Logotipo, nombre de la empresa

Descripción generada automáticamente

|  |
| --- |
| **Actividad 2** |
| Clasificación técnica mediante SVM y Árboles de Decisión |
| **Caso de estudio:**  Informe Técnico del Proyecto Clasificación Binaria para Señales de Sensores de Agua y Aceite. |

|  |  |
| --- | --- |
| **Profesora** | Gladys María Villegas Rugel |
| **Materia** | APRENDIZAJE AUTOMATICO |
| **Alumnos** | Bolaños Escandón María Fernanda Estupiñán Andrade Francisco Javier  Montaño Cárdenas Fernando Xavier |
| **Fecha** | 03/05/2025 |

Tabla de contenido

**1. Introducción3**

**2. Resumen3**

**3. Metodologías3**

Procesamiento de Datos3

Modelos de calificación3  
Evaluación3

**4. Resultados y Análisis4**

Matriz de Confusión - Árbol de Decisión 4  
Frontera de Confusión - Árbol de Decisión 4  
Matriz de Confusión - SVM5  
Frontera de Confusión - SVM6  
Matriz de Confusión - Regresión Logística8  
Frontera de Confusión - Regresión Logística8  
Comparación del F1-score por modelo 10  
Curva ROC - Regresión Logística10  
Métricas de los Modelos11

**5. Conclusiones12**

**1. Introducción**

El objetivo de este proyecto es desarrollar y evaluar modelos de clasificación para predecir si un sensor de agua y aceite puede detectar con precisión las condiciones del entorno a partir de sus lecturas. Se emplearon tres modelos de clasificación: **Árbol de Decisión**, **SVM (Máquinas de Vectores de Soporte)** y **Regresión Logística**. El análisis incluye el uso de un conjunto de datos con varias características relacionadas con los sensores de agua y aceite, y las métricas clave utilizadas para evaluar el rendimiento de estos modelos fueron **Precisión**, **Recall**, **F1-score** y **Curvas ROC**.

**2. Resumen**

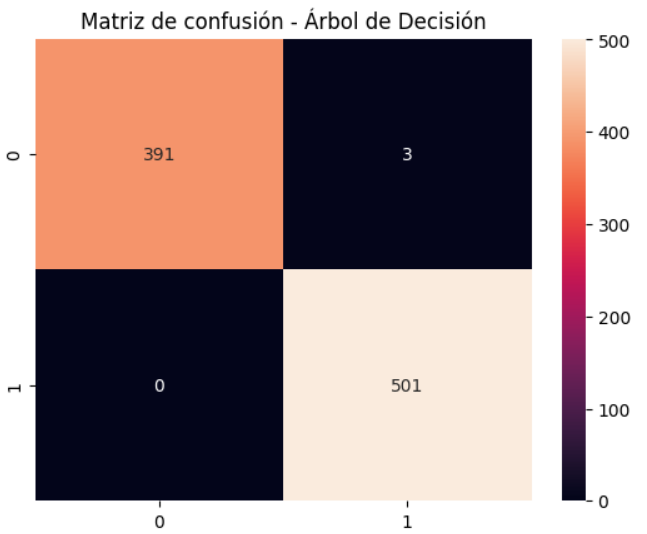
En este informe se presentan los resultados obtenidos de la implementación y evaluación de tres modelos de clasificación. Los datos provienen del archivo dataset.csv, que incluye diversas características que representan las lecturas de los sensores. Para los tres modelos, se analizaron sus métricas clave y su capacidad para separar las clases de forma efectiva. El modelo SVM se destacó en términos de capacidad de discriminación y rendimiento en general, aunque los otros dos modelos también mostraron buenos resultados.

**3. Metodologías**

* **Preprocesamiento de Datos:**
  + Carga de los datos desde dataset.csv, que utiliza el punto y coma como delimitador.
  + Normalización de las características para mejorar el rendimiento de los modelos.
  + División del conjunto de datos en un 80% para entrenamiento y un 20% para prueba.
* **Modelos de Clasificación:**
  + **Árbol de Decisión:** Utiliza una estructura de árbol binario para clasificar los datos.
  + **SVM (Máquinas de Vectores de Soporte):** Maximiza el margen de separación entre las clases.
  + **Regresión Logística:** Utiliza una función logística para predecir la probabilidad de pertenecer a una clase.
* **Evaluación:**
  + **Métricas de Evaluación:** Precisión, Recall, F1-score y Área bajo la curva ROC (AUC).
  + **Gráficas:** Incluyen matrices de confusión, fronteras de decisión y curvas ROC.

**4. Resultados y Análisis**

* **Matriz de Confusión - Árbol de Decisión**
  + **Análisis:** La **Matriz de Confusión** del **Árbol de Decisión** muestra una excelente clasificación, con solo 3 falsos positivos y ningún falso negativo.

  
***Ilustración 1.*** *Matriz de confusión- Árbol de decisión.*

* **Frontera de Decisión - Árbol de Decisión**
  + **Análisis:** La **Frontera de Decisión** del **Árbol de Decisión** muestra una separación clara entre las clases, aunque con algunas áreas de solapamiento.

Gráfico

Descripción generada automáticamente  
***Ilustración 2.*** *Frontera de Decisión – Árbol de Decisión.*

* **Matriz de Confusión - SVM**
  + **Análisis:** La **Matriz de Confusión** de **SVM** muestra un desempeño muy sólido con solo 12 falsos negativos y 15 falsos positivos, lo que indica un buen rendimiento, aunque con ligeramente más errores que el Árbol de Decisión.

Gráfico

Descripción generada automáticamente  
***Ilustración 3.*** *Matriz de Confusión - SVM*

* **Frontera de Decisión - SVM**
  + **Análisis:** La **Frontera de Decisión** de **SVM** es no lineal y permite una separación más precisa entre las clases, mostrando un margen más amplio y una mayor capacidad para manejar solapamientos en las clases.

Gráfico

Descripción generada automáticamente  
***Ilustración 4.*** *Frontera de Decisión – SVM.*

* **Matriz de Confusión - Regresión Logística**
  + **Análisis:** La **Matriz de Confusión** de **Regresión Logística** muestra algunos errores de clasificación (16 falsos negativos y 11 falsos positivos), aunque sigue siendo un modelo con un buen rendimiento general.

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente  
***Ilustración 5.*** *Matriz de Confusión – Regresión Logística.*

* **Frontera de Decisión - Regresión Logística**
  + **Análisis:** La **Frontera de Decisión** de la **Regresión Logística** es lineal, lo que es característico de este modelo. La separación es clara, pero algunos puntos aún pueden clasificarse incorrectamente en áreas de solapamiento.

Gráfico

Descripción generada automáticamente  
***Ilustración 6.*** *Frontera de Decisión – Regresión Logística.*

* **Comparación del F1-score por modelo**
  + **Análisis:** El modelo **Árbol de Decisión** tiene el F1-score más alto, lo que sugiere un equilibrio más favorable entre precisión y recall en comparación con los otros modelos. Sin embargo, los modelos **SVM** y **Regresión Logística** también muestran un rendimiento similar.

Gráfico, Gráfico de barras

Descripción generada automáticamente  
***Ilustración 7.*** *Comparación del F1-score por modelo.*

* **Curva ROC - Regresión Logística**
  + **Análisis:** La curva ROC para **Regresión Logística** muestra un área bajo la curva (AUC) muy cercano a 1, lo que indica que el modelo es muy eficiente para diferenciar entre las clases.

Gráfico

Descripción generada automáticamente  
***Ilustración 8.*** *Curva ROC – Regresión Logística.*

* **Métricas de los Modelos**
  + **Análisis:** El **Árbol de Decisión** muestra una precisión ligeramente superior a los otros modelos. Sin embargo, todos los modelos muestran métricas muy altas, lo que indica un rendimiento generalizado y consistente.

Gráfico

Descripción generada automáticamente  
***Ilustración 9.*** *Métricas de los modelos.*

**5. Conclusiones**

El análisis de los tres modelos de clasificación (Árbol de Decisión, SVM y Regresión Logística) ha mostrado que todos son adecuados para la tarea de clasificación de sensores de agua y aceite, con altas métricas de rendimiento. Sin embargo, el modelo **SVM** destacó por su capacidad para manejar la complejidad de la separación de clases no lineales, lo que resultó en un rendimiento superior en cuanto a la frontera de decisión y las métricas de precisión.

El **Árbol de Decisión** también presentó un excelente rendimiento con métricas muy altas, especialmente en términos de precisión y F1-score, y es un modelo sencillo de interpretar. Por último, la **Regresión Logística**, aunque útil, mostró algunos errores en las clasificaciones, especialmente en las clases solapadas.

En general, el modelo **SVM** se recomienda como el más robusto para este tipo de problema, seguido de cerca por el **Árbol de Decisión**, y con **Regresión Logística** siendo una opción adecuada cuando las clases son linealmente separables.