



Universidad **Ricardo Palma**

RECTORADO

**ESPECIALIZACIÓN EN CIENCIA DE DATOS**

Taller de Especialización en  
“**MACHINE LEARNING FOR DATA SCIENCE**” con PYTHON  
**Modalidad Virtual**

26 Enero-11 Febrero 2021

En este **taller de especialización** aprenderás la metodología de construcción de modelos predictivos, así como el uso de los algoritmos más avanzados de **Machine Learning**, poniéndote a prueba en competencias reales de **Analytics Vidhya y Kaggle** con el objetivo de sumergirte en el mundo del Data Science. Las clases tienen una fuerte componente teórica y casos prácticos desarrollados en Mini Data Challenge entre los participantes. El desarrollo del programa será en el ambiente de Google Colab - Python.



**INFORMES E INSCRIPCIÓN:**

Correo : [oroque@urp.edu.pe](mailto:oroque@urp.edu.pe) - 993 477 990



# TALLER DE ESPECIALIZACIÓN “MACHINE LEARNING FOR DATA SCIENCE” CON PYTHON

## MODALIDAD VIRTUAL

### (24 horas)

#### I. ENTENDIMIENTO DEL NEGOCIO

- Entendimiento del problema.
- Definición de la Variable Target.
- Definición y creación de Drivers.
- Elección y definición del Horizonte temporal ( Ventana de análisis).
- Criterios de Inclusión y Exclusión de la Información.

#### II. TRATAMIENTO Y PREPARACIÓN DE LA INFORMACIÓN

- Análisis Exploratorio de Datos. Tratamiento de Datos.
- LabelEncoder. OneHotEncoder. TargetEncoder.
- Imputación Paramétrica (Media, Mediana, Moda, Regresión).
- Imputación No paramétrica ( KNN, Random Forest).
- Transformación de Datos Paramétricamente ( Tasas, Deciles).
- Transformación de Datos No Paramétrico ( Árboles de Decisión).
- Ingeniería de Variables. PolynomialFeatures. ACP.

#### III. BALANCEO DE LOS DATOS Y SELECCIÓN DE DRIVERS

- Balanceo de la data mediante Oversampling, Undersampling, SMOTE, TomekLink, ReSampling, K-Centroides.
- Selección de variables mediante Técnicas Paramétricas.
- Selección de variables mediante Machine Learning (Random Forest, Boruta, XGBoost).

#### IV. MODELAMIENTO DE LA INFORMACIÓN

##### APRENDIZAJE SUPERVISADO RESPUESTA CATEGÓRICA

- Árboles de Clasificación ( CART).
- SVM ( Máquina de Soporte Vectorial).
- Random Forest.
- AdaBoost, GBM, XGBoost, LightGBM, CATBoost.
- Ensamble de Modelos.
- Stacking de Modelos.
- GridSearch de Parámetros. RandomGridSearch. GridSearch. BayesianOptimization.

##### APRENDIZAJE SUPERVISADO RESPUESTA NUMÉRICA

- Regresión Lineal Múltiple y Penalizadas.
- Regresión por Árboles de Regresión y Ensamblados de Árboles.
- Pronósticos con Series de Tiempo.

##### APRENDIZAJE NO SUPERVISADO

- Segmentación o Puntuación RFM.
- Análisis de Cesta de Canasta de Mercado.
- Segmentación con K-Means.

#### V. VALIDACIÓN DE MODELOS

- Validación Cruzada.
- Matriz de Confusión.
- GINI, K-S, ROC, Sensibilidad, Recall, Divergencia, F-Score.

#### VI. EVALUACIÓN Y PRESENTACIÓN DEL TALLER INTEGRADOR

- Público en general: S/. 380.00
- Miembro URP: S/. 350.00
- Estudiantes pre-grado: S/. 280.00

# TE MACHINE LEARNING FOR DATA SCIENCE CON PYTHON-MODALIDAD VIRTUAL

## PLANA DOCENTE

¡Hola! Soy,

# André Chávez

Lead Data Scientist Estadístico

Candidate MSc. Data Science

Me puedes encontrar como:



[andre.chavez@urp.edu.pe](mailto:andre.chavez@urp.edu.pe)



[\[www.linkedin.com/in/andr -ch vez-a101010/\]](https://www.linkedin.com/in/andr -ch vez-a101010/)



Conocimiento y experiencia laboral

- ☐ Horarios: Martes y Jueves  
De 19:00-22:00
- ☐ Duraci n: 3 semanas  
efectivas de clase

“Somos pioneros en la especializaci n de los profesionales en Data Science”