

Autor: Iván García Santillán

Predicción del cáncer de mama / supervivencia en el Titanic usando todos los algoritmos predictivos estudiados

Objetivo y alcance del trabajo

Esta práctica tiene el objetivo de realizar la predicción del tipo de cáncer de mama (maligno=1, benigno=0) utilizando todos los algoritmos de clasificación revisados de regresión logística, K-NN, árbol de decisión, bosque aleatorio y redes neuronales artificiales, así como realizar la evaluación del rendimiento de los algoritmos usando las métricas y gráficas respectivas. Al finalizar la práctica, los maestrantes podrán entender y manipular adecuadamente los diferentes parámetros e hiperparámetros de los algoritmos estudiados.

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. Adicionalmente, utilice el dataset del **Titanic** analizado en las primeras sesiones.

Fase I

Se pide al maestrante realizar las siguientes actividades:

- Descargue los datasets de Internet o cargue directamente desde alguna librería.
- Cargue el conjunto de datos en el entorno de Python
- Realizar el preprocesamiento de datos y Análisis exploratorio de datos (estadísticas y visualización) como se indicó en la unidad anterior.
- Implementar, entrenar y afinar los algoritmos estudiados usando Scikit-Learn y TensorFlow.

- Evaluar los algoritmos usando las métricas y gráficas de rendimiento básicas.
- Realice una comparativa de todos los algoritmos basado en el Accuracy usando un bar chart e interprete el rendimiento de ellos en cada dataset.
- Suba el notebook de Python y/o el informe Word al LMS (plataforma) indicando los integrantes.
- Realice la presentación en 10-15 min (un representante del grupo).

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.	
Entrenamiento del modelo	El entrenamiento del modelo de clasificación es adecuado incluyendo todos los algoritmos revisados, el afinamiento y selección de los mejores parámetros.	El entrenamiento del modelo de clasificación es adecuado incluyendo varios algoritmos revisados, el afinamiento y selección de los mejores parámetros.	El entrenamiento del modelo de clasificación es adecuado incluyendo algunos de los algoritmos revisados, pero podría mejorarse considerando el afinamiento y selección de los mejores parámetros.	El entrenamiento del modelo de clasificación no es adecuado con métricas de evaluación bajas.	
Elaboración de informe	El informe presentado es relevante, claro y conciso incluyendo todas las métricas y gráficas de evaluación del modelo entrenado.	El informe presentado es adecuado y conciso incluyendo varias métricas y gráficas de evaluación del modelo entrenado.	El informe presentado es adecuado incluyendo algunas métricas y gráficas de evaluación del modelo entrenado, pero podría	El informe presentado no es adecuado, faltando métricas y gráficas de evaluación del modelo entrenado.	

			complementarse con otras adicionales.	
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.