

Instrucciones para el Trabajo Práctico Final de Machine Learning II

Objetivo: Este trabajo práctico tiene como objetivo aplicar los conocimientos adquiridos en el curso para resolver un problema de Machine Learning mediante un enfoque estructurado. Utilizarás herramientas de análisis exploratorio, ingeniería de atributos, modelos de aprendizaje automático y optimización para presentar una solución completa en un Notebook (Colab o Jupyter).

Estructura del Notebook

Tu Notebook debe seguir la siguiente estructura y contener los elementos detallados:

1. **Motivación y audiencia**
 - Escribe una breve descripción sobre el conjunto de datos elegido y por qué decidiste analizarlo.
 - Identifica a quién podría beneficiar este análisis. Por ejemplo, ¿es útil para una empresa, una comunidad científica, o para decisiones personales?
 - Incluye la relevancia del problema y el impacto esperado de su solución.
2. **Planteamiento del problema**
 - Define las preguntas o problemas que deseas resolver.
3. **Análisis Exploratorio de Datos (EDA)**
 - Realiza un análisis descriptivo inicial de los datos utilizando herramientas estadísticas y visualizaciones gráficas.
 - Identifica patrones, distribuciones y posibles problemas en los datos.
 - Analiza valores faltantes o inconsistencias en el conjunto de datos y documenta cómo los abordarás.
4. **Ingeniería de atributos**
 - Genera nuevas variables relevantes o transforma las existentes.
 - Aplica técnicas como:
 - Normalización o estandarización de variables numéricas.
 - Codificación de variables categóricas (e.g., one-hot encoding, label encoding).
 - Cualquier otra técnica relevante para preparar los datos para los modelos.
5. **Entrenamiento y testeo de modelos**
 - Implementa y compara al menos **dos modelos diferentes de Machine Learning**.
 - Divide tus datos en conjuntos de entrenamiento y prueba utilizando un método adecuado (e.g., validación cruzada).
 - Explica brevemente las características de los modelos seleccionados y justifica su elección.
6. **Optimización de hiperparámetros**
 - Usa alguna técnica de optimización de hiperparámetros, como **GridSearchCV** o **RandomizedSearchCV**, para mejorar el rendimiento de al menos uno de los modelos.

- Documenta los resultados y el impacto de la optimización en el rendimiento del modelo.
7. **Selección del modelo**
- Evalúa los modelos utilizando métricas adecuadas para el problema planteado (e.g., AUC para clasificación, MSE para regresión).
 - Compara los modelos y selecciona el mejor basándote en los resultados de las métricas.
 - Justifica claramente tu selección del modelo final.
-

Requisitos Técnicos

1. Utiliza Python y bibliotecas comunes como `pandas`, `numpy`, `matplotlib`, `seaborn`, y `scikit-learn`.
 2. Incluye celdas de código y celdas explicativas para guiar al lector a través de tu análisis y conclusiones.
 3. El Notebook debe ser **limpio, bien documentado y organizado**, con explicaciones claras para cada paso realizado.
-

Entrega

- **Formato:** Notebook en formato `.ipynb`
 - **Criterios de evaluación:**
 - Estructura y claridad del Notebook.
 - Profundidad del análisis exploratorio y preparación de datos.
 - Correcta implementación de los modelos y la validación.
 - Uso adecuado de técnicas de optimización y selección del modelo.
 - Justificación y claridad de las conclusiones.
-

Recursos

- Revisa la documentación oficial de las herramientas mencionadas.
- Usa datasets públicos de plataformas como Kaggle, UCI Machine Learning Repository o datasets disponibles en `sklearn` para practicar.