

PROGRAMA DE LA ASIGNATURA

MINERÍA DE DATOS UTILIZANDO SISTEMAS INTELIGENTES (INFORMÁTICA UNLP)

CARRERA: GEOFÍSICA

CARGA HORARIA SEMANAL: 2 HORAS DE TEORÍA Y 2 HORAS DE PRÁCTICA

CARÁCTER: CUATRIMESTRAL

PROFESOR A CARGO: LAURA LANZARINI

CONTENIDO TEMÁTICO

·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. Medidas estadísticas básicas. Construcción y análisis de representaciones gráficas. Medidas de similitud entre atributos y entre ejemplos. 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.

·Técnicas de Agrupamiento. Métricas de calidad del agrupamiento. Tipos de agrupamiento: Jerárquico, particivo y probabilista. Medidas de distancia y de conectividad. Proceso de agrupamiento. Clustering particivo. Algoritmo k-medias. Algoritmos de clustering jerárquicos aglomerativos y divisivos. Dendrogramas. Algoritmo probabilista EM (Expectation – Maximization)

·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. Algoritmos de Boosting. AdaBoost.

·Clasificadores bayesianos. Teorema de Bayes. Hipótesis máxima a posteriori. Clasificador Naïve Bayes. Ejemplos.

·Reglas de clasificación. Partición vs cobertura. Métodos ZeroR, OneR, PRISM, PART y CN2. Métricas de una regla: soporte, cobertura, confianza, interés y convicción.

·Reglas de asociación. Calidad de las reglas. Algoritmo A priori. Concepto de ítem frecuente. Mejoras del algoritmo a priori: FP-Growth y FP-Tree.

·Redes Neuronales Feedforward. Descripción de la arquitectura. Regla delta generalizada. Algoritmo de entrenamiento backpropagation. Incorporación del término de momento. Capacidad de generalización de la red. Resolución de problemas de clasificación y predicción.

·Análisis y difusión del modelo obtenido. Evaluación de modelos. Comparación de técnicas de aprendizaje. Evaluación y mejora del modelo obtenido. Performance del modelo. Matriz de confusión. Sensibilidad, especificidad, precisión y recall. F measure. Visualización utilizando curvas ROC.