

Análisis y Reporte sobre el desempeño del modelo.

Alumno: José Antonio Moreno Tahuilán

Matrícula: A01747922

1. Introducción

El objetivo de este proyecto fue implementar y evaluar diferentes algoritmos de aprendizaje supervisado utilizando el framework scikit-learn en Python. Se buscó comparar el desempeño de tres modelos:

- Random Forest
- Support Vector Machine (SVM)
- Regresión Logística

Los experimentos se realizaron con tres datasets clásicos ampliamente usados en la literatura: Iris, Wine y Breast Cancer, los cuales permiten validar la capacidad de los modelos en tareas de clasificación multiclase y binaria.

2. Datasets Utilizados

1. Iris Dataset: Clasificación de 3 especies de flores basado en medidas de pétalos y sépalos.
 2. Wine Dataset: Clasificación de 3 tipos de vino con base en 13 características químicas.
- Breast Cancer Dataset: Clasificación binaria (benigno vs maligno) utilizando 30 características clínicas.

Cada dataset se dividió en train/test split para una primera evaluación, y posteriormente se aplicó validación cruzada con 10 folds para una medida más robusta del desempeño.

3. Metodología

- Framework: Python 3 + scikit-learn.
- Algoritmos: Random Forest, SVM, Logistic Regression.
- Validación: Comparación entre evaluación simple (train/test) y validación cruzada.
- Optimización: Se aplicó GridSearchCV para encontrar hiperparámetros óptimos.
- Métricas: Accuracy, Precision, Recall, F1-score y análisis con matriz de confusión.

4. Resultados

4.1 Iris Dataset

- Evaluación Simple: Los tres modelos alcanzaron 100% de accuracy.
- Validación Cruzada: Accuracy promedio entre 0.9533 y 0.9600, mostrando una ligera variación entre folds.
- Modelos Optimized con GridSearch: Mejora notable en Logistic Regression (hasta 0.98).

El dataset es relativamente sencillo y los tres algoritmos logran un desempeño excelente.

4.2 Wine Dataset

- Evaluación Simple: Accuracy de 100% en todos los modelos.
- Validación Cruzada: Accuracy promedio de 0.9778 a 0.9833.
- GridSearch: Mejoras menores, destacando Random Forest con 0.9944.

Todos los modelos se desempeñan muy bien. La optimización de hiperparámetros aporta pequeñas mejoras de precisión.

4.3 Breast Cancer Dataset

- Evaluación Simple: Accuracy entre 0.9649 y 0.9825.
- Validación Cruzada: Accuracy promedio de 0.9526 a 0.9772.
- GridSearch: Mejoras ligeras, con Logistic Regression alcanzando 0.9807.

SVM y Logistic Regression logran un rendimiento superior. Random Forest fue ligeramente menos consistente.

5. Matrices de Confusión y Métricas

En los tres datasets, las matrices de confusión confirmaron un excelente desempeño con muy pocos errores de clasificación. Las métricas Precision, Recall y F1-score se mantuvieron cercanas a 1.0 en la mayoría de los casos, reflejando un balance adecuado entre clases.

6. Diagnóstico de Bias, Varianza y Ajuste del Modelo

Bias (sesgo):

En los tres datasets el bias es **bajo**, ya que los modelos logran un desempeño cercano al 100% en entrenamiento y prueba, lo que indica que capturan bien las relaciones en los datos.

- **Iris:** En la evaluación simple (train/test split) los tres modelos tuvieron accuracy = 1.0. Esto significa que en entrenamiento los modelos clasificaron perfectamente, indicando bias muy bajo.
- **Wine:** También se obtuvo accuracy = 1.0 en train/test, por lo que el bias también es muy bajo.
- **Breast Cancer:** los resultados fueron ligeramente menores (RandomForest = 0.9649, SVM = 0.9825, LogisticRegression = 0.9737). Aunque no son perfectos, siguen siendo altos → bias bajo.

Varianza:

- En **Iris** y **Wine** la validación cruzada mostró fluctuaciones notables entre folds (rangos entre 0.8667 y 1.0), lo que refleja **varianza media**.
- En **Breast Cancer**, la desviación estándar entre folds fue menor (<0.05), indicando **varianza baja**.

Nivel de ajuste (underfit/fit/overfit):

Iris:

- Train/Test = 1.0
- Validación cruzada = 0.95–0.96
- Diferencia de ~0.04 a 0.05.

Ligero overfitting (el modelo se ajusta demasiado al conjunto de entrenamiento).

Wine:

- Train/Test = 1.0
- Validación cruzada = 0.97–0.98
- Diferencia de ~0.02 a 0.03.

También ligero overfitting, aunque menos pronunciado que en Iris.

Breast Cancer:

- Train/Test = 0.96–0.98
- Validación cruzada = 0.95–0.97
- Diferencia muy pequeña (0.003–0.007).

Modelo bien ajustado (fit), sin señales claras de underfitting ni overfitting.

Regularización y mejora con GridSearch:

La optimización de hiperparámetros permitió reducir la varianza y mejorar la estabilidad entre folds.

- **Logistic Regression** en Iris pasó de 0.9533 a 0.9800 (+2.8%).
- **Random Forest** en Wine pasó de 0.9778 a 0.9944 (+1.7%).
- **Logistic Regression** en Breast Cancer pasó de 0.9772 a 0.9807 (+0.36%).

Estas mejoras, aunque pequeñas, muestran que la regularización (penalizaciones L1/L2) y el ajuste de parámetros (profundidad de árboles, C en SVM, etc.) ayudan a mitigar el overfitting y a obtener modelos más robustos.

6. Análisis y Conclusiones

- Todos los algoritmos implementados alcanzaron un alto desempeño en los tres datasets.
- SVM y Logistic Regression mostraron mayor consistencia tras la validación cruzada.
- La optimización de hiperparámetros con GridSearch permitió mejoras pequeñas pero significativas, especialmente en Logistic Regression en Iris (0.98) y en Random Forest en Wine (0.9944).
- El uso de validación cruzada fue clave para observar variabilidad y evitar sobreajuste que no se percibe en la evaluación simple.

Los tres algoritmos son altamente efectivos para datasets clásicos de clasificación. Sin embargo, en escenarios más complejos y con datos menos balanceados, la elección del modelo y la correcta optimización de hiperparámetros se vuelve crítica.