**CUESTIONARIO METODOLOGÍA CRISP-DM**

**Computación Blanda – 2020-02**

**AGOSTO 31 DE 2020**

**NOMBRE: Genny Paola Rivera Becerra CÓDIGO: 1087561571**

**1. ¿Cuál es la fase que se considera probablemente la más importante de la metodología?**

R= La fase de comprensión del negocio o problema.

**2. Haga una lista de las tareas generales de la fase de comprensión del negocio**

R=

* **Determinar los objetivos del negocio:** Esta es la primera tarea a desarrollar y tiene como metas, determinar cuál es el problema que se desea resolver, porque la necesidad de utilizar Data Mining y definir los criterios de éxito.
* **Evaluación de la situación**: Esta tarea se debe calificar el estado de la situación antes de iniciar el proceso DM, considerando aspectos tales como: ¿cuál es el conocimiento previo disponible acerca del problema?, ¿se cuenta con la cantidad de datos requerida para resolver el problema?, ¿cuál es la relación coste beneficio de la aplicación de DM?, etc.
* **Determinación de los objetivos de DM**: Esta tarea tiene como objetivo representar los objetivos del negocio en términos de las metas del proyecto de DM.

**3. ¿En qué fase se establece el plan del proyecto?**

R= Se establece el plan del proyecto en la fase de comprensión del negocio.

**4. ¿En qué fase se define el coste y beneficios del negocio?**

R= El coste y beneficio se define en la fase de comprensión del negocio.

**5. Definir los criterios de éxito del negocio, ¿es una tarea general o específica?**

R= Evaluación del modelo: En esta tarea, los ingenieros de DM interpretan los modelos de acuerdo al conocimiento preexistente del dominio y los criterios de éxito preestablecidos. Es una tarea general ya que esta fase evalúa un modelo mediante el cual se puede emplear múltiples herramientas para la interpretación de los resultados.

**6. ¿Cuáles son las fases que demandan mayor esfuerzo y tiempo en la metodología DM?**

R= Las fases que demandan mayor esfuerzo y tiempo son: Comprensión de los datos, Preparación de los datos y Modelado.

**7. Enuncie las tareas específicas que deben realizarse en la verificación de la calidad de los datos.**

R= Determinar la consistencia de los valores de los datos individuales de los campos.

* Determinar la cantidad y distribución de los valores nulos
* Encontrar valores fuera de rango

**8. ¿En qué fase se realiza la limpieza de los datos?**

R= Fase de preparación de los datos

**9. ¿En qué fase se encuentra la necesidad (potencial) de la creación de nuevas estructuras de datos?**

R= Fase de compresión de los datos

Exploración de datos: Su fin es encontrar una estructura general para los datos. Esto involucra la aplicación de pruebas estadísticas básicas, que revelen propiedades en los datos recién adquiridos, se crean tablas de frecuencia y se construyen gráficos de distribución.

**10. ¿En qué fase se realiza la selección de la técnica de DM más apropiada?**

R= Fase de preparación de los datos

**11. Si se determina que un modelo produce resultados no satisfactorios en relación con los objetivos del negocio, ¿En qué fase se ha realizado dicho descubrimiento?**

R= Fase de comprensión del negocio o problema.

**12. ¿A qué tarea general corresponde la actividad específica: “Plan de monitorización y mantenimiento?**

R= Generalmente un proyecto de Data Mining, no concluye en la implantación del modelo, se deben documentar y presentar los resultados de manera comprensible para el usuario, con el objetivo de lograr un incremento del conocimiento.

Por otra parte, en la fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados.

Plan de implementación: Para implementar el resultado de DM en la organización, esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación. Monitorización y mantenimiento.