

UNIVERSIDAD AUTÓNOMA GABRIEL RENÉ MORENO

**FACULTAD DE INGENIERÍA Y CIENCIAS EN LA COMPUTACIÓN Y
TELECOMUNICACIONES**



PROYECTO 2

WEKA I CREDIT-G

INTEGRANTES:

- Mamani Rodriguez Kasandra
- Sejas Mamani Dennis
- Quispe Sabinas Brayan

DOCENTE: Ing. Peinado Pereira Miguel Jesus

ASIGNATURA: Soporte para la Toma de Decisiones

SANTA CRUZ – BOLIVIA

ÍNDICE GENERAL

1.OBTENCIÓN DE DATOS?	3
1.1.Recopilación de datos	3
1.2.Procesamiento inicial de los datos	4
1.3. Normalización y codificación	4
1.4. Etiquetado manual o automático	4
1.5. Validación y limpieza	5
2. ¿QUÉ SIGNIFICA CADA DATO?	5
1. checking Status	6
2. Duration	7
3. Credit history	8
4. Purpose	9
5. Credit amount (Numerico)	10
6. Savings status	11
7. Employment	12
8. Installment commitment	13
9. Personal status	13
10. Other Parties	14
11. Residence since	15
12. Property magnitude	16
13. Age (Numérico)	17
14. Other Payment Plans	18
15. Housing	19
16. Existing Credit (Numérico)	20
17. Job	20
18. Number Dependents (Numérico)	21
19. Own Telephone (Categórico)	22
20. Foreign worker	23
3. ¿CUÁL ES EL OBJETIVO DE ESOS DATOS?	24
4. ¿QUÉ ALGORITMOS SE PROBARON?	25
5.¿PORQUE SE ESCOGIERON ESTOS ALGORITMO?	25
6. IMÁGENES DE LAS PRUEBAS CON EL ALGORITMO DE NAIVE BAYES	33
7. CONCLUSIÓN DEL RESULTADO OBTENIDO	40

1.OBTENCIÓN DE DATOS?

Para entender cómo se obtuvieron los datos utilizados en el proyecto **Credit-G** del conjunto **German Credit Data**, es necesario analizar las características del dataset y el contexto en el que fue creado, a continuación se detalla el proceso general que puede haberse seguido para la generación de este tipo de dataset:

1.1.Recopilación de datos

Los datos del German Credit Dataset probablemente fueron recolectados por una institución financiera o un banco alemán. Este conjunto contiene información de clientes relacionada con su historial crediticio y datos demográficos, utilizados para evaluar la probabilidad de que un cliente cumpla con sus obligaciones de pago.

Fuentes de datos:

- **Historial crediticio:** Los datos fueron extraídos de los registros internos del banco, como:
 - Préstamos anteriores.
 - Información de pagos atrasados o incumplidos.
 - Relación con otras instituciones crediticias.
- **Información del cliente:** Recopilada al momento de solicitar el crédito, incluyendo:
 - Edad, género, ocupación.
 - Estado civil y número de dependientes.
 - Ingresos y propiedades (p. ej., casa o automóvil).
- **Comportamiento transaccional:** Datos sobre el uso de servicios bancarios, como cuentas de ahorro o inversión.

1.2. Procesamiento inicial de los datos

Los datos brutos se someten a un procesamiento previo para estructurarlos adecuadamente.

Esto incluye:

- **Conversión a formato tabular:** Los datos se organizan en filas (individuos) y columnas (atributos).
- **Definición de la clase objetivo:** En este caso, los registros se etiquetan como:
 - **Good (Buen riesgo):** Clientes que cumplen con los pagos.
 - **Bad (Mal riesgo):** Clientes que incumplen.
- **Eliminación de datos irrelevantes o confidenciales:** Información personal sensible se excluye para proteger la privacidad.

1.3. Normalización y codificación

Dado que los datos provienen de diferentes fuentes, se asegura la uniformidad mediante:

- **Normalización:** Atributos como montos crediticios o duración de préstamos son escalados a valores estándar.
- **Codificación de atributos categóricos:** Variables como "historial crediticio" o "estado civil" se convierten en valores discretos o numéricos. Ejemplo:
 - Historial crediticio: [1 = excelente, 2 = bueno, 3 = malo].
 - Estado civil: [1 = soltero, 2 = casado, 3 = divorciado].

1.4. Etiquetado manual o automático

La clase objetivo (Good/Bad) pudo haberse determinado de dos maneras:

- **Manual:** Expertos financieros evalúan el riesgo basado en políticas del banco y asignan etiquetas.
- **Automática:** Uso de reglas predefinidas, como:
 - Si un cliente tiene más de tres pagos atrasados → **Bad**.
 - Si tiene un ingreso estable y sin historial de incumplimientos → **Good**.

1.5. Validación y limpieza

Los datos pasan por una fase de validación para garantizar consistencia y calidad:

- **Eliminación de duplicados:** Clientes con múltiples solicitudes.
- **Manejo de valores faltantes:** Relleno con promedio, mediana o eliminación de registros incompletos.
- **Balanceo de clases:** Ajuste del número de ejemplos de "Good" y "Bad" para evitar sesgo en el análisis.

2. ¿QUÉ SIGNIFICA CADA DATO?

Este conjunto de datos tiene 20 atributos que representan características financieras y demográficas de los solicitantes de crédito, así como la clase objetivo:

Clase Objetivo

- **class:**
 - **Good (buen riesgo):** El solicitante tiene un bajo riesgo de incumplir con los pagos.
 - **Bad (mal riesgo):** El solicitante tiene un alto riesgo de incumplir con los pagos.

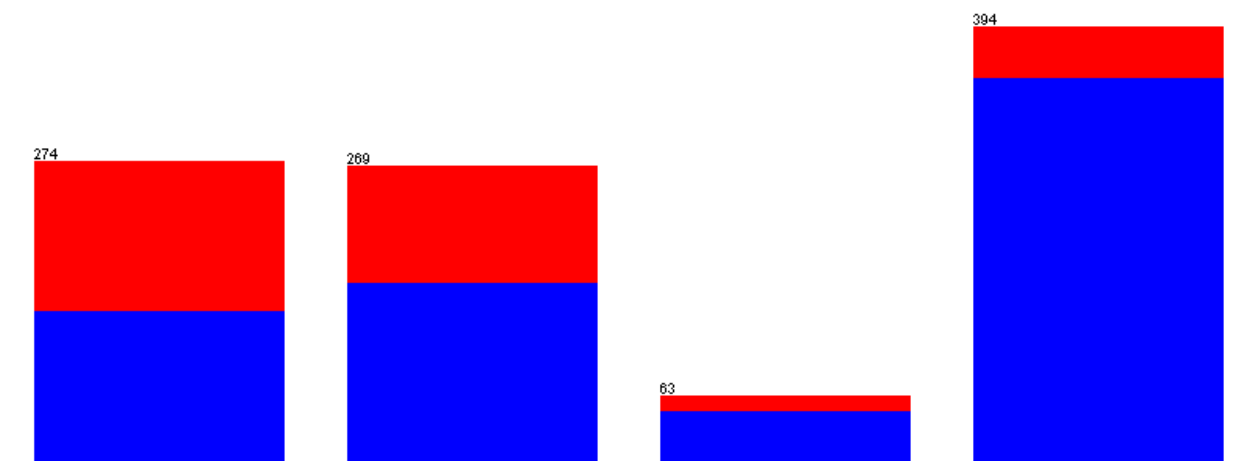
Atributos del dataset

1. checking Status

- Estado de la cuenta corriente del cliente:
 - < 0 : Saldo negativo
 - $0 \leq X < 200$: Saldo entre 0 y 200 DM
 - ≥ 200 : Saldo mayor o igual a 200 DM
 - NO checking: El cliente no tiene cuenta corriente

Selected attribute				
Name: checking_status		Type: Nominal		
Missing: 0 (0%)		Distinct: 4		Unique: 0 (0%)
No.	Label	Count	Weight	
1	< 0	274	274	
2	$0 \leq X < 200$	269	269	
3	≥ 200	63	63	
4	no checking	394	394	

Class: class (Nom) Visualize All

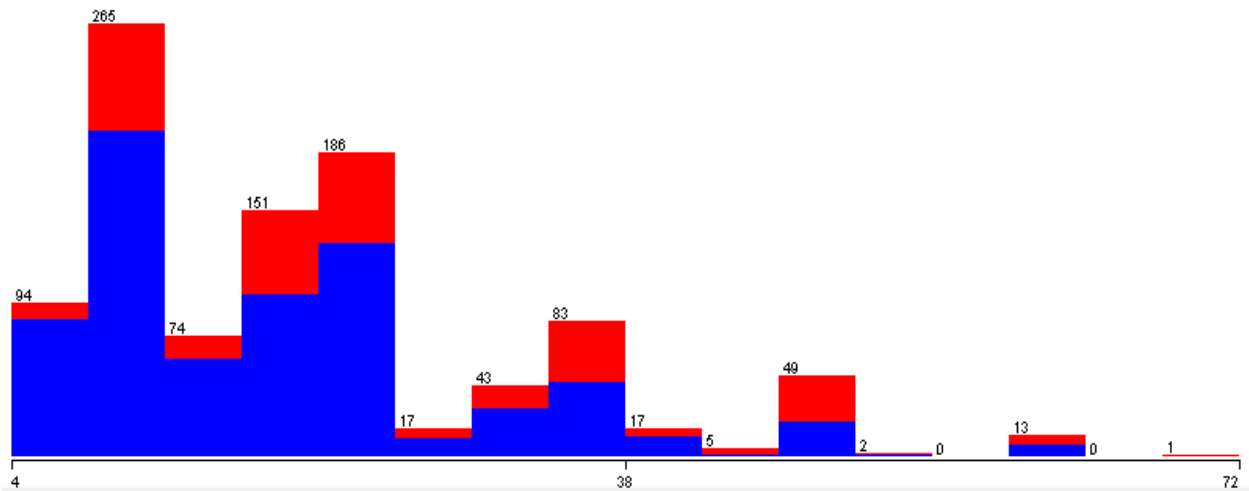


2. Duration

- Duración del crédito solicitado (en meses).
 - Min 4 meses
 - Max 72 meses

Selected attribute		
Name: duration		Type: Numeric
Missing: 0 (0%)	Distinct: 33	Unique: 5 (1%)
Statistic		Value
Minimum		4
Maximum		72
Mean		20.903
StdDev		12.059

Class: class (Nom) Visualize All

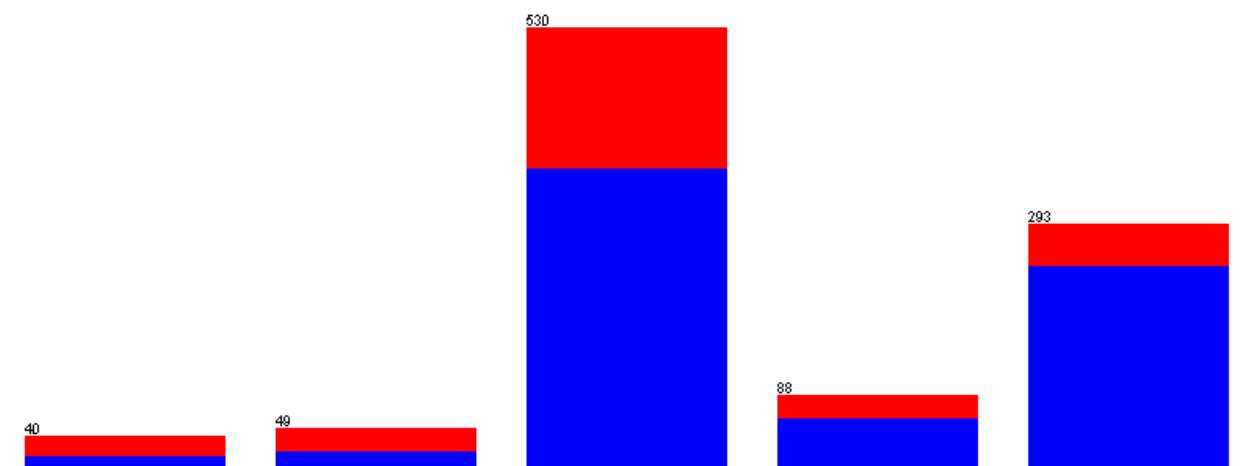


3. Credit history

- Historial crediticio del cliente:
 - no creadits/all paid: Sin créditos tomados o todos los créditos pagados a tiempo.
 - all paid: Todos los créditos pagados a tiempo.
 - existing paid: Tiene créditos existentes, pero los ha pagado puntualmente hasta ahora.
 - delayed previously: Ha tenido retraso en pagos en el pasado.
 - critical / other existing credit: Tiene cuentas criticas o créditos en otros bancos.

Selected attribute				
Name: credit_history		Type: Nominal		
Missing: 0 (0%)		Distinct: 5		Unique: 0 (0%)
No.	Label	Count	Weight	
1	no credits/all paid	40	40	
2	all paid	49	49	
3	existing paid	530	530	
4	delayed previously	88	88	
5	critical/other existing credit	293	293	

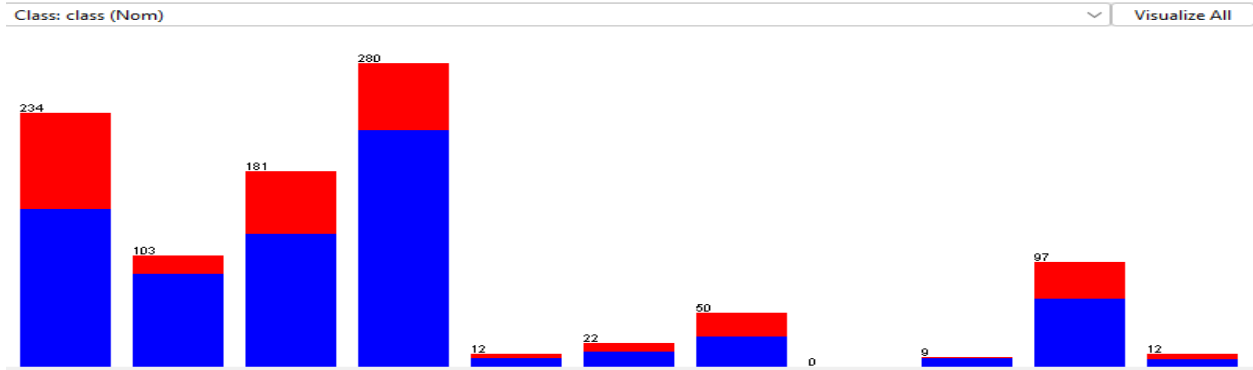
Class: class (Nom) Visualize All



4. Purpose

- Propósito del crédito:
 - new car: Compra de automóvil (nuevo).
 - User car: Compra de automóvil (usado).
 - Furniture/equipment: Muebles o equipos.
 - Radio/tv: Radio o televisión..
 - Domestic appliance: Electrodomésticos.
 - Repairs: Reparaciones.
 - Education: Educación.
 - Vacation: Vacaciones.
 - retraining: Capacidad o reentrenamiento
 - Business: Negocios.
 - other: Otros.

Selected attribute				
Name: purpose		Distinct: 10		Type: Nominal
Missing: 0 (0%)				Unique: 0 (0%)
No.	Label	Count	Weight	
1	new car	234	234	
2	used car	103	103	
3	furniture/equipment	181	181	
4	radio/tv	280	280	
5	domestic appliance	12	12	
6	repairs	22	22	
7	education	50	50	
8	vacation	0	0	
9	retraining	9	9	
10	business	97	97	
11	other	12	12	

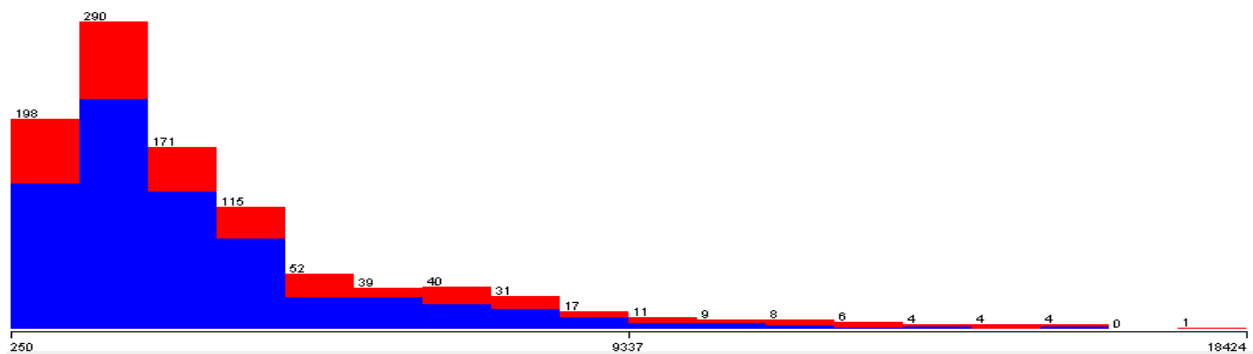


5. Credit amount (Numerico)

- Monto del crédito solicitado (en DM).
 - Min 250
 - Max 18424

Selected attribute		
Name: credit_amount		Type: Numeric
Missing: 0 (0%)		Unique: 847 (85%)
Distinct: 921		
Statistic	Value	
Minimum	250	
Maximum	18424	
Mean	3271.258	
StdDev	2822.737	

Class: class (Nom) Visualize All

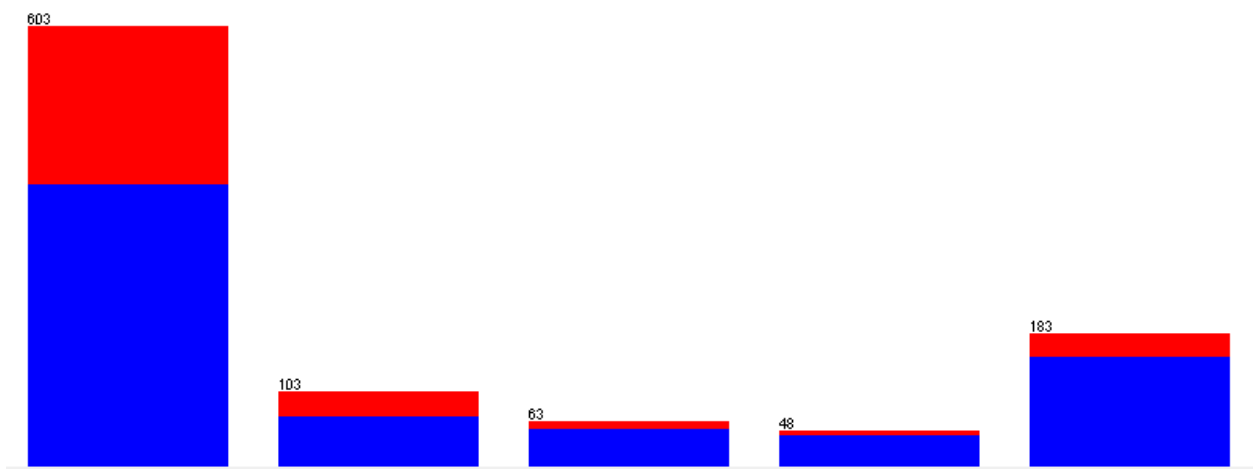


6. Savings status

- Ahorros o bonos del cliente:
 - <1000: Menos de 100 DM.
 - $100 \leq 500$: Entre 100 y 500 DM.
 - $500 \leq X < 1000$: Entre 500 y 1000 DM.
 - ≥ 1000 : Más de 1000 DM.
 - No known savings: No tiene ahorros conocidos

Selected attribute				
Name: savings_status		Distinct: 5		Type: Nominal
Missing: 0 (0%)				Unique: 0 (0%)
No.	Label	Count	Weight	
1	<100	603	603	
2	$100 \leq X < 500$	103	103	
3	$500 \leq X < 1000$	63	63	
4	≥ 1000	48	48	
5	no known savings	183	183	

Class: class (Nom) Visualize All

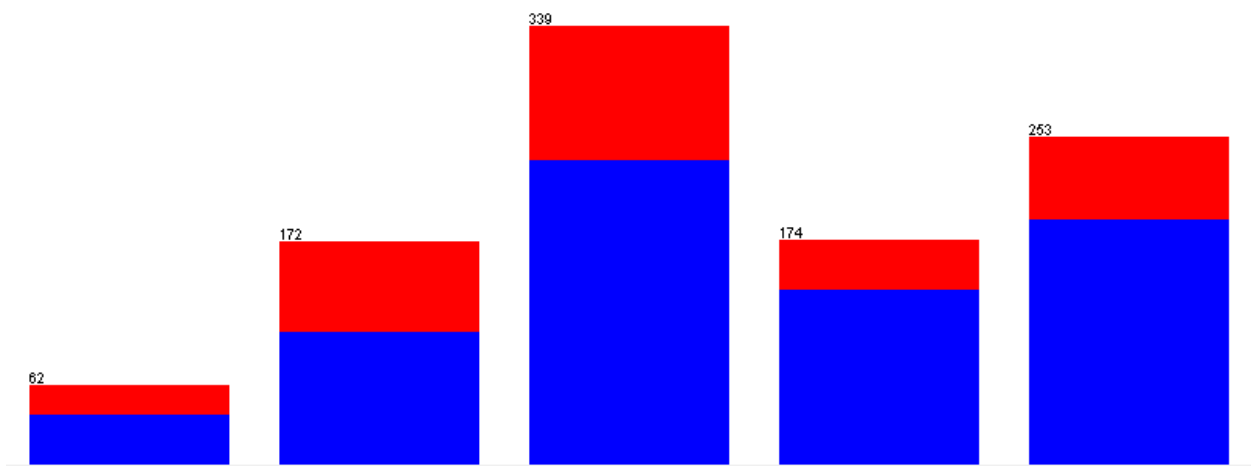


7. Employment

- Tiempo de empleo actual:
 - unemployed: Desempleado
 - < 1 : Menos de 1 año.
 - $1 \leq X < 4$: Entre 1 y 4 años.
 - $4 \leq X < 7$: Entre 4 y 7 años.
 - ≥ 7 : Más de 7 años.

Selected attribute				
Name: employment		Distinct: 5		Type: Nominal
Missing: 0 (0%)				Unique: 0 (0%)
No.	Label	Count	Weight	
1	unemployed	62	62	
2	<1	172	172	
3	$1 \leq X < 4$	339	339	
4	$4 \leq X < 7$	174	174	
5	≥ 7	253	253	

Class: class (Nom) Visualize All

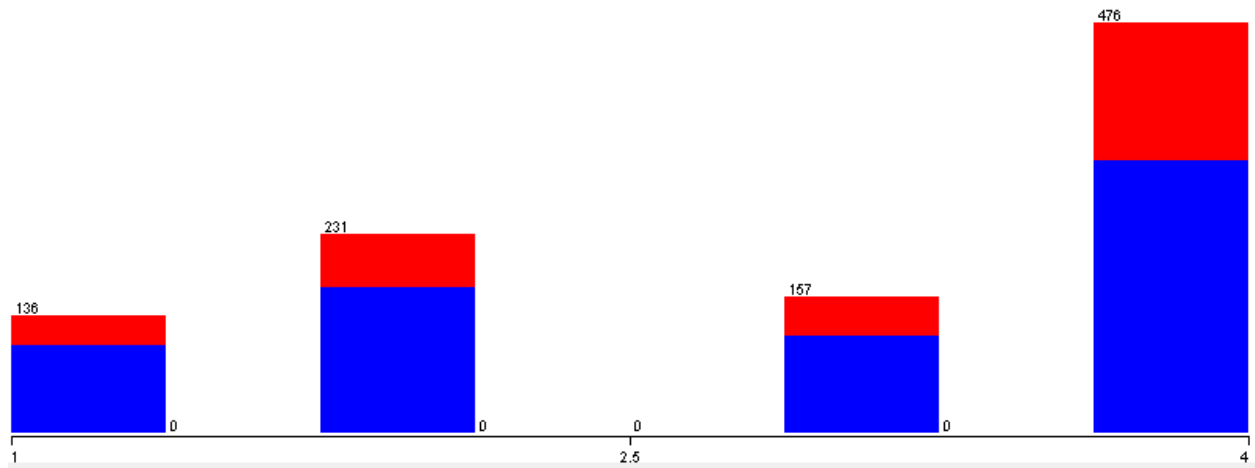


8. Installment commitment

- Porcentaje del ingreso disponible destinado a pagar el crédito (1-4).

Selected attribute		
Name: installment_commitment		Type: Numeric
Missing: 0 (0%)	Distinct: 4	Unique: 0 (0%)
Statistic		Value
Minimum		1
Maximum		4
Mean		2.973
StdDev		1.119

Class: class (Nom) Visualize All



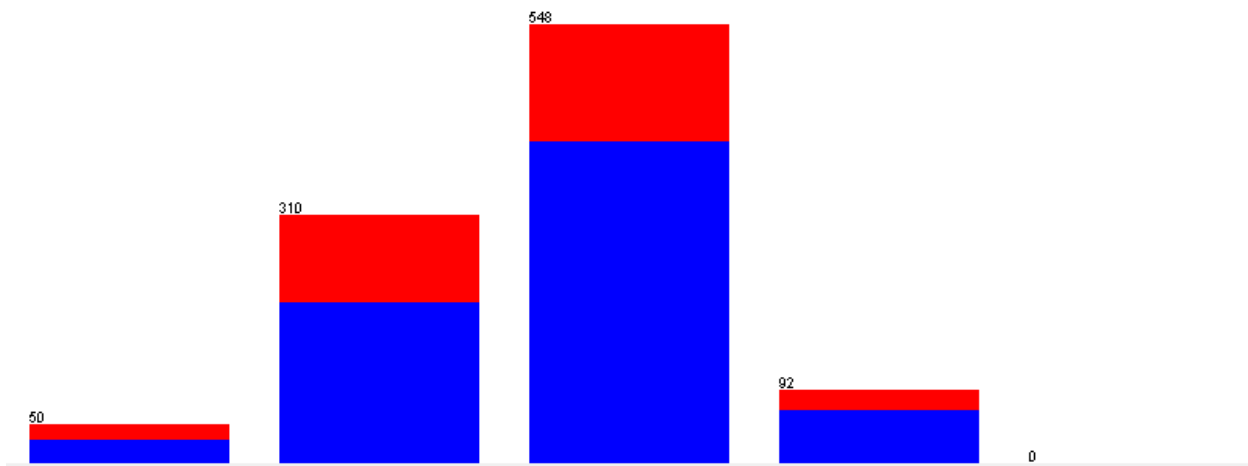
9. Personal status

- Estado civil y género del cliente:
 - male div/sep : Hombre, divorciado o separado
 - female div/dep/mar: Mujer, divorciada, dependiente o casada.
 - male single: Hombre, soltero

- male mar/wid: Hombre, casado o viudo.
- female single: Mujer, soltera.

Selected attribute				
Name: personal_status		Distinct: 4		Type: Nominal
Missing: 0 (0%)				Unique: 0 (0%)
No.	Label	Count		Weight
1	male div/sep	50		50
2	female div/dep/mar	310		310
3	male single	548		548
4	male mar/wid	92		92
5	female single	0		0

Class: class (Nom) Visualize All

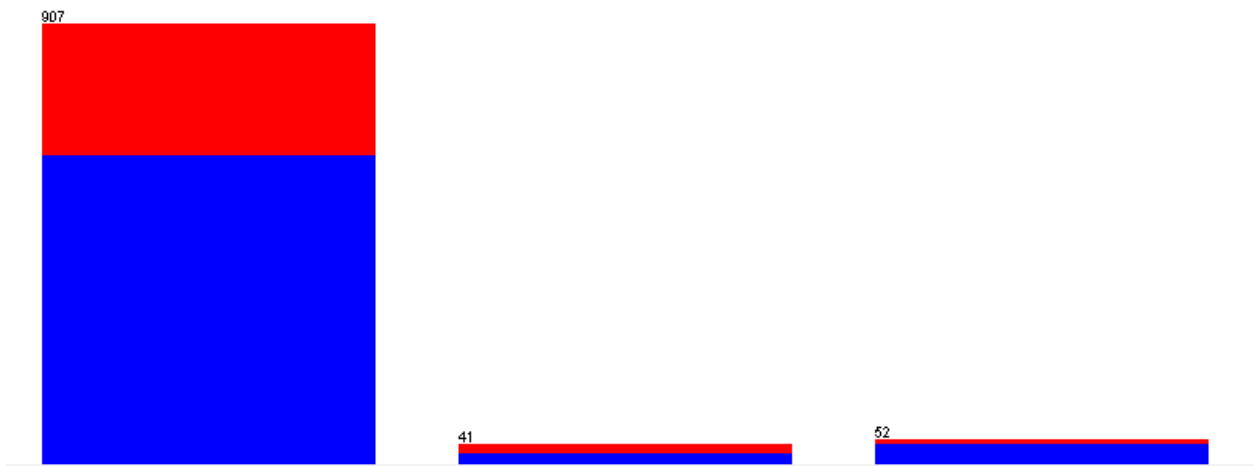


10. Other Parties

- Otras partes involucradas en el credito:
 - none : Ninguna
 - co applicant: Co-aplicante
 - guarantor: Avalista

Selected attribute				
Name: other_parties		Distinct: 3		Type: Nominal
Missing: 0 (0%)				Unique: 0 (0%)
No.	Label	Count	Weight	
1	none	907	907	
2	co applicant	41	41	
3	guarantor	52	52	

Class: class (Nom) Visualize All



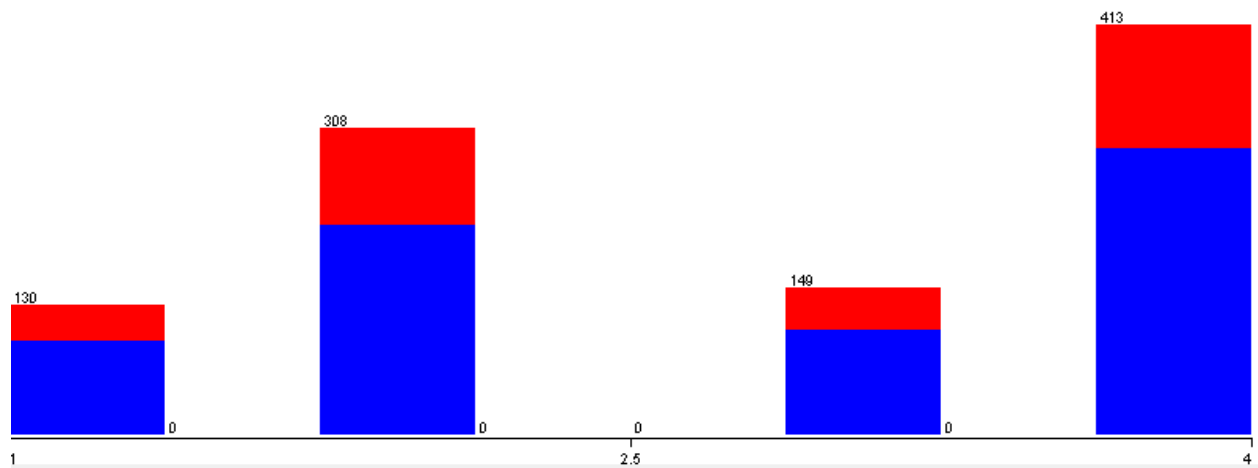
11. Residence since

- Tiempo de residencia actual (en años).
 - Min 1
 - Max 4

Selected attribute		
Name: residence_since		Type: Numeric
Missing: 0 (0%)		Unique: 0 (0%)
Distinct: 4		
Statistic		Value
Minimum		1
Maximum		4
Mean		2.845
StdDev		1.104

Class: class (Nom)

Visualize All

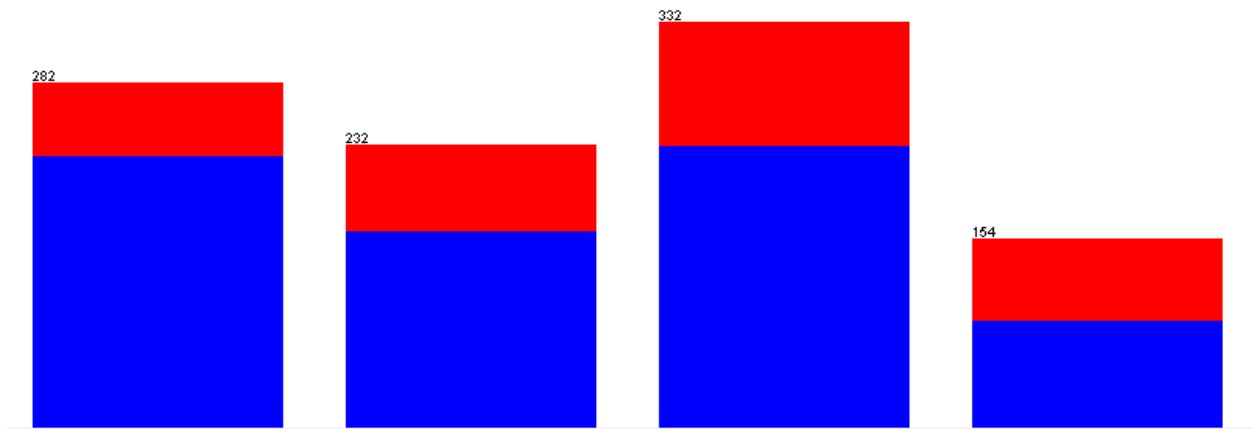


12. Property magnitude

- Tipo de propiedad del solicitante:
 - real estate: bienes raices
 - life insurance: seguro de vida
 - car: auto
 - no known property: ninguna propiedad conocida

Selected attribute				
Name: property_magnitude		Type: Nominal		
Missing: 0 (0%)		Distinct: 4		Unique: 0 (0%)
No.	Label	Count	Weight	
1	real estate	282	282	
2	life insurance	232	232	
3	car	332	332	
4	no known property	154	154	

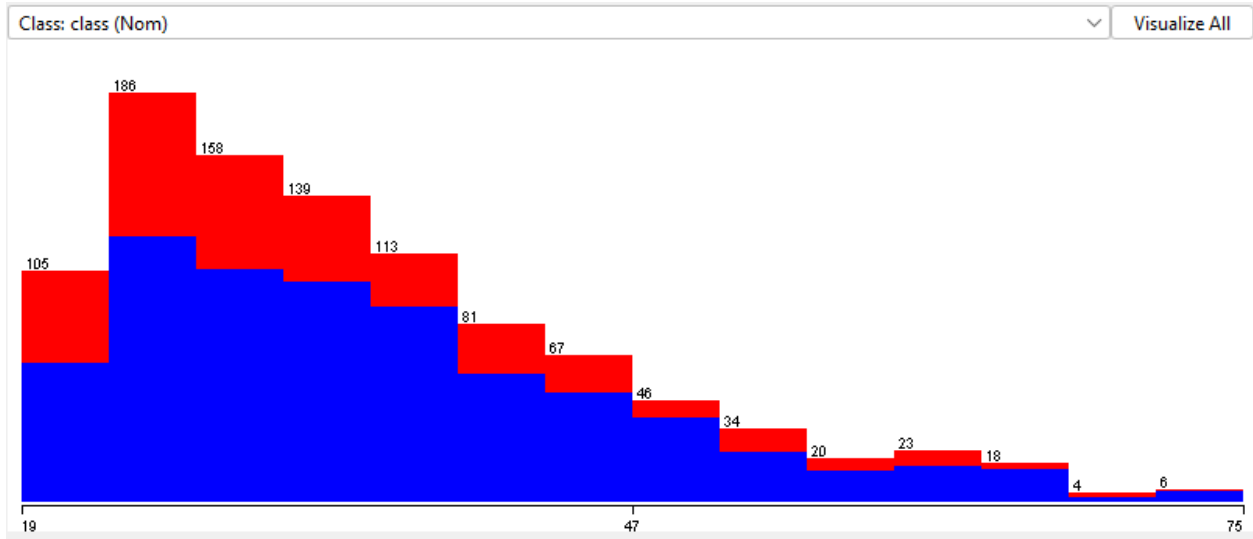
Class: class (Nom) Visualize All



13. Age (Numérico)

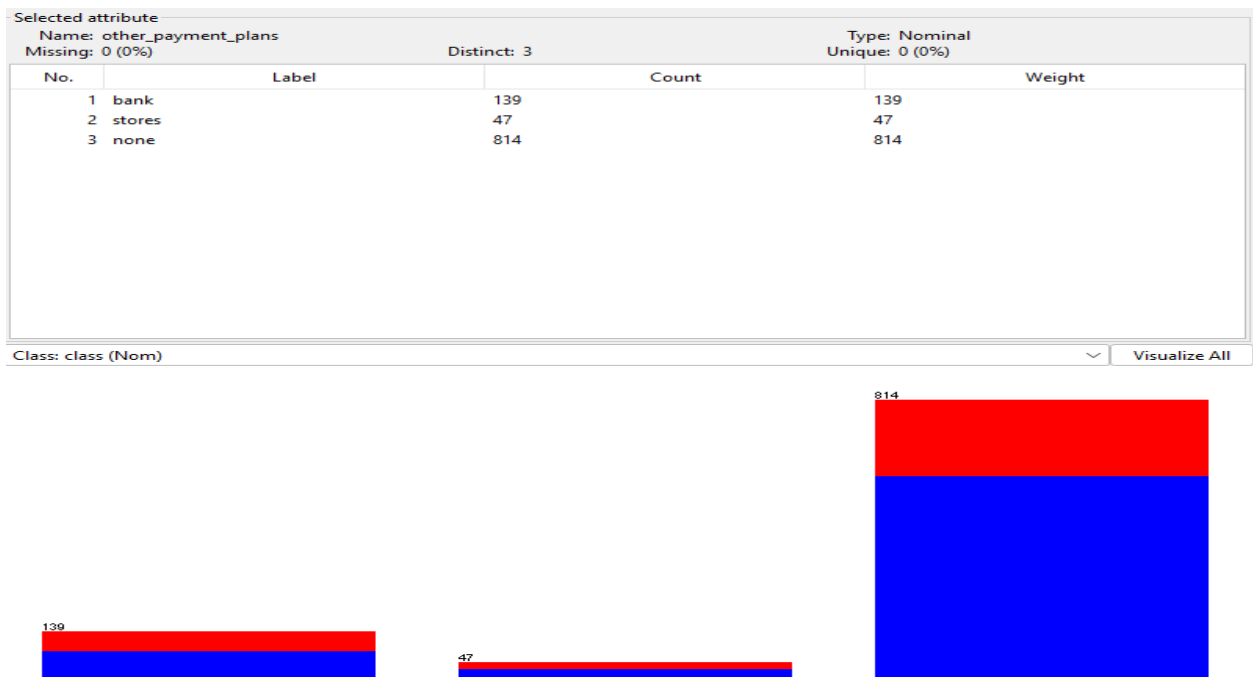
- Edad del cliente(en años).
 - Min 19
 - Max 75

Selected attribute		
Name: age		Type: Numeric
Missing: 0 (0%)		Unique: 1 (0%)
Distinct: 53		
Statistic	Value	
Minimum	19	
Maximum	75	
Mean	35.546	
StdDev	11.375	



14. Other Payment Plans

- Planes de pago adicionales:
 - blank: A través de un banco.
 - stores: A través de tiendas.
 - none: Ninguno.

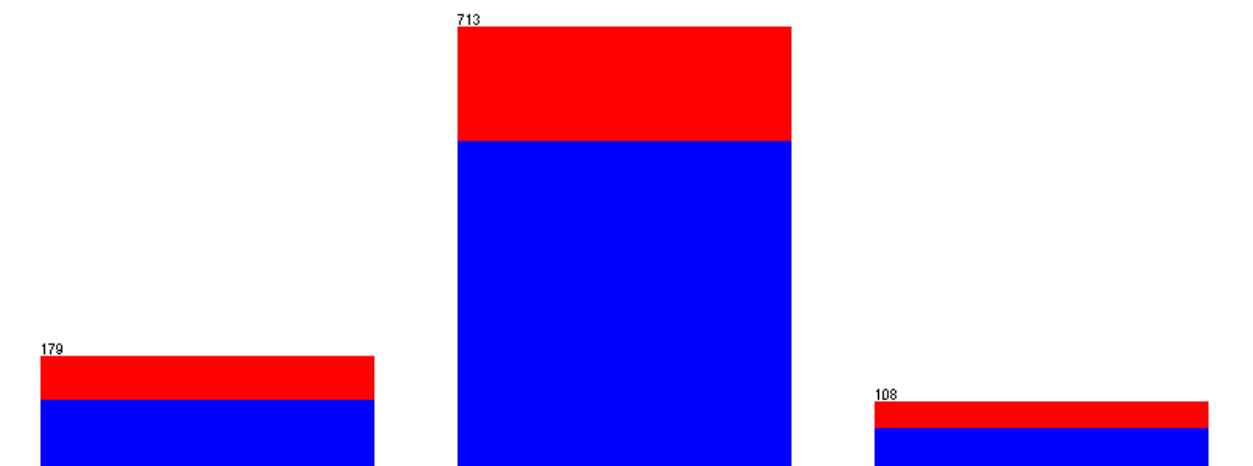


15. Housing

- Tipo de vivienda del cliente:
 - rent: vive en alquiler.
 - own: Posee la vivienda.
 - for free: Vive gratis (casa de familiar).

Selected attribute				
Name: housing			Type: Nominal	
Missing: 0 (0%)			Unique: 0 (0%)	
No.	Label	Count	Weight	
1	rent	179	179	
2	own	713	713	
3	for free	108	108	

Class: class (Nom) Visualize All

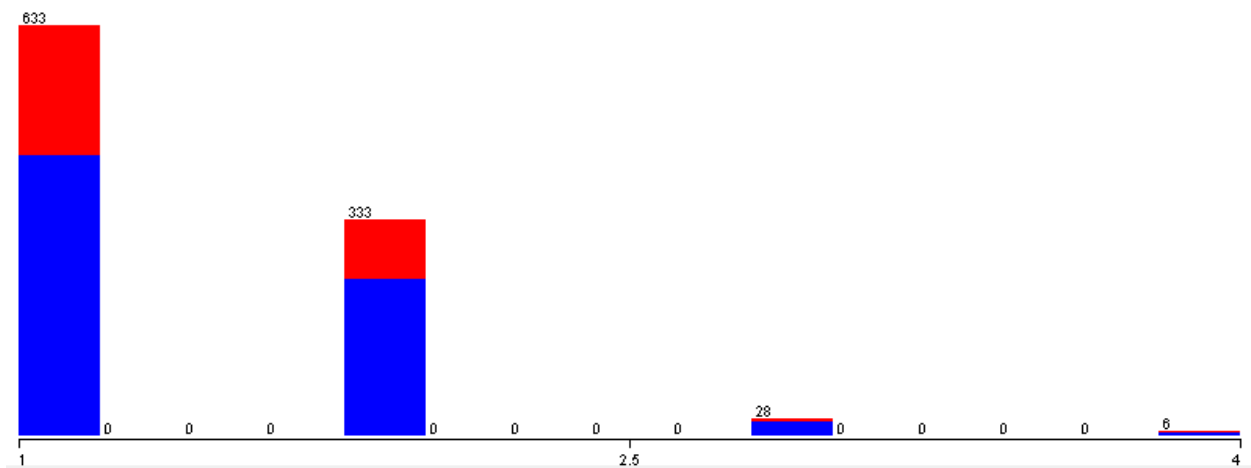


16. Existing Credit (Numérico)

- Número de créditos existentes en este banco.

Selected attribute		
Name: existing_credits		Type: Numeric
Missing: 0 (0%)	Distinct: 4	Unique: 0 (0%)
Statistic	Value	
Minimum	1	
Maximum	4	
Mean	1.407	
StdDev	0.578	

Class: class (Nom) Visualize All



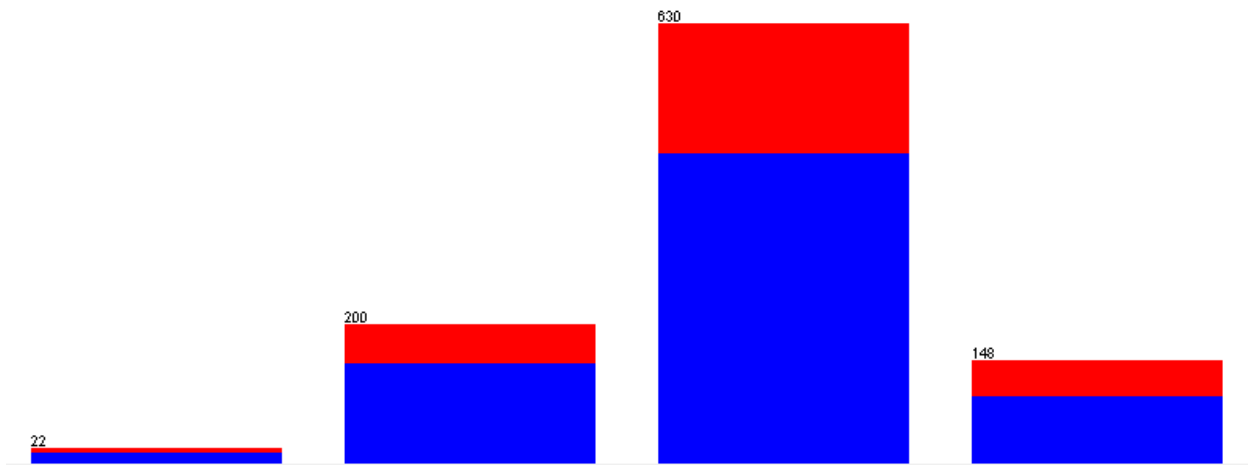
17. Job

- Categoría laboral del cliente:
 - unemp/unskilled non res: desempleado / no calificado
 - unskilled resident: residente no calificado
 - skilled: experto

- high qualif/self emp/ mgmt: alta calificación/autoempleo/administración.

Selected attribute				
Name: job		Type: Nominal		
Missing: 0 (0%)		Distinct: 4		Unique: 0 (0%)
No.	Label	Count	Weight	
1	unemp/unskilled non res	22	22	
2	unskilled resident	200	200	
3	skilled	630	630	
4	high qualif/self emp/mgmt	148	148	

Class: class (Nom) Visualize All



18. Number Dependents (Numérico)

- Número de personas a cargo del cliente(hijos o dependientes).

Selected attribute

Name: num_dependents

Missing: 0 (0%)

Distinct: 2

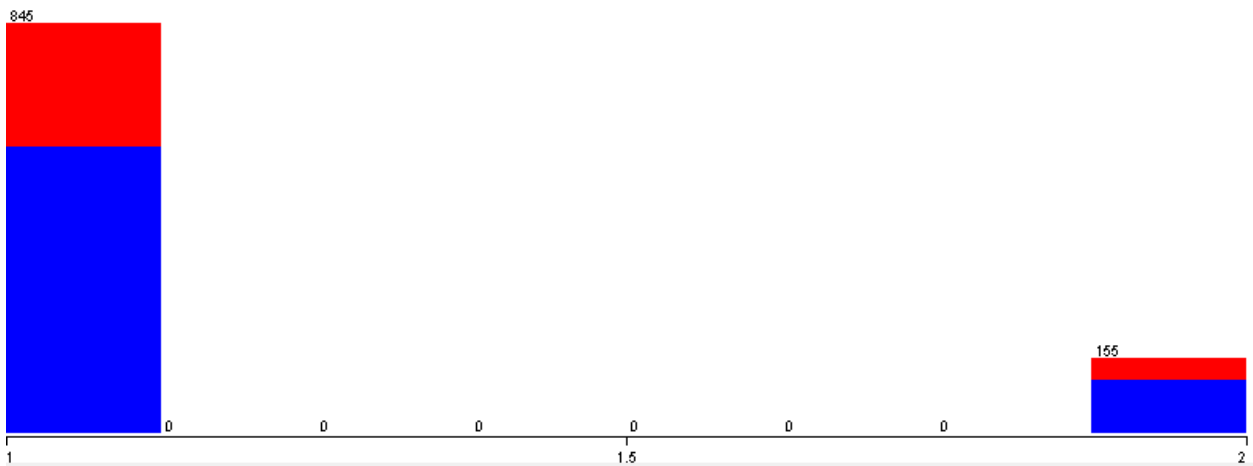
Type: Numeric

Unique: 0 (0%)

Statistic	Value
Minimum	1
Maximum	2
Mean	1.155
StdDev	0.362

Class: class (Nom)

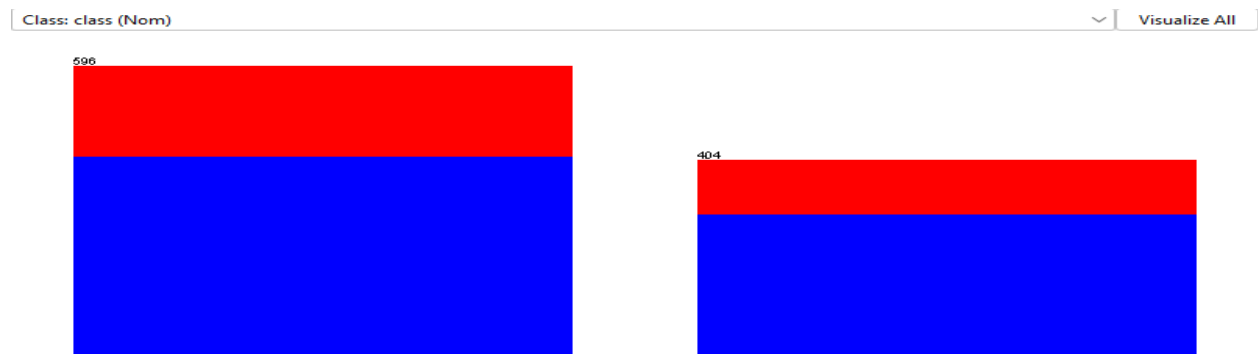
Visualize All



19. Own Telephone (Categórico)

- Teléfono propio:
 - none: No tiene teléfono.
 - yes: Tiene teléfono registrado a su nombre.

Selected attribute			
Name: own_telephone		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
Distinct: 2			
No.	Label	Count	Weight
1	none	596	596
2	yes	404	404



20. Foreign worker

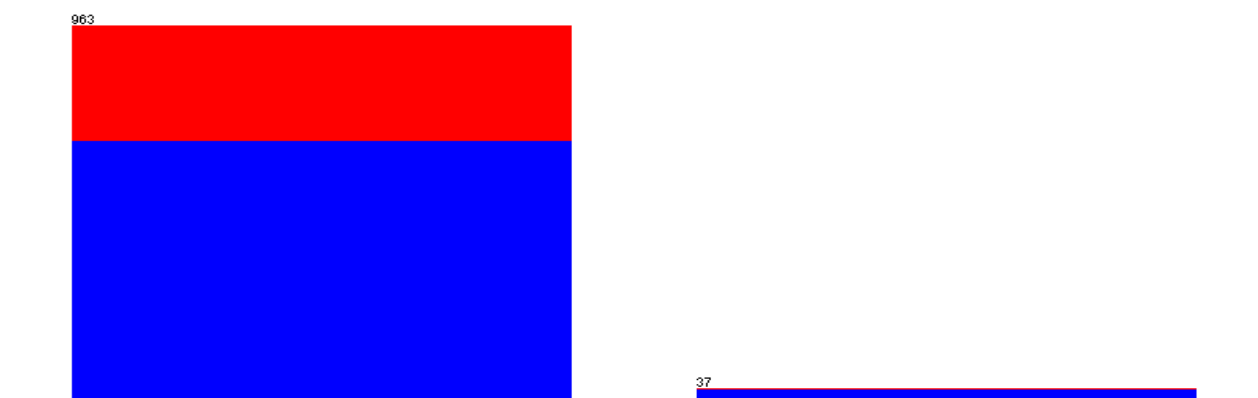
- Si el solicitante es un trabajador extranjero:
 - yes: Sí.
 - not: No.

Selected attribute

Name: foreign_worker Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

No.	Label	Count	Weight
1	yes	963	963
2	no	37	37

Class: class (Nom) Visualize All



3. ¿CUÁL ES EL OBJETIVO DE ESOS DATOS?

El **objetivo** de los datos en el conjunto **German Credit Dataset** (credit-g.arff) es proporcionar un caso práctico para **evaluar el riesgo crediticio** de los clientes de una entidad financiera. Este dataset sirve como base para la construcción y validación de modelos de aprendizaje automático que clasifiquen a los solicitantes en dos categorías principales:

Las dos categorías principales en el conjunto de datos **German Credit Dataset** (credit-g.arff) son:

1. **Good (Buen riesgo)**

- Representa a los solicitantes que tienen un **bajo riesgo de incumplimiento** en sus pagos.
- Estos solicitantes son considerados como confiables por la entidad financiera y tienen alta probabilidad de cumplir con las condiciones del crédito.

2. **Bad (Mal riesgo)**

- Representa a los solicitantes que tienen un **alto riesgo de incumplimiento** en sus pagos.
- Estos solicitantes son considerados como no confiables o de alto riesgo por la entidad financiera, lo que podría resultar en la negativa del crédito o en la aplicación de condiciones más estrictas (como tasas de interés más altas).

Objetivo principal

Predecir si un solicitante de crédito es un **"buen riesgo"** o un **"mal riesgo"**, basándose en sus características financieras, demográficas y comportamentales.

Contexto práctico

En la vida real, las instituciones financieras y bancos utilizan este tipo de modelos para tomar decisiones informadas sobre:

1. **Aprobación o rechazo de créditos.**
2. **Condiciones del crédito** (tasa de interés, monto permitido, plazo de pago).
3. **Gestión del portafolio de riesgo** para minimizar pérdidas y maximizar ganancias.

4. ¿QUÉ ALGORITMOS SE PROBARON?

- ALGORITMO DE NAIBE BAYES
- ALGORITMO DEL ARBOL J48
- ALGORITMO ADABOOSTM1

5.¿PORQUE SE ESCOGIERON ESTOS ALGORITMO?

Justificación del Algoritmo Naive Bayes

La decisión de utilizar el algoritmo **Naive Bayes** en este proyecto se fundamentó en un análisis cuidadoso de las características del problema de clasificación y las ventajas del algoritmo frente a otras alternativas. A continuación, justificación de su selección:

1. Naturaleza del Problema: Clasificación Binaria

El problema a resolver consiste en clasificar a los clientes en dos categorías:

- **Good** (clientes con alto potencial de cumplir con sus pagos).
- **Bad** (clientes con riesgo de incumplimiento).

Naive Bayes es un modelo estadístico probabilístico que se especializa en tareas de clasificación binaria, ofreciendo resultados confiables en problemas de evaluación de riesgo crediticio. Además, la interpretación probabilística de los resultados se alinea con la necesidad de justificar las decisiones de aprobación o rechazo de créditos.

2. Estructura del Conjunto de Datos

El *German Credit Dataset* contiene 20 atributos que combinan:

- **Variables categóricas**, como estado civil, historial crediticio, y tipo de vivienda.
- **Variables numericas**, como monto del crédito y duración del préstamo.

Naive Bayes tiene una ventaja particular porque:

- Puede manejar **variables categóricas** utilizando la **distribución categórica**, que calcula las probabilidades directamente para cada categoría.
- Es capaz de procesar **variables continuas** mediante la **distribución Gaussiana**, que modela estos datos según una curva normal.

Esta flexibilidad asegura que el modelo pueda utilizar de manera efectiva toda la información presente en el dataset.

3. Simplicidad y Velocidad de Implementación

Naive Bayes es conocido por ser:

- **Rápido de entrenar y clasificar:** Su bajo costo computacional permite generar resultados en tiempo real, lo cual es crucial para aplicaciones como evaluaciones de crédito, donde se requiere una respuesta inmediata.

- **Simple de implementar:** Su estructura matemática es sencilla, basada en el Teorema de Bayes y una independencia condicional que simplifica los cálculos.

4. Robustez frente a Suposiciones de Independencia

Aunque Naive Bayes asume independencia condicional entre las características (atributos), ha demostrado ofrecer resultados sólidos incluso en casos donde esta suposición no se cumple completamente. Dado que los atributos en el *German Credit Dataset* tienen cierta correlación (por ejemplo, historial crediticio y monto solicitado), esta propiedad lo convierte en una opción robusta para este caso.

5. Interpretabilidad Probabilística

Una de las ventajas más importantes de Naive Bayes es que:

- Proporciona una **probabilidad explícita** de que un cliente pertenezca a la categoría "Good" o "Bad".
- Esto permite **justificar la decisión final** de aceptar o rechazar una solicitud de crédito, aportando transparencia al proceso.

Por ejemplo, en las pruebas realizadas, el sistema calculó una **probabilidad de 97.06%** de que un cliente cumpliera con sus pagos, justificando claramente la decisión de aprobar su crédito.

LOAN APPROVAL CALCULATOR

<p>Status of existing checking account</p> <p>0<=X<200 ▼</p> <p>Duration in month.</p> <p>12 ▼</p> <p>Credit history.</p> <p>critical/other existing ... ▼</p> <p>Purpose.</p> <p>used car ▼</p> <p>Credit amount.</p> <p>1804</p> <p>Savings account/bonds.</p> <p>100<=X<500 ▼</p> <p>Present employment since.</p> <p><1 ▼</p> <p>Installment rate in percentage of disposable income .</p> <p>3 ▼</p> <p>Personal status and sex.</p> <p>male single ▼</p> <p>Other debtors / guarantors</p> <p>none ▼</p>	<p>Present residence since.</p> <p>4</p> <p>Property</p> <p>life insurance ▼</p> <p>Age in years</p> <p>44</p> <p>Other installment plans</p> <p>none ▼</p> <p>Housing</p> <p>own ▼</p> <p>Number of existing credits at this bank</p> <p>1 ▼</p> <p>Job</p> <p>skilled ▼</p> <p>Persons liable for maintenance</p> <p>1 ▼</p> <p>Telephone</p> <p>none ▼</p> <p>Foreign worker</p> <p>yes ▼</p>
---	--

Cliente aceptado, probabilidad de 97.06% de cumplir con sus pagos

Ejecutar

Cliente aceptado, probabilidad de 97.06% de cumplir con sus pagos

Ejecutar

ALGORITMO DEL ÁRBOL J48

- **Facilidad de interpretación:**

Los árboles de decisión generados por J48 representan gráficamente las reglas de clasificación en un formato fácil de entender. Esto es valioso para los analistas financieros, ya que pueden visualizar las decisiones del modelo mediante reglas como:

- "Si la duración del crédito es mayor a 36 meses y el historial crediticio es excelente, clasificar como *Good*".
- "Si el estado de cuenta corriente es menor a 200 y el cliente tiene pagos atrasados, clasificar como *Bad*".

Estas reglas son directamente interpretables y utilizables para justificar las decisiones ante clientes o reguladores.

- **Capacidad para manejar atributos mixtos:**

J48 puede trabajar con atributos categóricos y numéricos. Durante el entrenamiento, convierte automáticamente los datos numéricos en rangos que mejoran la precisión de la clasificación.

- **Identificación de interacciones entre características:**

J48 no asume independencia entre las características. Esto permite identificar relaciones complejas entre los atributos, como la interacción entre el monto del crédito solicitado y el historial crediticio, que son fundamentales para el análisis del riesgo crediticio.

- **Manejo de valores faltantes:**

J48 incluye técnicas internas para gestionar valores ausentes, como dividir nodos basándose en distribuciones ponderadas. Esto reduce la necesidad de preprocesamiento manual y asegura que el modelo utilice la mayor cantidad de información posible.

- **Control del sobreajuste mediante poda:**

J48 utiliza un proceso llamado **pruning** (poda) para reducir el riesgo de sobreajuste. Esto es crucial en datasets como este, donde un modelo muy ajustado a los datos de entrenamiento podría fallar en generalizar correctamente a nuevos clientes.

ALGORITMO ADABOOSTM1

AdaBoostM1 (Adaptive Boosting) fue seleccionado como el algoritmo principal debido a sus características avanzadas de mejora de clasificación, robustez frente a ruido y capacidad de generalización. Este enfoque es ideal para el problema de aprobación de préstamos, donde se busca clasificar con precisión entre clientes de bajo y alto riesgo crediticio (*Good vs Bad*) razones específicas de su elección:

1. Mejora del Desempeño

AdaBoostM1 sobresale al mejorar clasificadores débiles, como los árboles de decisión (J48, en este caso), mediante un método iterativo que ajusta los errores en cada paso.

Por qué es útil en este caso:

- **Clasificación de casos difíciles:** AdaBoostM1 se enfoca dinámicamente en las instancias más difíciles de clasificar ajustando sus pesos, permitiendo que el modelo corrija sus errores de forma progresiva.
- **Mayor precisión:** Al combinar múltiples clasificadores débiles, el modelo logra un rendimiento superior en comparación con el uso de un solo clasificador, lo que es crucial para un sistema financiero donde la precisión en la evaluación del riesgo crediticio es fundamental.

2. Robustez ante Ruido y Datos Desbalanceados

En problemas reales como el análisis de riesgo crediticio, los datos suelen ser ruidosos y desbalanceados, es decir, hay muchas más instancias de clientes "Good" que "Bad".

3. Capacidad de Generalización

AdaBoostM1 reduce el riesgo de **sobreajuste** (overfitting) al enfocarse en minimizar errores globales y no solo en memorizar patrones específicos de los datos de entrenamiento.

4. Versatilidad en Tipos de Datos

El dataset de evaluación crediticia contiene variables categóricas (como credit_history o purpose) y numéricas (como duration o credit_amount), lo que puede ser un desafío para algunos algoritmos.

5. Interpretabilidad y Explicabilidad

En el sector financiero, la interpretabilidad del modelo es clave para justificar las decisiones de aprobación o rechazo de préstamos.

Por qué AdaBoostM1 es adecuado:

- **Explicabilidad a través del clasificador base:** Usando J48 como clasificador base, las decisiones pueden explicarse mediante reglas claras, como:
"Si el historial crediticio es bueno y el monto del préstamo es menor a 10,000, clasificar como 'Good'."

6. IMÁGENES DE LAS PRUEBAS CON EL ALGORITMO DE NAIVE BAYES

LOAN APPROVAL CALCULATOR

Status of existing checking account	Present residence since.
0<=X<200	4
Duration in month.	Property
12	life insurance
Credit history.	Age in years
critical/other existing ...	44
Purpose.	Other installment plans
used car	none
Credit amount.	Housing
1804	own
Savings account/bonds.	Number of existing credits at this bank
100<=X<500	1
Present employment since.	Job
<1	skilled
Installment rate in percentage of disposable income .	Persons liable for maintenance
3	1
Personal status and sex.	Telephone
male single	none
Other debtors / guarantors	Foreign worker
none	yes

Cliente aceptado, probabilidad de 97.06% de cumplir con sus pagos

Ejecutar

LOAN APPROVAL CALCULATOR

<p>Status of existing checking account</p> <div style="border: 1px solid black; background-color: #007bff; color: white; padding: 2px;">current</div> <p>Duration in month.</p> <div style="border: 1px solid black; padding: 2px;">12</div> <p>Credit history.</p> <div style="border: 1px solid black; padding: 2px;">critical/other existing ...</div> <p>Purpose.</p> <div style="border: 1px solid black; padding: 2px;">used car</div> <p>Credit amount.</p> <div style="border: 1px solid black; padding: 2px;">1804</div> <p>Savings account/bonds.</p> <div style="border: 1px solid black; padding: 2px;">100<=X<500</div> <p>Present employment since.</p> <div style="border: 1px solid black; padding: 2px;"><1</div> <p>Installment rate in percentage of disposable income .</p> <div style="border: 1px solid black; padding: 2px;">3</div> <p>Personal status and sex.</p> <div style="border: 1px solid black; padding: 2px;">male single</div> <p>Other debtors / guarantors</p> <div style="border: 1px solid black; padding: 2px;">none</div>	<p>Present residence since.</p> <div style="border: 1px solid black; padding: 2px;">4</div> <p>Property</p> <div style="border: 1px solid black; padding: 2px;">life insurance</div> <p>Age in years</p> <div style="border: 1px solid black; padding: 2px;">44</div> <p>Other installment plans</p> <div style="border: 1px solid black; padding: 2px;">none</div> <p>Housing</p> <div style="border: 1px solid black; padding: 2px;">own</div> <p>Number of existing credits at this bank</p> <div style="border: 1px solid black; padding: 2px;">1</div> <p>Job</p> <div style="border: 1px solid black; padding: 2px;">skilled</div> <p>Persons liable for maintenance</p> <div style="border: 1px solid black; padding: 2px;">1</div> <p>Telephone</p> <div style="border: 1px solid black; padding: 2px;">none</div> <p>Foreign worker</p> <div style="border: 1px solid black; padding: 2px;">yes</div>
---	--

Cliente aceptado, probabilidad de 95.35% de cumplir.

Ejecutar

LOAN APPROVAL CALCULATOR

<p>Status of existing checking account</p> <p><input type="text" value=""/></p> <p>Duration in month.</p> <p><input type="text" value="12"/></p> <p>Credit history.</p> <p><input type="text" value="critical/other existing ..."/></p> <p>Purpose.</p> <p><input type="text" value="used car"/></p> <p>Credit amount.</p> <p><input type="text" value="1804"/></p> <p>Savings account/bonds.</p> <p><input type="text" value="100<=X<500"/></p> <p>Present employment since.</p> <p><input type="text" value="<1"/></p> <p>Installment rate in percentage of disposable income .</p> <p><input type="text" value="3"/></p> <p>Personal status and sex.</p> <p><input type="text" value=""/></p> <p>Other debtors / guarantors</p> <p><input type="text" value="none"/></p>	<p>Present residence since.</p> <p><input type="text" value="4"/></p> <p>Property</p> <p><input type="text" value="life insurance"/></p> <p>Age in years</p> <p><input type="text" value="44"/></p> <p>Other installment plans</p> <p><input type="text" value="none"/></p> <p>Housing</p> <p><input type="text" value="own"/></p> <p>Number of existing credits at this bank</p> <p><input type="text" value="1"/></p> <p>Job</p> <p><input type="text" value="skilled"/></p> <p>Persons liable for maintenance</p> <p><input type="text" value="1"/></p> <p>Telephone</p> <p><input type="text" value="none"/></p> <p>Foreign worker</p> <p><input type="text" value="yes"/></p>
--	--

Cliente aceptado, probabilidad de 94.54% de cumplir.

Ejecutar

Cliente aceptado, probabilidad de 94.54% de cumplir.

Ejecutar

LOAN APPROVAL CALCULATOR

<p>Status of existing checking account</p> <p><input type="text" value=""/></p> <p>Duration in month.</p> <p><input type="text" value="12"/></p> <p>Credit history.</p> <p><input type="text" value="critical/other existing ..."/></p> <p>Purpose.</p> <p><input type="text" value="used car"/></p> <p>Credit amount.</p> <p><input type="text" value="1804"/></p> <p>Savings account/bonds.</p> <p><input type="text" value="100<=X<500"/></p> <p>Present employment since.</p> <p><input type="text" value="<1"/></p> <p>Installment rate in percentage of disposable income .</p> <p><input type="text" value="3"/></p> <p>Personal status and sex.</p> <p><input type="text" value=""/></p> <p>Other debtors / guarantors</p> <p><input type="text" value="none"/></p>	<p>Present residence since.</p> <p><input type="text" value=""/></p> <p>Property</p> <p><input type="text" value="life insurance"/></p> <p>Age in years</p> <p><input type="text" value="44"/></p> <p>Other installment plans</p> <p><input type="text" value="none"/></p> <p>Housing</p> <p><input type="text" value="own"/></p> <p>Number of existing credits at this bank</p> <p><input type="text" value="1"/></p> <p>Job</p> <p><input type="text" value="skilled"/></p> <p>Persons liable for maintenance</p> <p><input type="text" value=""/></p> <p>Telephone</p> <p><input type="text" value=""/></p> <p>Foreign worker</p> <p><input type="text" value="yes"/></p>
--	--

Cliente aceptado, probabilidad de 94.86% de cumplir.


Ejecutar

Cliente aceptado, probabilidad de 94.86% de cumplir.

Ejecutar

DESPUÉS DE HACER LAS MODIFICACIONES SE OBTUVIERON LOS SIGUIENTES RESULTADOS

ANTES:



—

□

×

LOAN APPROVAL CALCULATOR

Status of existing checking account

0<=X<200

Duration in month.

12

Credit history.

critical/other existing ...

Purpose.

used car

Credit amount.

1804

Savings account/bonds.

100<=X<500

Present employment since.

<1

Installment rate in percentage of disposable income .

3

Personal status and sex.

male single

Other debtors / guarantors

none

Present residence since.

4

Property

life insurance

Age in years

44

Other installment plans

none

Housing

own

Number of existing credits at this bank

1

Job

skilled

Persons liable for maintenance

1

Telephone

none


Foreign worker

yes

Cliente aceptado, probabilidad de 97.06% de cumplir con sus pagos

Ejecutar

AHORA :

—□×

CALCULATOR OF PROBABILITY

Duration in month.	Property
12	life insurance
Credit history.	Age in years
critical/other existing ...	44
Purpose.	Other installment plans
used car	none
Credit amount.	Housing
1804	own
Savings account/bonds.	Number of existing credits at this bank
100<=X<500	1
Present employment since.	Job
<1	skilled
Installment rate in percentage of disposable income .	Foreign worker
3	yes
Other debtors / guarantors	
none	

Cliente aceptado, probabilidad de 94.91% de cumplir con sus cuotas.

Ejecutar

ANTES:

LOAN APPROVAL CALCULATOR

<p>Status of existing checking account</p> <p>0<=X<200 ▼</p> <p>Duration in month.</p> <p>24 ▼</p> <p>Credit history.</p> <p>existing paid ▼</p> <p>Purpose.</p> <p>used car ▼</p> <p>Credit amount.</p> <p>12579</p> <p>Savings account/bonds.</p> <p><100 ▼</p> <p>Present employment since.</p> <p>>=7 ▼</p> <p>Installment rate in percentage of disposable income .</p> <p>4 ▼</p> <p>Personal status and sex.</p> <p>female div/dep/mar ▼</p> <p>Other debtors / guarantors</p> <p>none ▼</p>	<p>Present residence since.</p> <p>2</p> <p>Property</p> <p>no known property ▼</p> <p>Age in years</p> <p>44</p> <p>Other installment plans</p> <p>none ▼</p> <p>Housing</p> <p>for free ▼</p> <p>Number of existing credits at this bank</p> <p>1 ▼</p> <p>Job</p> <p>high qualif/self emp/... ▼</p> <p>Persons liable for maintenance</p> <p>1 ▼</p> <p>Telephone</p> <p>yes ▼</p> <p>Foreign worker</p> <p>yes ▼</p>
---	--

Cliente rechazado, probabilidad de 99.16% de incumplir.

Ejecutar

AHORA:

CALCULATOR OF PROBABILITY

Duration in month. <input type="text" value="24"/>	Property <input type="text" value="no known property"/>
Credit history. <input type="text" value="existing paid"/>	Age in years <input type="text" value="44"/>
Purpose. <input type="text" value="used car"/>	Other installment plans <input type="text" value="none"/>
Credit amount. <input type="text" value="12579"/>	Housing <input type="text" value="for free"/>
Savings account/bonds. <input type="text" value="<100"/>	Number of existing credits at this bank <input type="text" value="1"/>
Present employment since. <input type="text" value="≥7"/>	Job <input type="text" value="high qualif/self emp/..."/>
Installment rate in percentage of disposable income . <input type="text" value="4"/>	Foreign worker <input type="text" value="yes"/>
Other debtors / guarantors <input type="text" value="none"/>	

Cliente rechazado, probabilidad de 98.57% de incumplir con sus cuotas.

Ejecutar

7. CONCLUSIÓN DEL RESULTADO OBTENIDO

El sistema ha determinado que el cliente tiene una **probabilidad del 98.57% de incumplir con sus cuotas**, por lo que se ha decidido rechazar la solicitud de crédito. Este resultado se explica de la siguiente manera:

Factores Relevantes Identificados

1. **Historial crediticio ("existing paid"):** Aunque el cliente ha realizado pagos previos, otros factores influyen negativamente en su evaluación general.
2. **Duración del crédito (24 meses):** Un plazo de 2 años puede aumentar la incertidumbre sobre la estabilidad del cliente para cumplir con los pagos.
3. **Monto del crédito solicitado (12,579):** El monto elevado del crédito aumenta el riesgo financiero.
4. **Ahorros (<100):** La falta de ahorros sustanciales (< 100) reduce la capacidad del cliente para manejar imprevistos financieros.
5. **Tasa de pago sobre ingresos disponibles (4%):** Aunque este porcentaje no es alto, en combinación con otros factores, no mejora la evaluación del cliente.
6. **Propiedad ("no known property"):** La ausencia de bienes como respaldo reduce la confianza de que el cliente pueda cumplir con sus obligaciones.
7. **Ocupación ("high qualified/self employed"):** A pesar de ser una ocupación calificada, trabajos autónomos pueden asociarse con ingresos variables o menos predecibles.

Justificación del Resultado

El sistema considera un conjunto de factores interrelacionados que incrementan el riesgo de incumplimiento, como el **monto solicitado elevado**, la **falta de respaldo financiero (ahorros y propiedades)** y el contexto laboral del cliente. Aunque el historial crediticio podría ser un indicador positivo, no es suficiente para compensar los demás factores desfavorables.

Decisión

El sistema concluye que aprobar el crédito representaría un riesgo financiero considerable, justificando la decisión de **rechazar** la solicitud para minimizar posibles pérdidas.

7. IMÁGENES DE LAS PRUEBAS CON EL ALGORITMO DE ADABOOSTM1

Loan Approval Calculator (Using Pretrained Model)

Loan Approval Calculator

Checking Status	<0
Duration (months)	24
Credit History	critical/other existing credit
Purpose	new car
Credit Amount	15000
Savings Status	no known savings
Employment Duration	unemployed
Installment Commitment (%)	3

Decision: REJECTED (Confidence: 97,0%)

Calculate Loan Approval



Loan Approval Calculator

Checking Status	<input type="text" value=">=200"/>
Duration (months)	<input type="text" value="24"/>
Credit History	<input type="text" value="all paid"/>
Purpose	<input type="text" value="new car"/>
Credit Amount	<input type="text" value="15000"/>
Savings Status	<input type="text" value="<100"/>
Employment Duration	<input type="text" value="unemployed"/>
Installment Commitment (%)	<input type="text" value="3"/>

Decision: REJECTED (Confidence: 70,6%)

Calculate Loan Approval



Loan Approval Calculator

Checking Status	<input type="text" value=">=200"/>
Duration (months)	<input type="text" value="36"/>
Credit History	<input type="text" value="all paid"/>
Purpose	<input type="text" value="furniture/equipment"/>
Credit Amount	<input type="text" value="10000"/>
Savings Status	<input type="text" value=">=1000"/>
Employment Duration	<input type="text" value=">=7"/>
Installment Commitment (%)	<input type="text" value="3"/>

Decision: APPROVED (Confidence: 100,0%)

Calculate Loan Approval



Loan Approval Calculator

Checking Status	<input type="text" value="≥200"/>
Duration (months)	<input type="text" value="36"/>
Credit History	<input type="text" value="all paid"/>
Purpose	<input type="text" value="vacation"/>
Credit Amount	<input type="text" value="15000"/>
Savings Status	<input type="text" value="≥1000"/>
Employment Duration	<input type="text" value="<1"/>
Installment Commitment (%)	<input type="text" value="4"/>

Decision: APPROVED (Confidence: 97,1%)

Calculate Loan Approval