

PRIMER PARCIAL

Requerimientos generales a ser cumplidos en el ejercicio de aplicación:

Etapas del proceso

- Análisis del problema planteado – contexto del problema.
- Identificación del tipo de problema (ej: supervisado / no supervisado, regresión, clasificación, agrupación)
- Análisis de los datasets y preparación de los datos (todas las actividades que sean necesarias y convenientes).
- Identificación de **dos (2) o más** enfoques de algoritmos de ML para resolver el problema.
 - Descripción de configuración de parámetros para cada modelo/algoritmo (justificación de cada una)
 - **NOTA: ¡DEBEN USARSE ALGORITMOS VISTOS DURANTE EL CURSO!**
- Descripción de los procesos y modelos creados, y ejecución de los mismos.
 - Descripción de las técnicas usadas para entrenamiento / prueba / validación
 - Ajuste de parámetros: aplicar técnicas para realizar ajuste de hiperparámetros de los modelos seleccionados, y documentar tanto estas técnicas como el resultado de la aplicación de las mismas (o sea, valores finales de los parámetros e impacto en los resultados de los modelos)
- Análisis de resultados
 - Matrices de confusión
 - Predicciones y estadísticas de confianza
 - Selección de mejor modelo / proceso y justificación.

Se pide documentar y remitir a la tarea correspondiente de la webasignatura:

- Documento con cada uno de estos requerimientos, análisis completos de alternativas, descripción de técnicas de ajuste de hiperparámetros, gráficos ilustrativos.
- Procesos completos de RapidMiner generados, completamente funcionales
- Planillas electrónicas eventualmente utilizadas

17 de Octubre de 2023

EJERCICIO - DIAGNÓSTICO DE ENFERMEDAD HEPATICA.

Los pacientes con enfermedad hepática han aumentado continuamente debido al consumo excesivo de alcohol, la inhalación de gases nocivos, la ingesta de alimentos contaminados, encurtidos y drogas. El objetivo de este trabajo es evaluar algoritmos de predicción en un esfuerzo por reducir la carga sobre los médicos.

Un dataset de ejemplo utilizado en múltiples trabajos es el que puede ser encontrado en el repositorio de UCI (<http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29#>).

En particular, este conjunto de datos contiene 416 registros de pacientes hepáticos y 167 registros de pacientes no hepáticos recopilados del noreste de Andhra Pradesh, India.

En la tarea podrás encontrar el dataset ya descargado.

TRABAJO A SER REALIZADO.

1. ANALISIS DEL PROBLEMA, CONTEXTO Y DOCUMENTACIÓN CORRESPONDIENTE

- a. Realiza una revisión sobre el problema de la enfermedad hepática, estadísticas (en diferentes países y regiones), el impacto económico y social asociado y otros problemas relacionados.

2. PREPARACIÓN DE LOS DATOS y ANÁLISIS PREVIO.

Describir todas las tareas pertinentes de preparación de los datos, y responder las siguientes preguntas (con las explicaciones y razonamientos correspondientes):

- a. ¿cómo es la distribución de la enfermedad en función de otros atributos relevantes?
- b. Describe las características de los atributos (físicas, estadísticas, correlaciones, etc.). ¿Cuál es el verdadero significado de cada atributo? ¿cuáles son los procedimientos y dificultades para obtener datos correctos de estos atributos?
- c. ¿cómo cambia la probabilidad de desarrollar la enfermedad en relación con otras enfermedades relacionadas (ej., diabetes, hipertensión), sexo o edad?

3. MODELOS PREDICTIVOS

- a. **Objetivo:** predecir el desarrollo de la enfermedad a partir de los atributos físicos existentes.
- b. Seleccionar 2 algoritmos apropiados para resolver el problema (entre los vistos en clase)
- c. Crear los procesos / modelos en RapidMiner correspondientes
- d. Identificar las técnicas de prueba y evaluación más apropiadas, describirlas
- e. Ejecutar los procesos y evaluaciones para cada modelo
- f. Consolidar los resultados comparativos de las evaluaciones, con las técnicas habituales, incluyendo todas las métricas que correspondan, y su interpretación
- g. Ilustrar con técnicas de visualización las relaciones entre atributos y los resultados obtenidos

NOTA IMPORTANTE: RECUERDA REMITIR PROCESOS Y DATASETS EN FORMATOS APROPIADOS PARA SER ABIERTOS EN RAPIDMINER

REMITIR A LA TAREA "PARCIAL 1" un archivo comprimido conteniendo todo lo realizado (documentos, modelos, planillas, etc.)