

**DIPLOMADO Ciencia de Datos****Examen 2**

**Nombre:** Ibarra                      Ramírez                      Sergio                      **Fecha:** Mayo de 2023  
Apellido Paterno                      Apellido Materno                      Nombre(s)

**Calificación:** \_\_\_\_\_

**Objetivo:** *El participante identificará las ventajas y desventajas de las metodologías KDD, Agile design y CRISP para un caso de estudio proporcionado por el participante.*

**Instrucciones:**

1. Llenar cada uno de los recuadros siguientes explicando el caso de estudio a tratar, así como sus objetivos y los objetivos de minería de datos.
2. Llenar cada uno de los recuadros explicando las ventajas y desventajas al utilizar cada una de las metodologías aplicadas al caso de estudio elegido por el participante.

**Explicar el caso de estudio:****Valor 2.5 puntos**

Históricamente se ha dado importancia al pronóstico de demanda de hidrocarburos ya que son base de la economía, transporte y manufactura de cada país, por ello resulta de vital importancia contar con estudios y proyecciones sobre la demanda de dichos bienes. En este sentido se propone generar un sistema de pronóstico y recomendación de demanda de gas natural para el caso de México.

El principal objetivo es contar con varios métodos que permitan llevar a cabo un pronóstico de demanda de gas natural en los sectores eléctrico, petrolero y residencial para los siguientes 12-18 meses con el menor error posible.

**Mencionar las ventajas y desventajas utilizando la metodología KDD en el caso de estudio:****Valor 2.5 puntos**

	Ventajas	Desventajas
KDD	<ul style="list-style-type: none"><li>- Nos permitirá entender de manera profunda las variables predictoras, las posibles relaciones entre ellas y asegurar su independencia lineal</li><li>- Nos lleva de la mano para hacer una minería de datos correcta, en especial me preocupa entender las variables más significativas en el pronóstico del gas natural (se pretende usar PCA como modelo de aprendizaje no supervisado)</li><li>- Seremos capaces de extraer patrones poco vistos o difíciles de observar</li></ul>	<ul style="list-style-type: none"><li>- Podría ser un estudio muy específico de un conjunto determinado de datos y por lo tanto sufrir de "sobreajuste"</li><li>- Podrían reducirse o descartarse variables que resulten ser significativas para el modelo</li><li>- Podría ser un paso no tan necesario si ya se cuenta con datos relativamente claros, limpios, normalizados y ajustados</li></ul>

Mencionar las ventajas y desventajas utilizando la metodología Agile design en el caso de estudio:

Valor 2.5 puntos

	Ventajas	Desventajas
Agile design	<ul style="list-style-type: none"> <li>- Nos permitiría obtener un análisis de los datos y resultados relativamente rápidos</li> <li>- Dará una guía sencilla para obtener valor e <i>insights</i> de los datos, así como para llegar a pronosticar u optimizar un proceso</li> <li>- Dará la oportunidad de llevar a cabo “pequeños <i>sprints</i>” o experimentos para probar ágilmente ideas y metodologías distintas</li> </ul>	<ul style="list-style-type: none"> <li>- Se podría perder un poco de estructura y formalidad en cada unos de los pasos</li> <li>- Podría generar falta de documentación y/o justificación correcta de los informes o resultados a presentar</li> <li>- Solo el creador del dashboard / modelo será el experto en el mismo, será complicado que una persona ajena puede entender al 100% el proyecto</li> </ul>

Mencionar las ventajas y desventajas utilizando la metodología CRISP en el caso de estudio:

Valor 2.5 puntos

	Ventajas	Desventajas
CRISP	<ul style="list-style-type: none"> <li>- Permitirá ser muy organizado y meticuloso para no perderse entre la gran cantidad de información y pasos</li> <li>- Asegurará que cualquier persona que en un futuro lea el desarrollo del proyecto sea capaz de entender ¿qué se hizo? Y ¿Cómo se hizo? para llegar a los resultados que se dieron</li> <li>- Será la base de una documentación sólida que permita tener una presentación adecuada con los tomadores de decisiones y que haga el enlace correcto con el cliente final</li> </ul>	<ul style="list-style-type: none"> <li>- Debido a la gran cantidad de pasos y a la granularidad de las instrucciones podría llegar a ser algo burocrático / poco ágil</li> <li>- Puede requerir de mucho tiempo y recursos para ser correctamente documentado</li> <li>- Requiere que al menos una persona en el equipo esté realmente familiarizada con la metodología (de preferencia certificada) y en este caso yo seré ese experto</li> <li>- Si el proyecto no es factible o “usable” en el futuro, habrá representado una perdida grande de recurso</li> </ul>

Valor total 10 puntos