

Programa Curso de Posgrado Data Mining (Minería de Datos)

Contenidos:

Introducción. Obtención de conocimiento a partir de los datos. El concepto de patrón. El proceso KDD. Fases del proceso de extracción del conocimiento. La Minería de Datos

como fase del proceso KDD. Relación con otras disciplinas.

- Recuperación de información vs. recuperación de datos. Proceso de recuperación de información.

- Preparación de Datos. Metadatos. Análisis de la información de entrada. Construcción y análisis de representaciones gráficas. Limpieza y transformación. Transformación y creación de atributos. Discretización y Numerización, Normalización de rango, escalado y centrado. Exploración mediante visualización y selección de datos.

- Técnicas de Minería de Datos. Extracción de Patrones. Introducción. Tareas y Métodos. Tareas predictivas y descriptivas. Aprendizaje supervisado y aprendizaje no supervisado. La Minería de Datos y el aprendizaje inductivo. Comparación de las técnicas de Minería de Datos.

- Árboles de decisión. Métricas de selección de atributos. Entropía. Ganancia de Información. Tasa de Ganancia. Índice Gini. Poda y Sobreajuste. Algoritmos Id3, C4.5 y Random Forest. Construcción de árboles para grandes volúmenes de datos.

- Reglas de clasificación. Partición vs cobertura. Métodos ZeroR, OneR, PRISM y PART. Métricas de una regla: soporte, cobertura, confianza, interés y convicción.