**UNIVERSIDAD FRANCISCO MARROQUÍN**

**FACULTAD DE CIENCIAS ECONÓMICAS**

**DATA SCIENCE FOR FINANCE**

**MARIA ISABEL AVILA RIGALT**

**CASO PRÁCTICO NO. 3**

**PREDICCIÓN DE CALIFICACIÓN CREDITICIA CON ALGORITMOS DE CLASIFICACIÓN**

**Instrucciones generales:**

Para este caso práctico trabajarán con información de clientes de productos de crédito de Bank of America. Los datos incluyen información demográfica de los clientes y de su historial crediticio. Desarrolle un algoritmo de regresión logística multinomial (polinomial) para predecir la calificación de riesgo de los clientes. Para entrenar y probar el algoritmo debe utilizar el set de “bank1.csv”. Al final, debe usar su modelo para predecir la calificación de riesgo de los clientes de “bank2.csv” (este no tiene las respuestas).

Nota: Pueden encontrar la descripción de variables en el ANEXO.

**Entregables:**

1. 1 código de R por grupo con su procedimiento.
2. Un archivo de csv con la columna de ID y la de su predicción (la calificación de riesgo de ese cliente) para el dataset de bank2.
3. 1 Word o Markdown con las conclusiones de su exploración de variables, los resultados de las métricas de sus modelos, su interpretación y justificación para elegir uno de los 2 modelos.

**Requerimientos específicos:**

Con bank1:

1. Utilizar los métodos vistos en clase para evaluar las variables categóricas y numéricas y dejar conclusiones sobre posibles variables útiles para la predicción en el Word.
2. Utilizar tanto variables categóricas como numéricas en el modelo.
3. Crear al menos 2 modelos y con diferente combinación de predictoras (las variables utilizadas deben ser consistentes con su exploración).
4. Interpretación de las métricas obtenidas con sus modelos al evaluarlos y por qué eligieron un modelo sobre el otro.

Con bank2:

1. Predecir la calificación de riesgo de los clientes con su mejor modelo y escribirlos en un csv junto con el ID de los clientes.

**ANEXO: Descripción de variables**

Edad: numérica

Ocupación: categórica (tiene blanks)

Ingreso anual: numérica

Ingreso mensual: numérica

Cuentas bancarias: numérica – cantidad de cuentas bancarias que tiene el cliente

Tasa interés: numérica – tasa de interés mensual del crédito (en porcentaje)

Meses atraso: numérica – meses de atraso en el pago (negativo es que tiene pago adelantado)

Cambio límite: numérica – en cuánto ha cambiado su límite crediticio desde que es cliente (en porcentaje)

Consulta crédito: numérica – cuántas veces ha consultado acerca de su crédito

Deuda: numérica – deuda actual

Utilización crédito: numérica – porcentaje de su crédito que está utilizando actualmente

Pago mínimo: categórica

0 – su último pago no fue del mínimo

1 – su último pago fue del mínimo

Inversión mensual: numérica

Tipo pago: categórica

1 – pocos pagos de valor pequeño

2 – pocos pagos de valor mediano

3 – pocos pagos de valor grande

4 – muchos pagos de valor pequeño

5 – muchos pagos de valor mediano

6 – muchos pagos de valor grande

Balance mensual: numérica

Calificación riesgo: categórica

0 – Bueno

1 – Regular

2 – Malo

Historial crediticio: numérica, en años

Summary:

* Age (Edad): The minimum age is 14, and the maximum age is 95. The mean age is approximately 33.46 years.
* Annual Income (Ingreso\_anual): The minimum annual income is $7,006, and the maximum is $24,198,062. The mean income is approximately $183,663.
* Monthly Income (Ingreso\_mensual): The minimum monthly income is $0, and the maximum is $21,167. The mean monthly income is approximately $4,179.
* Number of Bank Accounts (Cuentas\_bancarias): The minimum number of bank accounts is 0, and the maximum is 49. The mean number of bank accounts is approximately 5.385.
* Interest Rate (Tasa\_Interes): The minimum interest rate is 1%, and the maximum is 34%. The mean interest rate is approximately 14.55%.
* Months in Arrears (Meses\_atraso): The minimum number of months in arrears is -5, and the maximum is 67. The mean number of months in arrears is approximately 21.11.
* Credit Limit Change (Cambio\_limite): The minimum credit limit change is 0.01, and the maximum is 36.09. The mean credit limit change is approximately 10.61.
* Credit Inquiries (Consulta\_credito): The minimum number of credit inquiries is 0, and the maximum is 2,597. The mean number of credit inquiries is approximately 28.37.
* Debt (Deuda): The minimum debt is $0.23, and the maximum is $4,998.07. The mean debt is approximately $1,425.41.
* Credit Utilization (Utilizacion\_credito): The minimum credit utilization is 20%, and the maximum is 50%. The mean credit utilization is approximately 32.28%.
* Monthly Investment (Inversion\_mensual): The minimum monthly investment is $0, and the maximum is $1,977.33. The mean monthly investment is approximately $194.37.
* Payment Type (Tipo\_pago): The payment type ranges from 1 to 6. The mean payment type is approximately 3.247.
* Monthly Balance (Balance\_mensual): The minimum monthly balance is 0.0886, and the maximum is 1,602.0405. The mean monthly balance is approximately 402.5058.
* Credit History (Historial\_crediticio): The credit history ranges from 0 to 39.69. The mean credit history is approximately 17.95.

Correlation:

* Age (Edad) has a negative correlation with the number of bank accounts, interest rate, months in arrears, credit limit change, and debt. It has a positive correlation with monthly income and monthly balance.
* Annual Income (Ingreso\_anual) has a positive correlation with monthly income.
* Monthly Income (Ingreso\_mensual) has a negative correlation with interest rate, months in arrears, credit limit change, and debt. It has a positive correlation with the number of bank accounts, credit utilization, monthly investment, payment type, and monthly balance.
* Number of Bank Accounts (Cuentas\_bancarias) has a positive correlation with interest rate, months in arrears, credit limit change, and debt. It has a negative correlation with credit utilization and monthly balance.
* Interest Rate (Tasa\_Interes) has a negative correlation with credit utilization and credit history.
* Months in Arrears (Meses\_atraso) has a positive correlation with credit utilization