



TRABAJO PRÁCTICO NRO 3

MINERIA DE DATOS



FECHA DE ENTREGA: 15 DE NOVIEMBRE DE 2024

AUTOR: CADER FERNANDA

PROFESOR: REA OSCAR

Consignas: De acuerdo a las clases vistas:

1. ¿Cuáles con las medidas que se usan para conocer la calidad cuando se evaluación de reglas de asociación?
2. ¿Para que se usan la combinación de modelos en las técnicas de Minería de datos?
3. ¿Qué es minería de datos espaciales?
4. Explique cada una de las áreas de la minería web
5. Cuales consideran que son las claves de éxito de un programa de minería de datos?

Desarrollo

1. Medidas para evaluar la calidad de reglas de asociación:

- **Cobertura (o soporte):** Representa la cantidad de instancias donde la regla puede ser aplicada.
- **Confianza (o precisión):** Indica el porcentaje de veces que la regla se cumple cuando se puede aplicar.
- **Interés:** Mide la capacidad de una regla de atraer la atención del usuario.
- **Novedad:** Evalúa la capacidad de la regla de sorprender al usuario en relación a su conocimiento previo del problema.

2. Combinación de Modelos en Minería de Datos:

La combinación de modelos, o métodos multclasificadores, se utiliza para **mejorar la precisión de las predicciones**. Esto se logra construyendo un conjunto de hipótesis (ensemble) y combinando sus predicciones, usualmente a través de un sistema de votación. La precisión de esta combinación suele superar la de cada modelo individual.

Las claves para un modelo combinado exitoso son:

- **Precisión:** Los modelos individuales deben ser precisos.
- **Diversidad:** Los modelos deben ser diferentes entre sí para que sus errores no estén correlacionados.

3. Minería de Datos Espaciales:

La minería de datos espaciales se define como la extracción de conocimiento a partir de **bases de datos que contienen información sobre un espacio determinado**. Algunos ejemplos incluyen bases de datos sobre inmuebles en una ciudad, habitaciones de un hotel, o la distribución de moléculas en una proteína. Un concepto fundamental es la **dimensión espacio**, que establece el marco de referencia para ubicar los datos.

4. Áreas de la Minería Web:

- **Minería del Contenido:** Se centra en descubrir información útil a partir del contenido textual y gráfico de las páginas web. Sus raíces se encuentran en el procesamiento del lenguaje natural y la recuperación de información.
- **Minería de la Estructura:** Busca descubrir el modelo subyacente en la estructura de enlaces de la web. Analiza la topología de los hipervínculos para categorizar páginas web, determinar similitudes y relaciones entre sitios, detectar páginas de autoridad, y estudiar topologías.
- **Minería del Uso:** Analiza información sobre los accesos a la web disponible en los servidores. Examina datos como archivos de registro del servidor, del navegador y proxy, perfiles de usuario, cookies, y patrones de navegación.

5. Claves de Éxito de un Programa de Minería de Datos:

- **Orientación al Negocio:** El programa debe estar guiado por las necesidades y objetivos del negocio.
- **Especificación Clara de Problemas:** Definir problemas específicos de minería de datos con objetivos claros y la calidad de modelo requerida.
- **Integración y Apoyo:** Integrar el programa con otros sistemas y obtener apoyo de la dirección.
- **Calidad de los Datos:** Asegurar datos limpios y de calidad, e integrar información externa relevante.
- **Herramientas Adecuadas:** Utilizar herramientas integradas y fáciles de usar, e informatizar el proceso.
- **Equipo Capacitado:** Contar con un equipo diverso y formado en minería de datos, estadística, bases de datos, y el área de negocio.
- **Evaluación y Despliegue:** Evaluar los modelos de manera integral y desplegarlos en todos los niveles de la organización.