



Análisis y Procesamiento de Grandes Volúmenes de Datos (BigData)

MÓDULO 1 “INTRODUCCIÓN AL ANÁLISIS DE DATOS”

OBJETIVO:

1. En este módulo se ofrecerá al alumno una perspectiva del análisis de datos, herramientas y algoritmos utilizados para BigData en la industria.

TEMAS:

1. Ciclo de vida del análisis de datos
2. El rol del científico de datos.
3. Métodos estadísticos con lenguaje R.
4. Fuentes de datos.
5. Trabajando con datos categóricos y numéricos.
6. Introducción a la serie de tiempo.
7. Trabajando con datos de redes sociales.
8. Trabajando con sensores y equipos móviles.
9. Trabajando con datos de la nube.
10. Tratamiento y limpieza de datos con OpenRefine.

Duración: 30 horas.



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LA COSTA

MÓDULO 2

“MÉTODOS PARA EL APRENDIZAJE AUTOMÁTICO”

OBJETIVO:

1. En este módulo se explorarán los algoritmos para Maching Learning utilizados en clasificación, predicción y agrupamiento.

TEMAS:

1. Introducción al aprendizaje automático con Python.
2. Métodos de reducción de complejidad en datos complejos.
3. Análisis de componentes principales.
4. Clasificación con árboles de decisión.
5. Implantando regresión lineal.
6. Predicción en series de tiempo financieras con ANOVA.
7. Agrupación con K-Means
8. Clasificación de textos con Naive Bayes.
9. Desarrollando un sistema de recomendación.

Duración: 30 Horas.



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LA COSTA

MÓDULO 3

“BASES DE DATOS NO RELACIONALES (NOSQL).”

OBJETIVO:

1. En este módulo se trabajará con un modelo de acceso y persistencia de datos basado en documento, así como su implementación en proyectos de data warehousing.

TEMAS:

1. Modelo de datos no relacionales basado en documentos.
2. Introducción a MongoDB
3. Agrupación y agregación de datos con MongoDB
4. Procesamiento de datos con MapReduce en MongoDB
5. Data Warehousing con MongoDB
6. Consultas Avanzadas y cubos de datos.
7. Replicación y sharding de datos.

Duración: 30 Horas.



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LA COSTA

MÓDULO 4

“VISUALIZACIÓN DE DATOS”

OBJETIVO:

1. Este módulo explorará el análisis de datos a través de graficación y visualización para encontrar patrones, proporciones y distribuciones con herramientas Web y de escritorio.

TEMAS:

1. Introducción análisis de datos exploratorio a través de la graficación.
2. Visualización de datos unidimensionales y multidimensionales.
3. Visualización de datos con D3js.
4. Visualización de datos con Python.
5. Desarrollo de un motor de búsqueda visual.

Duración: 30 Horas.



UNIVERSIDAD DE GUADALAJARA

CENTRO UNIVERSITARIO DE LA COSTA

Módulo 5

“Procesamiento Avanzado de Datos”

OBJETIVO:

1. Este módulo introducirá al participante al procesamiento avanzado de BigData con Hadoop y su ecosistema (Hive y Mahout).

TEMAS:

1. Introducción a Hadoop.
2. Conociendo el ecosistema Hadoop.
3. Instalación de la infraestructura.
4. Modelo de programación MapReduce.
5. Consulta avanzada de datos con Hive.
6. Implementando algoritmos de Machine Learning con Mahout.

Duración: 30 Horas.