

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN  
SECRETARÍA DE EXTENSIÓN UNIVERSITARIA  
Y VINCULACIÓN INSTITUCIONAL  
CENTRO DE EDUCACIÓN CONTINUA



## DIPLOMADO CIENCIA DE DATOS

Coordinador académico:  
**Mtro. José Gustavo Fuentes Cabrera**

Agosto 2018

# ESTRUCTURA DEL DIPLOMADO

## Objetivo general:

Aplicar las técnicas y conocimientos de ciencia de datos en el estado del arte con el propósito de desarrollar en el alumno la capacidad de generar valor en las organizaciones, a través del análisis profundo de grandes volúmenes de datos.

## Módulo I Introducción, Manipulación, Exploración y Visualización de Datos

**Duración:** 48 horas

**Objetivo del módulo:** Aplicar las diferentes técnicas de manipulación, exploración y visualización de datos mediante el uso de distintas librerías gráficas para interpretar visualmente datos digitales.

## Temas:

- 1.1 La ciencia de datos en el mundo actual y perspectivas a futuro
- 1.2 Habilidades necesarias del científico de datos
- 1.3 Tendencias de software para manipulación, visualización y explotación de datos
- 1.4 Calidad de datos.
- 1.5 Conversión de estructuras OLAP y OLTP a TAD, ingeniería de características
- 1.6 Visualización de datos
- 1.7 Análisis exploratorio univariante de variables continuas y discretas
- 1.8 Técnicas de imputación de valores ausentes
- 1.9 Reducción de dimensionalidad
  - 1.9.1 Análisis de multicolinealidad
  - 1.9.2 Análisis de componentes principales
  - 1.9.3 Análisis de varianza explicada
  - 1.9.4 Importancia de variables
  - 1.9.5 Poder predictivo de características: Transformación entrópica

## **Módulo II**

### **Modelación Supervisada**

**Duración:** 48 horas

**Objetivo del módulo:** Emplear técnicas de modelación supervisada para desarrollar modelos estadísticos predictivos para tareas de clasificación y regresión comparando sus ventajas y desventajas, a través de software especializado.

#### **Temas:**

- 2.1 Metodología de modelación supervisada y métricas de ajuste
- 2.2 Modelos lineales
  - 2.2.1 Regresión Lineal
  - 2.2.2 Regresión Logística
  - 2.2.3 Regresión LARS
  - 2.2.4 Red elástica
  - 2.2.5 Regresión Lazo
  - 2.2.6 Regresión de cresta bayesiana
  - 2.2.7 Regresión bayesiana
- 2.3 Gradiente estocástico descendiente
- 2.4 Análisis Discriminante
- 2.5 Regresión de cresta Kernel
- 2.6 Máquinas Vector Soporte
- 2.7 Vecinos más cercanos
- 2.8 Bayes Ingenuo
- 2.9 Árboles de decisión
- 2.10 Redes Neuronales
- 2.11 Ensamblados
  - 2.11.1 Bosque Aleatorio
  - 2.11.2 ADABOOST
  - 2.11.3 Impulso de Árboles Gradiente
  - 2.11.4 Clasificador Votante
  - 2.11.5 Impulso gradiente extremo
- 2.12 Reportes de estabilidad y desempeño de modelos supervisados
  - 2.12.1 Estabilidad de la población
  - 2.12.2 Desempeño de poder predictivo
  - 2.12.3 Estabilidad de características
  - 2.12.4 Desempeño de características

## **Módulo III**

### **Modelación no supervisada**

**Duración:** 48 horas

**Objetivo del módulo:** Construir modelos estadísticos de clasificación a partir de datos no etiquetados que permitan perfilar segmentos subyacentes en grandes volúmenes de datos.

#### **Temas:**

- 3.1 Metodología de modelación no supervisada y métricas de ajuste
- 3.2 Visualización de grupos en dimensión reducida
  - 3.2.1 Análisis de componentes principales
  - 3.2.2 Escalamiento multidimensional
- 3.3 Clustering jerárquico
  - 3.1 Método de promedio
  - 3.2 Método de centroide
  - 3.3 Método de Ward
- 3.4 Clustering de optimización
  - 3.4.1 K-Medias
- 3.5 Clustering de densidad
  - 3.5.1 Modelos Gaussianos Mixtos
  - 3.5.2 Clustering espacial basado en densidad de aplicaciones con ruido (DBSCAN)
- 3.6 Perfilamiento de grupos
- 3.7 Reportes de estabilidad

## **Módulo IV**

### **Técnicas cognitivas e introducción a BigData**

**Duración:** 48 horas

**Objetivo del módulo:** Examinar las técnicas tendencia en ciencia de datos (aprendizaje profundo y bigdata), su futuro y uso a corto plazo, a través de la revisión de sus fundamentos, características y funciones, aplicables a análisis de grandes volúmenes de datos e inteligencia artificial.

**Temas:**

- 4.1 Introducción
- 4.2 Redes profundas alimentadas hacia adelante
- 4.3 Redes neurales convolucionales
- 4.4 Modelado secuencial
  - 4.4.1 Redes recurrentes
  - 4.4.2 Redes recursivas
- 4.5 Introducción a Bigdata
- 4.6 Herramientas para procesamiento y explotación de BigData.
- 4.7 Herramientas de procesamiento paralelo de datos.

**Evaluación:**

- |  |     |
|--|-----|
| • Examen práctico                              | 20% |
| • Reporte de solución de casos prácticos       | 40% |
| • Exposición de avances de proyecto individual | 40% |

## **Módulo V**

### **Módulo Integrador**

**Duración:** 48 horas

**Objetivo del módulo:** Desarrollar una aplicación mediante los conocimientos adquiridos durante el diplomado con el objeto de dar solución a un problema real dentro de la industria.

**Temas:**

5.1 Proceso preliminar del desarrollo de un proyecto de ciencia de datos

5.1.1 Determinación de datos digitales disponibles

5.1.2 Delimitación del problema

5.1.3 Elección de tratamientos previos

5.1.4 Elección de la técnica de modelación

5.2 Extracción de datos

5.3 Análisis exploratorio de datos

5.4 Modelación estadística

5.5 Comunicación de resultados

5.6 Implementación productiva

5.7 Mantenimiento y recalibración