

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Un modelo de predicción para analizar el riesgo de crédito en MiBanco S.A.

INFORME DE PASANTÍA PARA OPTAR POR EL TÍTULO DE MATEMÁTICO
PROGRAMA ACADÉMICO DE MATEMÁTICAS

Alisson Camila Romero Gutierrez

Jefe inmediato
Roderick Luna

Profesor director
Luis Fernando Villarraga Poveda

Entidad
Mibanco S.A.

Facultad de Ciencias Matemáticas y Naturales
Bogotá D.C.
2023

Resumen

En Colombia la mayor parte del comercio se debe a microempresas que aportan producción y brindan trabajo. Para que dichas microempresas continúen en funcionamiento de la mejor manera existen instituciones financieras que basan su modelo operativo en la oferta de soluciones de financiamiento y ahorro a través de microcrédito, para poder realizar la inclusión financiera es necesario realizar un análisis de riesgo de crédito para tener información acertada, además calcular los indicadores que muestren la solución de crédito más conveniente tanto para la microempresa como para la institución financiera.

En este trabajo, se propone utilizar bases de datos de microempresas que por razones de seguridad de los datos, los modelos y resultados presentados en la entrega final se basarán en datos simulados. Se utilizan estos datos para llevar a cabo un análisis de caracterización y predicción del acceso al crédito. Se seleccionarán las variables significativas para el estudio y se realizarán ajustes de curvas utilizando el método de mínimos cuadrados, en particular rectas de regresión, ajuste potencial y linealización de los datos. El objetivo es encontrar los tipos de curvas que mejor se ajusten al comportamiento de los datos. Para concluir, se presentarán los modelos óptimos encontrados y se realizará un análisis comparativo. Para evaluar los modelos y validar las técnicas de análisis utilizadas, se llevará a cabo la práctica en la entidad Mibanco S.A.

Palabras clave: Análisis, datos, modelos, predicción.

Índice

1. Objetivo de la pasantía **5**

1.1. Objetivo general 5

1.2. Objetivos específicos 5

2. Preliminares **6**

2.1. Conceptos del riesgo de crédito 6

2.1.1. Riesgo de crédito 6

2.1.2. Elementos del riesgo de crédito 6

2.2. Análisis de regresión 7

2.2.1. Rectas de regresión con mínimos cuadrados 7

2.2.2. Ajuste potencial 9

2.2.3. Método de linealización de los datos para $y = Ce^{Ax}$ 10

3. Modelo para predicción **12**

3.1. Datos 12

3.2. Modelo de regresión con mínimos cuadrados 17

3.2.1. Metodología 17

3.2.2. Evaluación del modelo 18

3.2.3. Resultados 19

3.3. Modelo de regresión con ajuste potencial 19

3.3.1. Metodología 20

3.3.2. Evaluación del modelo 21

3.3.3. Resultados 22

3.4. Modelo de regresión con linealización de los datos 22

3.4.1. Metodología 23

3.4.2. Evaluación del modelo 23

3.4.3. Resultados 25

4. Análisis comparativo de los modelos **26**

5. Conclusiones **28**

5.1. Conclusiones proyecto 28

5.2. Conclusiones pasantía 29

5.3. Recomendaciones 29

Referencias **30**

1. Objetivo de la pasantía

1.1. Objetivo general

Desarrollar un modelo de predicción para una entidad financiera que otorga créditos a microempresas. El propósito de este modelo es estimar la probabilidad de aprobación de crédito para una microempresa específica. Basándose en diversas características y variables relacionadas con la microempresa, como su historial crediticio, ingresos, tiempo de operación, entre otros, el modelo proporcionará una estimación de la probabilidad de que la microempresa cumpla con los criterios de crédito y, por lo tanto, sea considerada para el proceso de financiamiento.

1.2. Objetivos específicos

- Realizar un análisis a través del uso de bases de datos creadas con el fin de tener a disposición una simulación basada en datos reales, estos datos serán usados para realizar un modelo de predicción que pueda ser aplicado para medir la viabilidad financiera de las microempresas.
- Predecir y evaluar el acceso al crédito para las microempresas en Colombia, utilizando técnicas de análisis de riesgo crediticio, con el fin de proporcionar información acertada sobre las posibilidades de obtener crédito.
- Aplicar 3 técnicas de ajuste de curvas para encontrar los modelos óptimos que se ajusten al comportamiento de los datos y, de esta manera, determinar las soluciones de crédito más convenientes tanto para las microempresas como para las instituciones financieras.
- Crear modelos de predicción para que sean aplicados en la entidad financiera donde se realiza la práctica.
- Realizar un análisis sobre los modelos encontrados y los errores que presentan respecto a los datos.
- Presentar los resultados del análisis de riesgo crediticio y los modelos óptimos encontrados en gráficas que permitan una interpretación clara y concisa de los resultados obtenidos.

2. Preliminares

2.1. Conceptos del riesgo de crédito

2.1.1. Riesgo de crédito

El riesgo de crédito se refiere a la posibilidad de que, al llegar la fecha de vencimiento, un deudor no cumpla total o parcialmente con su obligación de reembolsar una deuda o el rendimiento acordado en un instrumento financiero, ya sea debido a una situación de quiebra, falta de liquidez u otras razones. La evaluación del riesgo de crédito se fundamenta en la probabilidad de que el prestatario o el emisor del bono no cumpla con sus responsabilidades, es decir, que ocurra un incumplimiento. Un factor crucial en la ocurrencia del incumplimiento está relacionado con los ciclos económicos, ya que durante las fases de expansión económica, generalmente se registra un menor número de incumplimientos, con tasas de impagos totales bajas. En contraste, en periodos de contracción económica, la incidencia de incumplimientos tiende a aumentar. Además, las circunstancias particulares de una empresa que ha emitido deuda generan lo que se conoce como riesgo específico de crédito, el cual no guarda relación con los ciclos económicos, sino que se relaciona con eventos particulares en la actividad comercial o industrial de las empresas. Estos eventos pueden afectar la capacidad de la empresa para cumplir con sus compromisos adquiridos al financiarse mediante la emisión de bonos u obligaciones[2].

2.1.2. Elementos del riesgo de crédito

El riesgo de crédito puede analizarse en tres aspectos básicos [3]:

- *Probabilidad de incumplimiento:* se refiere a la posibilidad de que no se cumpla con una obligación de pago, se rompa un acuerdo estipulado en un contrato de crédito o se produzca un incumplimiento económico. En este contexto, es común que las autoridades establezcan períodos de gracia antes de poder declarar formalmente el incumplimiento de pago.
- *Exposición:* la falta de certeza en relación con los montos futuros que pueden estar en riesgo es un aspecto relevante a considerar. En términos generales, los créditos se deben pagar según un calendario de pagos preestablecido, lo que permite calcular con anticipación el saldo pendiente en una fecha determinada. Sin embargo, es importante destacar que no todos los créditos cuentan con esta característica esencial que proporciona información crucial para evaluar el monto en riesgo.
- *Perdida dado el incumplimiento:* se deriva de la posibilidad de un incumplimiento y no puede ser anticipado, ya que depende de la naturaleza de la garantía que se haya recibido y de su condición en el momento en que se produce el incumplimiento. La presencia de una garantía puede reducir el riesgo de crédito, siempre y cuando esta

garantía sea fácilmente convertible en efectivo y tenga un valor suficiente para cubrir la deuda pendiente. Sin embargo, incluso en el caso de garantías, existe incertidumbre, ya que no solo implica la transferencia del riesgo en caso de incumplimiento por parte del deudor principal, sino que también existe la posibilidad de que el garante (aval) incumpla al mismo tiempo, lo que podría dar lugar a una probabilidad conjunta de incumplimiento.

2.2. Análisis de regresión

El método de mínimos cuadrados emerge como una elección óptima y fundamental en el análisis de datos y modelado estadístico, tanto para ajustes de regresión lineal como no lineal, como el caso de regresión potencial. El método de mínimos cuadrados ofrece estimaciones precisas de los parámetros del modelo, como los coeficientes de regresión en el caso de la regresión lineal, los exponentes y coeficientes en el caso de la regresión no lineal [4]. Para garantizar que las curvas que obtenemos son la mejor aproximación hay que considerar los errores, también llamados desviaciones:

- Error máximo: $E_\infty(f) = \max\{|f(x_k) - y_k : 1 \leq k \leq N|\}$
- Error medio: $E_1(f) = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N}|f(x_k) - y_k|$
- Error cuadrático medio: $E_2(f) = \left(\dfrac{1}{N}\displaystyle\sum_{k=1}^{N}|f(x_k) - y_k|^2\right)^{1/2}$

Un ajuste óptimo es aquel que minimiza las desviaciones dadas anteriormente.

2.2.1. Rectas de regresión con mínimos cuadrados

Sea $\{(x_k, y_k)\}_{k=1}^N$ un conjunto de N puntos cuyas abscisas $\{x_k\}$ son todas distintas. La recta óptima en el sentido de los mínimos cuadrados es la recta de ecuación $y = f(x) = Ax + B$ que para este caso minimiza el error cuadrático medio $E_2(f)$. El valor $E_2(f)$ será mínimo si, y sólo si, lo es el valor $N(E_2(f))^2 = \displaystyle\sum_{k=1}^{N}(Ax_k + B - y_k)^2$, dicho valor puede verse geométricamente como la suma de los cuadrados de las distancias verticales de los puntos hasta la recta[4].

Teorema 2.1. *Suponamos* $\{(x_k, y_k)\}_{k=1}^N$ *un conjunto de N puntos cuyas abscisas* $\{x_k\}$ *son todas distintas. Entonces los coeficientes de la recta de regresión*

$$y = Ax + B$$

son la solución del siguiente sistema lineal, conocido tambien como las ecuaciones normales de Gauss:

$$\left(\sum_{k=1}^{N} x_k^2\right)A + \left(\sum_{k=1}^{N} x_k\right)B = \sum_{k=1}^{N} x_k y_k,$$
$$\left(\sum_{k=1}^{N} x_k\right)A + NB = \sum_{k=1}^{N} y_k$$

Demostración. Si tomamos la recta $y = Ax + B$ vamos a tener que la distancia vertical d_k desde el punto (x_k, y_k) hasta el punto $(x_k, Ax_k + B)$ de la recta es $d_k = |Ax_k + B - y_k|$ como se puede ver:

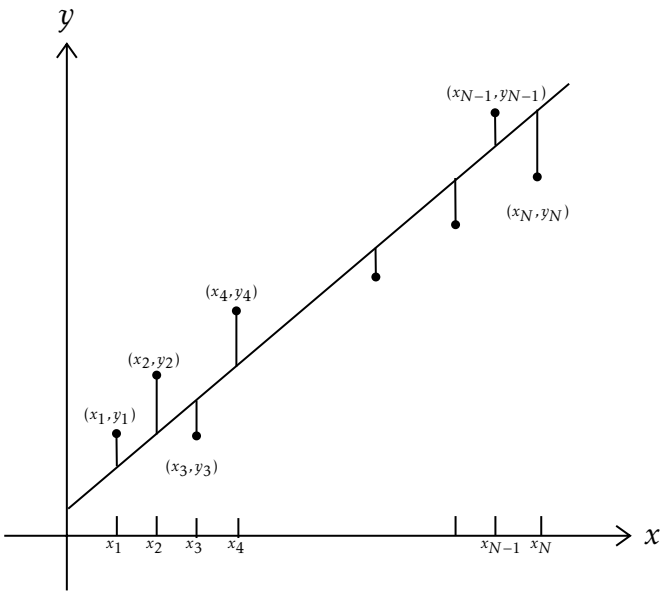


Figura 1: Distancias verticales entre los puntos $\{(x_k, y_k)\}$
y la recta $y = Ax + B$

Dado esto debemos minimizar la suma de los cuadrados de las distancias verticales d_k:

$$E(A, B) = \sum_{k=1}^{N} (Ax_k + B - y_k)^2 = \sum_{k=1}^{N} d_k^2$$

El valor mínimo de la función $E(A, B)$ lo podemos encontrar igualando a cero las derivadas parciales $\partial E/\partial A$ y $\partial E/\partial B$ luego resolviendo las ecuaciones que resultan en A y B. Cabe

notar que $\{x_k\}$ y $\{y_k\}$ son constantes, con esto obtenemos

$$\frac{\partial E(A,B)}{\partial A} = \sum_{k=1}^{N} 2(Ax_k + B - y_k)(x_k) = 2\sum_{k=1}^{N}(Ax_k^2 + Bx_k - x_k y_k)$$
$$\frac{\partial E(A,B)}{\partial B} = \sum_{k=1}^{N} 2(Ax_k + B - y_k) = 2\sum_{k=1}^{N}(Ax_k + B - y_k)$$

Igualando a cero las derivadas parciales obtenemos

$$\sum_{k=1}^{N}(Ax_k^2 + Bx_k - x_k y_k) = A\sum_{k=1}^{N} x_k^2 + B\sum_{k=1}^{N} x_k - \sum_{k=1}^{N} x_k y_k = 0$$
$$\sum_{k=1}^{N}(Ax_k + B - y_k) = A\sum_{k=1}^{N} x_k + NB - \sum_{k=1}^{N} y_k = 0$$

Reescribiendo las ecuaciones llegamos a

$$A\sum_{k=1}^{N} x_k^2 + B\sum_{k=1}^{N} x_k = \sum_{k=1}^{N} x_k y_k$$
$$A\sum_{k=1}^{N} x_k + NB = \sum_{k=1}^{N} y_k$$

Así llegamos a las ecuaciones normales de Gauss y este sistema se puede resolver por cualquier método para resolver ecuaciones lineales. □

2.2.2. Ajuste potencial

Algunas situaciones como pasa en este caso se modelan mediante una función del tipo $f(x) = Cx^A$, para el ajuste potencial es necesario determinar dos parámetros [4].

Teorema 2.2. *Suponamos $\{(x_k, y_k)\}_{k=1}^{N}$ un conjunto de N puntos cuyas abscisas $\{x_k\}$ son todas distintas. Entonces el coeficiente C de la curva potencial óptima en mínimos cuadrados $y = Cx^A$ viene dado por*

$$C = \frac{\left(\sum_{k=1}^{N} x_k^A y_k\right)}{\left(\sum_{k=1}^{N} x_k^{2A}\right)}$$

Demostración. Usando la forma de los mínimos cuadrados debemos hallar el mínimo de la función:

$$E(C) = \sum_{k=1}^{N} (Cx_k^A - y_k)^2$$

Lo que es igual en este caso a resolver $E'(C) = 0$. La derivada es

$$\begin{aligned} E'(C) &= 2\sum_{k=1}^{N} (Cx_k^A - y_k)(x_k^A) \\ &= 2\sum_{k=1}^{N} (Cx_k^{2A} - x_k^A y_k) \end{aligned}$$

Igualando a cero obtenemos

$$\begin{aligned} 2\sum_{k=1}^{N} (Cx_k^{2A} - x_k^A y_k) &= 0 \\ \sum_{k=1}^{N} (Cx_k^{2A} - x_k^A y_k) &= 0 \\ C\sum_{k=1}^{N} x_k^{2A} - \sum_{k=1}^{N} x_k^A y_k &= 0 \end{aligned}$$

Despejando C se llega a que

$$C = \frac{\left(\sum_{k=1}^{N} x_k^A y_k\right)}{\left(\sum_{k=1}^{N} x_k^{2A}\right)}$$

□

2.2.3. Método de linealización de los datos para $y = Ce^{Ax}$

Para esto supongamos que vamos a ajustar una curva exponencial de la forma

$$y = Ce^{Ax}$$

a un conjunto de puntos $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$ dados. Para encontrar la solución a este problema podemos tomar logaritmos a ambos lados así

$$\begin{aligned} y &= Ce^{Ax} \\ \ln(y) &= \ln(Ce^{Ax}) \\ \ln(y) &= \ln(C) + \ln(e^{Ax}) \\ \ln(y) &= \ln(C) + Ax \end{aligned}$$

haciendo un cambio de variable apropiado se tiene

$$\begin{aligned} Y &= \ln(y) \\ X &= x \\ B &= \ln(C) \end{aligned}$$

y así se obtiene $Y = AX + B$ que es una relación lineal sobre las nuevas variables. El conjunto de puntos originales que son de la forma (x_k, y_k) con el cambio de variable, se han transformado en $(X_k, Y_k) = (x_k, \ln(y_k))$, este proceso que se realizo lo llamamos método de linealización de los datos.

Ahora para hallar la recta de regresión $Y = AX + B$ para los puntos (X_k, Y_k), planteamos como se ha dicho anteriormente las ecuaciones normales de Gauss

$$\begin{gathered} \left(\sum_{k=1}^{N} X_k^2\right) A + \left(\sum_{k=1}^{N} X_k\right) B = \sum_{k=1}^{N} X_k X_k, \\ \left(\sum_{k=1}^{N} X_k\right) A + NB = \sum_{k=1}^{N} Y_k \end{gathered}$$

luego de obtener los valores de A y B, hallamos el valor de C con

$$C = e^B$$

3. Modelo para predicción

Los modelos de predicción son usados para anticipar eventos futuros en función de datos históricos y patrones observados. En este estudio se pretende buscar tres diferentes modelos que puedan explicar el comportamiento de clientes en el sector financiero y de esta manera generar curvas que puedan predecir si el cliente es apto o no para acceder a un crédito, cabe destacar que los modelos están pensados para clientes con microempresas que desempeñan una labor específica.

3.1. Datos

Los datos que se usaron para realizar los siguientes modelos y el análisis final son datos generados buscando representar posibles escenarios donde los clientes tienen acceso o no a la inclusión financiera. Dichos datos se generaron a partir de una simulación teórica con el propósito de proteger la información de personas reales incluidas en las bases de datos de la entidad financiera en cuestión.

Para identificar los modelos adecuados, se han considerado variables significativas que generen un valor importante al evaluar la elegibilidad para un crédito. En este estudio, se han seleccionado dos variables principales:

$$x_1 = \text{Tiempo}$$
$$x_2 = \text{Ingresos}$$

Se toman sólo estas variables para mostrar de mejor manera la representación gráfica. Es importante destacar que la entidad financiera puede utilizar cualquier cantidad de variables necesarias en su modelo.

El siguiente paso es asignar valores adecuados a estas variables representativas. En este contexto, se han propuesto dos modelos que asignan valores en un rango de 0 a 1 a cada variable, esto con el fin de realizar un manejo más adecuado y poder ingresar dichos valores a una función que asigne un valor de comportamiento final.

Para la primera variable x_1 = Tiempo, se plantea un modelo de intervalos de tiempo donde sea posible darle un puntaje dependiendo de los años en los que la empresa lleve funcionando. Este puntaje cambia para los primeros 20 años, después de que la empresa lleve 20 años en servicio el puntaje siempre será igual a 1.

Para la segunda variable x_2 = Ingresos, se plantea un modelo similar donde dependiendo de la solicitud se le asigne un valor entre 0 y 1 a los ingresos anuales que genera la empresa, es decir, la solicitud del crédito tiene que ser consecuente a los ingresos generados mensualmente por la empresa, de esta forma la entidad asigna el valor que crea correspondiente. La variable ingresos esta en el rango entre 0 a 800 ya que las microempresas en Colombia se clasifican por tener ingresos menores a 800 millones de pesos anuales.

Para entender mejor se realizan los siguientes códigos y sus respectivas gráficas, se resalta que estos modelos se pueden ajustar a las condiciones que defina la entidad financiera para otorgar un crédito.

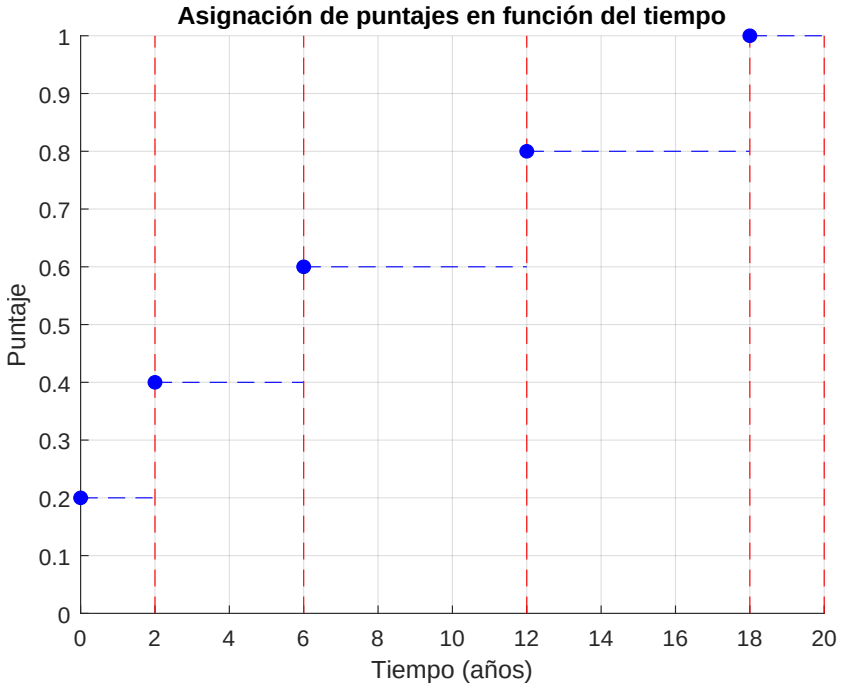


Figura 2: Asignación de puntaje al tiempo de funcionamiento de la empresa

Para obtener este modelo se realiza el siguiente código:

```
% Generar datos de tiempo
tiempo = [0, 2, 6, 12, 18, 20, 22];  % Intervalos de tiempo

% Asignar puntajes entre 0 y 1 para cada intervalo de tiempo
puntajes = [0.2, 0.4, 0.6, 0.8, 1, 1, 1];  % Puntajes

% Graficar los puntajes en funcion del tiempo
figure;
hold on;

for i = 1:length(tiempo) - 1
    if tiempo(i) == 20
        plot(tiempo(i), puntajes(i), 'b--');
    else
        plot(tiempo(i), puntajes(i), 'b-o', 'MarkerFaceColor', 'b');
    end
    if i < length(tiempo) - 1
        plot([tiempo(i), tiempo(i+1)], [puntajes(i), puntajes(i)], 'b--')
   ; % Linea horizontal
        plot([tiempo(i+1), tiempo(i+1)], [0, 1], 'r--'); % Linea vertical
    end
```

```
end

xlabel('Tiempo');
ylabel('Puntaje');
title('Asignacion de puntajes en funcion del tiempo');
grid on;

hold off;
```

Para la variable ingresos se tiene el siguiente modelo que sigue una función logística, se toma esta función ya que cumple las características de ser acotada y en sus límites inferior y superior tienden a un valor específico que se ajusta a los valores que se necesitan. Se toma como punto de inflexión 350 y parámetro 0,013 para mantener la curva suave.

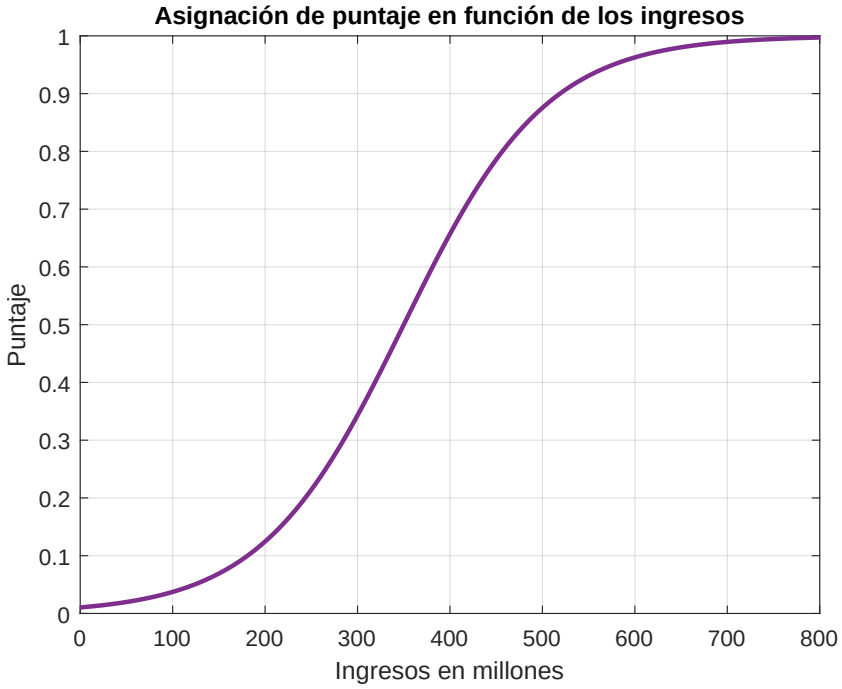


Figura 3: Asignación de puntaje a los ingresos anuales de la empresa

Este modelo se obtuvo con el siguiente código:

```
% Definir la funcion logistica para asignar valores entre 0 y 1 a
    ingresos en millones de pesos
ingresos_en_millones = 0:1:800;  % Ingresos en millones de pesos
y = 1 ./ (1 + exp(-0.013 * (ingresos_en_millones-350)));  % Funcion
    logistica

% Graficar la funcion
figure;
plot(ingresos_en_millones, y,'Color',[[0.4940 0.1840 0.5560]],'LineWidth'
    ,2);
xlabel('Ingresos en millones');
```

```
ylabel('Puntaje');
title('Asignacion de puntaje en funcion de los ingresos');
grid on;
```

Este enfoque proporciona una forma de medir cada una de las variables y asignarles un valor, lo cual es fundamental para mantener un rango general en todas las variables. Con estos valores asignados, ahora podemos desarrollar un modelo de regresión lineal múltiple que combine las variables significativas en una función única y ofrezca una evaluación inicial del comportamiento financiero.

Ahora con estos valores se realiza la siguiente regresión multilineal que a partir de los valores ingresados para x_1 y x_2 entrega un valor entre 0 y 1 que se interpreta como el comportamiento financiero, dicho valor va a ser de gran utilidad ya que representa el acceso a crédito que puede tener la microempresa en cuestión.

Para este caso la ecuación de comportamiento es la siguiente

$$y = 0{,}0033 + 0{,}2827x_1 + 0{,}7555x_2$$

Se obtuvo con el siguiente código y datos simulados

```
% Datos de entrada
x1 = [0.1, 0.3, 0.5, 0.7, 0.9, 0.2, 0.4, 0.6, 0.8, 0.3, 0.5, 0.7, 0.1,
    0.4, 0.6, 0.8, 0.2, 0.9, 0.3, 0.7];
x2 = [0.2, 0.4, 0.6, 0.8, 0.1, 0.3, 0.5, 0.7, 0.9, 0.4, 0.6, 0.8, 0.3,
    0.5, 0.7, 0.9, 0.1, 0.2, 0.6, 0.8];
Y = [0.15, 0.45, 0.65, 0.75, 0.3, 0.35, 0.55, 0.85, 0.95, 0.4, 0.6, 0.8,
    0.3, 0.5, 0.7, 0.9, 0.1, 0.4, 0.3, 0.75];

% Crear una matriz de diseno con las variables independientes (x1 y x2)
X = [x1', x2'];

% Realizar la regresion lineal multiple
mdl = fitlm(X,Y);

% Obtener los coeficientes del modelo
coefficients = mdl.Coefficients.Estimate;

%Mostrar la ecuacion de la regresion
fprintf('La ecuacion de la regresion es: Y = %.4f + %.4f*x1 + %.4f*x2\n',
    coefficients(1), coefficients(2), coefficients(3));


x1_values = linspace(min(x1), max(x1), 100);
x2_values = linspace(min(x2), max(x2), 100);
[X1, X2] = meshgrid(x1_values, x2_values);
Y_pred = coefficients(1) + coefficients(2) * X1 + coefficients(3) * X2;

%Graficar la superficie y los puntos reales
figure;
scatter3(x1, x2, Y, 'filled', 'MarkerFaceColor', 'r');
```

```
hold on;
mesh(X1, X2, Y_pred);
colormap('parula');
xlabel('x1');
ylabel('x2');
zlabel('Y');
title('Comportamiento');

hold off;
```

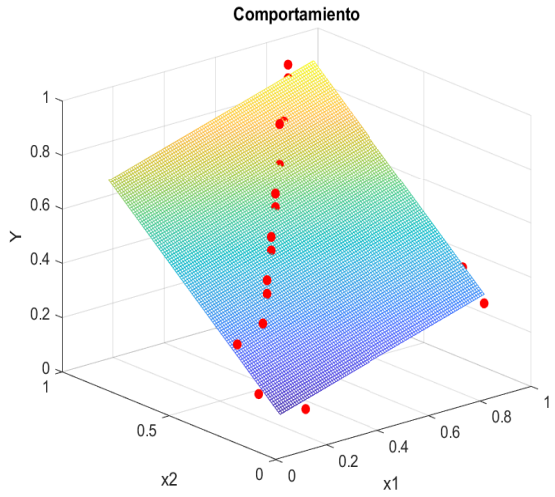


Figura 4: Vista 1-Modelo de comportamiento

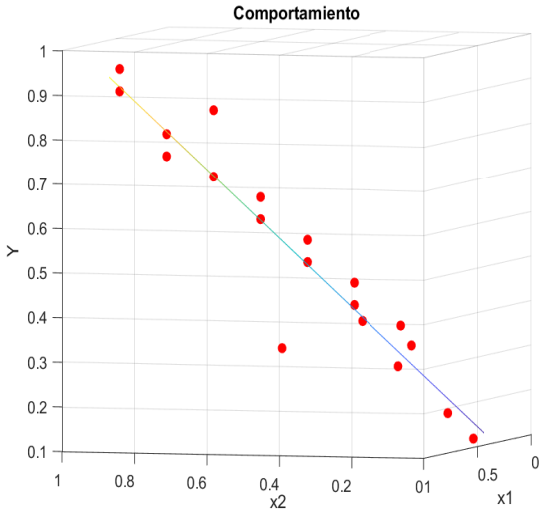


Figura 5: Vista 2-Modelo de comportamiento

3.2. Modelo de regresión con mínimos cuadrados

Dado los valores de comportamiento se busca un modelo que ajuste los valores de comportamiento a un score o puntaje que se entienda de manera general en toda la entidad financiera y permite aprobar o no el crédito solicitado por la microempresa.

El primer modelo que se abordo es el de mínimos cuadrados sin linealizar ninguna de las variables, este método como se explico anteriormente lo que intenta es buscar la recta de regresión que mejor se aproximé a los datos ingresados y de esta manera tomar datos en el futuro para analizar sus posibles valores de salida.

3.2.1. Metodología

Se tienen datos que asignan un puntaje de crédito a cada comportamiento, medido en un rango de 150 a 950. Estos datos se modelan utilizando una regresión básica de mínimos cuadrados.

Después de haber realizado la regresión para los datos, el objetivo es desarrollar un nuevo modelo que represente la probabilidad de obtener el crédito solicitado en función del puntaje de crédito obtenido.

```
function[T] = minc(X,Y)
% X es el vector con las abscisas
% Y es el vector con las ordenadas
xm=mean(X); %Media de los valores en X
ym=mean(Y); %Media de los valores en Y1
n=length(X); %Largo del vector X

sumx2=(X-xm)*(X-xm)'; %Sumatoria de las X^2
sumxy=(Y-ym)*(X-xm)'; %Sumatoria de los XY

A=sumxy/sumx2; %Valor de A
B=ym-A*xm; %Valor B

x=linspace(min(X),max(X));
z=A*x+B; %Curva
for i=1:n %Iteracion que evalua los valores de X en la curva y calcula el
    error cuadratico medio
 fxi=A*(X(i))+B;
 error=sum((fxi-Y(i)).^2); %Error
 errorcm=sqrt(sum((fxi-Y(i)).^2)/n); %Error cuadratico medio
end

plot(X,Y,'o'); grid on; title('Curva f(x)=Ax+B','Color',[0 0 1]); xlabel(
   ''); ylabel('') %Grafica de los puntos
hold on
plot(x,z,'LineWidth',1, 'Color',[0.4940 0.1840 0.5560]); %Grafica de la
   curva
hold off
```

```
T=table (A, B, error, errorcm); %Tabla con el valor de A, B, el error y
    el error cuadratico medio
end
```

3.2.2. Evaluación del modelo

El modelo que relaciona el puntaje con el comportamiento se presenta en la Figura 6.

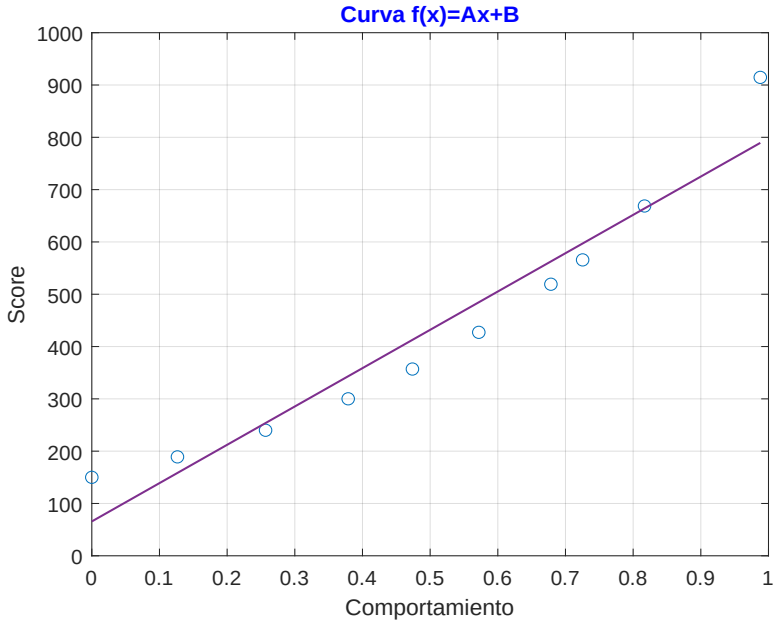


Figura 6: Comportamiento vs Score

Se obtuvo para esta curva los siguientes valores:

$$A = 732{,}6$$
$$B = 65{,}645$$
$$error = 15679$$
$$errorcm = 39{,}597$$

Al comparar los datos del puntaje con la probabilidad de aprobación del crédito solicitado, se obtienen los siguientes valores para la curva. El modelo correspondiente se muestra en la Figura 7.

$$A = 0{,}11343$$
$$B = -8{,}3335$$
$$error = 0{,}16922$$
$$errorcm = 0{,}13009$$

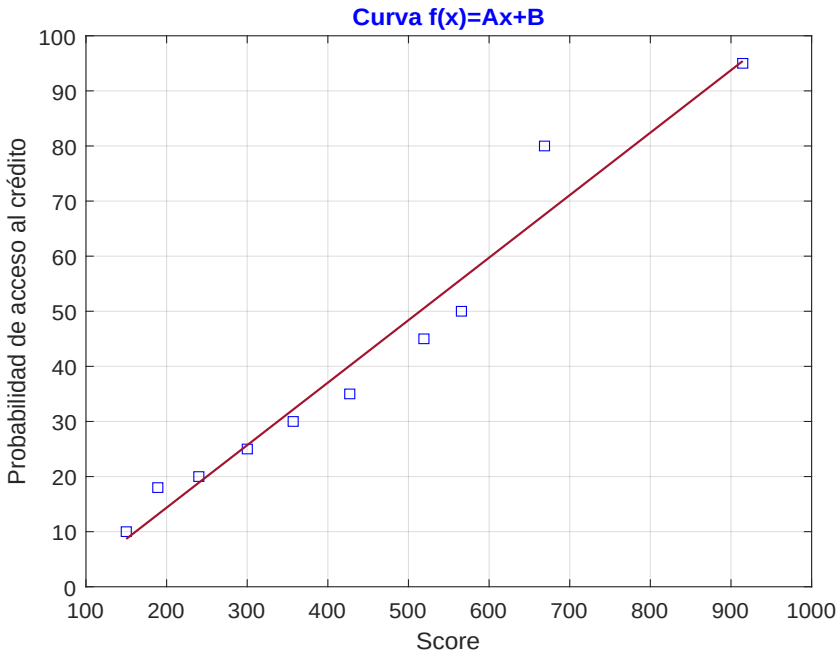


Figura 7: Score vs Probabilidad

3.2.3. Resultados

La primera función que se modela para el *Comportamiento vs Score* es la siguiente:

$$f(x) = 732{,}6x + 65{,}645$$

este modelo se evalúa mediante dos medidas de error, el error medio y el error cuadrático medio. La línea de ajuste óptimo es aquella que minimiza uno de estos errores. Sin embargo, observamos que este modelo tiene un error cuadrático medio de $39,597$, lo cual es un error considerablemente alto, lo que sugiere que este modelo no es el más adecuado.

El modelo *Score vs Probabilidad* tiene como función

$$f(x) = 0{,}11343x - 8{,}3335$$

este modelo presenta un error cuadrático medio en la regresión de $0{,}13009$, lo que indica que es un modelo adecuado para los valores disponibles. Sin embargo, es necesario evaluar otros modelos para determinar cuál es el más óptimo.

3.3. Modelo de regresión con ajuste potencial

Este modelo permite obtener curvas del tipo $f(x) = Cx^A$ que se asemejen al comportamiento de los datos, el modelo anterior al ser lineal produce limitaciones ya que en muy poco casos los datos van a tener una tendencia a la linealidad. En el ajuste potencial, se

pueden emplear métodos tanto lineales como no lineales. El teorema 2.2 menciona un método lineal; sin embargo, para lograr un ajuste más preciso, se realizarán cambios de variables y se optará por el método no lineal. Es importante señalar que el método no lineal deriva del método lineal.

3.3.1. Metodología

De igual forma se busca el modelo de la forma $f(x) = Cx^A$, usando el cambio de variable $X = \ln(x)$, $Y = \ln(y)$ y $C = e^B$ para linealizar los datos.

Se realiza la regresión para obtener un score a partir del comportamiento y con base en dicho en score obtener la probabilidad de acceso al crédito.

```
function[T] = regre_2(X,Y)
% X es el vector con las abscisas
% Y es el vector con las ordenadas
% T es la tabla de salida con los valores de A, C y el error cuadratico
    medio
X1=log(X); %Cambio de variable de X en log(X) para linealizar los datos
Y1=log(Y); %Cambio de variable de Y en log(Y) para linealizar los datos
xm=mean(X1); %Media de los valores en X1
ym=mean(Y1); %Media de los valores en Y1
n=length(X); %Largo del vector X

sumx2=(X1-xm)*(X1-xm)'; %Sumatoria de las X1^2
sumxy=(Y1-ym)*(X1-xm)'; %Sumatoria de los X1Y1

A=sumxy/sumx2; %Valor de A
B=ym-A*xm; %Valor B
C=exp(B); %Cambio de variable para hallar el valor de C

x=linspace(min(X)-1,max(X)+1);
z=C*x.^A; %Curva
for i=1:n %Iteracion que evalua los valores de X en la curva y calcula el
     error cuadratico medio
 fxi=C*X(i).^A;
 error=sum((fxi-Y(i)).^2); %Error
 errorcm=sqrt(sum((fxi-Y(i)).^2)/n); %Error cuadratico medio
end

plot(X,Y,'*','Color',[1 0 0]); grid on; title('Curva f(x)=Cx^{A}','Color'
    ,[0 0 1]); xlabel(''); ylabel('') %Grafica de los puntos
hold on
plot(x,z,'Color',[0 1 0]); %Grafica de la curva
hold off
T=table (A, C, error, errorcm); %Tabla con el valor de A, C, el error y
    el error cuadratico medio
end
```

3.3.2. Evaluación del modelo

El modelo de regresión con ajuste potencial que relaciona el puntaje con el comportamiento se muestra en la Figura 8.

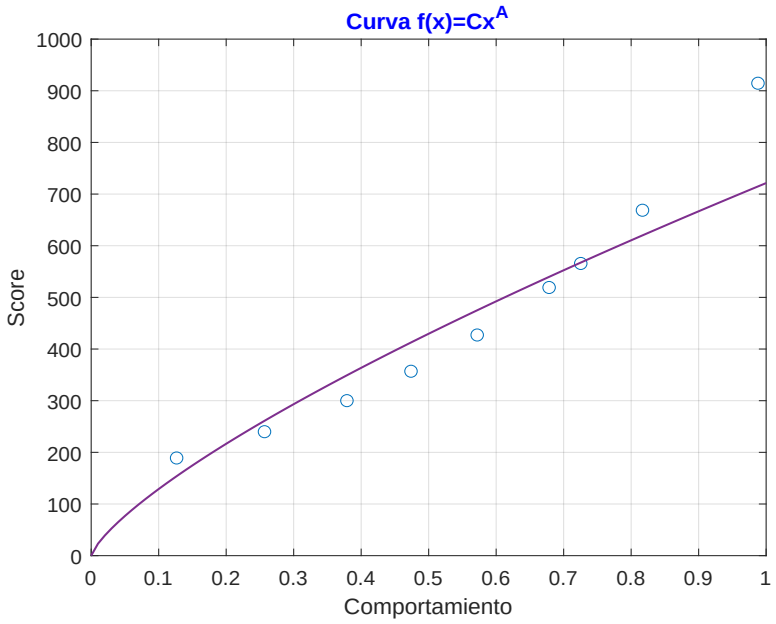


Figura 8: Comportamiento vs Score

Se obtuvo para esta curva los siguientes valores:

$$A = 0{,}74759$$
$$C = 721{,}32$$
$$error = 39925$$
$$errorcm = 66{,}604$$

Para los datos que relacionan el puntaje con la probabilidad de aprobación del crédito solicitado, se tiene la Figura 9. A continuación, se presentan los valores asociados con la curva

$$A = 1{,}1853$$
$$C = 0{,}02944$$
$$error = 0{,}06613$$
$$errorcm = 0{,}08132$$

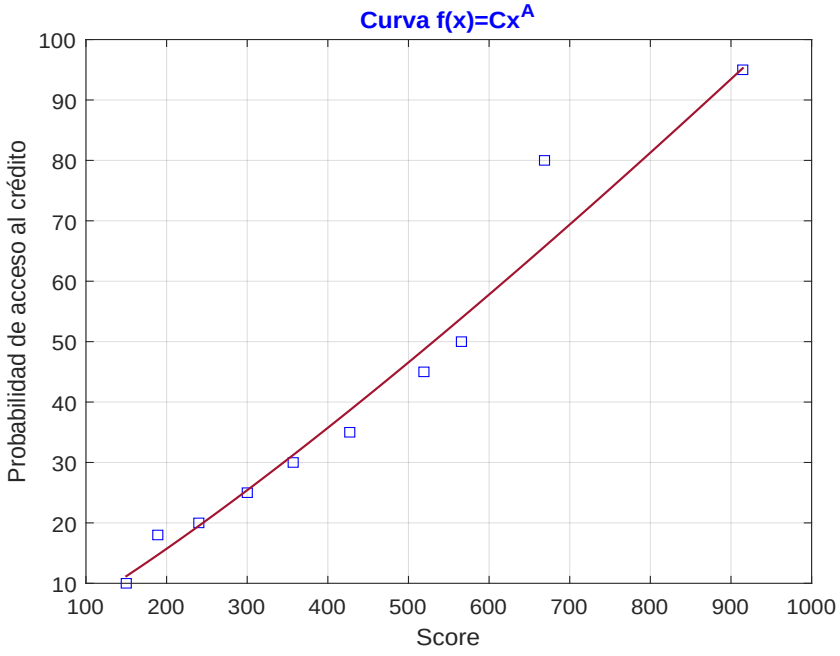


Figura 9: Score vs Probabilidad

3.3.3. Resultados

El modelo para el *Comportamiento vs Score* es

$$f(x) = 721{,}32x^{0{,}74759}$$

este modelo también se evalúa mediante dos medidas de error, el error medio y el error cuadrático medio. Se observa que el error cuadrático medio es de 66,604, ya que el error medio se desvía considerablemente, se considera el error cuadrático medio. A pesar de esto, el modelo todavía presenta un error considerable, lo que sugiere que no es el más óptimo en este caso.

El modelo *Score vs Probabilidad* tiene como función

$$f(x) = 0{,}02944x^{1{,}1853}$$

y presenta un error cuadrático medio en la regresión de 0,08132 lo cual indica que el modelo es bueno para los valores que se tienen pero es necesario verificar los demás modelos para encontrar el más óptimo.

3.4. Modelo de regresión con linealización de los datos

Este modelo busca a partir de un conjunto de datos dados encontrar la recta de regresión que mejor se ajuste, para esto usa un cambio de variable que permite disminuir el error en la curva encontrada.

3.4.1. Metodología

$f(x) = Ce^{Ax}$, usando el cambio de variable $X = x$, $Y = \ln(y)$ y $C = e^B$ para linealizar los datos.

Se realiza la regresión para obtener un score a partir del comportamiento y con base en dicho en score obtener la probabilidad de acceso al crédito.

```
function[T] = regre_1(X,Y)
% X es el vector con las abscisas
% Y es el vector con las ordenadas
% T es la tabla de salida con los valores de A, C y el error cuadratico
    medio
Y1=log(Y); %Y1 es el cambio de variable de Y en log(Y)
xm=mean(X); %Media de los valores en X
ym=mean(Y1); %Media de los valores en Y1
n=length(X); %Largo del vector X

sumx2=(X-xm)*(X-xm)'; %Sumatoria de las X^2
sumxy=(Y1-ym)*(X-xm)'; %Sumatoria de los XY

A=sumxy/sumx2; %Valor de A
B=ym-A*xm; %Valor B
C=exp(B); %Cambio de variable para hallar el valor de C

x=linspace(min(X)-1,max(X)+1);
z=C*exp(A*x); %Curva
for i=1:n %Iteracion que evalua los valores de X en la curva y calcula el
     error cuadratico medio
 fxi=C*exp(A*(X(i)));
 error=sum((fxi-Y(i)).^2); %Error
 errorcm=sqrt(sum((fxi-Y(i)).^2)/n); %Error cuadratico medio
end

plot(X,Y,'*','Color',[1 0 0]); grid on; title('Curva f(x)=Ce^{Ax}','Color
    ',[0 0 1]); xlabel('Coordenadas en x'); ylabel('Coordenadas en y') %
    Grafica de los puntos
hold on
plot(x,z,'Color',[0 1 0]); %Grafica de la curva
hold off
T=table (A, C, error, errorcm); %Tabla con el valor de A, C, el error y
    el error cuadratico medio
end
```

3.4.2. Evaluación del modelo

El primer modelo que se busca es aquel que ajuste al igual que en los anteriores un valor de score para diferentes valores de comportamiento, esto se puede ver en la Figura 10.

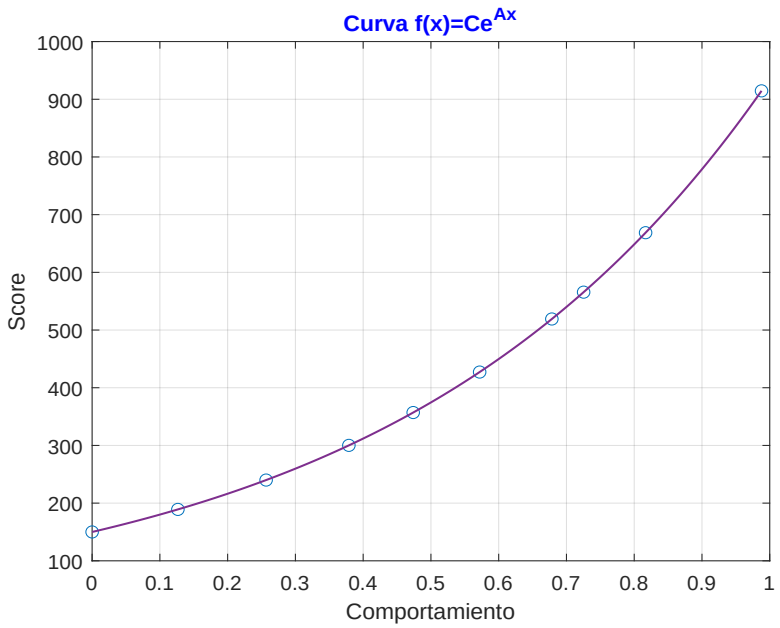


Figura 10: Comportamiento vs Score

Se obtuvo para esta curva los siguientes valores:

$$A = 1{,}83$$
$$C = 150$$
$$error = 6{,}5236e-06$$
$$errorcm = 0{,}00080769$$

Ahora comparando los datos de score con la probabilidad de aprobación de crédito solicitado se obtiene los siguientes valores para la curva y el modelo se puede ver en la Figura 11.

$$A = 0{,}0027806$$
$$C = 9{,}9487$$
$$error = 994{,}89$$
$$errorcm = 9{,}9744$$

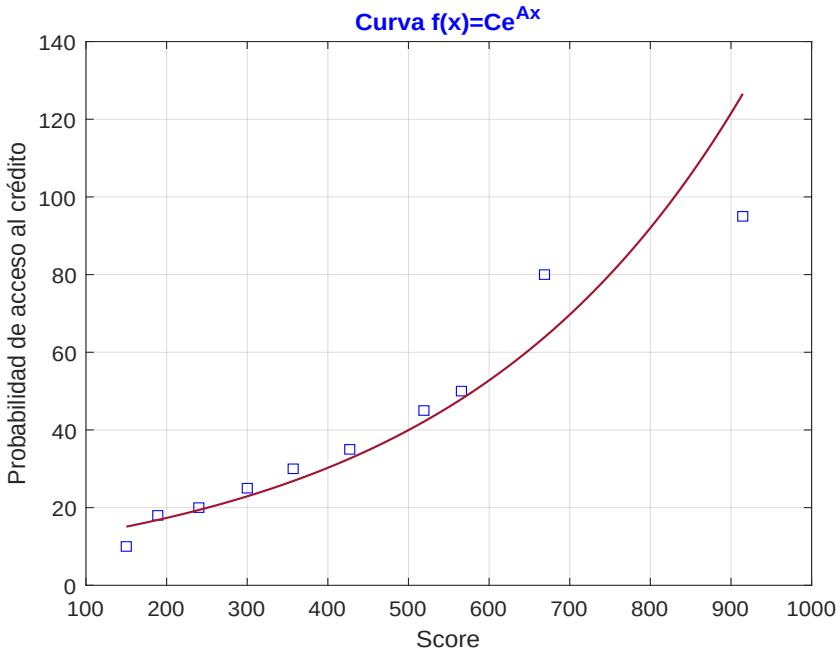


Figura 11: Score vs Probabilidad

3.4.3. Resultados

Para el modelo de *Comportamiento vs Score* se tiene:

$$f(x) = 150e^{1,83x}$$

se puede ver que el modelo tiene un error cuadrático medio de 0,00080769 y un error medio de $6,5236e-06$ lo cual indica que el modelo que se obtuvo es óptimo y adecuado para los datos que se han estado tratando.

El modelo *Score vs Probabilidad* tiene como función

$$f(x) = 9,9487e^{0,0028x}$$

y presenta un error cuadrático medio en la regresión de 9,9744, lo que indica que el error no se acerca a cero. Esto sugiere que existe una mejor curva de regresión para predecir el comportamiento de estos datos.

4. Análisis comparativo de los modelos

En este estudio, se han desarrollado varios modelos de regresión con el propósito de analizar y predecir a partir de los datos de comportamiento que se obtuvieron nuevos datos de puntuación de crédito y su relación con la probabilidad de aprobación de crédito, para así llegar al objetivo deseado que es la inclusión financiera. Cada modelo presenta características únicas y resultados específicos. Con cada uno de los métodos utilizados se encontraron dos modelos donde se resalta el error con los datos trabajados. A continuación, se realiza un análisis comparativo de los modelos para determinar cuál de ellos es el más adecuado en cada caso y con que método se pudo determinar.

El primer modelo que se buscó fue el *Modelo Comportamiento vs. Score*, para este caso se va a analizar el error cuadrático medio dado que se obtiene con mayor precisión.

- **Mínimos cuadrados**: $f(x) = 732,6x + 65,645$ con un error cuadrático medio de $39,597$.
- **Ajuste potencial**: $f(x) = 721,32x^{0,74759}$ con un error cuadrático medio de $66,604$.
- **Linealización de los datos**:$f(x) = 150e^{1,83x}$ con un error cuadrático medio de $0,00080769$.

Con base en los datos analizados, el modelo más óptimo para el *Comportamiento vs. Score* es la linealización de los datos, ya que muestra el error cuadrático medio más bajo con un valor de $0,00080769$. Esto sugiere una mayor capacidad de predicción en comparación con los otros modelos considerados.

El siguente modelo que se busco es el *Modelo Score vs. Probabilidad*, de igual forma se toma el error cuadrático medio para determinar cual modelo es e más óptimo.

- **Mínimos cuadrados**: $f(x) = 0,11343x - 8,3335$ con un error cuadrático medio de $0,13009$.
- **Ajuste potencial**: $f(x) = 0,02944x^{1,1853}$ con un error cuadrático medio de $0,08132$.
- **Linealización de los datos**: $f(x) = 9,9487e^{0,0028x}$ con un error cuadrático medio de $9,9744$.

El análisis de estos resultados revela que el modelo encontrado con el ajuste potencial es el más óptimo dentro de este contexto, ya que presenta el error cuadrático medio más bajo con un valor de $0,08132$, indicando una mayor precisión en la predicción en comparación con los otros modelos considerados.

En resumen, después de comparar los modelos, podemos concluir que el modelo encontrado por el método de linealización de los datos, que relaciona el comportamiento con

el score bancario, es el más adecuado para predecir el comportamiento de los datos. Por otro lado con el método de ajuste potencial se encontró el modelo óptimo que compara el score bancario con la probabilidad de obtener un crédito solicitado. Cabe resaltar que estos modelos se ajustan al trabajo propuesto con los datos generados y que si una entidad bancaria maneja datos con tendencias similares los modelos se ajustarán a lo que se necesite.

5. Conclusiones

5.1. Conclusiones proyecto

- En el proceso de modelado, se identificaron variables clave que desempeñan un papel fundamental en la evaluación de la elegibilidad para un crédito. Para este trabajo se tomaron inicialmente dos variables esto para poder representar de manera adecuada pero para un desarrollo más preciso es necesario tener en cuenta desde el inicio del análisis más variables que ayuden a precisar el tipo de cliente que se esta presentando a la entidad financiera.
- Se destacó la importancia de considerar las características específicas de los datos y los objetivos del análisis al elegir el modelo adecuado. La elección del modelo óptimo puede variar según el contexto y los datos disponibles.
- Para el trabajo se tomo como método principal los mínimos cuadrados y se tomaron los otros métodos por separado pero en general son casos derivados de los mínimos cuadrados, se optó por este método ya que es eficiente y además tiene un costo computacional menor a otros métodos de regresión.
- Los modelos que alcanzaron un menor rango de error pueden ser usados, asegurando que la tendencia de los datos y su manejo sea similar a como hizo en este trabajo de otro forma los modelos no se ajustarán y será necesario hacer un nuevo análisis.
- Se evidencio que al minimizar el error hay mayor ajuste del modelo lineal generalizado, con esto es posible encontrar una función que se ajuste a la necesidad que se tiene y de esta forma predecir el comportamiento del modelo.
- Los modelos de regresión desempeñan un papel fundamental en la toma de decisiones financieras y crediticias. Los resultados de este trabajo resaltan su utilidad para evaluar y predecir la elegibilidad de crédito.
- Este trabajo ha proporcionado un análisis desde la conceptualización teórica hasta la aplicación práctica. Considerando la simulación de datos y los modelos óptimos obtenidos, se sugiere a la entidad financiera la implementación gradual de estos modelos. Se propone comenzar con un conjunto inicial de datos homogéneo y representativo para poner a prueba cada modelo. Luego, conforme se obtengan resultados confiables, se recomienda aumentar progresivamente la cantidad de datos. Esto, ya que si se manejan datos muy dispersos al aumentar la cantidad de datos el modelo con el método utilizado no podrá aprender el comportamiento e imposibilitará la predicción de resultados.

5.2. Conclusiones pasantía

Las matemáticas generan una base fundamental en distintos aspectos de la vida, para este caso, la pasantía abordo el aspecto financiero, algo que afecta a la población en general. Desde el inicio de la pasantía fue posible encontrar herramientas computacionales que hacen el trabajo del análisis matemático mucho más sencillo y esta es una forma de entender las matemáticas desde las nuevas tecnologías. En este trabajo se logró explicar desde el punto de vista matemático como es posible llegar a acceder a un crédito, esto muestra que si se estudia a profundidad, la matemática va a dar el principio de diversas actividades financieras que podrán ser manejadas de mejor manera.

Con la pasantía se lograron implementar conocimientos matemáticos de modelado, lógica, probabilidad, manejo y análisis de datos, en general, esta experiencia fue enriquecedora tanto profesional como académicamente ya que se dio una visión de la importancia de la matemática en todos los ámbitos, lo fundamental de mantenerse actualizado y la importancia de continuar buscando problemas que solucionar.

5.3. Recomendaciones

Universidad

- Se sugiere generar oportunidades educativas al abrir espacios electivos en áreas de gran relevancia actual, como el análisis de datos. Esto permitiría a los estudiantes adquirir habilidades valiosas en la aplicación de las matemáticas en el mundo laboral.
- Se propone incluir la asignatura estadística en el plan de estudios, ya sea como una materia obligatoria o electiva. Dado que la estadística desempeña un papel esencial en numerosos campos laborales, proporcionar a los estudiantes una base sólida en esta disciplina contribuirá a su preparación para el mercado laboral.
- Se recomienda ofrecer cursos de formación y capacitación que abarquen el aprendizaje y dominio de herramientas ampliamente utilizadas en el ámbito laboral, tales como Microsoft Excel, Microsoft Power BI, entre otras.

Empresa

- Mantener y expandir programas de pasantías para brindar a los estudiantes la oportunidad de adquirir experiencia laboral valiosa y desarrollar relaciones con futuros profesionales.
- Sostener una relación cercana con instituciones educativas para identificar talento joven y ayudar a preparar a los estudiantes para el entorno laboral.
- Buscar retroalimentación de pasantes y empleados sobre sus experiencias para mejorar continuamente los programas y prácticas.

Referencias

[1] R.L. Burden and J.D. Faires. *Análisis numérico.* Matemáticas Thomson. INTERNATIONAL THOMSON EDITORES, 1998.

[2] D. Chofaras. *Managing credit risk, analysing rating and pricing the probability of the default.* Londres: Euromoney Institutional Investor PLC., 2000.

[3] M. Galicia. *Los enfoques del riesgo de crédito.* México: Instituto del Riesgo Financiero, 2003.

[4] John H Mathews, Kurtis D Fink, et al. *Numerical methods using MATLAB*, volume 4. Pearson prentice hall Upper Saddle River, NJ, 2004.