

MACHINE LEARNING CON R Y PYTHON

Grupo Lambda 2020

1 Información General

1.1 Sobre el docente

MSc Aldo M. Lezama Benavides

MSc en Finanzas Corporativas por la Universidad del Pacífico y MSc(c) Estadística Aplicada por la UNALM. Asimismo, es especialista de Modeling & Scoring en Visanet Peru, y cuenta con experiencia en el sistema financiero, desarrollando modelos analíticos en consultorías internacionales como Instructor en temas de analytics para ejecutivos y Machine Learning.

1.2 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

El curso tiene una duración de 24 horas

1.3. Dirigido

El curso está dirigido a estudiantes de Economía, Economistas o personas interesados en formarse en la materia de Machine Learning.

1.4 Requisitos

De preferencia conocimientos básicos en cálculo, estadística y programación.

2 Contenido del Curso

2.1 Introducción al Machine Learning y uso de las herramientas open source

- Minería de Datos/Machine Learning
- Metodología CRISP
- Manejo del R y R Studio: Tratamiento de Datos, Funciones, Visualización
- Spyder & Jupyter para Python: librerías Pandas, Numpy, Sci Kit Learn, Matplotlib

2.2. Análisis Supervisado y No Supervisado

- Regresión lineal, regresión logística.
- Árboles de decisión.
- Random Forest/SVM
- XG Boost, Adaboost
- Redes Neuronales
- Componentes Principales
- KMeans

2.3. Evaluación de Modelos

- Principales Indicadores de Discriminación GINI, KS y ROC
- F1 Score, Accuracy, Precision
- Matriz de Confusión

2.4 Técnicas de Balanceo y Tuning de Algoritmos

- Undersampling,
- Oversampling,
- SMOTE.
- Tuning de Hiperparametros

2.5 Métodos de Validación

- Resustitución
- Validación Cruzada
- Train Test

2.6 Aplicación / Casos Prácticos

- Modelo de Fuga de Clientes para una empresa comercial/Telecom
- Modelo de Credit Scoring para una entidad financiera

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