

- Modelos Regresión vs Modelos de Clasificación, Casos de Uso.
- Diferencia entre Aprendizaje Supervisado vs No Supervisado.
- Uso de R & Python local: Instalación software y paquetes del CRAN/ PyPi, Conda.
- Uso Google Colab para procesamiento en la nube.

II. Métodos de validación de Modelos y Métricas de Evaluación.

- Resustitución, Validación Cruzada y Train Test: Ventajas y Desventajas.
- Modelos Clasificación: Matriz de Confusión, Gini, Accuracy, f1-score, recall, otros.

III. Feature Engineer y Componentes Principales.

- Transformación de Variables.
- Reducción de la dimensionalidad.

IV. Modelo de regresión Lineal y Modelo Logístico.

- Aspectos Teóricos.
- Lógica del Modelo.
- Aplicación.

V. Técnicas de Balanceo de Datos.

- Problema y consecuencias.
- Undersampling.
- Oversampling.

VI. Modelo de Árbol de Clasificación.

- Aspectos Teóricos.
- Lógica del Modelo.
- Aplicación.

VII. Modelo Random Forest y Gradient Boosting.

- Aspectos Teóricos.
- Lógica del Modelo.
- Aplicación.

VIII. Modelo de Redes Neuronales.

- Aspectos Teóricos.
- Lógica del Modelo.
- Aplicación.

6. Referencias

[1] ROBERT TIBSHIRANI Y TREVOR HASTIE, The Elements of Statistical Learning Data Mining, Inference and Prediction. Springer 2da Edición

[2] ROBERT TIBSHIRANI Y TREVOR HASTIE, An Introduction of Statistical Learning with Applications in R. Springer 2da Edición

[3] WES MCKINNEY, Python for Data Analysis. O'Reilly 2da Edición