

Autor: Iván García Santillán

Semana (1)

Preparando y entendiendo los datos para la Predicción del cáncer de mama

Objetivo y alcance del trabajo

Esta práctica tiene el objetivo de realizar un preprocesamiento de datos y un Análisis exploratorio de datos (estadísticas y visualización) de las variables más relevantes que considere para el propósito de predicción de cáncer de mama. Al finalizar la práctica, los maestrantes podrán manipular adecuadamente un dataset de datos estructurados (alfanuméricos).

Explicación del Trabajo

Utilice el dataset de cáncer de mama (Diagnostic Wisconsin Breast Cancer Database, WBCD) donde tenemos los datos de 569 mujeres. Cada mujer está descrita por 32 atributos. El primero es un identificador, el segundo el tipo de cáncer (Maligno o Benigno) y el resto son el resultado de otros análisis clínicos. Se pretende aprender el tipo de cáncer de mama (maligno=1, benigno=0). En este conjunto de datos la distribución de clases es: 357 benignos y 212 malignos.

Fase I

Se pide al maestrante realizar las siguientes actividades:

- Descargue el dataset:
<http://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>
- Cargue el conjunto de datos en el entorno de Python
- Realizar el preprocesamiento de datos:
 - Sustituir las M's por 1 y las B's por 0.
 - Descarte la primera columna (el identificador) ya que no aporta nada al problema:
 - Verifique valores perdidos.

Fase II

Se pide al maestrante realizar las siguientes actividades:

- Realizar un Análisis exploratorio de datos (estadísticas y visualización) de las variables más relevantes que considere para el propósito de predicción de cáncer de mama.
- Interprete los resultados obtenidos en un informe conciso en Word o en dentro del mismo notebook de Python.
- Suba el notebook de Python y/o el informe Word al LMS (plataforma).

ANEXOS

Rúbrica de Evaluación de la práctica

Rúbrica de calificación de la práctica:

Criterio	Excelente	Bueno	Aceptable	Bajo
Análisis exploratorio de datos	El análisis exploratorio de datos es relevante, detallado y adecuado para la comprensión de los datos y el negocio incluyendo un preprocesamiento de datos, estadísticas y visualización de datos.	El análisis exploratorio de datos es adecuado para la comprensión de los datos y el negocio incluyendo un preprocesamiento de datos, estadísticas y visualización de datos.	El análisis exploratorio de datos es adecuado para la comprensión de los datos y el negocio, pero podría mejorarse en la variedad de técnicas, estadísticas y visualizaciones.	El análisis exploratorio de datos no es adecuado para la comprensión de los datos y el negocio y debe mejorarse significativamente.
Elaboración de informe / notebook	El informe presentado es relevante, claro y conciso incluyendo todas las estadísticas y visualizaciones de datos.	El informe presentado es adecuado y conciso incluyendo varias estadísticas y visualizaciones de datos.	El informe presentado es adecuado incluyendo algunas estadísticas y visualizaciones de datos, pero podría complementarse con otras.	El informe presentado no es adecuado, faltando estadísticas y visualizaciones de datos.
Presentación del caso práctico	La presentación del caso práctico es clara, completa y concisa utilizando material visual en el tiempo asignado demostrando un alto nivel de creatividad.	La presentación del caso práctico es clara y concisa utilizando material visual en el tiempo asignado.	La presentación del caso práctico es clara utilizando material visual en el tiempo asignado, pero podría profundizarse en aspectos relevantes.	La presentación del caso práctico no es clara, no utiliza material visual adecuado y no respeta el tiempo asignado.