Estrategia:

1. Se utilizaron técnicas de búsqueda de hiperparámetros como **GridSearchCV** o **RandomizedSearchCV** para encontrar los mejores hiperparámetros para cada algoritmo.
2. Se busco optimizar el modelo de **random forest** a través del cambio de sus parámetros, ya que mostraba una precisión del 100% en entrenamiento y solo una precisión del 77,62% en prueba, lo cual demuestra un problema de overfitting.
3. Se probo con el algoritmo de **Support Vector Machine** para mirar si ofrecía un grado de precisión más alto en entrenamiento y prueba, que regresión lineal y bosques aleatorios.

Resultados:

Mejores parámetros para Regresión Logística: {'C': 1.623776739188721, 'solver': 'liblinear'}

Mejores parámetros para Random Forest: {'max_depth': 10, 'max_features': 'log2', 'min_samples_leaf': 2, 'min_samples_split': 5, 'n_estimators': 500}

Mejores parámetros para SVM: {'C': 5, 'gamma': 0.01, 'kernel': 'rbf'}

Logistic Regression

- Train Accuracy: 77.56% | Test Accuracy: 76.05%

Random Forest:

- Train Accuracy: 89.04% | Test Accuracy: 76.95%

SVM

- Train Accuracy: 82.57% | Test Accuracy: 76.38%

Cohen's Kappa (Logistic Regression vs Random Forest): 0.79

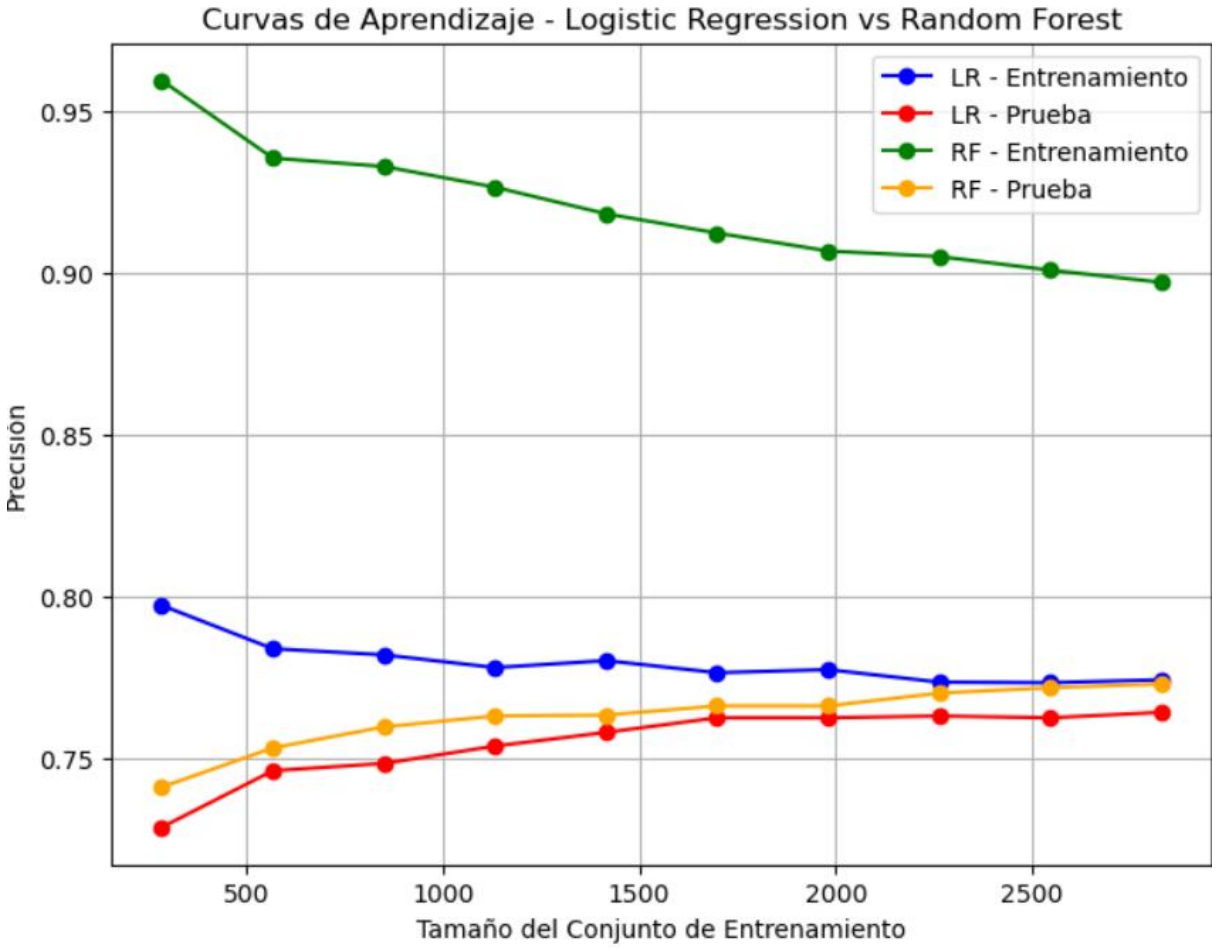
Cohen's Kappa (Logistic Regression vs SVM): 0.82

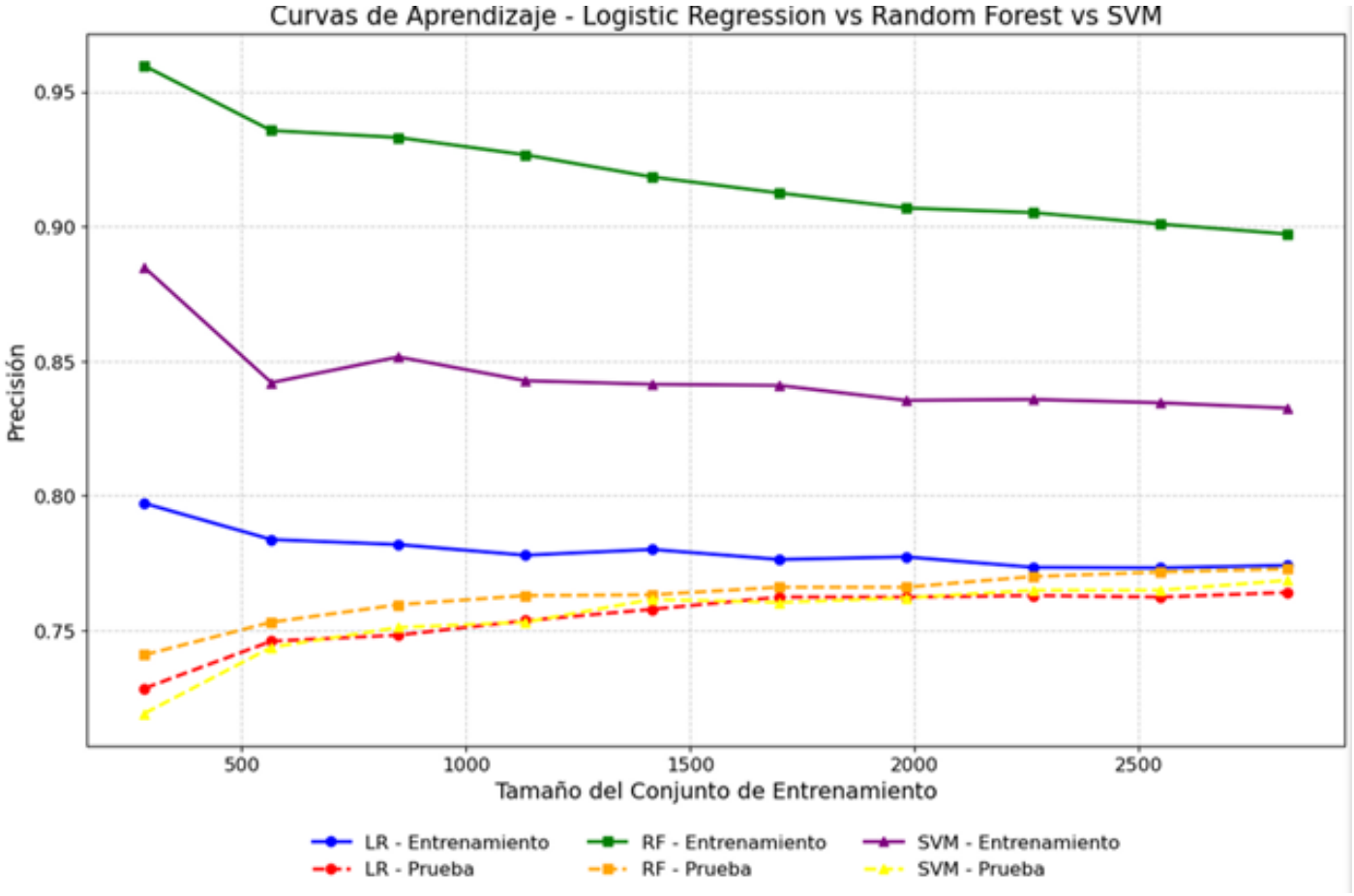
Cohen's Kappa (Random Forest vs SVM): 0.78

Pearson Correlation (Logistic Regression vs Random Forest): 0.87

Pearson Correlation (Logistic Regression vs SVM): 0.90

Pearson Correlation (Random Forest vs SVM): 0.87


Curvas de Aprendizaje - Logistic Regression vs Random Forest
LR - Entrenamiento
LR - Prueba
RF - Entrenamiento
RF - Prueba
Precisión
0.95
0.90
0.85
0.80
0.75
500
1000
1500
2000
2500
Tamaño del Conjunto de Entrenamiento


Curvas de Aprendizaje - Logistic Regression vs Random Forest vs SVM
Precisión
0.95
0.90
0.85
0.80
0.75
500
1000
1500
2000
2500
Tamaño del Conjunto de Entrenamiento
LR - Entrenamiento
LR - Prueba
RF - Entrenamiento
RF - Prueba
SVM - Entrenamiento
SVM - Prueba

Conclusiones:

1. **Random Forest** es el modelo más adecuado para este conjunto de datos, con una mayor precisión tanto en entrenamiento como en prueba.

2. **Regresión Logística** muestra limitaciones y no es la mejor opción para este problema en particular.

3. Después de la optimización de lo tres modelos, se evidencia un balance adecuado entre la precisión del entrenamiento y prueba, eliminando el sobreajuste (**overfitting**), no hay **underfitting**, ya que el modelo sigue logrando mostrar un rendimiento alto tanto en entrenamiento como en prueba.

4. Se presenta una **alta concordancia** y **correlación** entre todos los modelos ya que producen resultados similares y consistentes.