

## TEMARIO

### **Clase 1: Introducción General y Proceso de Extracción de Conocimiento**

- Presentación del curso y cuerpo docente. Presentación de alumnos. Calendario. Forma de aprobación
- Motivación del Data Mining y Evolución de los sistemas para la toma de decisiones
- Minería de datos y descubrimiento de conocimiento. Definiciones y conceptos.
- Dominios de aplicaciones y ejemplos.
- Presentación de las diferentes técnicas de Data Mining. Aprendizaje supervisado y no supervisado.
- El proceso de extracción de conocimiento. Etapas.
- Presentación del Trabajo Práctico

### **Clase 2: Exploración y preparación de datos (feature engineering /data cleansing)**

- Tipos de atributos (categóricos, ordinales, continuos).
- Análisis exploratorio de datos. Medidas estadísticas resumen (media, mediana, ...). Visualizaciones (Histogramas, box plot, matriz de dispersión, coordenadas paralelas, diagramas estrella, etc)
- Calidad de datos. Problemas (Ruido, outliers, valores faltantes)
- Preparación de datos (Agregaciones, Muestreo, Reducción de la dimensionalidad, Selección de atributos, Creación de atributos, Discretización, Transformación de atributos. Escalar atributos. Variables dummies.
- Ejemplos y Ejercicios.

### **Clase 3: Métodos de Clasificación. Árboles de decisión**

- Definición. Características.
- Ejemplo de aplicación del modelo.

- Mecanismo de construcción.
- Condiciones de corte según el tipo de atributo.
- Homogeneidad e Impureza. Entropía. Criterios de parada.
- Sobreajuste. Poda.
- Métricas de evaluación
- Ejemplos y Ejercicios

#### **Clase 4: Métodos de Clasificación.**

- Bayes Ingenuo
  - Teorema de Bayes. Ejemplo de aplicación.
  - Clasificador Bayes Ingenuo. Características.
- Vecinos más cercanos
  - Características y consideraciones.
  - Elección del K.
  - Distancias.
  - Diagramas de Voronoi.
- Comparación de las técnicas
- Práctica. Ejemplos y Ejercicios

#### **Clase 5. Práctica sobre TP**

- Evaluación de modelos. Matriz de confusión, métricas, costos. Crossvalidation. Curvas ROC, etc
- Trabajo sobre el caso de negocio con las diferentes herramientas de SW.
- Objetivos. Consultas de avance. Creación de variables.

#### **Clase 6: Estadísticas Descriptivas y Análisis de componentes principales (PCA)**

- Repaso conceptual: Media. Mediana. Desvío. Varianza. Covarianza. Correlación.
- Análisis de componentes principales
- Práctica. Ejemplos y Ejercicios

## **Clase 7: Predicción Numérica. Regresión lineal simple y múltiple**

- Regresión lineal simple y múltiple.
- Función de costo y error cuadrático medio.
- Estimación por Cuadrados Mínimos.
- Residuos. Medidas de ajuste
- Variables dummy.
- Interacciones.
- Ridge regression.
- Vecinos más cercanos y árboles para regresión.
- Práctica. Ejemplos, Ejercicios y TP

## **Clase 8: Regresión Logística**

- Modelo de Regresión logística
- Cross-entropy
- Interpretación de los coeficientes. Significación de variables. Intervalo de Confianza. Selección de variables. Comparación de modelos
- Práctica. Ejemplos, Ejercicios y TP

## **Clase 9: Redes Neuronales**

- Definición. Similitudes entre neuronas biológicas y artificiales. Estructura de procesamiento de una RNA.
- Algoritmo de aprendizaje. Funciones de activación.
- Arquitecturas. Perceptrón. Backpropagation.
- Características. Dificultades
- Deep Learning
- Práctica. Ejemplos, Ejercicios y TP

## **Clase 10: Ensamblados y selección de modelos**

- Tradeoff entre sesgo y varianza
- Votación por mayoría.
- Boosting. Bagging. Random forest. Stacking. Ensamblados personalizados.
- Comparación de modelos. Model selection.
- Optimización. Hiperparámetros
- Implementación del modelo.
- Práctica. Ejemplos, Ejercicios y TP

### **Clase 11: Reglas de Asociación**

- Definiciones.
- Construcción de las reglas. Algoritmo Apriori.
- Evaluación (Soporte, Confianza, Lift, ...)
- Distintos tipos de asociaciones.
- Ejemplos de trabajos realizados
- Práctica. Ejemplos y Ejercicios.

### **Clase 12: Métodos de Aglomeramiento (Clustering)**

- Métodos jerárquicos y no jerárquicos (Kmeans).
- Distancias.
- Construcción de los clusters.
- Ventajas y limitaciones de cada técnica.
- Evaluación.
- Ejemplos de agrupamientos óptimos y problemáticos.
- Ejemplos de trabajos realizados utilizando clustering.
- Práctica. Ejemplos y Ejercicios.

### **Clase 13: Presentación de otras herramientas de DM.**

- Presentación de diferentes herramientas del mercado (IBM SPSS Modeler, SAS, etc). Ejemplos de modelos desarrollados.

- Consultas sobre TP

#### **Clase 14: Tópicos relacionados**

- Ejemplos de aplicaciones y tópicos relacionados (Social Network Analysis, Sistemas de Recomendación, Series Temporales, Text Mining, Sentiment Analysis, Visualización, etc)

#### **Clase 15: Exposición de los Trabajos Prácticos**

- Presentación y defensa del trabajo práctico integrador
- Repaso de temas para el examen

#### **Clase 16: Examen**

- Examen múltiple choice.
- Revisión de respuestas.
- Cierre.

### **Cronograma de Clases**

1. Introducción a DM y KDD. Presentación del Trabajo Práctico
2. Exploración y preparación de datos
3. Árboles de decisión
4. Bayes Ingenuo y Vecinos más cercanos
5. Trabajo sobre TP
6. Repaso estadísticas descriptivas. PCA
7. Regresión lineal simple y múltiple
8. Regresión Logística
9. Redes Neuronales
10. Ensamble. Evaluación. Implementación
11. Reglas de Asociación
12. Métodos de Aglomeramiento (Clustering)
13. Otras herramientas de DM
14. Tópicos relacionados
15. Presentación de TP
16. Examen