**ANÁLISIS DE MODELOS DE APRENDIZAJE AUTOMATICO PARA LA DETECCIÓN DE CÁNCER**

El presente proyecto analiza el desempeño de tres modelos de aprendizaje automático para la detección de cáncer utilizando un conjunto de datos especifico. Para este proyecto se realizó el análisis de los modelos Random Forest, SVM y Regresión Logística.

**DESCRIPCIÓN**

El objetivo principal de este proyecto es comparar y evaluar la eficacia de diferentes algoritmos de aprendizaje automático en la detección de cáncer. Se utilizan tres modelos principales: SVM, Random Forest y Regresión Logística, los cuales son entrenados y evaluados utilizando un conjunto de datos previamente recolectado.

**CONJUNTO DE DATOS**

El conjunto de datos utilizado en este análisis es Breast Cancer Wisconsin (Diagnostic).

Se proporciona un enlace al conjunto de datos en https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic.

**DEPENDENCIAS**

Para ejecutar el código proporcionado, se requiere tener instaladas las siguientes bibliotecas de Python:

* scikit-learn
* numpy
* pandas
* matplotlib
* seaborn

**RESULTADOS**

A continuación, se presentan los resultados obtenidos para cada modelo:

***Random Forest***

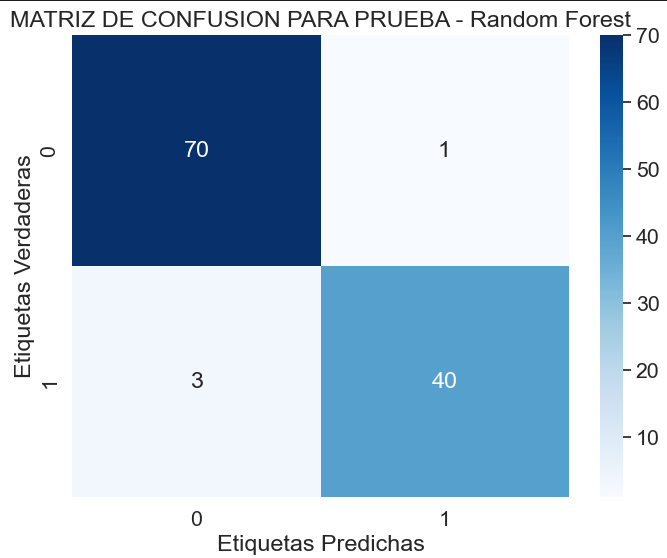
**Métricas de evaluación en el conjunto de prueba:**

**Precisión:** 0.975609756097561

**Recuerdo:** 0.9302325581395349

**F1-score:** 0.9523809523809524

**Precisión global:** 0.9649122807017544



**Interpretación de la matriz de confusión:**

Verdaderos Positivos (VP): 40

Verdaderos Negativos (VN): 70

Falsos Positivos (FP): 1

Falsos Negativos (FN): 3

***SVM***

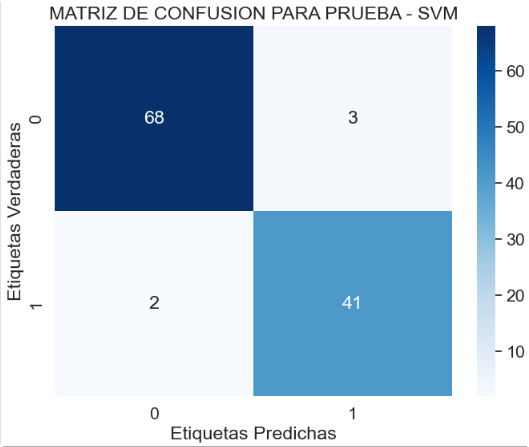
**Métricas de evaluación en el conjunto de prueba:**

**Precisión:** 0.9318181818181818

**Recuerdo:** 0.9534883720930233

**F1-score:** 0.942528735632184

**Precisión global:** 0.956140350877193



**Interpretación de la matriz de confusión:**

Verdaderos Positivos (VP): 41

Verdaderos Negativos (VN): 68

Falsos Positivos (FP): 3

Falsos Negativos (FN): 2

***Regresión Logística***

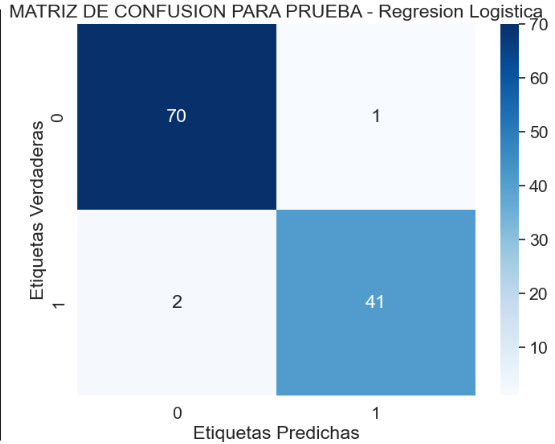
**Métricas de evaluación en el conjunto de prueba:**

**Precisión:** 0.9761904761904762

**Recuerdo:** 0.9534883720930233

**F1-score:** 0.9647058823529412

**Precisión global:** 0.9736842105263158



**Interpretación de la matriz de confusión:**

Verdaderos Positivos (VP): 41

Verdaderos Negativos (VN): 70

Falsos Positivos (FP): 1

Falsos Negativos (FN): 2

**INTERPRETACION**

En virtud a los datos obtenidos y tomando en cuenta los aspectos de mencionados como los falsos positivos y negativos que indican que un falso positivo no es tan preocupante como un falso negativo podemos determinar que los modelos de SVM y Regresión Logística son los modelos con menor valor obtenido en este aspecto.

|  |  |  |
| --- | --- | --- |
| **Random Forest** | **SVM** | **Regresion Logistica** |
| Falsos Positivos (FP): 1 | Falsos Positivos (FP): 3 | Falsos Positivos (FP): 1 |
| Falsos Negativos (FN): 3 | Falsos Negativos (FN): 2 | Falsos Negativos (FN): 2 |

Si la metrica que queremos evaluar es la precisión de cada modelo para determinar que modelo es el mas adecuado podemos concluir que el modelo de Regresion Logistica es el mejor.

***Random Forest***

**Métricas de evaluación en el conjunto de prueba:**

**Precisión:** 0.975609756097561

**Recuerdo:** 0.9302325581395349

**F1-score:** 0.9523809523809524

**Precisión global:** 0.9649122807017544

***SVM***

**Métricas de evaluación en el conjunto de prueba:**

**Precisión:** 0.9318181818181818

**Recuerdo:** 0.9534883720930233

**F1-score:** 0.942528735632184

**Precisión global:** 0.956140350877193

***Regresión Logística***

**Métricas de evaluación en el conjunto de prueba:**

**Precisión:** 0.9761904761904762

**Recuerdo:** 0.9534883720930233

**F1-score:** 0.9647058823529412

**Precisión global:** 0.9736842105263158

En general podemos determinar que el mejor modelo para detección de cáncer es el modelo de regresión logística, ya que, posee una precisión global mayor que los demás y sus valores tanto en falsos positivos como negativos es menor que los otros.