



BCIT in
Data Science



Objetivos

El objetivo general de la formación consiste en preparar a los alumnos de forma intensiva en alguna de las **principales herramientas utilizadas en el área de Data Science**, llegando a conocerlas y a ser capaces de desenvolverse en éstas de forma fluida consiguiendo ser productivos de forma inmediata tan pronto ingresan en un equipo “data-driven” o también denominados orientado al dato.

Los objetivos concretos de la formación son:

- Adquirir conocimientos sobre programación y los procesos de captura, extracción, manipulación y conversión de datos en diferentes entornos.
- Conocer los principales métodos estadísticos para el análisis de datos descriptivo y predictivo.
- Dominar los métodos de visualización de datos, tanto estáticos como dinámicos.
- Manejar modelos y algoritmos para extraer conocimiento de los datos.
- Aplicar los conocimientos en problemas reales.
- Desarrollar la capacidad de transmitir sus conclusiones de manera clara a un público variado, especializado o no.

Requisitos

Es necesario tener **nociones de básicas de programación en Python**, conocimiento del lenguaje de consultas SQL y alguna herramienta de control de versiones, como git.

Metodología

En esta formación, el aprendizaje se llevará a cabo de **una forma dinámica** con una enseñanza teórica y práctica compaginando ambas mediante la realización de ejercicios que permitan hacer uso de todo lo aprendido buscando asentar el conocimiento.

Programa

El programa se estructura en los bloques que se describen a continuación:

Bloque 1: Conceptos básicos Generales

- SQL / Base de Datos Relacionales / No Relacionales.
- Gestión de Entornos y Control de Versiones
- Python
- Bibliotecas de terceros de Python: pandas/numpy, scikit-learn, scipy
- Metodología de trabajo Agile y Herramientas trabajo: Jira, Trello,...

Bloque 2: Fundamentos básicos de Estadística

- Álgebra: Variables, Ec. Lineales, Logaritmos, Función sigmoide.
- Álgebra lineal: Tensor, multiplicación de matrices.
- Trigonometría: Función Tanh (o de activación).
- Estadística: Media, mediana, valores atípicos, desviación estándar, histograma.
- Cálculo (opcional para temas avanzados): derivada, gradiente.
- Correlaciones

Bloque 3: Tratamiento de los datos

- Minería Datos: preparación, estandarización y normalización de los datos
- Visualización: Matplotlib, Plotly, Dash, Ggplot, Seaborn, geoploly.

Bloque 4: Modelado

- Machine Learning – Curso en 10 pasos
 - Regression Algorithms: Simple Linear Regression model, Lasso Regression, Logistic regression, Support Vector Machines (SVM).
 - Classification Algorithms: Logistict Regression, K-nearest neighbours, Decision tree, Random forest, SVM
 - Clustering Algorithms: K-Means, DBSCAN.

- Forecasting Algorithms: AR, Arima, XGBoost, Prophet.
- Machine Learning – Learning by Doing
 - Titanic, problema de clasificación
 - Anomalías, problema datos desbalanceados
 - NLP, tratamiento de textos
 - Series Temporales y predicción
 - Clustering
 - Predicción con árboles de decisión
 - Grafos

Bloque 5: Programación en Distribuido

- PySpark

Bloque 6: ¿Qué hacer cuando tienes ya el modelo?

- MLOps:
 - Model Fine-Tuning y Optimización
 - Deployment of the ML model

Bloque 7: Especialización

En este bloque se llevará a cabo una especialización en una de las siguientes herramientas:

- AWS
- Azure
- Snowflake



binaia.es