



**BROCHURE**

**CECAT**

**4.0**

# **MACHINE LEARNING**

**Curso**

Encuéntranos en:



[datascience.pe](http://datascience.pe)



## ¿Quiénes somos?

Somos una organización educativa que nace para difundir, descentralizar, democratizar, la educación en ciencia de datos e inteligencia artificial en el país. Hoy el mundo ha cambiado y la transformación digital juega un rol importante en las organizaciones, los datos son la nueva mina de oro y el científico de datos es la profesión del futuro.



## ¿Cómo lo hacemos?



**Difundir**



**Democratizar**



**Descentralizar**

Los conocimientos de Ciencias de Datos e Inteligencia Artificial en favor de las personas, para desarrollar sus talentos y maximizar sus oportunidades, contribuyendo con la construcción de sociedades más justas.



# Machine Learning

---

El presente curso es de naturaleza teórico-práctico tiene como propósito profundizar el conocimiento del alumno en técnicas para el modelamiento predictivo de Machine Learning. Se abordan modelos basados en boosting: XGBoost, LightGBM; y modelos predictivos basados en stacking.

Adicionalmente, el curso explica el problema de la baja explicabilidad (cajas negras) en Machine Learning y se enseñarán métodos contemporáneos para su solución: Shapley y Permutation analysis.

## Modalidad

Online en vivo

## Duración

27 horas cronológicas

# Módulos

## 1.- Limpieza de datos, e ingeniería de variables (4h)

Detección de vacíos en los datos: casuísticas y criterios. Métodos de imputación de datos.

Variables categóricas sus técnicas de encoding. Detección de outliers y discusión de distintos criterios. Reducción dimensional con PCA.

## 3.- Cajas negras y la explicabilidad de algoritmos (8h)

Análisis de resultados de modelo lineal y de árboles: cumplimiento de supuestos, análisis del error (residual), gráfico del árbol, técnicas de importancia de variables.

El problema de las cajas negras en modelos de boosting. Técnicas de explicabilidad de cajas negras: shap values, permutation importances.

## 2.- Modelos predictivos de boosting (6h)

Definición y lógica de Boosting. Boosting vs bagging. LightGBM, Adaboost, XGBOOST.

Hiperparámetros en modelos de boosting. Introducción al método bayesiano de elección de hiperparámetros.

## 4.- Aplicaciones diversas de machine learning en industria (6h)

Introducción al pronóstico y al uso de datos temporales. Técnicas tradicionales: media móvil y exponencial. Introducción a modelo ARIMA.

Aplicación de clustering: ejercicio de clasificación de clientes de una empresa. Presentación gerencial de proyecto de clustering.



## Docentes



**Jorge Guevara Diaz**

Perú

[in /jorgeguevaradiaz](#)

Doctor en Ciencia de la Computación, Research Scientist at IBM Research, con más de 13 años de experiencia en docencia nacional e internacional.



**Wilma Pairo Huaynoca**

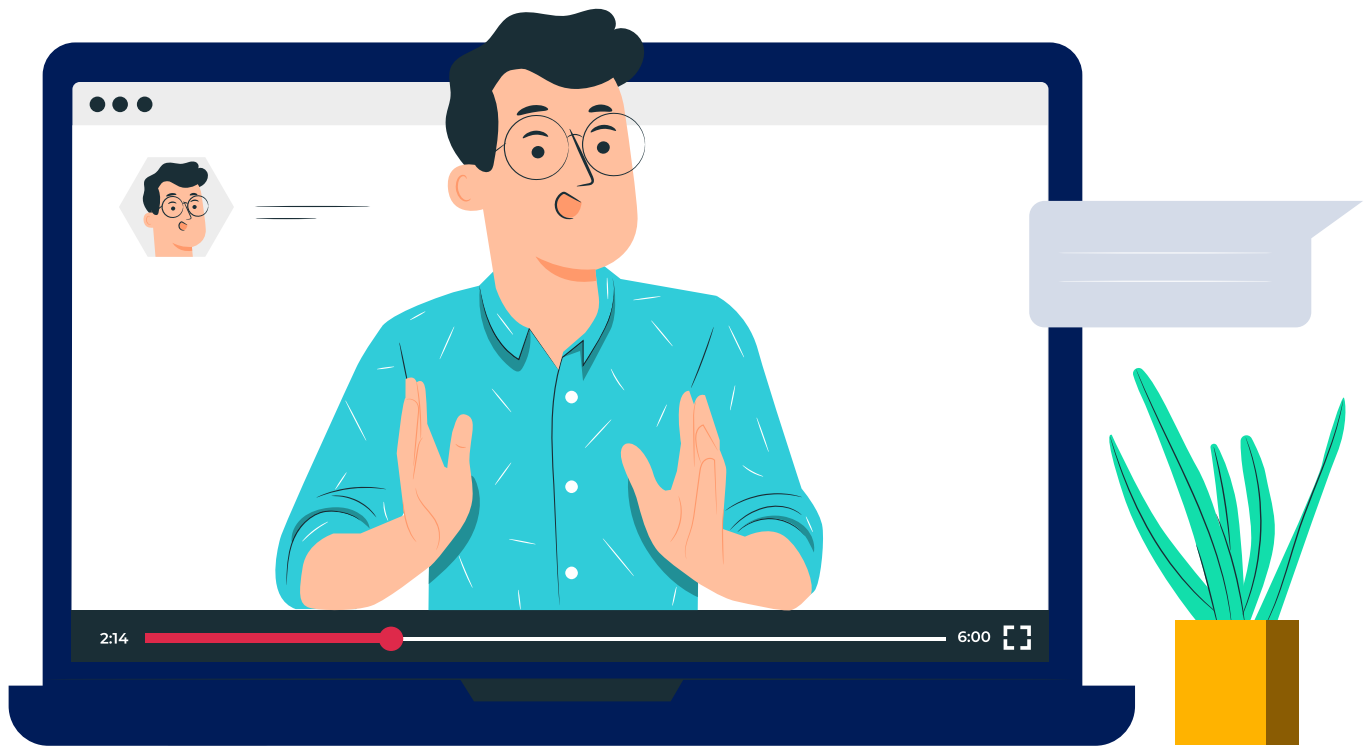


Chile



[/wilma-pairo](#)

Magister en Administración en la Universidad Rey Juan Carlos, con un Micro degree de Data Scientist en el MIT y con más de 5 años de experiencia liderando proyectos para empresas privadas y públicas.



## Horarios

Martes y Jueves de 20:00 a 22:00

Inicio: Martes 16 de Marzo

Sábado y Domingo de 10:00 a 13:00

Inicio: Sábado 13 de Marzo



## Metodos de pago

Precio normal: s/. 600.00



Cuenta de Ahorros Soles

:

2003131111482

Cuenta Interbancario

:

00320001313111148234



BBVA Cuenta sueldo

:

0011-0057-0239209883

BBVA Cuenta interbancario

:

011-057-000239209883-72



BCP Ahorros soles

:

19101387048057

BCP CCI

:

00219110138704805753



<https://www.paypal.me/DataScienceResearch>

En caso de pagar por Paypal en dólares se agregará un adicional del 10% del valor de su cuota.

DNI: 47487919

Propietario: Javier Cuicapuza (CEO DSRP)

Contacto: [informes@datascience.pe](mailto:informes@datascience.pe)