

Pautas para la evaluación

- El examen es individual y según el tema asignado en planilla.
- El examen se desarrolla de 16:00 a 18:30 horas, con 15 minutos adicionales para preparar el documento final en formato pdf y cargarlo al Aula Virtual.
- **Prever:**
 - o **Cámara del equipo deberá estar encendida sí o sí todo el tiempo.**
 - o **Tener a mano una identificación para mostrarla a la cámara cuando se le pida.**
 - o **Colocar su Nombre y Apellido en el ZOOM (no apodos, ni alias).**
 - o **Asegurarse en cuanto a la calidad de la/las fotos que envíen para realizar la evaluación (que sea legible).**
- Durante todo el tiempo, los alumnos deberán estar conectados a una videoconferencia que se habilitará mediante Zoom (asegurarse de contar con el equipamiento y conexión de internet adecuados).
- Desarrollo completo, agregar copias de planillas electrónicas y fotos al documento que remitan.
- En el plazo establecido, convertir el archivo a formato **PDF**, renombrarlo con el formato **BD2.Apellido.nombre.SegundoParcial.pdf**
- Cada alumno, deberá cargar al Aula Virtual de la cátedra, en el plazo de tiempo establecido el documento final del examen parcial elaborado, quien no cumpla con este paso quedará fuera de la evaluación y deberá de recuperar en la siguiente instancia de ese parcial.

Marcar con una **X** la opción u opciones correctas, en el casillero ():

- 1- Las principales formas de validación cruzada de la calidad predictiva de un árbol de decisión, son:
 - A) **Validación cruzada en dos mitades** ()
 - B) Validación cruzada ad hoc ()
 - C) **Validación cruzada en v partes** ()
 - D) **Validación cruzada global** ()
 - E) Ninguna es correcta ()
- 2- Respecto a las herramientas para el trabajo con árboles de decisión, podemos mencionar:
 - A) **Familia Cart** ()
 - B) **Método de tipo discriminante** ()
 - C) Método de tipo múltiple-complejo ()
 - D) **Modelos híbridos** ()
 - E) Ninguna es correcta ()
- 3- Para la creación y aplicación de una red neuronal a un problema concreto se deben distinguir los siguientes pasos:
 - A) Conceptualización del modelo para el estudio del modelo concreto. ()
 - B) **Adecuación de la información de que se dispone a la estructura de la red a crear.** ()
 - C) Se constituirán los patrones de aprendizaje. ()
 - D) **Se constituirán los patrones de validación** ()
 - E) Ninguna es correcta. ()
- 4- Las Redes Neuronales:
 - A) **Tratan de emular el sistema nervioso** ()
 - B) Reflejan todas las características fundamentales del comportamiento del cerebro ()
 - C) **Intenta modelizar una de las estructuras fisiológicas de soporte del cerebro, la neurona** ()
 - D) Construyen sistemas que presentan un alto grado de inteligencia ()
 - E) Ninguna es correcta ()
- 5- Respecto a los sistemas ROLAP:
 - A) **Se implementa sobre tecnología relacional** ()
 - B) Disponen de técnicas de compactación de datos ()
 - C) Disponen de estructuras de almacenamiento específicas como Arrays ()

- D) Almacena físicamente los datos en estructuras multidimensionales de forma que la representación externa e interna coincidan. ()
 E) Ninguna es correcta. ()

6- Entre las áreas de aplicación de la minería de datos, podemos mencionar:

- A) Medicina ()
 B) Astronomía ()
 C) Políticas demográficas ()
 D) Economía ()
 E) Ninguna es correcta ()

7- En el proceso KDD, algunas fases son:

- A) Estudio de la Minería ()
 B) Evaluación y validación ()
 C) Normalización de datos ()
 D) Interpretación y difusión ()
 E) Ninguna es correcta ()

8- Respecto a la fase de selección en KDD:

- A) Se implementan criterios de evaluación de hipótesis ()
 B) Comprende la recolección e integración de la información ()
 C) Uso de ROLAP o MOLAP ()
 D) Se utiliza análisis exploratorio y grafico de los datos ()
 E) Ninguna es correcta ()

9- Respecto a las técnicas descriptivas y predictivas de clasificación de minería de datos, entre las técnicas del análisis cluster, podemos mencionar:

- A) Métodos solapados-exclusivos ()
 B) Métodos directo-iterativos ()
 C) Métodos ponderados-no ponderados ()
 D) Métodos indirectos-iterativos ()
 E) Ninguna es correcta ()

10- Respecto a las técnicas descriptivas y predictivas de clasificación de minería de datos, entre las técnicas del análisis cluster no jerárquico, podemos agrupar los métodos como:

- A) Designación ()
 B) Reducción de dimensiones ()
 C) Indirectos ()
 D) Búsqueda de la densidad ()
 E) Ninguna es correcta ()

11- Respecto a las reglas de parada directa para detener la construcción de un árbol de decisión, podemos mencionar:

- A) Extensión máxima del árbol ()
 B) Máximo número de casos en un nodo ()
 C) Mínima fracción de objetos ()
 D) Determinación a priori de coeficiente ()
 E) Ninguna es correcta ()

12- Respecto a las técnicas predictivas de clasificación de minería de datos, incluye (entre otros):

- A) Series temporales ()
 B) Árboles de decisión ()
 C) Regresión ()
 D) Redes neuronales ()
 E) Ninguna es correcta ()

13- Describa el trabajo de monografía: "Fase de Limpieza y Transformación de Datos".

14) Considerando la siguiente consulta para el análisis de datos:

| Producto | Trimestre | Kg. vendidos |
|-----------|-----------|--------------|
| Cubiertas | T1 | 30000 |
| Cubiertas | T2 | 60000 |
| Cubiertas | T3 | 12000 |
| Cubiertas | T4 | 14000 |
| Baterías | T1 | 80000 |

| | | |
|----------|----|-------|
| Baterías | T2 | 12000 |
| Baterías | T3 | 16000 |
| Baterías | T4 | 20000 |

- a) Disgregar en 2 nuevos grupos cada Producto/Trimestre de la consulta original, considerando que las sucursales del Centro y Shopping conforman el universo estudiado, y las unidades vendidas se distribuyen en igual proporción en dichas sucursales.
- b) Presente matricialmente los datos seleccionados.
- c) Realice las operaciones necesarias para expresar la consulta obtenida en el punto a) en semestres.
- d) Represente el punto b) luego de utilizar el operador tipo Pivot.

15) Sea la siguiente definición del esquema anidado Comunas:

Comunas = (Nombre, Dirección, Teléfono, Funcionarios, Secretarías, Organigrama)

Funcionarios = (Nombre autoridad, Función)

Secretarías = (Nombre secretaría, Oficinas, Correo electrónico)

Oficinas = (Jefe, Subjefe)

Organigrama = (Jerarquías)

- a) Represente el esquema anidado Comunas con el formato de una tupla
- b) Realice la representación en árbol del esquema anidado Comunas

Puntajes:

| | | |
|--------------|-----------------|-------------|
| Item 1 a 12 | Bien contestada | 0,50 Puntos |
| | No contestada | 0 |
| Item 13 | 1 | Punto |
| Item 14 y 15 | 1,5 | Puntos c/u |

Respuestas:

13.

En la minería de datos el proceso de extracción del conocimiento incluye la fase de limpieza de datos, dado que la información extraída puede contener datos o valores faltantes o erróneos, incluso atípicos(outliers).

Luego se conoce la transformación de datos a través de técnicas de reducción.

Un valor *outlier*(atípico) es una puntuación extrema dentro de una variable y esto afecta gravemente en los análisis en los cuales se esté usando dicha variable.

Transformación de datos: Cuando el análisis exploratorio lo indique, los datos originales pueden necesitar ser remplazados. Suelen considerarse cuatro tipos de transformaciones:

1) *Transformaciones Lógicas*: Se unen características del campo de definición de variable para reducir así su amplitud.

2) *Transformaciones Lineales*: Se obtiene de sumar, restar, multiplicar o dividir las observaciones originales por una constante para mejorar su interpretación. Estas transformaciones no cambian la forma de la distribución ni la distancia entre los valores ni el orden, y por tanto no provocan cambios considerables entre las variables.

3) *Transformaciones Algebraicas*: Se obtiene de aplicar transformaciones no lineales monotónicas a las observaciones originales por una constante para mejorar su interpretación. Cambian la forma de distribución al cambiar las distancias entre los valores, pero mantienen el orden.

4) *Transformaciones no Lineales no Monotónicas*: Cambian las distancias y el orden entre los valores. Pueden cambiar demasiado la información original.

Las técnicas de **pareamiento o matching** persiguen la comparabilidad de grupos utilizando características comunes de todos ellos. Aunque los grupos difieren respecto a algunas de sus variables, es posible compararlos mediante un procedimiento de ajuste o estandarización.

Este procedimiento consiste en igualar ambos grupos con relación a alguna(s) característica(s), haciéndola homogénea en ambos grupos.

Un efecto importante de matching es el aumento en la eficiencia del estudio, ya que permite circunscribir la población a estudiar a aquella en la cual la exposición es más representativa.

15.

a) Comunas

| Nombre | Dirección | Teléfono | Funcionarios | | Secretarías | | | Organigrama |
|--------|-----------|----------|------------------|---------|-------------------|----------|--------------------|-------------|
| | | | Nombre autoridad | Función | Nombre secretaría | Oficinas | Correo Electrónico | |
| | | | Jefe | Subjefe | | | | |

b)



