

## Documentación del Repositorio

### Conclusión General del Trabajo Práctico Integrador

El trabajo práctico integrador de la asignatura de Aprendizaje Automático (AA1) realizado por Gauto y GraviFiorino demuestra una sólida aplicación de técnicas de Machine Learning en el análisis de datos meteorológicos. A continuación se detallan los principales puntos y hallazgos del trabajo:

#### Objetivos del Trabajo

El objetivo principal de este proyecto fue aplicar diversas técnicas de aprendizaje automático para predecir eventos meteorológicos utilizando un conjunto de datos reales de Australia. Los subobjetivos incluyeron la limpieza y preprocesamiento de datos, el análisis exploratorio de datos (EDA), el entrenamiento de modelos tanto de clasificación como de regresión, y la evaluación y despliegue de dichos modelos.

#### Preprocesamiento de Datos

El preprocesamiento de datos es una etapa crítica en cualquier proyecto de Machine Learning. En este trabajo, se llevaron a cabo las siguientes tareas:

- Limpieza de datos: Eliminación de valores nulos y tratamiento de datos faltantes.
- Transformación de datos: Conversión de variables categóricas a numéricas mediante técnicas como el One-Hot Encoding.
- Normalización: Escalado de las características para mejorar el rendimiento de los algoritmos de aprendizaje.

#### Análisis Exploratorio de Datos (EDA)

El EDA permitió una comprensión profunda del conjunto de datos. Se exploraron las distribuciones de las variables, se identificaron patrones y se visualizaron las relaciones entre diferentes

## Documentación del Repositorio

características. Este análisis fue esencial para guiar la selección de modelos y la ingeniería de características.

### Entrenamiento de Modelos

Se entrenaron y compararon múltiples modelos de Machine Learning para abordar dos tareas principales: clasificación y regresión.

- Modelos de Clasificación: Se utilizaron algoritmos como el Árbol de Decisión, Random Forest y Support Vector Machine (SVM) para predecir si lloverá al día siguiente. El modelo Random Forest mostró el mejor rendimiento con una alta precisión y recall.
- Modelos de Regresión: Se entrenaron modelos para predecir la cantidad de lluvia utilizando algoritmos como la Regresión Lineal, Ridge y Lasso. La Regresión Ridge mostró el mejor equilibrio entre sesgo y varianza, proporcionando predicciones precisas y estables.

### Evaluación de Modelos

Los modelos se evaluaron utilizando métricas estándar como la precisión, recall, F1-score para clasificación, y el error cuadrático medio (MSE) y R-cuadrado para regresión. La evaluación demostró que los modelos seleccionados eran robustos y capaces de generalizar bien a nuevos datos.

### Despliegue del Modelo

Finalmente, los modelos entrenados se desplegaron en una aplicación Python utilizando Flask. Esta aplicación permite a los usuarios ingresar nuevas características meteorológicas y obtener predicciones en tiempo real tanto de la probabilidad de lluvia como de la cantidad esperada de lluvia.

## Documentación del Repositorio

### Hallazgos Clave

- El preprocesamiento adecuado de los datos es fundamental para el rendimiento de los modelos.
- La combinación de técnicas de EDA y la selección adecuada de características mejora significativamente la capacidad predictiva de los modelos.
- Los algoritmos de ensemble como Random Forest son muy efectivos para tareas de clasificación en contextos meteorológicos.
- La regularización (como en Ridge y Lasso) es crucial para modelos de regresión al manejar datos con muchas características correlacionadas.

### Conclusión

Este trabajo práctico integrador demuestra una implementación efectiva y rigurosa de técnicas de Machine Learning en un problema del mundo real. Los modelos desarrollados ofrecen herramientas útiles para la predicción meteorológica, con aplicaciones potenciales en agricultura, gestión de recursos hídricos y planificación de eventos. El proyecto no solo resalta la importancia de cada etapa del flujo de trabajo de Machine Learning, sino que también subraya la necesidad de una evaluación continua y un despliegue adecuado para asegurar que los modelos proporcionen valor en aplicaciones prácticas.

Este trabajo es un ejemplo de cómo los métodos de Machine Learning pueden ser aplicados para resolver problemas complejos y proporcionar insights valiosos a partir de datos.