



**Programa de Extensión Universitaria y Vinculación
(PEUVI), Departamento de Matemáticas.**

DIPLOMADO EN MINERÍA DE DATOS.

Coordinadora académica: Dra. Amparo López Gaona.

**MODULO I: INTRODUCCIÓN A LAS BASES DE DATOS Y
MODELADO DE DATOS (40 horas).**

Software: Diagramador DIA.

Temas.

- Bases de datos relacionales.
Introducción.
Bases de datos relacionales.
Concepto de Modelo de datos.
Reglas de Codd para un sistema relacional.
Aplicación práctica.
- Análisis del modelo relacional y los elementos que lo integran.
Introducción.
Concepto de relaciones y sus propiedades.
Concepto de claves en el modelo relacional.
Teoría de normalización y sus objetivos.
Aplicación práctica.
- Descripción y aplicación del modelo entidad-relación (E-R) para el modelado de datos.
Introducción.
Proceso de realización de diagramas entidad-relación y cómo aplicarlo.
Elementos principales de los diagramas E-R.

Diagrama entidad-relación entendido como elemento para resolver las carencias de los diagramas entidad relación simples.

Elementos detallados del diagrama E-R.

Paso a tablas de una agregación.

Aplicación práctica.

- Lenguaje para definición de datos (DDL).
 - Creación de tablas.
 - Tipos de datos.
 - Restricciones de integridad básicas en SQL.
- Actualización de datos.
 - Altas.
 - Bajas.
 - Cambios.
- Conclusiones generales.

MÓDULO II.-SQL: EL LENGUAJE PARA DESARROLLOS OLTP (40 horas).

Software: Postgres.

Temas.

- Introducción al lenguaje SQL.
- Consulta de datos.
 - Estructura básica.
 - Operaciones básicas.
 - Operaciones de conjuntos.
 - Trabajo con valores nulos.
 - Funciones de agregación.
 - Subconsultas.
- Actualización de datos.
 - Altas.
 - Bajas.
 - Cambios.
- SQL intermedio.
 - Expresiones Join.
 - Vistas en el modelo relacional.
 - Transacciones.
 - Programación en bases de datos (SP).
 - Disparadores.
- Temas avanzados.
 - Funciones de agregación personalizadas.
 - Funciones de ventana.
 - Optimización de consultas.

MÓDULO III.- ALMACENES DE DATOS (45 horas).

Software: Pentaho, SQLServer.

Temas.

- Conceptos generales de los almacenes de datos (DWH).
- Arquitectura básica.
- Estrategias para extracción, transformación y carga de datos (ETL).
- Diseño de un almacén de datos.
- Análisis OLAP.
- Explotación del DWH utilizando el lenguaje MDX.
- Generación de reportes.

MÓDULO IV.- ANÁLISIS EXPLORATORIO DE DATOS (35 horas).

Software: R.

Temas.

- Escala de medición de las variables.
 - Escala nominal.
 - Escala ordinal.
 - De intervalo o de razón.
 - Numérica de conteo.
- Estadística Descriptiva.
 - Resumen mediante medidas numéricas.
 - Medidas de tendencia central: Media, mediana, moda.
 - Medidas de dispersión: Varianza, desviación estándar, rango, rango intercuartil.
 - Medidas de posición: Cuartiles, deciles, percentiles.
 - Medidas de forma: Sesgo y curtosis.
 - Resumen mediante gráficas.
 - Histogramas.
 - Diagrama de tallo y hoja.
 - Gráficas de barras.
 - Gráficas circulares o de pie.
 - Box-plot.
 - Gráficas de dispersiones.
 - Medidas de asociación.
 - Variables numéricas: Correlación de Pearson, correlación de Spearman, Tau de Kendall.
 - Variables categóricas: Tablas de contingencia.
 - Coefficiente de contingencia, coeficiente ϕ , V de Cramer.
- Análisis Multivariado.
 - Introducción al análisis multivariado.
 - Resumen mediante descripciones gráficas.

Gráficas de la matriz de datos.

Diagramas de estrellas.

Caritas de Chernoff.

Análisis de Componentes Principales (C.P.).

C.P. utilizando la matriz de correlación.

C.P. utilizando la matriz de varianza-covarianza.

El biplot.

Análisis discriminante.

Funciones lineales discriminantes de Fisher.

Discriminante clásico.

Discriminante logístico.

MÓDULO V.- MINERÍA DE DATOS I (30 horas).

Software: Rapidminer, Weka, R.

Temas.

- Introducción.
- Metodología CRISP_DM.
- Procesamiento de datos.
- Reglas de asociación.
- Árboles de decisión.

MÓDULO VI.- MINERÍA DE DATOS II (35 horas).

Software: Rapidminer, Weka, R.

Temas.

- Redes neuronales.
- Agrupación (Clusterin).
- Vecinos más cercanos.
- Detección de datos atípicos.
- Evaluación de Modelos.

MÓDULO VII.- MINERÍA DE TEXTOS (30 horas).

Software: R.

Temas.

- Introducción a la minería de textos. Áreas de aplicación de la minería de textos
- Recuperación de información. Medidas básicas para evaluar un conjunto de resultados recuperados (precisión y relevancia). Máquinas de búsqueda. Retroalimentación.
- Procesamiento de lenguaje natural. Representación de textos.
- Minado y análisis de asociaciones de palabras.
- Análisis de tópicos.
- Minado de opiniones y análisis de sentimientos.
- Predicción con base en textos.
- Visualización.