

Data Mining y Big Data

Contenido Académico

Introducción a la minería de datos

¿Qué es y qué no es Data Mining? Aplicaciones de la minería de datos en las organizaciones: predicciones bursátiles, marketing, detección de fraudes, patrones de fuga, producción, mantenimiento, logística, ciencia. Introducción a las herramientas de minería de datos: métodos predictivos y descriptivos. Tareas de Data Mining: clasificación, agrupamiento (*clustering*), descubrimiento de reglas de asociación, descubrimiento de patrones secuenciales, regresión, detección de anomalías.

Selección y pre-procesamiento de datos

Introducción a los proyectos de inteligencia de negocios (Business Intelligence). Manejo de datos e información. Bases de datos (Data Warehousing). Selección e integración de datos. Pre-procesamiento de datos: limpieza, integración, transformación, reducción, y discretización de datos.

Clasificación

Ejemplos de problemas de clasificación. Árboles de decisión. Clasificadores basados en reglas. Métodos bayesianos: Naïve Bayes. Máquinas de soporte vectorial (*Support Vector Machines*). Redes Neuronales: perceptrón multicapa. Vecinos más cercanos. Implementación de clasificadores con paquete de software libre WEKA (Universidad de Waikato, Nueva Zelanda).

Clustering

Aplicaciones del análisis de clustering: descriptiva y resumida. Clasificación de clusters: jerárquico/particional, exclusivos/no-exclusivos, fuzzy/no-fuzzy, parcial/completo,

heterogéneo/homogéneo. Tipo de clusters: bien separados, basados en centros, contiguos, basados en densidad, conceptuales, descritos por una función objetivo. Medidas de similaridad. Algoritmos de clustering: vecinos más cercanos, clustering jerárquico y basado en densidad. Análisis de casos.

Aspectos prácticos

Aspectos prácticos al implementar clasificadores: sobreajuste, subajuste, valores faltantes, expresividad, evaluación y comparación de modelos (métricas y métodos). Reducción de la dimensionalidad. Desbalanceo de clases.

Multimedia Data Mining

Aplicación de herramientas de Data Mining en datos provenientes de imágenes, videos, audio y habla.

Casos de aplicación de Data Mining

Introducción al paquete de software libre "R" (*R Foundation for Statistical Computing*). Aplicación de los métodos de regresión múltiple y regresión LASSO. Aplicación de redes neuronales a un proceso de negocios. Determinación de las causas de falla empleando la metodología de Clustering. Aplicación de la reducción de dimensionalidad en un proceso industrial.

Big Data

¿Qué es Big Data? Gestión de grandes volúmenes de datos ¿Cómo trabaja Map-Reduce? Entorno Open-Source de software para procesamiento y almacenamiento distribuido: Apache Hadoop. Algoritmos de Data Mining en Big Data. Computación iterativa distribuida. Entorno para el procesamiento en cluster de computadoras: Apache Spark.

Trabajo final

El trabajo final se realiza en forma grupal y tiene como finalidad que los participantes apliquen los conocimientos adquiridos duran

te el curso en la resolución de problemas de casos reales de sus organizaciones. Algunos trabajos presentados en ediciones anteriores:

- Optimización del diseño de productos siderúrgicos para el cumplimiento de especificaciones de atributos cualitativos.
- Optimización en el proceso de resolución de siniestros de daños materiales en automotores mediante el Sistema CLEAS.
- Análisis de datos antropométricos del cráneo humano: estrategias de clasificación y modelado.
- Determinación de la flota de equipos mineros más eficiente para la extracción de material desde un determinado PIT en Santa Cruz.
- Determinación de áreas inundadas a través del uso de imágenes satelitales y fotografías aéreas.
- Aplicación de técnicas de Data Mining para la detección de operaciones sospechosas de Lavado de Dinero.
- Reducción de tiempos en el proceso de cobranza de una empresa de venta de equipos de informática.
- Reducción de siniestros en una empresa de transporte público de pasajeros.
- Segmentación de compradores de camas náuticas en una guardería de lanchas.
- Optimización de parámetros de diseño de recubrimientos comestibles de alimentos.
- Detección de Inicios de Calidad en Venta Directa
- Mantenimiento de flota logística
- Parámetros de Inyección de pieza plástica vs pieza no conforme de calidad
- Predicción de compra de clientes