**EVALUACIÓN 2**

**MINERÍA DE DATOS MDY7101**

**OBJETIVOS**

Desarrollo, documentación y análisis de los resultados de un programa con la aplicación de modelos de Machine Learning en Python, para extraer conocimiento sobre el dominio el caso Climático de la Entrega 1.

**INSTRUCCIONES:**

* Trabajo grupal 3-5 integrantes.
* Lee atentamente la especificación del caso entregado por el docente.
* **Importar los datos depurados del clima de Australia que obtuviste como resultado de la Unidad 1**
* Realizar la exploración de los datos para verificar y validar que las conjeturas realizadas en la fase de limpieza de datos fueron implementadas correctamente, y en congruencia con las decisiones tomadas en el preprocesamiento de los datos. Si producto de este análisis exploratorio debes hacer cambios, deberán quedar documentados en el informe.
* Una vez que los datos estén depurados y validados, realizar la inicialización y separación de las variables dependientes (X) de la variable dependiente (Y). Considerar para sus modelos de regresión **un dato continuo como su variable objetivo, por ejemplo, temperatura o cantidad de lluvia.**
* Implementa una rutina que permita entrenar un **modelo de regresión** con los algoritmos de machine learning (de los que vimos hasta el momento). Fundamentar el análisis y explicar los parámetros utilizados para su aplicación.
* Confeccionar una tabla con las métricas de rendimiento obtenidas de los algoritmos de regresión seleccionados. De la misma tabla obtener las conclusiones y explicar los resultados en lenguaje natural.

|  |  |  |  |
| --- | --- | --- | --- |
| **Algoritmo** | **R2** | **RMSE** | **Validación Cruzada** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

* Una vez que el algoritmo está entrenado realiza la predicción en base al dataset proporcionado. Realizar un análisis de las predicciones en relación de los datos de testeo, apoyándose en gráficos comparativos o tablas cruzadas.

Presenta las conclusiones (al menos 5) obtenidas en función de las predicciones y los hallazgos del modelo supervisado obtenido, todo integrado para que los Directivos puedan entender la información y apoyarlos en la toma de decisiones.

**Entregables:**

* Cuaderno desarrollado en Jupyter Notebook con el código Python ejecutado y debidamente documentado
* Documento Word en formato DuocUC con evidencia documentada por cada paso solicitado en esta pauta y con todas las decisiones debidamente fundamentadas.