|  |
| --- |
| **Nombre del proyecto** |
| **Clasificación Supervisada - Human Activity Recognition** |

|  |  |
| --- | --- |
| **Profesor** | **Gladys María Villegas Rugel** |
| **Materia** | **Aprendizaje Automático** |
| **Alumno** | **- MARÍA AUGUSTA FLORES**  **- MARCELO ISMAEL ANDRADE**  **- JORGE ANDRÉS ORELLANA**  **- GALO VLADIMIR GONZÁLEZ** |
| **Fecha** | **Mayo 02 del 2025** |

|  |
| --- |
| **Resumen del Proyecto** |
| **Introducción**  El presente proyecto tiene como objetivo implementar, entrenar y comparar modelos de clasificación supervisada aplicados a un conjunto de datos técnico real: el Human Activity Recognition Dataset, el cual contiene medidas recolectadas de sensores en dispositivos móviles mientras usuarios realizan distintas actividades físicas.  **Dataset Utilizado**   * **Nombre:** Human Activity Recognition Using Smartphones Dataset (UCI Machine Learning Repository) * **Observaciones:** 7352 muestras de entrenamiento con 561 variables predictoras * **Variable objetivo:** tipo de actividad física (categórica: 1 a 6) * **Ubicación del dataset:** Carpeta data/UCI HAR Dataset   **Objetivos de la Actividad**   * Implementar modelos de clasificación supervisada (SVM, Árbol de Decisión, Regresión Logística). * Evaluar el rendimiento de los modelos mediante F1-score, matriz de confusión y reporte de clasificación. * Comparar gráficamente el desempeño de los clasificadores   **Documentar el proceso completo con código reproducible Dataset Utilizado**   * **Nombre:** Human Activity Recognition Using Smartphones Dataset (UCI Machine Learning Repository) * **Observaciones:** 7352 muestras de entrenamiento con 561 variables predictoras * **Variable objetivo:** Tipo de actividad física (categórica: 1 a 6) * **Ubicación del dataset:** Carpeta data/UCI HAR Dataset  **Metodología Aplicada** El desarrollo se dividió en tres notebooks:  **1. EDA\_actividad1.ipynb:** Análisis exploratorio, correlaciones y visualización. **2. Preprocesamiento\_actividad1.ipynb:** Limpieza, escalado, división de datos. **3. Modelado\_actividad1.ipynb:** Entrenamiento y evaluación de modelos supervisados. **Resultados Obtenidos** A continuación, se presentan las matrices de confusión para los tres modelos evaluados:  **📌 Árbol de Decisión**    **📌 SVM (Kernel RBF)**    **📌 Regresión Logística**    **📌 Comparación de F1-Score Macro entre modelos**    **Resumen de F1-score (macro):**   |  |  | | --- | --- | | Modelo | F1-score (Macro) | | Regresión Logística | 0.9812 | | SVM (RBF) | 0.9798 | | Árbol de Decisión | 0.9399 |  **Conclusiones**  * El modelo de Regresión Logística obtuvo el mejor rendimiento, aunque SVM también mostró una precisión muy alta. * El Árbol de Decisión fue menos preciso, pero más interpretable. * El preprocesamiento (escalado, limpieza y partición) fue fundamental para el rendimiento. * Este ejercicio permite visualizar el poder de los modelos supervisados aplicados a datos reales de sensores.  **📚 Estructura del Proyecto** actividad1\_semana2/ ├── data/ → Dataset original (UCI HAR) ├── notebooks/ → Notebooks del desarrollo (EDA, preprocesamiento, modelado) ├── results/ → Gráficos y matrices de confusión generadas ├── src/ → Código modular adicional (si aplica) ├── docs/ → Documentación del proyecto └── README.md → Descripción general para GitHub |