



UNIVERSIDAD
DE PIURA

En convenio con



PROGRAMA ESPECIALIZADO

DATA SCIENCE

DECISIONES BASADAS EN DATOS



ONLINE
SINCRÓNICA

**POSGRADOS &
FORMACIÓN CONTINUA**

Ingenierías

Objetivos

El Programa Data Science - Decisiones permitirá al alumno potenciar su visión de la importancia del uso de los datos para el negocio, incorporando la aplicación de nuevas técnicas y herramientas de analítica con el fin de encontrar valor en la información.



Desarrollará una sólida comprensión de los pasos claves involucrados en cómo resolver un problema usando Ciencia de Datos.



Comprenderá el negocio y comprender los datos: su preparación, modelamiento y despliegue.



Analizará y comprenderá qué pasó con la empresa y cómo puede actuar en el futuro.



Conocerá los alcances, limitaciones y aspectos vinculados a la interpretación de diferentes metodologías analíticas que se aplican en el proceso de Ciencia de Datos.



Aprenderá diversas herramientas y estadísticas utilizadas por la Ciencia de Datos, incluyendo el lenguaje de programación Python, para realizar análisis y entendimiento de los datos.



Conocer las principales técnicas de Machine Learning usadas para crear y entrenar modelos de funciones que realicen labor predictiva y permitan tomar decisiones.

Dirigido a:

Ejecutivos y profesionales de diferentes áreas de negocio, Marketing, Comercial, IT, BI o Finanzas, investigadores y personal científico que deseen entender y desarrollar proyectos basados en Ciencia de Datos.

Así como los que deseen comprender y utilizar la Ciencia de Datos para resolver problemas o proveer información para tomar decisiones inteligentes.



Plan de estudios

⌚ 3h

INTRODUCCIÓN

Proporcionar a los participantes de un entendimiento sobre la aplicación estratégica de la ciencia de datos en contextos empresariales. Enfatizando la importancia de identificar necesidades de negocio, desarrollar casos de uso pertinentes y emplear diferentes tipos de analíticas.

- Necesidades del negocio
- Casos de Usos
- Analítica: descriptiva, predictiva y prescriptiva
- Entendimiento de los datos
- Planificación de proyectos
- Comunicación de resultados
- Data Business Transformation

Proporcionar a los participantes de un entendimiento sobre la aplicación estratégica de la ciencia de datos en contextos empresariales. Enfatizando la importancia de identificar necesidades de negocio, desarrollar casos de uso pertinentes y emplear diferentes tipos de analíticas.



1.1. Conceptos fundamentales de la Ciencia de Datos

- Definición y avances de Ciencia de Datos.
- Python y lenguajes de programación para la Ciencia de Datos.
- IDEs actuales para el desarrollo de la Ciencia de Datos.
- Python y Colab.



1.2 Fundamentos de Python

- Syntax, tipos de datos, variables.
- Creación de funciones.
- Funciones repetitivas.
- Estructuras de control: Condicionales, bucle.



1.3 Librerías en Python

- Librería Matplotlib.
- Librería Pandas.
- Librería Seaborn.
- Librería Scikit-learn.
- Otras Librerías.



1.4 Preparación de datos

- Definición de variables.
- Listas y estructuras.
- Gestión de datos con librerías Numpy y Pandas.



1.5 Modelado y visualización de datos

- Selección del tipo de modelo (librerías).
- Librerías para diferentes tipos de análisis: NumPy, Matplotlib, Seaborn, SciPy.
- Librerías para Machine Learning: Scikit-learn, Tensorflow.
- Análisis de datos.
- Visualización de datos y gráficos estadísticos con Matplotlib y Seaborn.

Caso práctico

Utilizar Python y sus librerías para realizar un análisis exploratorio de datos que incluye carga de datos, inspección inicial, manipulación de datos y finalmente la visualización.

Desarrollar un entendimiento profundo de los conceptos estadísticos fundamentales y su aplicación en la ciencia de datos. Este módulo está diseñado para enseñar a los participantes cómo preparar, analizar y interpretar datos utilizando métodos estadísticos, con el objetivo de apoyar la toma de decisiones empresariales efectivas. Se enfocará en técnicas descriptivas e inferenciales, con ejemplos prácticos usando Python.



1.1. Estadística descriptiva

- Introducción a la estadística: Población y Muestra.
- Tipos de variables y escalas de medición.
- Datos agrupados y tabla de frecuencias.



1.2 Preparación de datos

- Transformación de datos y discretización.
- Limpieza de datos.
- Identificación de Outliers.



1.3 Análisis descriptivo

- Medidas de tendencia central: Media, Mediana y Moda.
- Medidas de dispersion: Rango, Cuantiles, Varianza y desviación estándar.
- Elaboración de histogramas, gráficos de caja.
- ¿Cuándo usar? ¿Qué preguntas responde?



1.4 Inferencia estadística

- Relación entre variables cuantitativas: coeficiente de correlación.
- Relación entre variables cualitativas: Prueba chi cuadrado.
- Pruebas de hipótesis y su aplicación en la toma de decisiones.
- ¿Cuándo usar? ¿Qué preguntas responde?

Caso práctico

Desarrollo de análisis estadístico completo desde la recolección de datos hasta las conclusiones, utilizando Python para manejar datos reales y simular decisiones basadas en análisis estadísticos.

Desarrollar habilidades en la selección, implementación y evaluación de modelos analíticos avanzados, con un enfoque en cómo estos modelos pueden ser aplicados para resolver problemas reales de negocio y mejorar la toma de decisiones basada en datos. Los participantes aprenderán a equilibrar la complejidad del modelo con la interpretabilidad y la relevancia para los objetivos comerciales.



1.1 Introducción a las metodologías para el desarrollo de modelos

- ¿Qué son las metodologías para el desarrollo de modelos?
- Ventajas y desventajas de cada metodología.
- Ejemplo de metodologías: CRISP-DM.



1.2 Regresión lineal simple

- Definición y supuestos de la regresión lineal simple.
- Interpretación de los coeficientes de regresión.
- Evaluación de la bondad de ajuste del modelo.



1.3 Regresión lineal múltiple

- Definición y supuestos de la regresión lineal múltiple.
- Selección de variables y ajuste del modelo.
- Evaluación de la bondad de ajuste del modelo.



1.4 Estimación de modelos

- Selección de variables explicativas.
- Estimación del modelo de regresión.
- Análisis de Residuos.



1.5 Validación de modelos

- Validación cruzada y particionamiento de datos.
- Métricas de evaluación de modelos: R2, MAE, RMSE, etc.
- Interpretación de las métricas de evaluación de modelos.



1.6 Regularización en modelos de regresión

- Regresión ridge y lasso.
- Selección del parámetro de regularización.
- Efectos de la regularización en los coeficientes del modelo.



1.7 Modelos de regresión no lineal

- Definición y ejemplos de modelos de regresión no lineal.
- Métodos de ajuste de modelos no lineales.
- Evaluación de la bondad de ajuste en modelos no lineales.



1.8 Selección de modelos

- Métodos de selección de modelos: MSE, MAPE, AIC, BIC, etc.
- Comparación de modelos y selección del mejor modelo.
- Interpretabilidad vs. complejidad del modelo.



1.9 Implementación de modelos

- Implementación de modelos.
- Integración de modelos en los proceso de negocio.
- Consideraciones prácticas para la implementación de modelos.



1.10 Monitoreo de modelos

- Monitoreo del rendimiento de modelos en producción.
- Detección y diagnóstico de problemas en modelos.
- Actualización y reentrenamiento de modelos en producción.

Caso práctico

Desarrollo de un caso que abarca el proceso de modelado, desde la recopilación de datos hasta la implementación y monitoreo del modelo en un contexto de negocio realista.



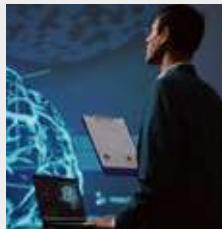
Capacitar a los participantes en los principios y técnicas de Machine Learning e Inteligencia Artificial, con un enfoque en la aplicación práctica y la integración de estos modelos para mejorar la toma de decisiones en entornos empresariales. Los participantes aprenderán a elegir, implementar y evaluar modelos de Machine Learning usando Python.



1.1. Inteligencia Artificial y Machine Learning

- Introducción. Definiciones. Conceptos. Procesos y Aplicaciones. Tipos de Aprendizaje.

1.2 Aprendizaje Supervisado



1.3 Regresión Logística

- Introducción y Definición.
- Supuestos del modelo.
- Formulación del modelo.
- Estimación de parámetros.
- Aplicación.

1.4 Validación de modelos de Aprendizaje Supervisado

- Matriz de Confusión. GINI. K-S. ROC. Sensibilidad. Especificidad.

1.5 Aprendizaje Supervisado

- Árboles de Clasificación. Random forest. SVM (Máquinas de Soporte Vectorial). Naive Bayes. Aplicaciones.



1.6 Redes neuronales

- La Neurona. Funcionamiento. Arquitectura. Perceptrón. Funciones de Activación. Aplicación.



1.7 Aprendizaje No Supervisado

- Agrupamiento K-means. Componentes Principales. Aplicaciones.

1.8 Fundamentos de Deep Learning: Perceptrón Multicapa

1.9 Aplicación Global del Mejor Modelo

Caso práctico

Desarrollo de un proyecto de Machine Learning desde el análisis de datos hasta la implementación y evaluación de un modelo seleccionado, utilizando los conceptos aprendidos.

Docentes:



Richard Pérez

Trayecto

MSc en Finanzas en la Universidad Pacífico y BSc en Ingeniería Económica en la UNI. Actualmente, se desempeña como Gerente de Data & Analytics en Alicorp, y posee una sólida experiencia en las áreas de Modelación y Validación de Riesgo de Crédito y Riesgo de Mercado.



Denis Palomino

Trayecto

Ingeniero Estadístico – UNALM, Estudios de Maestría en Estadística Aplicada – UNALM, Master en Marketing Digital – Universidad de la Rioja UNIR México, Master in Business Administration Executive – Escuela Europea de Negocios España. Actualmente se desempeña como Senior Audit Data Scientist en Scotiabank.



Carlos Barragán

Trayecto

Ingeniero Industrial – UNI, MBA – ESAN, Magister en Dirección de Tecnologías de Información – ESAN, Doctorando en Ingeniería de Sistemas e Informática en la UNMSM, Maestría en Gestión de Hidrocarburos – CENTRUM. Actualmente se desempeña como asociado en BTA Company.



Marco Barrientos

Trayecto

MBA de ESAN e Ingeniero de Sistemas de la Universidad Nacional de Ingeniería. Actualmente es socio fundador y director comercial de BTA Company, empresa que colabora con sus clientes en su proceso de transformación digital.

Evaluación

El proceso de evaluación se llevará a cabo combinando diferentes métodos y técnicas:

- Ejercicios individuales o grupales durante la sesión que permitirán monitorear el avance los estudiantes.
- Casos prácticos por módulo.
- Evaluación de Trabajo integrador a presentar al final del Programa.

Cálculo de la nota final

Modulo 1:	CP1 - Caso práctico 1	15%
Modulo 2:	CP2 - Caso práctico 2	15%
Modulo 3:	CP3 - Caso práctico 3	15%
Modulo 4:	CP4 - Caso práctico 4	15%

Trabajo Final: TF 40%

$$\text{Nota Final} = (\text{CP1} * 15 + \text{CP2} * 15 + \text{CP3} * 15 + \text{CP4} * 15 + \text{TF} * 40) / 100$$

Software

*No se requiere la instalación de software

*Se usará el Google Colab, Python, las librerías Numpy, Scilearn, Matplotlib, Pandas, entre otras.

*Se utilizará Mentimeter o reuniones virtuales por Zoom.

Duración y Horario

- **Duración:**
80 horas lectivas
- **Horario:**
Martes y jueves de 7:00 p.m. a 10:00 p.m.



ONLINE
SINCRÓNICO

INICIO
MARTES
01
ABRIL

Inversión

La inversión del programa es de: **S/ 5,500 soles**

Certificado

- Se entregará a los participantes que cumplan con más del 80% de asistencia a clases.



UNIVERSIDAD
DE PIURA

CAMPUS LIMA
Calle Mártir Olaya 162 Miraflores

969 262 897
yusahara.huaman@udep.edu.pe

