Alumno: Johan Callomamani Buendia

Seccion: A

Pregunta 01:  
- ¿Qué resultados se han encontrado de la exploración y pre-procesamiento de los datos? Explicar.

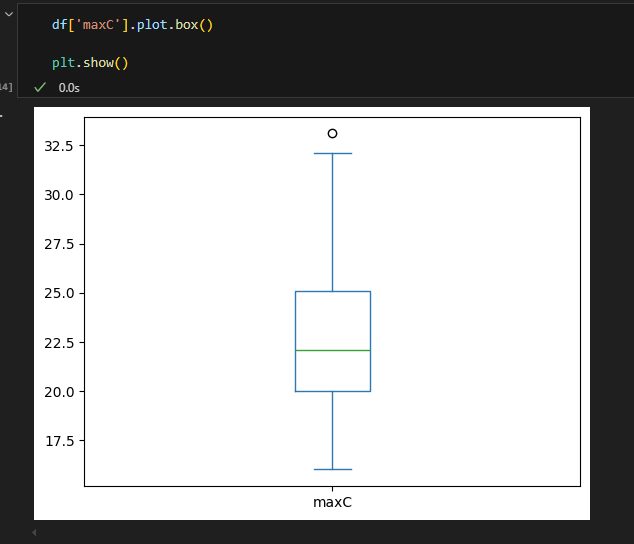
Al procesar los datos encontramos que tenemos 2 variables input que no agregan valor al modelo, “name” y “datetime”, para el caso de name y datetime se eliminaron las variables.

A screenshot of a computer program

Description automatically generated

Se graficaron histogramas de todas las variables restante, generando un entendimiento de las variables y tomando la decisión de quitar los outlayers solo de las variables de maxC y minC.

Antes



Ahora:

A screen shot of a computer

Description automatically generated

- ¿Qué es una matriz de confusión? Y expresar los resultados obtenidos.

Una matriz de confusión nos permite evaluar el rendimiento de un modelo, realizando una combinación de cada posible predicción y valores reales.

Para nuestro caso trabajamos con 2 modelos:

* Árbol de decisión

A screenshot of a computer

Description automatically generated

* Radom forest

A screenshot of a computer

Description automatically generated

- Aplicar los modelos de clasificación revisados en el curso y explicar el mejor modelo obtenido con los mejores parámetros (explicar el resultado de las métricas obtenidas)

- ¿Qué criterios se puede considerar para comparar los modelos obtenidos? Seleccionar el mejor modelo.

Para el criterio se puede considerar la precisión y la exactitud.

Para nuestro caso sabemos que el random forest tiene accuracy = 0.69 en comparación a decisión tree con accuracy = 0.58, demostrando que radom forest es el mejor modelo para esta data y objetivo de predicción.

- Identificar las variables que tienen mayor relevancia para predecir el clima.