

UNIVERSIDAD AUTÓNOMA DE NUEVO  
LEÓN  
FACULTAD DE CIENCIAS FÍSICO  
MATEMÁTICAS  
TAREA 6

Maestría en Ciencia de Datos  
ACT. ERICK ADRIÁN GARZA TAMEZ

Marzo 2023

## 1. Introducción

Este reporte busca generar un **aprendizaje supervisado** para predecir si un cliente va a caer en morosidad o no. Esta es una base de datos la cual contiene las siguientes características:

- *Monto del Préstamo*: Es el monto del préstamo que dispuso el cliente.
- *Term*: Es el plazo del préstamo.
- *Interest Rate*: Tasa de interés del préstamo.
- *Grade*: Es el nivel de riesgo asignado por el banco, siendo 1 el mejor y 7 el peor.
- *Loan Status*: Esta es nuestra variable de interés y nos indica si el cliente es o no moroso.

El aprendizaje supervisado es un tipo de algoritmo de aprendizaje automático que utiliza datos etiquetados para entrenar un modelo y hacer predicciones o clasificaciones precisas. En este tipo de aprendizaje, se proporcionan al algoritmo tanto los datos de entrada como las respuestas deseadas (o etiquetas) correspondientes a esas entradas. El objetivo del algoritmo es aprender una función que pueda mapear las entradas a las etiquetas de salida.

Para seleccionar el mejor modelo haremos pruebas de precisión para los siguientes modelos:

- Clasificación logística

- KNN o k-vecinos más cercanos
- Máquinas de vectores de soporte
- Clasificador Bayesiano Ingenuo
- Árbol de decisión
- Bosque aleatorio
- Perceptrón
- Red neuronal artificial
- Máquina de vector de relevancia (RVM)

Sin embargo, es importante tener en cuenta que el aprendizaje automático no es una solución infalible y siempre es necesario evaluar los resultados y tomar medidas adicionales para mitigar los riesgos identificados.

## 2. Modelos

Para realizar estas pruebas de precisión se utilizaron las siguientes variables de nuestros datos (primeras 10 filas):

ID	Loan Amount	Term	Interest Rate	Grade	Loan Status
65087372	10000	59	11.135007	2	0
1450153	3609	59	12.237563	3	0
1969101	28276	59	12.545884	6	0
6651430	11170	59	16.731201	3	0
14354669	16890	59	15.008300	3	0
50509046	34631	36	17.246986	2	0
32737431	30844	59	10.731432	3	0
63151650	20744	58	13.993688	1	0
4279662	9299	59	11.178457	7	0
4431034	19232	58	5.520413	3	0

Dicho lo anterior, se realizaron pruebas de precisión con cada uno de los modelos y nos arrojaron los siguientes resultados:

Model	Score
Random Forest	100.00
Decision Tree	100.00
KNN	91.50
Support Vector Machines	90.75
Logistic Regression	90.75
Naive Bayes	90.75
Perceptron	90.75
Linear SVC	90.75
Stochastic Gradient Decent	9.25

### **3. Conclusiones**

De lo anterior podemos concluir que nuestro modelo de aprendizaje supervisado ‘Bosque Aleatorio’ y ‘Árbol de Decisión’ son los mejores modelos con 100 % de precisión.

Por lo tanto, se buscará realizar pruebas para determinar que no se trate de un sobre ajuste de los datos.