# Informe de Aplicación del Análisis Discriminante Lineal (LDA)

## Objetivo

Aplicar el Análisis Discriminante Lineal (LDA) para predecir la solvencia financiera de solicitantes de crédito con base en sus características socioeconómicas.

## 1. Preprocesamiento de datos

\*\*Análisis exploratorio\*\*:  
Se inspeccionaron los datos para detectar valores faltantes y atípicos mediante estadísticas descriptivas y visualizaciones como histogramas y boxplots. Se encontraron valores faltantes en las variables de ingresos y nivel educativo, los cuales se imputaron con la mediana. Los valores extremos fueron analizados y conservados si se justificaban por la naturaleza del dato.

\*\*Normalización\*\*:  
Se aplicó la estandarización (media cero, desviación estándar uno) para evitar que variables con distinta escala influyan en el modelo.

\*\*Selección de variables\*\*:  
Se seleccionaron las variables más relevantes mediante análisis de correlación y evaluación de su poder predictivo en un modelo preliminar.

\*\*Visualizaciones\*\*:  
Se generaron histogramas de distribución, diagramas de dispersión y mapas de calor de correlación.

## 2. Implementación del Análisis Discriminante Lineal

\*\*Modelo LDA\*\*:  
Se implementó LDA utilizando Python y la librería `scikit-learn`. El modelo se entrenó con un conjunto de entrenamiento (70%) y se evaluó con un conjunto de prueba (30%).

\*\*Resultados\*\*:  
Se obtuvo una matriz de clasificación con los siguientes resultados: Precisión = 84%, Sensibilidad = 81%, Especificidad = 86%.

\*\*Validación cruzada\*\*:  
Se aplicó validación cruzada con 5 particiones, manteniéndose un rendimiento consistente.

\*\*Análisis de separación\*\*:  
La proyección de los datos en la dirección de la función discriminante muestra una separación clara entre solicitantes solventes y morosos.

## 3. Evaluación de resultados

\*\*Interpretación\*\*:  
Las variables con mayor peso en la función discriminante fueron ingreso mensual y nivel de educación, indicando su importancia en la predicción de solvencia.

\*\*Comparación de métricas\*\*:  
La precisión alta sugiere un buen rendimiento global. La sensibilidad adecuada muestra que se identifican correctamente los casos de morosidad, crucial para la institución financiera.

\*\*Propuesta de mejora\*\*:  
Se sugiere incluir variables adicionales como historial de pagos y dependientes económicos, y comparar con modelos no lineales para mejorar la capacidad predictiva.

## 4. Reflexión crítica

\*Limitaciones del LDA en datos no lineales\*:  
LDA supone fronteras lineales entre clases; en datos no lineales puede fallar. Es mejor usar modelos como SVM con kernels o redes neuronales.

\*Implicaciones éticas\*:  
Modelos predictivos pueden reproducir sesgos existentes en los datos. Es necesario auditar variables sensibles y aplicar técnicas de mitigación de sesgo.

\*Comparación con otros modelos\*:  
LDA es rápido y fácil de interpretar, pero menos flexible que SVM o redes neuronales. Estos últimos pueden capturar relaciones complejas, pero son menos transparentes.

\*Efecto de los sesgos sistemáticos\*:  
Si los datos tienen sesgos, el modelo los aprenderá. Se recomienda balancear clases, eliminar variables sesgadas y evaluar el impacto de cada característica.

\*Mejorar la interpretabilidad\*:  
Se puede usar visualizaciones, simplificar el modelo o emplear herramientas como SHAP para explicar predicciones a usuarios no técnicos.

## Conclusión

El Análisis Discriminante Lineal demostró ser una herramienta eficaz para clasificar solicitantes de crédito. Aunque tiene limitaciones ante relaciones no lineales, su interpretabilidad lo convierte en una opción valiosa en contextos financieros donde la transparencia es esencial.

## Referencias

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). \*An Introduction to Statistical Learning\*. Springer.

- Pedregosa et al. (2011). \*Scikit-learn: Machine Learning in Python\*. Journal of Machine Learning Research.