

**Evaluación del riesgo crediticio en entidades bancarias en el área de microfinanzas  
utilizando árboles de decisión**

**Yeltsin Sleyther Gil Hoyos**

**Trabajo de grado presentado para optar el  
título de Ingeniera industrial**

**Universidad Santo Tomas**

**Facultad de Ingeniería Industrial**

**División de Ingenierías y Arquitectura**

**Bucaramanga**

**2017**

**Evaluación del riesgo crediticio en entidades bancarias en el área de microfinanzas  
utilizando árboles de decisión**

**Yeltsin Sleyther Gil Hoyos**

**Trabajo de grado presentado para optar el  
título de Ingeniera industrial**

**Director**

**Javier Hernández Cáceres**

**Universidad Santo Tomas**

**Facultad de Ingeniería Industrial**

**División de Ingenierías y Arquitectura**

**Bucaramanga**

**2017**

**Contenido**

1	Descripción del problema.....	6
1.1	Formulación del problema.....	10
2	Justificación.....	11
3	Alcance.....	12
4	Objetivos.....	13
4.1	Objetivo general.....	13
4.2	Objetivos específicos.....	13
5	Marco referencial.....	14
5.1	Marco teórico.....	14
5.1.1	Algoritmo j48.....	19
5.1.2	Croos validation.....	21
5.1.3	Selección de atributos para microempresarios.....	21
5.2	Marco conceptual.....	22
5.3	Marco legal y normativo.....	24
5.4	Marco histórico.....	26
5.5	Estado del arte.....	27
6	Marco metodológico.....	33
6.1	Fundamentos epistemológicos.....	33

6.2	Diseño de la investigación.....	34
7	Desarrollo del sistema .....	35
7.1	Desarrollo de la muestra.....	36
7.1.1	Formula de la muestra infinita.....	36
7.1.2	Definición de variables.....	36
7.1.3	Resultados .....	37
7.2	Preparación y resultados de los atributos numéricos .....	37
7.2.1	Análisis de los resultados por atributo.....	38
7.3	Cumplimiento al objetivo específico número 1. ....	40
7.3.1	Resultados de los atributos y/o variables que intervienen en la aprobación de una solicitud de crédito .....	40
7.3.2	Análisis de los resultados por atributo.....	41
7.3.3	Análisis de instancias o registros correctos o incorrectos .....	43
7.3.4	Valoración de los atributos de mayor relevancia .....	45
7.3.5	Árbol j48.....	46
7.3.6	Predicción de aprobación a la solicitud de crédito .....	47
7.4	Cumplimiento a los objetivos específicos número 2 y 3.....	49
7.4.1	Resultados de los atributos y/o variables .....	50
7.4.2	Análisis de los resultados por atributo de los 92 créditos activos .....	51
7.4.3	Análisis de instancias o registros correctos o incorrectos .....	54

- 7.4.4 Valoración de los atributos de mayor relevancia: ..... 54
- 7.4.5 Predicción del posible comportamiento futuro de pago del nuevo cliente:... 55
- 8 Conclusiones..... 57
- 9 Bibliografía..... 58
- 10 Apéndices ..... 61
  - 10.1 Carta de validación de la información..... 61
  - 10.2 Base microfinanzas..... 62

### Lista de figuras

<i>Figura 1.</i> Descripción del flujo del dinero en Entidades Bancarias, Coporaciones y Compañías de financiamiento, por Zapata, 2005.....	6
<i>Figura 2.</i> Componentes y estructura del árbol de decisión j48, S. Antonelli, «Árboles de clasificación, algoritmo J48,» 2010. ....	20
<i>Figura 3.</i> Resultados de las instancias correctas e incorrectas y su participación. ....	43
<i>Figura 4.</i> Resultados de las instancias correctas e incorrectas y su participación. ....	44
<i>Figura 5.</i> Dispersión de las instancias incorrectas.....	45
<i>Figura 6.</i> Resultados de los atributos con mayor relevancia. ....	45
<i>Figura 7.</i> Árbol j48 de la base de microempresarios para la toma de decisión de aprobación de crédito.....	46
<i>Figura 8.</i> Predicción y probabilidad de la nueva solicitud de crédito. ....	48
<i>Figura 9.</i> Resultados de las instancias correctas e incorrectas y su participación.....	54
<i>Figura 10.</i> Resultados de los atributos con mayor relevancia. ....	54
<i>Figura 11.</i> Predicción y probabilidad del comportamiento de pago del nuevo cliente. ...	56

**Lista de tablas**

Tabla 1..... 34

Tabla 2..... 38

Tabla 3..... 40

Tabla 4..... 47

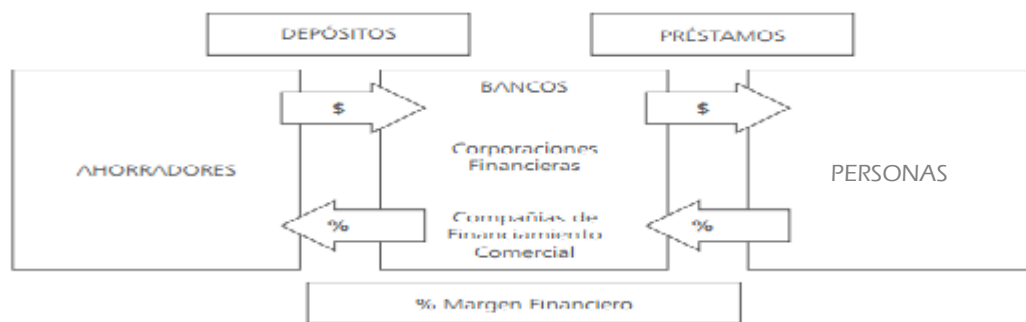
Tabla 5..... 49

Tabla 6..... 50

Tabla 7..... 55

## 1 Descripción del problema

Para comenzar, la palabra crédito bancario según “el dinero depositado por el cliente de la institución bancaria se supone que está, al mismo tiempo, en el propio banco a disposición del depositante y fuera del banco en forma de préstamo realizado por el depositario”. [1] El dinero está en constante flujo donde las entidades bancarias lo ofrecen a tasas de créditos mayores a las tasas de rentabilidad que les dan a los inversionistas. Siendo esta la principal actividad de ingreso de los bancos.



*Figura 1.* Descripción del flujo del dinero en Entidades Bancarias, Corporaciones y Compañías de financiamiento, por Zapata, 2005.

Las entidades financieras utilizan diferentes métodos de aprobación de créditos donde algunos sistemas dependen de los criterios que se le evalúan a cada persona, los criterios que evalúan en su mayoría las entidades financieras son: “1. Si es un cliente que cumple con sus compromisos de pago en la fecha. 2. Cuál es la actividad económica de la persona. 3. Se tomará en cuenta la capacidad de pago, que es la cantidad de deuda que puede manejar de acuerdo a los ingresos y gastos. 4. En algunos casos se tendrá en cuenta si se dispone de capital adicional, como

cuentas de ahorros, cdt's, o alguna propiedad que pueda funcionar de garantía para el préstamo. 5. Apoyarse e en centrales de información financiera y crediticia de cada país, donde le permite conocer información secuencial y cronológica del historial crediticio de todas las personas que ingresan a dichas bases". [3] Los criterios que se toman en cuenta en el momento de realizar una aprobación de crédito pueden variar con respecto a la entidad financiera, el tipo de crédito que se solicite, el tipo de persona que lo solicite y los diferentes sistemas de aprobación que utiliza cada entidad para valorar el riesgo del préstamo; y de esta manera aprobar una solicitud de crédito o no.

El riesgo crediticio al que se expone un banco al prestarles cierta cantidad de dinero a personas que lo soliciten resulta ser determinante, ya que es su principal fuente de ingresos. "El riesgo crediticio es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia del incumplimiento de las obligaciones de un deudor o contraparte". [4] Las técnicas para medir el riesgo de crédito son hoy en día muy variadas, y abarcan procedimientos que van desde simples cálculos, hasta sofisticadas metodologías con simulaciones dinámicas del futuro más próximo. Algunos modelos para la evaluación de riesgo de crédito son: sistemas expertos (Galicia, 2003), sistemas de calificación, modelo KMV, modelo de evaluación de Merton, modelo de credimetrics (J. P. Morgan, 1997), modelo de Credit Risk + (Morgan, 1997), modelo de retorno sobre capital ajustado al riesgos (Falkenstein, 1997), modelo CyRCE, Z – Score etc.

El desarrollo de cualquiera de estos modelos y de la interpretación subjetiva del gerente financiero o analista de la aprobación del crédito en dicha entidad, brinda un mar de caminos donde se pueden quedar criterios importantes de la persona sin evaluar o por lo contrario criterios que no tienen margen de importancia para el objetivo principal, algunos son más eficientes que otros, otros son más difíciles de desarrollar ya sea porque no se cuenta con la información necesaria o el

método que se utiliza no sea el adecuado y otros no brinda la suficiente seguridad para la toma de decisiones.

Todas estas series de incertidumbres en el proceso de aprobación de una solicitud de crédito en un banco acarrearán una serie de molestias, demoras y problemas tanto para el solicitante como para la entidad financiera. Por qué en primera parte lo que busca el solicitante es una respuesta rápida de si la entidad le va a prestar el dinero o si se lo busca por otra parte. El proceso de aprobación de un crédito de consumo en Colombia para una persona natural en bancos demora de 4 a 5 días hábiles aproximadamente, tiempo que el banco ocupa estudiando la solicitud con los diferentes modelos de aprobación que este utilice, donde en un porcentaje mayor estos créditos resultan siendo negados sea por problemas financieros de la persona o de los modelos de aprobación. En otro contexto por la necesidad de reducir los días de espera para la posible aprobación o en la necesidad de las entidades financieras por vender y poder cumplir sus metas comerciales las entidades financieras caso Colombia como Banco Av. Villas, Colpatria, Credifacil etc. Crean estrategias de márketing y mercados, éstas con posibilidad de prestar dinero en cuestión de horas donde los requisitos son la presencia física del solicitante y la fotocopia de las cedula. Estos tipos de estrategias de mercado trabajan en modelos de aprobación que sustentan información financiera histórica y actual de las personas, por lo tanto dan una respuesta inmediata donde muestra una calificación positiva o negativa. En Colombia las principales centrales de riesgo que existen son Data crédito y Cifin, y dependiendo de esta calificación se aprueba o no.

La creación de estas entidades que con el solo número de cedula y una previa autorización de cada persona se puede saber su comportamiento financiero, han ayudado mucho a la aprobación de créditos dando una rápida respuesta, pero no puede ser el único modelo que se tenga en cuenta al momento de una aprobación ya que es un sistema que trabaja con un porcentaje de error alto, el

proceso de actualización de la información es de cada dos meses y donde las personas son reportados por deudas mínimas. Además este modelo solo brinda información histórica donde se refleja el comportamiento de pago de cada persona, las fechas exactas que lo hizo y las deudas presentes; siendo este uno de los criterios principales de los bancos al momento de la aprobación; pero este modelo no brinda un comportamiento futuro en un 100% donde puede variar esta información sea de positiva a negativa o de negativa a positiva.

Debido a la utilización de estos sistemas las carteras vencidas en los bancos se han incrementado proporcionalmente en los últimos 5 años según información reportada por la Superintendencia financiera.

El tener información financiera historia verídica, no sirve si no se usa de una manera eficiente y si no se utilizan metodologías que evalúan criterios determinantes para dar una respuesta con un porcentaje de error mínimo y poder mostrar un comportamiento futuro real de pago de cada persona que solicite un crédito.

Por lo anterior esta investigación tendrá de base la minería de datos el término minería de datos hacer referencia a un “campo interdisciplinario que nace a partir de áreas como los sistemas de bases de datos, almacenamiento de datos, estadística, visualización de datos, recuperación de información y programación avanzada. Otras áreas que han contribuido al desarrollo de la minería de datos son las redes neuronales, el reconocimiento de patrones, el análisis de datos espaciales, las bases de datos de imágenes, el procesamiento de señales y muchos otros campos de aplicación como son los negocios, la economía y la bioinformática.

La minería de datos utiliza diferentes técnicas, técnicas que se aplican con referente a lo que se quiere lograr con nuestro conjunto de datos; técnicas descriptivas que proporcionan información entre los diferentes datos y sus correspondientes características, técnicas predictivas las cuales se aplican a un conjunto de datos con el objetivo de obtener conocimiento a futuro como las ventas que se van a producir en una empresa comercial en un año y están las que se van a manejar en esta investigación que son la técnicas de modelado las cuales se emplean en la comprensión de sistemas, hace referencia a la obtención de una representación de un sistema que permite imitar comportamientos. Para ello el modelado emplea agrupamiento, árboles de decisión y análisis de secuencia/asociaciones.

### **1.1 Formulación del problema.**

Con base en lo anterior, nos podríamos hacer la siguiente pregunta:

¿Cómo podríamos evaluar el riesgo crediticio al que se expone una entidad bancaria al momento del estudio y aprobación de un crédito bancario?

## 2 Justificación

Evaluando las metodologías utilizadas actualmente en Colombia para el análisis de crédito para las personas naturales por parte de entidades financieras, se evidencia la gran influencia de la subjetividad en el proceso y la falta de aplicación de herramientas que permitan hacer un análisis más objetivo en cuanto a la aprobación de un crédito. La preponderancia de la subjetividad y poca utilización de amplias herramientas financieras para el análisis hace necesario la aplicación de nuevos métodos e indicadores que guíen a los analistas de crédito sobre la situación financiera de cada uno de los solicitantes.

Por esta razón, vemos la necesidad de proponer la utilización de un indicador y/o metodología adicional a los usados generalmente en las entidades financieras colombianas, el cual podría servir como una herramienta bastante útil al momento de realizar análisis de créditos para este segmento. Sin bien es cierto, la parte cuantitativa es fundamental para los análisis, la parte cualitativa también juega un papel primordial en operaciones de crédito. El segmento al cual queremos enfocar la investigación son la microfinanzas donde están las personas naturales y/o jurídicas las cuales sus ventas anuales no superan los 100 millones de pesos; ya que este segmento son personas que no se tienen en su mayoría un historial creditico, por ende resulta difícil para los bancos tomar un decisión acertada con referente si presta o no presta el dinero al solicitante. La investigación será enfocada a la creación de una metodología utilizando árboles de decisión con el algoritmo j48 del software libre Weka de inteligencia artificial.

### 3 Alcance

La técnica de Árbol de Decisión será utilizada en un set de datos de una entidad financiera para la evaluación del riesgo crediticio de sus clientes, utilizando el algoritmo j48 del software Libre Weka de inteligencia Artificial. Por ejemplo, a partir de los comportamientos de los clientes antiguos con respecto a la morosidad o no de sus pagos del crédito concedido, podemos inferir qué nuevos clientes pueden ser los más convenientes para la concesión de un crédito, es decir, cuáles de ellos tienen más probabilidad de hacer frente al pago del mismo y cuáles más probabilidad de dejarlo sin pagar.

Por medio de un árbol de decisión se representaran las diferentes alternativas y estrategias para la empresa y se evaluarán cada una de las alternativas para determinar las decisiones más óptimas.

La investigación estará enfocada en el área de microfinanzas en las entidades bancarias, específicamente en los bancos. El segmento de microfinanzas están las personas naturales y/o jurídicas las cuales sus ventas anuales no superan los 100 millones de pesos; ya que este segmento son personas que no se tienen en su mayoría un historial crediticio, por ende es el segmento donde se presenta la mayor cantidad de cartera vencida en los bancos, y resulta necesario crear alternativas de evaluación del riesgo para disminuir la cartera vencida y para tomar la mejor decisión al momento de recibir una solicitud de crédito.

## **4 Objetivos**

### **4.1 Objetivo general**

Caracterizar los clientes de una entidad bancaria mediante la identificación de atributos tomados de bases de datos utilizando inteligencia artificial.

### **4.2 Objetivos específicos**

1. Identificar el riesgo crediticio que se expone una entidad bancaria al recibir una solicitud de crédito con la metodología de árboles de decisión para establecer si se le presta o no el dinero al solicitante.
2. Evaluar similitudes de clientes antiguos con posibles clientes nuevos como ingresos, actividad económica, calificación de pago, activos etc. Mediante el logaritmo J48, y de esta manera definir un posible comportamiento de pago futuro del cliente nuevo.
3. Determinar de manera asertiva el posible comportamiento de pago del nuevo cliente mediante la comparación de las características y comportamiento de pago del historial de antiguos clientes.

## 5 Marco referencial

### 5.1 Marco teórico

La superintendencia bancaria caso Colombia regula y reglamenta la gestión de valoración del riesgo por parte de las entidades financieras, y como tal, se orienta a que se fortalezca esta gestión del riesgo crediticio, generando una cultura y prácticas de alto nivel técnico en la administración de estos sistemas.

Por lo anterior con la circular externa 11 y la carta circular de 31 de marzo 5 de 2002, la superintendencia bancaria de Colombia, modifico de manera radical el capítulo II de la circular externa 100 de 1999. Donde esta nueva circular estable que cada entidad financiera está en la obligación de mantener permanentemente una adecuada administración del riesgo crediticio realizando hincapié el estudio del proceso de aprobación de créditos, de tal modo deberá crear y mantener un sistema de administración de los riesgos crediticios (SARC). [5] El Sistema de Administración de Riesgo Crediticio es el conjunto de políticas, procedimientos, normas y metodologías de medición de los riesgos que rigen y controlan los procesos de crédito y cobranza de las entidades vigiladas por la Superintendencia Financiera de Colombia. El SARC contiene políticas y procedimientos claros y precisos que definen los criterios y la forma mediante la cual la entidad evalúa, asume, califica, controla y cubre su riesgo crediticio. El SARC es reglamentado para todas las entidades financieras bajo vigilancia de la Superintendencia Financiera, con el fin de que todas las instituciones tengan una infraestructura tecnológica y los sistemas necesarios para garantizar la adecuada administración de crédito. [6]

En un proceso de análisis crediticio se valoran elementos tanto cualitativos como cuantitativos, relacionados especialmente con la capacidad de pago del sujeto de crédito, la coyuntura del negocio y del sector; al igual que las posibles garantías y el historial financiero del

sujeto. [7]Conocimiento del sujeto de crédito: este se refiere a la solvencia moral, reputación y la disposición de cumplimiento de compromisos con terceros, así como el conocimiento del historial crediticio. De esta forma, se podría afirmar que el tiempo que tiene la empresa en el mercado es un buen indicador de la reputación en materia de crédito. En Colombia, se han implantado los denominados Data crédito y la Central de Información Financiera (CIFIN) como sistemas de información o centrales de riesgo que registran, con información proveniente de todo el sistema financiero, a los acreditados que incumplieron con alguna obligación crediticia. Capacidad de pago: esta se calcula mediante un análisis financiero exhaustivo del solicitante, de tal manera, que refleje la volatilidad de las utilidades generadas históricamente. El flujo de efectivo debe garantizar la capacidad de pago de la firma; actualmente, se han desarrollado modelos y herramientas estadísticas que permiten calcular la probabilidad de incumplimiento, la provisión de cartera y la migración del crédito. La evaluación de la capacidad de pago esperada de un deudor o proyecto a financiar es fundamental para determinar la probabilidad de incumplimiento del respectivo crédito. Para estos efectos, debe entenderse que el mismo análisis debe realizarse a los codeudores, avalistas, deudores solidarios y, en general, a cualquier persona natural o jurídica, que resulte o pueda resultar directa o indirectamente obligada al pago de los créditos. Para evaluar esta capacidad de pago, la entidad prestamista debe analizar, al menos, la siguiente información:

- Los flujos de ingresos y egresos, así como el flujo de caja del deudor o del proyecto financiado o a financiar.
- La solvencia del deudor, a través de variables como el nivel de endeudamiento y la calidad y composición de los activos, pasivos, patrimonio y contingencias del deudor o del proyecto.
- Información sobre el cumplimiento actual y pasado de las obligaciones del deudor. La atención oportuna de todas las cuotas. Adicionalmente, la historia financiera y crediticia proveniente de

centrales de riesgo, calificadoras de riesgo, del deudor o de cualquier otra fuente que resulte relevante.

- El número de veces que el crédito ha sido reestructurado y las características de la(s) respectiva(s) reestructuración(es). Se entenderá que entre más operaciones reestructuradas tenga un mismo deudor, mayor será el riesgo de no pago de la obligación.
- Los riesgos de contagio, legales, operacionales y estratégicos, a los que puede estar expuestas la capacidad de pago del deudor o el proyecto a financiar. En este contexto, es necesario evaluar, entre otros, la información relacionada con el conglomerado económico al que pertenece el deudor. [8]

Donde uno de los principios y criterios generales para la evaluación del riesgo crediticio en la cartera de créditos son las metodologías de evaluación la cual soporta resultados claros para una toma de decisión futura. La evaluación del riesgo crediticio de cada modalidad de crédito o contrato se realizará de acuerdo con una metodología que fije el respectivo organismo de dirección de la entidad vigilada, atendiendo para ello los parámetros mínimos establecidos por la entidad rectora en Colombia por la superintendencia financiera colombiana. [9]

La evolución y el mejoramiento continuo de la tecnología han servido para el desarrollo de sistemas los cuales gozan de una serie de beneficios dependiendo en el área que se trabajen. La minería de datos según Beatriz Beltrán debe su nombre a la analogía entre una montaña y la gran cantidad de datos almacenados en cualquier empresa. Dentro de la montaña, ocultos entre piedras y tierra, se encuentran diamantes de gran valor que mediante actividades de minería son encontrados y aprovechados. La Minería de Datos descubre relaciones, tendencias, desviaciones, comportamientos atípicos, patrones y trayectorias ocultas, con el propósito de soportar los

procesos de toma de decisiones con mayor conocimiento. La Minería de Datos se puede ubicar en el nivel más alto de la evolución de los procesos tecnológicos de análisis de datos. [10]

La minería de datos es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos con el objetivo de encontrar patrones que nos puedan aportar información valiosa en la toma de futuras decisiones. El proceso de minería de datos tiene normalmente cuatro etapas principales:

- Determinación de los objetivos
- Procesamiento de los datos
- Determinación del modelo
- Análisis de los resultados

En el primero de los pasos se trata el tipo de información que el cliente desea extraer de la base de datos. La segunda etapa es la que requiere más trabajo ya que se tiene de seleccionar, limpiar, enriquecer, reducir y transformar la base de datos que nos ha facilitado el cliente. Una vez la hemos procesado y está lista para implementarle el algoritmo de inteligencia artificial, tenemos que escoger qué algoritmo nos va a dar mejores resultados.

La elección del mejor algoritmo para una tarea analítica específica es un gran desafío y depende del problema a resolver. Hay básicamente cinco problemas diferentes en la minería de datos: clasificación, regresión, segmentación, asociación y análisis de secuencias.

Para resolver estos problemas hay muchos algoritmos a utilizar, los principales son: los de

asociación, los clústeres, los de árbol de decisión, los de regresión lineal, el clasificador Bayesiano ingenuo, los de red neuronal, los clústeres de secuencia y los de serie temporal. [11]

En la actualidad esta herramienta es pieza clave en el sector financiero, debido a sus características y funcionamiento. Los árboles de decisión son una de las tres técnicas de procesamiento de la información que maneja la minería de datos, que según William T. Greenwood explica que la toma de decisiones básicamente es una resolución de problemas, y como tal, se debe buscar la mejor alternativa, aplicar un diagnóstico adecuado y realizar una búsqueda minuciosa, para elegir la mejor alternativa basados en un proceso de ramificación. Este proceso de ramificación es el árbol de decisiones que permite estimar alternativas ante la solución de problemas. [12]

Los árboles de decisión basan su metodología en separar estimados, donde su principal objetivo es emplear distintos panoramas y poder seleccionar uno bajo números y propuestas objetivas, evitando la combinación de emociones subjetivas. Las principales características de un árbol de decisión son:

- Plantea el problema desde distintas perspectivas de acción.
- Permite analizar de manera completa todas las posibles soluciones.
- Provee de un esquema para cuantificar el costo del resultado y su probabilidad de uso.
- Ayuda a realizar las mejores decisiones con base a la información existente y a las mejores suposiciones.
- Su estructura permite analizar las alternativas, los eventos, las probabilidades y los resultados.

El proceso de la creación de un árbol de decisión y probar su efectividad consta de una serie de pasos:

1. Identifica todas las variables del problema central
2. Priorizar
3. Identifica la importancia de los criterios
4. Comienzan a ver variables
5. Se seleccionan la una alternativa de acuerdo a sus beneficios
6. Implementa la alternativa
7. Se evalúa la efectividad de la decisión.

### **5.1.1 Algoritmo j48**

El algoritmo J48 de WEKA es una implementación del algoritmo C4.5, uno de los algoritmos de minería de datos más utilizado. Se trata de un refinamiento del modelo generado con OneR2. Supone una mejora moderada en las prestaciones, y podrá conseguir una probabilidad de acierto 2 Es un clasificador de los más sencillos y rápidos. Sus resultados pueden ser muy buenos en comparación con algoritmos mucho más complejos. Selecciona el atributo que mejor “explica” la clase de salida. Ligeramente superior al del anterior clasificador. El parámetro más importante que deberemos tener en cuenta es el factor de confianza para la poda “confidence level”, que influye en el tamaño y capacidad de predicción del árbol construido. Para cada operación de poda, define la probabilidad de error que se permite a la hipótesis de que el empeoramiento debido a esta operación es significativo. A probabilidad menor, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor por defecto es del 25%. Según baje este valor, se permiten más operaciones de poda. El algoritmo J48 se basa en la utilización del criterio ratio de ganancia (gain ratio). De esta manera se consigue evitar que las variables con mayor número de posibles valores salgan

beneficiadas en la selección. Además el algoritmo incorpora una poda del árbol de clasificación una vez que éste ha sido inducido. [13]

Para tareas de clasificación, el algoritmo J48 resulta muy simple y potente, este algoritmo representa a su vez una evolución del algoritmo ID3. El procedimiento para generar el árbol consiste en seleccionar un atributo como raíz, y crear una rama con cada uno de los valores posibles de dicho atributo; con cada rama resultante se realiza el mismo proceso. En cada nodo se debe seleccionar un atributo para seguir dividiendo, y para ello se selecciona aquel que mejor separe los ejemplos de acuerdo a la clase. [14] El algoritmo selecciona los atributos nominales como nodo raíz los cuales son los que ofrecen la caracterización en un conglomerado de datos, donde el nodo más importante será la cabeza del árbol por ende estará el atributo que se desea evaluar o de mayor relevancia.

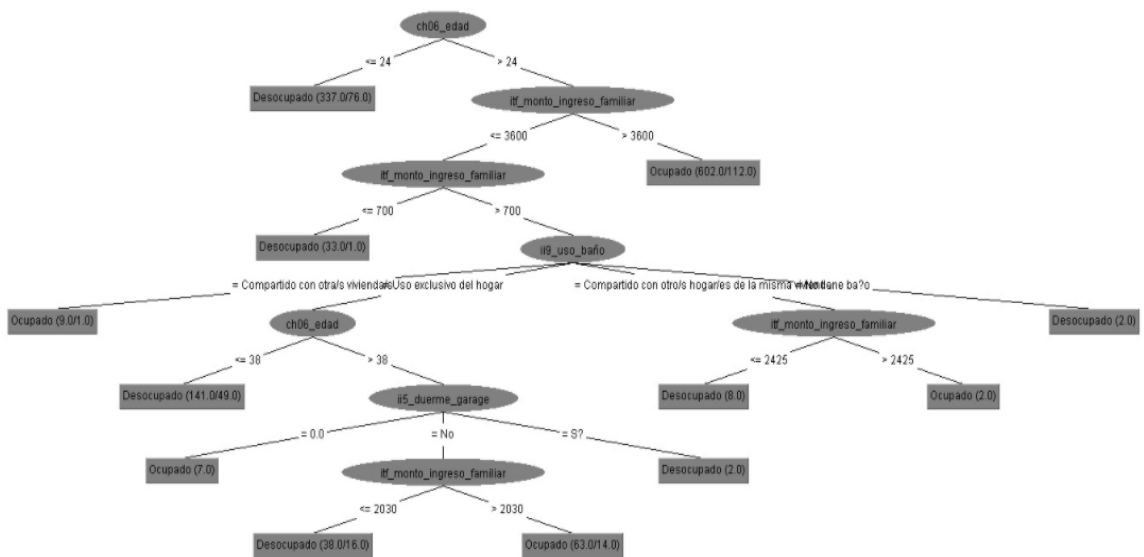


Figura 2. Componentes y estructura del árbol de decisión j48, S. Antonelli, «Árboles de clasificación, algoritmo J48,» 2010.

[15]

### 5.1.2 Croos validation

Evaluación con validación cruzada. Se dividirán las instancias en tantas carpetas como indica el parámetro “Folds”, y en cada evaluación se toman las instancias de cada carpeta como datos de test, y el resto como datos de entrenamiento para construir el modelo. Los errores calculados serán el promedio de todas las ejecuciones. [16] Técnica que trabaja el algoritmo j 48 en la ejecución de un árbol de decisión o clasificación.

En la validación cruzada de K iteraciones o K-fold cross-validation los datos se dividen en K subconjuntos (folds). Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante K iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. El error se calcula como la media aritmética de los errores de cada iteración para obtener un único resultado.

### 5.1.3 Selección de atributos para microempresarios

- ✓ Edad: número de años de cada persona a evaluar
- ✓ Género: si es masculino o femenino
- ✓ Estado civil: si la persona se encuentra sin ningún tipo de relación, en unión libre, comprometida o separada
- ✓ Hijos a cargo: cantidad de hijos que la persona tiene a cargo
- ✓ Monto de pasivos: deudas registradas en las centrales de riesgo
- ✓ ¿En qué tipo de casa vive?: si vive en casa arrendada, propia o familiar
- ✓ Sector y/o tipo de empresa: actividad comercial
- ✓ Promedio de ventas: promedio de ventas que efectúan en el periodo de 12 meses

- ✓ Costo totales de operación: El costo mensual el cual necesita para operar en negocio o pequeña empresa
- ✓ Aprobado: si el crédito le fue aprobado o rechazado

Después de haber aprobado el crédito resulta necesario investigar el comportamiento de pago que ha tenido el usuario.

- ✓ ¿Cuál es su comportamiento actual?: comportamiento de pago de crédito actual

Los datos almacenados son clientes microempresarios los cuales todos poseen los mismos atributos anteriormente mencionados, base que fue proporcionado por el departamento de microfinanzas de una entidad bancaria.

## 5.2 Marco conceptual

**MINERIA DE DATOS:** Es el proceso de descubrir conocimiento interesante de grandes cantidades de datos almacenadas en bases de datos, data warehouses u otro repositorio de información. [17]

**ÁRBOL DE DECISIÓN:** Se considera que un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (sucesos) que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión “más acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Permite desplegar visualmente un problema y organizar el trabajo de cálculos que deben realizarse. [18]

**ÁBOLES DE CLASIFICACIÓN:** Un clasificador es una partición del espacio de clasificación  $X$  en  $M$  subconjuntos disjuntos  $A_1, A_2, \dots, A_M$ , siendo  $X$  la unión de todos ellos y para todo  $x$  perteneciente a  $A_m$  la clase predicha es  $C_m$ . [19]

**BASE DE DATOS:** Es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. [20]

**RIESGO CREDITICIO:** El riesgo crediticio es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia del incumplimiento de las obligaciones de un deudor o contraparte. [21]

**PRESTAMO BANCARIO:** es el crédito que concede un banco. Por lo general, esta operación comienza cuando una persona acude a la entidad bancaria para solicitar dinero prestado. Al recibir el pedido, el banco analizará la capacidad de pago y aprobará la entrega de un cierto monto bajo determinadas condiciones. La ganancia del banco estará en que, al devolver el dinero, la persona tendrá que entregar un adicional en concepto de intereses. [22]

**ENTIDAD BANCARIA:** es un intermediario del mercado financiero. Las entidades financieras pueden ser bancos, cajas de ahorros o cooperativas de crédito, es decir, intermediarios que administran y prestan dinero; o empresas financieras, un tipo distinto de intermediarios financieros que, sin ser bancos, ofrecen préstamos o facilidades de financiamiento en dinero. [23]

**RIESGO OPERATIVO:** es la posibilidad de pérdida como resultado de deficiencias a causa de fallas en los sistemas de información, fallas en procesos, control de atribuciones y cupos en tiempo real, errores humanos, mala fe de los funcionarios y fallas en el control gerencial.

**RIESGO DE MERCADO:** el riesgo de mercado o precio es la contingencia de pérdida o ganancia de una posición de la entidad financiera, como resultado de un cambio en el nivel o la volatilidad de las tasas de interés, tasas de cambio o precios.

MICROFINANZAS: las microfinanzas son préstamos que se dirigen a personas o a grupos con pocos medios económicos y que normalmente están excluidos del sistema financiero tradicional, y también define los negocios que crecieron alrededor de estas actividades.

### 5.3 Marco legal y normativo

- **Ley estatutaria 1581 del 2102:** protección de datos, esta ley tiene como objeto desarrollar el derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos o archivos, y los demás derechos, libertades y garantías constitucionales a que se refiere el artículo 15 de la Constitución Política; así como el derecho a la información consagrado en el artículo 20 de la misma. Los principios y disposiciones contenidas en la presente ley serán aplicables a los datos personales registrados en cualquier base de datos que los haga susceptibles de tratamiento por entidades de naturaleza pública o privada.

La presente ley aplicará al tratamiento de datos personales efectuado en territorio colombiano o cuando al Responsable del Tratamiento o Encargado del Tratamiento no establecido en territorio nacional le sea aplicable la legislación colombiana en virtud de normas y tratados internacionales.

El régimen de protección de datos personales que se establece en la presente ley no será de aplicación:

- a) A las bases de datos o archivos mantenidos en un ámbito exclusivamente personal o doméstico.

Cuando estas bases de datos o archivos vayan a ser suministrados a terceros se deberá, de manera previa, informar al Titular y solicitar su autorización. En este caso los Responsables y Encargados de las bases de datos y archivos quedarán sujetos a las disposiciones contenidas en la presente ley;

b) A las bases de datos y archivos que tengan por finalidad la seguridad y defensa nacional, así como la prevención, detección, monitoreo y control del lavado de activos y el financiamiento del terrorismo;

c) A las Bases de datos que tengan como fin y contengan información de inteligencia y contrainteligencia;

d) A las bases de datos y archivos de información periodística y otros contenidos editoriales;

e) A las bases de datos y archivos regulados por la Ley 1266 de 2008;

f) A las bases de datos y archivos regulados por la Ley 79 de 1993.

Los principios sobre protección de datos serán aplicables a todas las bases de datos, incluidas las exceptuadas en el presente artículo, con los límites dispuestos en la presente ley y sin reñir con los datos que tienen características de estar amparados por la reserva legal. En el evento que la normatividad especial que regule las bases de datos exceptuadas prevea principios que tengan en consideración la naturaleza especial de datos, los mismos aplicarán de manera concurrente a los previstos en la presente ley. [24]

- **Licencia weka:** WEKA está licenciado bajo la licencia pública general GNU (GPL 2.0 para Weka 3.6 y GPL 3.0 para Weka > 3.7.5). Cualquier trabajo derivado obtenido bajo esta licencia debe ser licenciado bajo la GPL si este trabajo derivado se distribuye a un tercero.

Para proyectos comerciales que requieren la capacidad de distribución del código WEKA como parte de un programa que no puede ser distribuido bajo la licencia GPL, puede ser posible comprar una licencia apropiada de los titulares de derechos de autor que figuran en las correspondientes clases de Java. [25]

#### 5.4 Marco histórico

La idea de Minería de Datos no es nueva. Ya desde los años sesenta los estadísticos manejaban términos como Data Fishing, Data Mining (DM) o Data Archaeology con la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con ruido.

A principios de los años ochenta, Rakesh Agrawal, Gio Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro entre otros, empezaron a consolidar los términos de Minería de Datos y KDD.

Esta tecnología ha sido un buen punto de encuentro entre personas pertenecientes al ámbito académico y al de los negocios.

La evolución de sus herramientas en el transcurso del tiempo puede dividirse en cuatro etapas principales:

- Colección de Datos (1960).

- Acceso de Datos (1980).
- Almacén de Datos y Apoyo a las Decisiones (principios de la década de 1990).
- Minería de Datos Inteligente. (finales de la década de 1990). [26]

## 5.5 Estado del arte

**z – score (z2) como herramienta de análisis en la otorgación de créditos al segmento pyme en entidades financieras en Colombia.** z – core (z2) una de las metodologías actualmente utilizadas en el sistema financiero Colombiano para la otorgación de créditos en este caso para la pequeñas empresas o empresas que apenas están surgiendo llamas pyme, esta metodología creada por Altman se basa en esta fórmula:  $Z = V1X1 + V2X2 + V3X3 + V4X4 + V5X5$ .  $X1 = \text{Capital de trabajo} / \text{activos totales}$ ,  $X2 = \text{Utilidades retenidas} / \text{activos totales}$ ,  $X3 = \text{Utilidades retenidas antes de intereses e impuestos} / \text{activos totales}$ ,  $X4 = \text{Valor en libros del patrimonio} / \text{Valor pasivos totales}$ ,  $X5 = \text{Ventas} / \text{activos totales}$   $V_i = \text{Coeficientes o pesos asignados a cada una de las variables}$   $Z = \text{Índice o puntaje total}$ .

Donde se busca eliminar en gran proporción la subjetividad en el proceso y la falta de aplicación de herramientas que permitan hacer un análisis más objetivo en cuando a la parte financiera de las empresas Pyme o personas con establecimientos de comercio. La preponderancia de la subjetividad y poca utilización de amplias herramientas financieras para el análisis hace necesario la aplicación de nuevos métodos e indicadores que guíen a los analistas de crédito sobre la situación financiera de una empresa. [27]

**Clasificación de riesgo en carteras de crédito aplicando metodología de redes neuronales.** A medida del paso del tiempo y gracias a la evolución significativa de la tecnología se buscan alternativas más idóneas para valorar el riesgo en las entidades financieras, debido a que es un factor fundamental para no estar afectados por los cambios macroeconómicos que están en continuo cambio. Esta investigación se basa en la utilización de las redes neuronales como herramienta para aplicarla en base de datos comerciales, donde el principal objetivo del análisis es encontrar algunas relaciones para grupos determinados de la población, de acuerdo con sus características particulares, con el fin cuantificar el riesgo para una cartera comercial, calculando las probabilidades de impago de las personas morosas y establecer el capital mínimo requerido y las provisiones de cartera.

Las redes neuronales emulan la estructura y el comportamiento del cerebro, utilizando los procesos de aprendizaje para buscar una solución a diferentes problemas; son un conjunto de algoritmos matemáticos que encuentran las relaciones no lineales entre conjuntos de datos; suelen ser utilizadas como herramientas para la predicción de tendencias y como clasificadoras de conjuntos de datos. Se denominan neuronales porque están basadas en el funcionamiento de una neurona biológica cuando procesa información. Un ejemplo que presenta la investigación es el uso de una tarjeta de crédito, donde suelen acumularse datos sobre patrones de consumo de la persona. Sobre la base de los pagos efectuados en dicha tarjeta de crédito, de esta manera las entidades e instituciones de crédito irán elaborando un historial del usuario, el cual se utilizará para autorizar una transacción, para decidir cuándo extender el crédito y para detectar fraudes. Este tipo de procesos requiere de chequeos que suelen resultar bastante complejos, además del uso de criterios variables para poder tomar una decisión final en torno a la autorización de ciertas transacciones.

[28]

**Valoración y riesgo crediticio en Colombia,** El ejercicio de valorar el riesgo crediticio en Colombia es una tarea ardua, toda vez que los elementos del contexto que deben ser considerados en el mencionado proceso son diversos. Esta situación conduce necesariamente a la combinación de métodos, tanto cuantitativos como cualitativos, en aras de mejorar los criterios para la toma de decisiones. En este sentido, esta análisis inicia con la revisión de las fuentes de financiación empresarial más comunes en el sistema financiero colombiano, a partir de las cuales, identifica los principales elementos a considerar en el momento de adelantar un proceso crediticio. Enseguida, se examina la metodología de análisis financiero tradicional, se procede a destacar algunas experiencias a nivel de Latinoamérica con nuevos modelos probabilísticos de análisis financiero y, finalmente, se señalan las bondades de dichos modelos, a partir de sus ventajas y desventajas.

Todas las metodologías de aprobación crediticia existentes en las entidades bancarias surgen de una metodología tradicional, el cual consiste en herramientas propias de reclasificación de estados financieros, evaluación de indicadores y conocimiento del cliente mediante visitas, presentación de documentos; identificación de las debilidades y fortalezas en manera cuantitativa de una empresa o de una persona natural. [29]

**Líneas de investigación en minería de datos en aplicaciones en ciencia e ingeniería:**  
**Estado del arte y perspectivas,** según (José A. Garcia Gutiérrez) el aumento de la información digitalizada y la disponibilidad de los datos de todo tipo de naturaleza, unido al desarrollo de redes de comunicaciones de alta velocidad, la intercomunicación de los centros de datos y la aparición de la computación de alto desempeño, hace imprescindible el desarrollo de técnicas de minería de datos que permitan procesar y analizar grandes volúmenes de datos y extraer de ellos información

de valor. Este artículo habla sobre la evolución que ha tenido este campo en las últimas décadas, así como de su potencialidad y aplicabilidad presente en las diferentes ramas de la investigación científica. Así mismo, trataremos de repasar de forma breve las diferentes familias de algoritmos que se engloban en el campo de la minería de datos, su escalabilidad cuando aumenta la dimensionalidad de los datos de entrada y de cómo se puede abordar y cuál es el comportamiento de los diferentes métodos en un escenario en el que la información se encuentra distribuida o se procesa de manera descentralizada o paralela de forma que se pueda optimizar el rendimiento en entornos heterogéneos.

La minería de datos se podría definir de forma general como el proceso de descubrir, extraer almacenar información relevante de amplios conjuntos de datos a través de técnicas que permiten lidiar con la alta dimensionalidad de los mismos. Sus objetivos son diversos como lo son sus áreas de aplicabilidad, estando ya inserto en muchos de los casos de forma transparente en todo tipo de software que contemple tareas como el descubrimiento de patrones de comportamiento, encontrar interrelaciones sistemáticas entre variables, construir modelos predictivos, y la extracción de información no evidente a partir de un dataset de datos recogidos en bruto. [30]

**Evaluación de inversiones bajo incertidumbre: teoría y aplicaciones a proyectos en Chile,** Este documento, en sus aspectos teóricos se inicia con la presentación y análisis de los elementos y tópicos relevantes en materia de cuantificación y valoración del riesgo y/o incertidumbre, a partir de los distintos desarrollos disponibles a la fecha en materia de medición de riesgos en proyectos. Se incluye el desarrollo de los siguientes temas: Análisis probabilístico, Análisis de Sensibilidad y de Escenarios, Ajuste simple en la tasa de descuento, Equivalencia a la certidumbre, Simulación, Capital Assets Pricing Model (CAPM), opciones reales y árboles de

decisión. Para cada uno de ellos se presentan aplicaciones a proyectos. También se presenta el Valor en Riesgo (VeR). El marco teórico y conceptual de este último, por ser la herramienta de más reciente aplicación, se profundiza y complementa en mayor medida que los restantes temas.

Unas de las aplicaciones que se mejan en esta investigación son los arboles de decisión. Las decisiones más importantes de un proyecto no se toman de una vez en un momento del tiempo. En ocasiones, se toman por etapas, sobre todo cuando el resultado de la decisión es aleatorio, eso ya se analizó con el método de opciones reales. En algunos casos es útil el uso de árboles de decisión para encontrar el conjunto de decisiones óptimas. [31]

**Aplicación de árboles de decisión en modelos de riesgo crediticio,** En este artículo se presentan algunos puntos generales del marco teórico de los riesgos a los que se enfrenta una institución financiera, su clasificación y definición, centrándose específicamente en el riesgo crediticio, para el que se presenta el marco legal: los enunciados básicos del Acuerdo de Basilea II y la reglamentación del sistema de administración de riesgo crediticio de la Superintendencia Bancaria en Colombia. Dentro de este marco, se ilustrará cómo la estadística juega un papel importante en el cumplimiento de esta normatividad. Específicamente se presenta la utilización de los árboles de decisión como herramienta para el cálculo de probabilidades de incumplimiento en crédito, mostrando sus ventajas y desventajas.

Los árboles de decisión se presentan como una herramienta efectiva para la predicción de probabilidades de incumplimiento, no solo a nivel de capacidad de discriminación (potencia), estabilidad a través del tiempo, sino como una herramienta de fácil entendimiento que permite

potencializar sus usos y servir además de la predicción, para la planeación de estrategias comerciales de venta de servicios, estrategias de cobranza entre muchas otras. [32]

**Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de weka (waikato environment for knowledge analysis),**

Teniendo en cuenta el gran avance en los sistemas de minería de datos desde el último siglo, las entidades educativas y empresariales han buscado maneras de explotar al máximo la información existente en sus sistemas de información, esto basándose en técnicas y software especializados que permiten interpretación fácil y real de los resultados. Es así como para dar apoyo en la toma de decisiones a niveles administrativos o gerenciales, se crean metodologías especializadas y técnicas de extracción adecuada de la información, haciendo que el usuario final pueda ver los resultados en un solo clic o con pocos pasos, por tanto y teniendo en cuenta lo anterior, se crea el presente manual de usuario basado en la tecnología de información y software especializado WEKA (Waikato Environment for Knowledge Analysis) de la universidad de Waikato en Nueva Zelanda, este utiliza técnicas de minería de datos basándose en diferentes reglas y tipos de clasificación de información tales como árboles de decisión, reglas de clasificación, agrupamiento, etc. Al ser un software especializado brindara apoyo suficiente para interpretar resultados de manera matemática y estadística y por medio de visualización de gráficos o árboles que agregarán valor a los resultados obtenidos. Finalmente al usuario final se deja el trabajo profundo de investigación de teoremas o teorías si así lo desea para complementar sus interpretaciones, pero se deja por parte del autor conceptos que facilitaran la comprensión de funcionamiento de la herramienta para la generación de resultados adecuados y continuar en la mejora constante de la misma. [33]

## **6 Marco metodológico**

Esta investigación se orienta en el proceso de análisis en el otorgamiento de un crédito bancario, valorando el riesgo que una entidad se pueda exponer frente a una situación de solicitud de crédito en una entidad bancaria, para este proceso se utilizara CRISP – DM, la cual es una metodología que maneja la minería de datos en los procesos de transformación de la información y posteriormente una aprovechamiento de la misma.

### **6.1 Fundamentos epistemológicos**

El enfoque principal de la investigación es de tipo cuantitativo, ya que busca mediante los arboles de decisión determinar si aquella persona que presenta la solicitud es apta o no para tomar el crédito, también esta investigación es de tipo correlacional ya que busca medir el grado de relación existente de nuevos clientes con el historial crediticio de antiguos clientes.

## 6.2 Diseño de la investigación

Tabla 1.

*Fases y actividades de la investigación.*

<b>FASES</b>	<b>ACTIVIDADES</b>
Comprensión de la investigación	Definir el objetivo de la investigación Definir las variables a investigar Definir las variables principales
Compresión de los datos	Obtención de los datos Descripción de los datos Exploración de los datos
Preparación de los datos	Verificación de la calidad de los datos Selección de los datos Limpiar datos Construcción de los datos Integración de los datos
Modelamiento	Selección de la técnica de modelado Generar el diseño de prueba Construir y evaluación del modelo
Evaluación	Evaluar los resultados Revisar el proceso
Despliegue	Desplegar el modelo Monitoreo

[AUTORÍA PROPIA]

## 7 Desarrollo del sistema

El programa propuesto fue desarrollado en un lenguaje de programación JAVA, para poder utilizar el algoritmo j48 se trabajó con la licencia libre Weka GUI chooser 3.6.12. La investigación consiste en una base histórica y actual de clientes de una entidad financiera del sector de microfinanzas, la cual goza de una serie de atributos de cada uno de los clientes; donde se analizarán y clasificarán por el modelo de árboles de decisión que encuentra similitudes en cada uno de los registros históricos de créditos aprobados o rechazados por la entidad. Se realizará una nueva base con los datos de un nuevo cliente con los mismos atributos antes mencionados y estos resultados se compararán con los de la base histórica, de esta forma evidenciar si se le aprueba o no el crédito. De esta misma forma se evidenciará el posible comportamiento de pago futuro de los nuevos clientes aceptados con la base de datos de los usuarios que ya están activos y tienen cierta calificación referente a su comportamiento de pago. Vale aclarar que el sistema no arroja un resultado definitivo para la toma de la decisión de que si se le aprueba o rechaza la solicitud de crédito al cliente, sino que en base a patrones encontrados y en estadísticas históricas se ofrecen a los encargados de tomar la decisión información con un valor añadido para reducir la incertidumbre de las decisiones y poder mejorar la calidad del servicio, con respecto de mejorar el tiempo en la toma de decisión y disminuir la cartera vencida de la entidad bancaria.

## 7.1 Desarrollo de la muestra

### 7.1.1 Formula de la muestra infinita

$$n = \frac{k^2 * p * q * N}{(e^2 * (N - 1)) + k^2 * p * q}$$

### 7.1.2 Definición de variables

**n** = Es el tamaño de la muestra (número de registros o personas a investigar).

**N** = Es el tamaño de la población o universo (número total de posibles encuestados). En la actualidad la ciudad de Bucaramanga cuenta con una población de 521.857 personas, la cual se tomará de base como el número total de posibles personas a investigar ya que el 95% de los créditos que presenta la entidad son de la ciudad de Bucaramanga.

**K** = Es una constante que depende del nivel de confianza que asignemos. El nivel de confianza indica la probabilidad de que los resultados de nuestra investigación sean ciertos. Se utilizara un K igual a 1,6 el cual es un nivel de confianza del 90%.

**e** = Es el error muestral deseado. El error muestral es la diferencia que puede haber entre el resultado que obtenemos preguntando a una muestra de la población y el que obtendríamos si preguntáramos al total de ella. Para determinar la muestra de la investigación se manejará un error porcentual del 6%.

**p** = Es la proporción de individuos que poseen en la población la característica de estudio. En Colombia el porcentaje de bancarización está en 77,3%. [34] Donde Bucaramanga por ser unas de las 5 ciudades más importantes en Colombia llega a un porcentaje de bancarización del 80% aproximadamente, y donde el 80% tomado como un 100% se encuentra el sector microempresario

el cual se encuentra en un rango del 22% al 30%. Se tomara un 30% como la proporción característica a trabajar.

$q$  = Es la proporción de individuos que no poseen esa característica, es decir, es  $1-p$ .

### 7.1.3 Resultados

$$n = \frac{(1,65^2 * 0,3 * 0,7 * 521857)}{(0,06^2 * (521857 - 1)) + 1,65^2 * 0,3 * 0,7}$$

$n$  = el desarrollo de la formula arroja un resultado **159** personas a tomar en cuenta para el desarrollo de la investigación

## 7.2 Preparación y resultados de los atributos numéricos

En la base histórica con la información financiera de las 159 personas se presentan atributos de tipo numérico sin ningún orden nominal, donde resulta necesario clasificarlos en rangos o nuevas etiquetas para poder ejecutar la base y de esta forma encontrar similitudes entre todos los atributos. Los resultados y la nueva etiqueta de los atributos numéricos como: edad, hijos a cargo, pasivos, promedio de ventas y costos totales de operación entre los 10 atributos que se manejan en la investigación se muestran en la siguiente tabla:

Tabla 2.

*Resultados de cada uno de los atributos y las nuevas etiquetas.*

<b>ATRIBUTO</b>	<b>RESULTADOS</b>		<b>NUEVA ETIQUETA</b>
Edad	Min	18 años	18-34
	Max	65 años	35-51
	Media	38,25	>52
Hijos a cargo	Min	0	
	Max	4	0-2
	Media	1,8	>=3
Pasivos	Min	\$ 1.500.000	0-5000000
	Max	\$ 20.640.262	5000001-15000000
	Media	\$ 6.027.364	>15000000
Promedio de ventas	Min	\$ 2.196.206	1500000-7000000
	Max	\$ 24.901.666	7000001-14000000
	Media	\$ 8.335.287	14000001-25000000
Costo total de operación	Min	\$ 1.670.051	0-5000000
	Max	\$ 21.168.378	5000001-15000000
	Media	\$ 6.589.806	>15000000

[AUTORÍA PROPIA]

### 7.2.1 Análisis de los resultados por atributo

- **EDAD:** La edad de los microempresarios mínima registrada en la base es de 18 años y máxima 65 años con una media de 38.252 años. La nueva etiqueta para este atributo será en rangos de 18-34, 35-51 y >52 años.
- **HIJOS A CARGO:** En la base se registra dentro de los 159 registros un mínimo de 0 hijos a cargo; personas que no tienen hijos o quizá sí, pero que no dependen económicamente de

ellos y un máximo de 4 hijos a cargo con una media de 1,8. La nueva etiqueta para este atributo será en rangos de 0-2 y mayores o iguales a 3 hijos.

- **MONTO DE PASIVOS:** Los montos mínimos de pasivos o deudas registrados por los microempresarios que tienen con el sistema bancario son de \$1.500.000 y un máximo de \$20.640.262 con una media de \$6.027.364. Estas deudas que se registran por los microempresarios es debido a la necesidad que tienen cada uno de ellos tener una base económica el cual le permitió comenzar con el negocio o también para invertir, remodelar, o para capital de trabajo etc. La nueva etiqueta para este atributo será en rangos de 0-\$5.000.000, \$5.000.001-\$15.000.000 y mayores a \$15.000.000.
- **PROMEDIO DE VENTAS MENSUALES:** Las ventas mensuales registradas por los microempresarios mínimos son de \$2.196.206 y un máximo de \$24.901.666 con una media de ventas de \$8.335.287. La nueva etiqueta para este atributo será en rangos de \$1.500.000-\$7.000.000, \$7.000.001-\$14.000.000 y dentro de \$14.000.001-\$25.000.000; estos rangos debido a que la entidad bancaria clasifica a los micro empresarios a nivel general aquellos que presenten ventas mensuales mínimas de 2 salarios mínimos legales vigentes y máximos a \$300.000.000 anuales.
- **COSTO TOTAL DE OPERACIÓN MENSUAL:** Los 159 microempresarios tomados como muestran registran costos totales de producción, es decir los costos y/o gastos que requieren poner en funcionamiento el negocio, mínimos de \$1.670.051 y máximos \$21.168.378 con una media de \$6.589.806. La nueva etiqueta para este atributo será en rangos de 0-\$5.000.000, \$5.000.001-\$15.000.000 y mayores a \$15.000.000.

### 7.3 Cumplimiento al objetivo específico número 1.

“Identificar el riesgo crediticio que se expone una entidad bancaria al recibir una solicitud de crédito con la metodología de árboles de decisión para establecer si se le presta o no el dinero al solicitante”.

#### 7.3.1 Resultados de los atributos y/o variables que intervienen en la aprobación de una solicitud de crédito

Mediante la nueva clasificación de los atributos numéricos y la información obtenida de los atributos faltantes de cada uno de los 159 registros que se tiene en la base histórica, al ser ingresados al sistema operativo WEKA podemos diferir los siguientes resultados:

Tabla 3.

*Resultados de atributos, de sus rangos y su participación.*

ATRIBUTO	ETIQUETA	RESULTADOS	PORCENTAJE
Edad	18-34	61	38%
	35-51	74	47%
	>52	24	15%
Género	M	66	42%
	F	93	58%
Estado civil	Unión libre	69	43%
	Soltero	38	24%
	Casado	36	23%
Hijos a cargo	Separado	16	10%
	0-2	113	71%
¿Vive en casa?	>=3	46	29%
	Arrendada	56	35%
	Propia	54	34%
Tipo de empresa	Familiar	49	31%
	Comercio	95	60%
	Producción	16	10%

	Servicio	29	18%
	Rentista Capital	7	4%
	Agro	12	8%
	0-5000000	83	52%
Pasivos	5000001-15000000	51	32%
	>15000000	25	16%
	1500000-7000000	94	59%
Promedio de ventas	7000001-14000000	38	24%
	14000001-25000000	27	17%
	0-5000000	83	52%
Costos totales operación	5000001-15000000	50	31%
	>15000000	26	16%
	Si	92	58%
Aprobado	No	67	42%

[AUTORÍA PROPIA]

### 7.3.2 Análisis de los resultados por atributo

- **EDAD:** De los 159 registros ingresados al sistema en el rango de 18-34 años se obtuvieron 61 personas, 35-51 años 74 personas y mayores a 52 años 24; con un porcentaje de 38%, 47% y 15% respectivamente. Podemos concluir que en el sector microempresario la menor participación la tienen las personas de la tercera edad con un 15%.
- **GÉNERO:** De los 159 registros ingresados al sistema 66 son hombres y 93 mujeres; con un porcentaje de 42% y 58% respectivamente. Podemos concluir que en el sector microempresario las mujeres tienen mayor presencia que los hombres en un 58%.
- **ESTADO CIVIL:** De los 159 registros ingresados al sistema en el rango de unión libre hay 60 microempresarios, solteros 38, casados 36, separados 16; con un porcentaje de 43%, 24%, 23% y 11% respectivamente. Podemos concluir que en el sector microempresario la unión libre sin un compromiso formal es predominante y el 33% no cuenta con una relación estable actualmente.

- **HIJOS A CARGO:** De los 159 registros ingresados al sistema en el rango de 0-2 hijos a cargo se obtuvieron 113 y mayores o iguales a 3 hijos 46 personas; con un porcentaje de 71% y 29% respectivamente. Podemos concluir que en el sector microempresario en una porcentaje de 71% de participación no tienen responsabilidad económica con respecto a la manutención de hijos o presentan una responsabilidad menor de 1 a 2 hijos a cargo.
- **TIPO DE CASA:** De los 159 registros ingresados al sistema en el rango de que los microempresarios viven en propiedades arrendadas hay 56 personas, propia 54 y familiar 49; con un porcentaje de 35%, 34% y 31% respectivamente. Podemos concluir que en el sector microempresario el 65% no cuenta con una vivienda propia.
- **TIPO DE EMPRESA:** De los 159 registros ingresados al sistema en el rango de los microempresarios que tienen un tipo de negocio en comercio hay 95 personas, producción 16, servicio 29, rentista capital 7 y Agro 12; con un porcentaje de 60%, 10%, 18%, 4% y 8% respectivamente. Podemos concluir que en una gran participación los microempresarios están dedicados al comercio.
- **MONTO DE PASIVOS:** De los 159 registros ingresados al sistema en el rango de 0-\$5.000.000 hay 83 personas, \$5.000.001-\$15.000.000 51 y mayores a \$15.000.000 25; con un porcentaje de 52%, 32% y 16% respectivamente. Podemos concluir que en el sector el microempresario común con una participación del 52% accede a créditos menores o iguales a \$5.000.000.
- **PROMEDIO DE VENTAS MENSUALES:** De los 159 registros ingresados al sistema en el rango de \$1.500.000-\$7.000.000 hay 94 personas, \$7.000.001-\$14.000.000 38 personas y de \$14.000.001-\$25.000.000; con un porcentaje de 59%, 24% y 17% respectivamente.

Podemos concluir que en el sector microempresario en un 60% aproximadamente tiene unas ventas mensuales no mayores a \$7.000.000.

- **COSTO TOTAL DE OPERACIÓN MENSUAL:** De los 159 registros ingresados al sistema en el rango de 0-\$5.000.000 hay 83 personas, \$5.000.001-\$15.000.000 50 y mayores a \$15.000.000 26; con un porcentaje de 52%, 31% y 16% respectivamente. Podemos concluir que en el sector microempresario el costo de operación para empresas micro en 52% asciende hasta los \$5.000.000.
- **APROBADO:** De los 159 registros ingresados al sistema 92 de las solicitudes fueron aprobados y 67 negados; con un porcentaje de 58% y 42% respectivamente. Podemos concluir que en el sector el microempresario el porcentaje de aceptación aprobación crediticia es del 58%.

### 7.3.3 Análisis de instancias o registros correctos o incorrectos

Al ingresar la base de datos histórica con la información financieras de los 159 registros resulta necesarios analizar la calidad de la información:

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      87      54.717 %
Incorrectly Classified Instances    72      45.283 %
Kappa statistic                    0.0442
Mean absolute error                 0.4698
Root mean squared error             0.5272
Relative absolute error             96.2777 %
Root relative squared error         106.7367 %
Total Number of Instances          159
```

Figura 3. Resultados de las instancias correctas e incorrectas y su participación.

Con un número de 10 Folds y un porcentaje de 80% para datos de entrenamiento 20% datos de prueba, las 159 instancias o registros el 55% están clasificadas de forma correcta y siguen un mismo patrón mientras que el 45% presentan ruido o dispersión.

Mediante la herramienta Cross validation da la opción de cambiar el número de Folds y de esta forma buscar un mayor porcentaje de instancias correctas para que la información trabaje en rangos similares y no dispersa para que de esta forma se procese de manera efectiva, en este caso con un número de 15 Folds se obtuvo el siguiente resultado:

Correctly Classified Instances	95	59.7484 %
Incorrectly Classified Instances	64	40.2516 %
Kappa statistic	0.1323	
Mean absolute error	0.451	
Root mean squared error	0.5015	
Relative absolute error	92.4218 %	
Root relative squared error	101.5215 %	
Total Number of Instances	159	

*Figura 4.* Resultados de las instancias correctas e incorrectas y su participación.

Se mejora el número de instancias correctas a un 60% y un 40% de incorrectas, en este caso disminuye el rango de dispersión de la información

- La dispersión de la información se puede evidenciar en la siguiente gráfica:

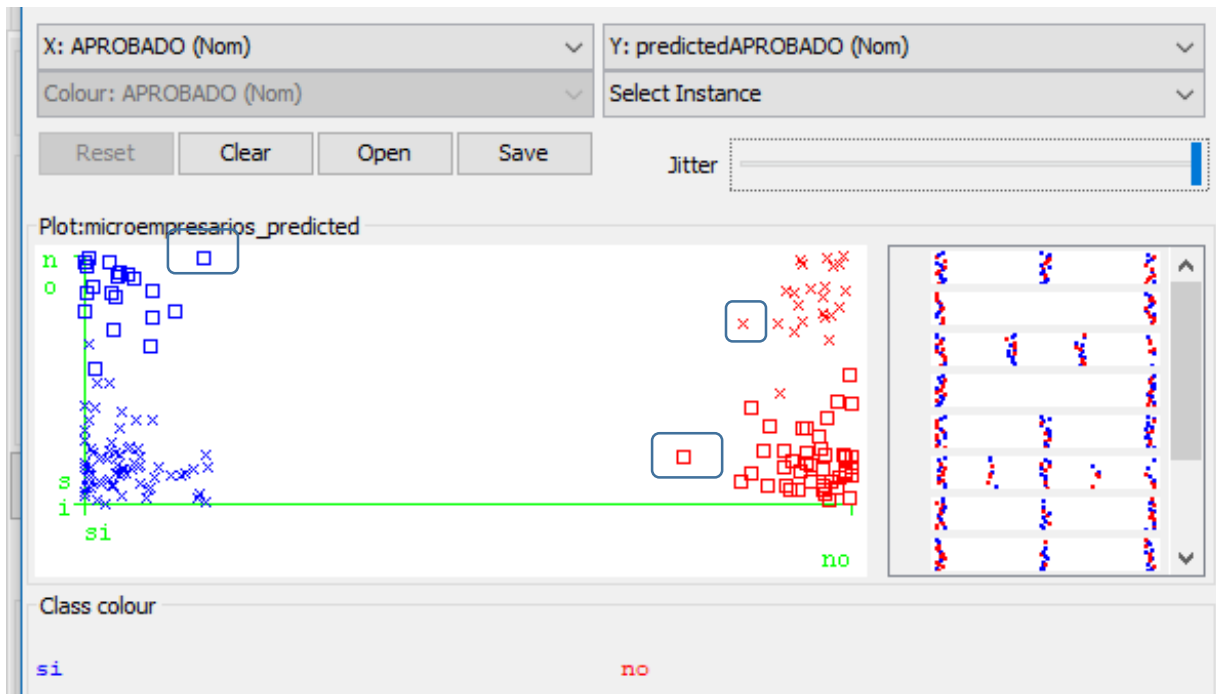


Figura 5. Dispersión de las instancias incorrectas.

### 7.3.4 Valoración de los atributos de mayor relevancia

Mediante la aplicación del logaritmo, el logaritmo j48 toma los atributos que más se relacionan entre los mismos atributos y los califica.

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

number of folds (%)  attribute
                    5( 50 %)   1 EDAD
                    0(  0 %)   2 GENERO
                    1( 10 %)   3 ESTADOCIVIL
                    4( 40 %)   4 HIJOSACARGO
                    1( 10 %)   5 VIVEENCASA
                    3( 30 %)   6 TIPODEEMPRESA
                    3( 30 %)   7 PASIVOS
                    1( 10 %)   8 PROMEDIOVENTAS
                    10(100 %)  9 COSTOSTOTALESOPERACION
```

Figura 6. Resultados de los atributos con mayor relevancia.

Como se puede evidenciar para la toma de decisión de si se aprueba o no el crédito los atributos a tomar en cuenta son los costos totales de operación, seguido de la edad, los hijos a cargo, el tipo de empresa, los pasivos, el promedio de ventas, estado civil y en qué tipo de casa vive; también se puede evidenciar que el género no toma parte en la decisión; no importa si es mujer o hombre.

### 7.3.5 Árbol j48

En el árbol de decisión J48 se refleja la clasificación de la base de datos general donde los atributos de mayor relevancia serán los nodos raíz debido a su relación con los demás atributos.

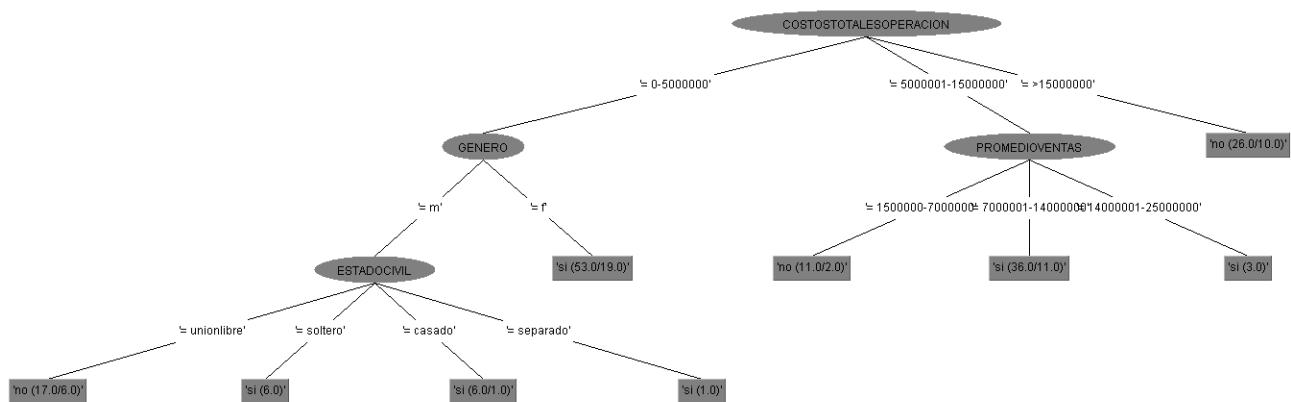


Figura 7. Árbol j48 de la base de microempresarios para la toma de decisión de aprobación de crédito.

### 7.3.6 Predicción de aprobación a la solitud de crédito

Para dar cumplimiento al objetivo específico “Identificar el riesgo crediticio que se expone una entidad bancaria al recibir una solitud de crédito con la metodología de árboles de decisión para establecer si se le presta o no el dinero al solicitante” se debe determinar si se concede o no el préstamo, el sistema analizará los 159 registros y con los resultados obtenidos se procederá analizar la base de datos donde se encuentra la información del nuevo caso; es decir, del nuevo cliente quien solicita el crédito. Este método es conocido como “supplied test set”, donde se proporciona un nuevo archivo de datos (en formato ARFF y con los mismos atributos) sobre el que se realizará la clasificación. A continuación se relaciona la información del nuevo solicitante:

Tabla 4.

*Información de la nueva solicitud de crédito.*

<b>ATRIBUTO</b>	<b>ETIQUETA</b>
Edad	35-51
Género	M
Estado civil	Soltero
Hijos a cargo	>=3
¿Vive en casa?	Propia
Tipo de empresa	Servicio
Pasivos	0-5000000
Promedio de ventas	14000001-25000000
Costos totales operación	5000001-15000000

Como resultado de la ejecución el sistema responderá y mostrara la sugerencia para dicho cliente como se muestra en la siguiente figura:

```
=== Predictions on test set ===
```

```
inst#,    actual, predicted, error, probability distribution
  1      1:si    1:si      *1      0
```

*Figura 8.* Predicción y probabilidad de la nueva solicitud de crédito.

En comparación de la información del nuevo cliente con la base histórica se aprueba la solicitud de crédito con una probabilidad del 100%.

#### 7.4 Cumplimiento a los objetivos específicos número 2 y 3

“Evaluar similitudes de clientes antiguos con posibles clientes nuevos como ingresos, actividad económica, calificación de pago, activos etc. Mediante el logaritmo J48, y de esta manera definir un posible comportamiento de pago futuro del cliente nuevo”.

“Determinar de manera asertiva el posible comportamiento de pago del nuevo cliente mediante la comparación de las características y comportamiento de pago del historial de antiguos clientes”.

En la bases histórica donde se refleja la información financiera de las 159 personas evaluadas anteriormente, dio como resultado la aprobación de crédito 92 personas y 67 fueron rechazadas, donde resulta necesario evaluar y analizar y calificar el comportamiento de pago de cada de uno de los 92 créditos ya activos que cuenta con cierto tiempo después de la aprobación.

La entidad bancaria en el área de microfinanzas califica a los usuarios con créditos activos con referencia a la información que se presenta en la siguiente tabla:

Tabla 5.

*Clasificación y descripción del comportamiento de pago.*

<b>NUMERAL</b>	<b>DESCRIPCIÓN</b>
A	Crédito normal sin ningún tipo de retraso.
B	Crédito aceptable, pero existen debilidades potenciales que afecten transitoriamente o permanentemente la capacidad de pago del usuario, entran a esta categoría los créditos que tengan de 1 a 3 meses vencidos.
C	Crédito deficiente, son los que presentan insuficiencias en la capacidad de pago del deudor o codeudores, entran a esta categoría los créditos que tengan de 3 a 6 meses vencidos.

- D Crédito de difícil cobro, es aquél que tiene cualquiera de las características del deficiente, pero en mayor grado, entran a esta categoría los créditos que tengan de 6 a 12 meses vencidos.
- E Crédito incobrable, es aquél que se estima irrecuperable, entran a esta categoría los créditos que tengan más de 12 meses vencidos.

[AUTORÍA PROPIA]

#### 7.4.1 Resultados de los atributos y/o variables

El ingreso de la base de datos de los 92 registros aprobados al sistema WEKA, arrojó los siguientes resultados:

Tabla 6.

*Consolidado de los resultados de los atributos de cada uno de los rangos y su participación.*

ATRIBUTO	ETIQUETA	RESULTADOS	PORCENTAJE
Edad	18-34	38	41%
	35-51	43	47%
	>52	11	12%
Género	M	38	41%
	F	54	59%
Estado civil	Unión libre	40	43%
	Soltero	22	24%
	Casado	22	24%
	Separado	8	9%
Hijos a cargo	0-2	68	74%
	>=3	24	26%
¿Vive en casa?	Arrendada	31	34%
	Propia	32	35%
	Familiar	29	32%
Tipo de empresa	Comercio	55	60%
	Producción	10	11%
	Servicio	15	16%

	Rentista Capital	4	4%
	Agro	8	9%
	0-5000000	52	57%
Pasivos	5000001-15000000	30	33%
	>15000000	10	11%
	1500000-7000000	54	59%
Promedio de ventas	7000001-14000000	25	27%
	14000001-25000000	13	14%
	0-5000000	52	57%
Costos totales operación	5000001-15000000	30	33%
	>15000000	10	11%
	A	79	86%
	B	6	7%
Comportamiento de pago	C	4	4%
	D	1	1%
	E	2	2%

[AUTORÍA PROPIA]

#### 7.4.2 Análisis de los resultados por atributo de los 92 créditos activos

- **EDAD:** De los 92 registros ingresados al sistema en el rango de 18-34 años se obtuvieron 38 personas, 35-51 años 43 personas y mayores a 52 años 11; con un porcentaje de 41%, 47% y 12% respectivamente. Podemos concluir que las solicitudes aprobadas en el sector microempresario en casi la mitad son de personas en un rango de edad de 35 a 51 años.
- **GÉNERO:** De los 92 registros ingresados al sistema 38 son hombres y 54 mujeres; con un porcentaje de 41% y 59% respectivamente. Podemos concluir que en el sector microempresario las mujeres tienen una mayor presencia.
- **ESTADO CIVIL:** De los 92 registros ingresados al sistema en el rango de unión libre hay 40 microempresarios, solteros 22, casados 22, separados 08; con un porcentaje de 43%, 24%, 24% y 9% respectivamente. Podemos concluir que en el sector microempresario la

unión libre sin un compromiso formal es predominante y el 33% no cuenta con una relación estable actualmente.

- **HIJOS A CARGO:** De los 92 registros ingresados al sistema en el rango de 0-2 hijos a cargo se obtuvieron 68 y mayores o iguales a 3 hijos 24 personas; con un porcentaje de 74% y 26% respectivamente. Podemos concluir que en el sector microempresario en una porcentaje de 71% de participación no tienen responsabilidad económica con respecto a la manutención de hijos o presentan una responsabilidad menor de 1 a 2 hijos a cargo.
- **TIPO DE CASA:** De los 72 registros ingresados al sistema en el rango de que los microempresarios viven en propiedades arrendadas hay 31 personas, propia 32 y familiar 29; con un porcentaje de 34%, 35% y 32% respectivamente. Podemos concluir que en el sector microempresario el 65% no cuenta con una vivienda propia.
- **TIPO DE EMPRESA:** De los 92 registros ingresados al sistema en el rango de los microempresarios que tienen un tipo de negocio en comercio hay 55 personas, producción 10, servicio 15, rentista capital 4 y Agro 8; con un porcentaje de 60%, 11%, 16%, 4% y 9% respectivamente. Podemos concluir que en una gran participación los microempresarios están dedicados al comercio.
- **MONTO DE PASIVOS:** De los 92 registros ingresados al sistema en el rango de 0-\$5.000.000 hay 52 personas, \$5.000.001-\$15.000.000 30 y mayores a \$15.000.000 10; con un porcentaje de 57%, 33% y 11% respectivamente. Podemos concluir que en el sector el microempresario común con una participación del 52% accede a créditos menores o iguales a \$5.000.000.
- **PROMEDIO DE VENTAS MENSUALES:** De los 92 registros ingresados al sistema en el rango de \$1.500.000-\$7.000.000 hay 54 personas, \$7.000.001-\$14.000.000 25 personas y

de \$14.000.001-\$25.000.000 13; con un porcentaje de 59%, 27% y 14% respectivamente. Podemos concluir que en el sector microempresario en un 60% aproximadamente tiene unas ventas mensuales no mayores a \$7.000.000.

- **COSTO TOTAL DE OPERACIÓN MENSUAL:** De los 92 registros ingresados al sistema en el rango de 0-\$5.000.000 hay 52 personas, \$5.000.001-\$15.000.000 30 y mayores a \$15.000.000 10; con un porcentaje de 57%, 33% y 11% respectivamente. Podemos concluir que en el sector microempresario el costo de operación para empresas micro en 52% asciende hasta los \$5.000.000.
- **COMPORTAMIENTO DE PAGO:** De los 92 créditos aprobados 79 tienen una calificación de comportamiento de pago de A, 6 de B, 4 de C, 1 de D y 2 de E; con un porcentaje de 86%, 7%, 4%, 1% y 2% respectivamente. Podemos concluir que los microempresarios activos con crédito en la entidad financiera tienen un porcentaje de 86% de buen comportamiento de pago es decir de 100 microempresarios 86 pagan a tiempo sus obligaciones financieras.

### 7.4.3 Análisis de instancias o registros correctos o incorrectos

Al ingresar la base de los 92 créditos aprobados resulta necesario analizar la calidad de la información:

Correctly Classified Instances	79	85.8696 %
Incorrectly Classified Instances	13	14.1304 %
Kappa statistic	0	
Mean absolute error	0.1028	
Root mean squared error	0.2273	
Relative absolute error	89.2566 %	
Root relative squared error	99.6945 %	
Total Number of Instances	92	

Figura 9. Resultados de las instancias correctas e incorrectas y su participación

Con un número de 10 Folds se obtuvo un porcentaje en un 86% de instancias correctas y en un 14% incorrectas las cuales presentan dispersión en la información.

### 7.4.4 Valoración de los atributos de mayor relevancia:

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

number of folds (%)  attribute
      9( 90 %)      1 EDAD
    10(100 %)      2 GENERO
    10(100 %)      3 ESTADOCIVIL
      1( 10 %)      4 HIJOSACARGO
    10(100 %)      5 VIVEENCASA
    10(100 %)      6 TIPODEEMPRESA
      9( 90 %)      7 PASIVOS
      1( 10 %)      8 PROMEDIOVENTAS
      0(  0 %)      9 COSTOSTOTALESOPERACION
```

Figura 10. Resultados de los atributos con mayor relevancia.

Como se puede evidenciar para conocer el comportamiento de pago los atributos a tomar en cuenta son: el tipo de empresa, el tipo de casa, el estado civil, genero, edad, pasivos y en una segunda los hijos a cargo y el promedio de ventas; también se puede evidenciar que el costo total de operación no intervine en el comportamiento de pago de cada cliente.

#### 7.4.5 Predicción del posible comportamiento futuro de pago del nuevo cliente:

Para dar cumplimiento a los objetivos 2 y 3 con respecto a predecir el comportamiento de pago futuro del nuevo crédito aceptado, el sistema analizará los 92 registros y con los resultados obtenidos se procederá a analizar la base de datos con la información financiera de nuevo cliente.

A continuación se relaciona la información del nuevo solicitante:

Tabla 7.

*Información del nuevo crédito aprobado.*

<b>ATRIBUTO</b>	<b>ETIQUETA</b>
Edad	35-51
Género	M
Estado civil	Soltero
Hijos a cargo	>=3
¿Vive en casa?	Propia
Tipo de empresa	Servicio
Pasivos	0-5000000
Promedio de ventas	14000001-25000000
Costos totales operación	5000001-15000000

Como resultado de la ejecución el sistema responderá y mostrara la sugerencia para dicho cliente como se muestra en la siguiente figura:

```
inst#,    actual, predicted, error, probability distribution
1         1:a  1:a      *0.859 0.065 0.043 0.011 0.022
```

*Figura 11.* Predicción y probabilidad del comportamiento de pago del nuevo cliente.

En comparación de la información del nuevo cliente con la base histórica se predice un comportamiento de pago futuro A con una probabilidad de 86%. Hace referencia que con el pasar de las cuotas que el nuevo cliente tiene que cancelar él no va quedar en mora.

## 8 Conclusiones

- ✓ La tasa de aprobación de créditos en el área de microfinanzas es del 58% y la de buen comportamiento de pago es del 89%.
- ✓ En los resultados obtenidos por el desarrollo del sistema podemos concluir que al momento de la aprobación de un crédito el atributo que no genera ninguna influencia en la decisión final es el género, no importa si es hombre o mujer.
- ✓ Las variables y/o atributos que más relevancia presentan al momento de discernir un posible comportamiento de pago futuro son: el tipo de empresa, el tipo de casa, el estado civil, genero.
- ✓ No es un análisis que permite suplir todos los riesgos al que se expone una entidad bancaria al aprobar un crédito, sin embargo reduce la incertidumbre y ofrece un valor agregado para tomar una correcta decisión y en menor tiempo.
- ✓ La utilización de los árboles de decisión nos permite determinar patrones de comportamiento en los datos de la muestra analizada.
- ✓ WEKA ofrece una plataforma de análisis de datos de procesamiento sencillo y de fácil interpretación.

## 9 Bibliografía

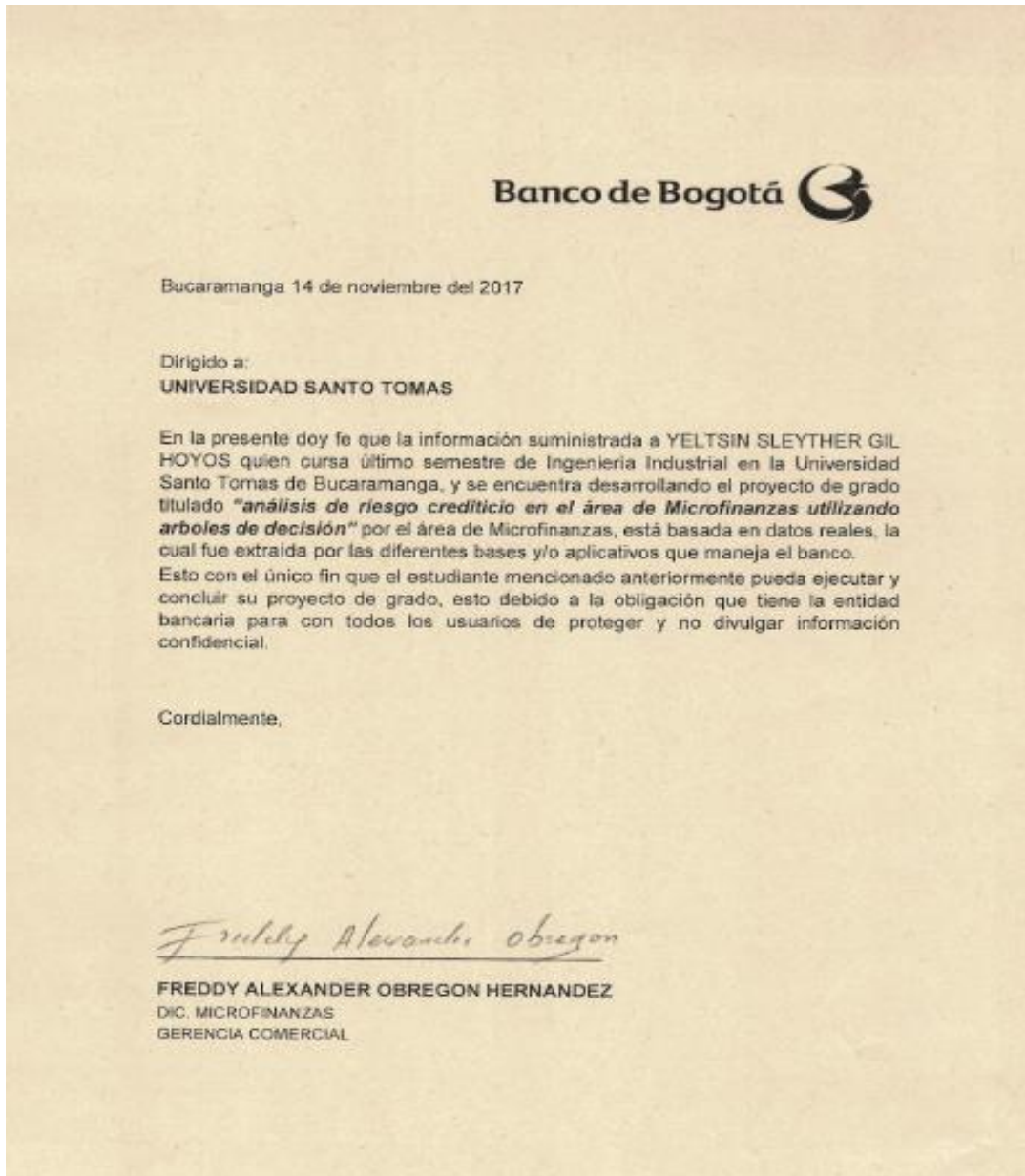
- [1] J. Huerta, Dinero credito bancario y ciclos economicos, 2009.
- [2] zapata, 2005.
- [3] L. Julio, 2016.
- [4] M. & R. M. Luís, 2011.
- [5] J. B. Salgar, EL SARC: UN CAMBIO CULTURAL, 2003.
- [6] L. & O. Álvarez, 2011.
- [7] M. & B. Hernández, 2005.
- [8] L. A. M. C. & R. A. MACUACÉ, Valoración y riesgo crediticio en colombia, 2011.
- [9] s. b. d. Colombia, Gestión del riesgo, 2002.
- [10] G. Martínez, Cómo hallar una aguja en un pajar. Ingenierías, 2011.
- [11] R. Brunet, Que es la minería de datos, 2010.
- [12] W. T. G. Wood, Teoría de decisiones y sistemas de información, 2005.
- [13] I. C. Hidalgo, «Aprobación de creditos bancarios utlizando inteligencia artificial,» 2010.

- [14] M. B. M. G. y. M. C. V. André Apuero, «Minería de datos aplicada a la formación de Equipos de Proyectos de Software,» 2010.
- [15] S. Antonelli, «Árboles de clasificación, algoritmo J48,» 2010.
- [16] M. G. J. - A. Álvarez, «Análisis de dato en WEKA».
- [17] J. H. & M. Kamber, 2001.
- [18] Y. Liz, 2012.
- [19] S. A. -. Basilio, «Aprendizaje automático: conceptos básicos y avanzados,» 2006.
- [20] R. G.R., Minería de datos para la toma de decisiones e inteligencia de negocios: aplicaciones en la mercadotecnia, Universidad nacional autónoma de México., 2012.
- [21] L. A. M. C. & R. A. Macuacé, Valoración y riesgo crediticio en Colombia, 2011.
- [22] [www.definición.com](http://www.definición.com), 2107.
- [23] I. C. Maldonado, 2004.
- [24] C. D. L. REPUBLICA, Ley estatutaria 1581, 2012.
- [25] <http://weka.wikispaces.com>, 2017.
- [26] <http://www.dataprix.com>, 2017.

- [27] M. A. R. O. y. J. M. S. G. z., 2012.
- [28] F. O. P. R. & H. F. Castaño, 2007.
- [29] L. Á. M. C. & R. A. MACUACE, Valoración y riesgo crediticio en Colombia, 2011.
- [30] J. A. G. Gutiérrez, Lineas de investigación en minería de datos en aplicaciones en ciencia e ingeniería: Estado del arte y perspectivas, 2016.
- [31] E. Contreras, Evaluación de inversiones bajo incertidumbre: teórica y aplicaciones a proyectos Chile, 2010.
- [32] P. A. C. Hernández, Aplicación de árboles de decisión en modelos de riesgo crediticio, 2004.
- [33] P. A. V. GARZON, APLICACIÓN DE TÉCNICAS DE INDUCCIÓN DE ÁRBOLES DE DECISIÓN A PROBLEMAS DE CLASIFICACIÓN MEDIENTA EL USO DE WEKA, 2008.
- [34] «Serca de siete millones de colombianos están sin bancarizar,» *El Tiempo*, 2017.
- [35] Orallo, Introduccion a la minería de datos, 2004.

## 10 Apéndices

### 10.1 Carta de validación de la información



## 10.2 Base microfinanzas

EDAD	GENERO	ESTADO CIVIL	HIJOS A CARGO	VIVE EN CASA	TIPO DE EMPRESA	MONTO DE PASIVOS	VENTAS MENSUALES	COSTOS TOTALES DE OPERACION	APROBADO	COMPORTEMENTO DE PAGO
35	m	unión libre	2	arrendada	Comercio	5000000	7004793	6385291	no	
18	f	soltero	0	propia	Producción	2500000	3710241	2548557	si	a
65	m	casado	0	propia	Comercio	1700000	2196206	2268829	no	
42	f	separado	3	propia	Comercio	3000000	4800606	4550606	no	
28	m	unión libre	2	arrendada	Comercio	5000000	8490843	5345251	si	a
24	m	unión libre	1	arrendada	Servicio	1500000	3471617	2263767	si	a
34	f	casado	2	propia	Comercio	2200000	4082818	2953452	si	a
56	f	unión libre	3	familiar	Comercio	2500000	5185416	3150375	si	a
45	f	soltero	3	familiar	Comercio	1700000	4217441	2486222	si	a
23	m	casado	0	familiar	Servicio	9070376	12076883	8040951	si	a
26	f	soltero	0	familiar	Comercio	15722338	21693464	13249125	si	a
36	f	separado	0	propia	Producción	2200000	3488150	2139920	si	a
43	m	unión libre	4	propia	Comercio	2500000	5289403	3577813	si	a
42	f	unión libre	3	propia	Comercio	1700000	4044612	2536914	si	a
45	f	soltero	2	arrendada	Servicio	1500000	3836893	2755306	si	a
36	f	separado	3	arrendada	Comercio	13500000	19348570	12712291	si	a
27	m	soltero	2	familiar	Comercio	5000000	7522638	5390137	si	a
19	m	casado	1	familiar	rentista capital	3000000	4704250	3397602	si	a
27	m	soltero	3	arrendada	comercio	6000000	9853502	6722451	si	b
37	f	unión libre	2	familiar	producción	5000000	7975833	4893350	si	a
62	m	casado	4	propia	comercio	6000000	10339963	6100777	si	a
20	f	unión libre	3	arrendada	comercio	1500000	4274531	2889681	si	c
32	f	unión libre	2	arrendada	comercio	1700000	3957017	2992361	si	b
53	m	casado	1	propia	servicio	3000000	4900128	3082748	si	a
24	f	soltero	2	familiar	agro	2500000	4442291	3300468	si	a
53	m	separado	2	propia	producción	8069544	11852650	7714330	si	c
60	f	separado	0	propia	comercio	1500000	2441111	1807744	si	b
35	m	casado	2	propia	agro	2200000	4127511	2775541	si	a
37	f	unión libre	3	familiar	comercio	2200000	4734966	2912946	si	a
40	m	unión libre	2	arrendada	producción	16759472	22653647	15438283	si	a
29	f	casado	2	arrendada	servicio	3000000	5926301	3635556	si	a
23	f	soltero	0	familiar	comercio	15629249	19400035	14034758	si	a
38	m	unión libre	0	arrendada	comercio	2500000	4558916	2914890	si	c
45	m	casado	4	propia	servicio	2500000	5277059	3750857	si	b
38	f	soltero	3	familiar	comercio	1700000	4079741	2547924	si	a
28	f	unión libre	2	arrendada	comercio	5000000	7636396	5998006	si	d
35	f	casado	1	propia	comercio	6000000	8116087	5962065	si	a
32	m	soltero	2	familiar	producción	3000000	5089833	3704810	si	a
27	f	unión libre	3	arrendada	agro	1700000	4475942	2979898	si	a
45	f	unión libre	1	propia	comercio	1700000	3262724	1993611	si	a
47	m	casado	4	propia	servicio	1500000	4030072	2719198	si	a
26	m	unión libre	0	familiar	comercio	18134289	22095569	16288328	si	c
38	f	soltero	2	familiar	producción	17500000	24901666	14503500	si	a
31	f	unión libre	3	arrendada	comercio	5000000	8524680	6037770	si	a
46	f	casado	0	arrendada	comercio	8758826	12428350	7983540	si	a
32	m	soltero	2	familiar	agro	2500000	4724444	2979200	si	a
45	m	unión libre	1	propia	rentista capital	5000000	7127584	4930975	si	b

27	f	soltero	0	familiar	comercio	5000000	6947638	4698037	si	a
31	m	unión libre	2	arrendada	producción	10000000	24300000	15722666	si	b
47	f	casado	1	arrendada	agro	1700000	3666370	2583006	si	a
27	f	unión libre	0	familiar	comercio	1500000	2775875	1670051	si	a
38	m	unión libre	3	propia	servicio	15500000	21010457	14200132	si	a
39	f	unión libre	3	familiar	comercio	2500000	4919444	3492000	si	a
36	f	unión libre	2	familiar	servicio	17000000	22975333	15141560	si	a
46	f	unión libre	2	propia	comercio	6000000	9137744	5678028	si	a
45	m	casado	1	arrendada	servicio	18126768	24830329	15698058	si	a
34	m	soltero	3	familiar	comercio	3000000	5825000	3465000	si	a
29	m	unión libre	0	arrendada	comercio	11000000	14524480	10379815	si	a
57	m	soltero	1	propia	comercio	1500000	2934552	1952995	si	a
54	f	casado	2	propia	producción	1700000	3442443	2521199	si	a
63	f	casado	0	propia	servicio	6854188	9591874	5781359	si	a
48	m	casado	3	propia	agro	5000000	8164178	5216438	si	a
56	f	unión libre	3	propia	comercio	2500000	4846315	3355962	si	a
28	f	unión libre	1	arrendada	servicio	2500000	4868944	3144753	si	a
43	f	unión libre	3	propia	comercio	6000000	9495223	5970276	si	a
52	m	separado	2	propia	rentista capital	2500000	4600991	3012872	si	a
35	f	soltero	1	familiar	comercio	1500000	3072541	1909451	si	a
31	f	unión libre	2	arrendada	producción	13400000	18738742	12544482	si	a
26	m	soltero	1	propia	comercio	1700000	3207380	2061776	si	e
24	f	unión libre	2	arrendada	comercio	13200000	18941804	11946754	si	a
39	f	casado	3	propia	comercio	3000000	5370667	3884086	si	a
22	m	soltero	0	familiar	servicio	6000000	8423000	5310720	si	a
46	f	unión libre	1	arrendada	comercio	1700000	3777622	2507278	si	a
36	m	separado	2	familiar	comercio	16000000	20572666	14823380	si	a
23	m	unión libre	1	arrendada	comercio	2292093	4573889	2993554	si	a
46	f	unión libre	0	propia	comercio	5000000	6563746	4795505	si	a
54	m	unión libre	1	arrendada	servicio	6000000	9377811	5659464	si	a
35	f	soltero	0	propia	comercio	11000000	14819654	9192007	si	a
37	f	casado	2	arrendada	comercio	3000000	5619634	4011082	si	a
28	m	unión libre	2	familiar	agro	2500000	4701319	3010718	si	e
26	f	separado	1	familiar	comercio	1700000	3394541	2114368	si	a
37	f	unión libre	2	arrendada	comercio	1500000	4036453	2436820	si	a
42	f	soltero	1	familiar	comercio	5000000	7929166	4657750	si	a
45	f	casado	3	arrendada	comercio	6000000	9948301	6499878	si	a
43	m	unión libre	2	propia	comercio	5000000	7662019	5361561	si	a
29	f	unión libre	1	arrendada	agro	1700000	3956980	2384952	si	a
38	m	casado	1	arrendada	comercio	8146621	11629311	7930538	si	a
47	m	separado	4	propia	comercio	6000000	9727690	6736213	si	a
31	f	unión libre	2	familiar	servicio	2500000	4793819	2881868	si	a
24	f	casado	1	arrendada	comercio	2200000	4560103	2843990	si	a
26	m	soltero	0	familiar	comercio	1700000	2948383	1810431	si	a
24	f	separado	1	arrendada	servicio	1500000	2989802	2875857	no	
45	f	unión libre	2	propia	comercio	2500000	4523093	3028481	si	a
36	f	unión libre	1	arrendada	servicio	20100948	22437523	19910039	no	
37	m	unión libre	2	arrendada	agro	6000000	7596499	7806148	no	
42	f	casado	3	propia	comercio	1700000	4035336	2418346	si	a
25	f	soltero	0	familiar	comercio	3000000	4296500	3142410	si	a
37	m	unión libre	1	arrendada	rentista capital	2200000	4386519	3006397	si	a
36	m	unión libre	2	familiar	comercio	6000000	6938000	7111520	no	
24	m	unión libre	0	familiar	comercio	15800000	14459133	15331464	no	

49	m	unión libre	2	arrendada	comercio	2500000	4281532	4525510	no	
39	m	unión libre	1	propia	comercio	6000000	7130996	6476826	no	
19	m	unión libre	0	arrendada	comercio	17350000	18634334	17176503	no	
51	m	unión libre	2	propia	comercio	1500000	2972578	2762150	no	
34	m	unión libre	2	arrendada	comercio	5000000	6457779	6824402	no	
40	m	unión libre	3	familiar	comercio	12670187	14225482	13542675	no	
44	m	unión libre	4	arrendada	comercio	2500000	5343398	4746558	no	
61	m	unión libre	3	arrendada	comercio	2200000	4436769	4356101	no	
51	m	unión libre	4	arrendada	comercio	1700000	4304581	4276085	no	
57	m	unión libre	3	familiar	comercio	6000000	7151500	7545105	no	
24	m	unión libre	2	arrendada	comercio	3000000	4789974	4787373	no	
60	m	unión libre	4	propia	comercio	18125000	19749325	18863824	no	
33	m	unión libre	2	familiar	comercio	19400000	19141366	19486688	no	
32	m	unión libre	3	arrendada	comercio	1500000	3771028	3791869	no	
36	m	unión libre	0	familiar	comercio	2500000	3242916	3009175	no	
51	m	unión libre	2	arrendada	comercio	5000000	7070116	6450076	no	
56	m	unión libre	2	propia	comercio	6000000	7627872	6789697	no	
26	m	unión libre	0	familiar	comercio	2200000	2683311	2831976	no	
29	m	unión libre	1	arrendada	comercio	17350000	17442515	17596350	no	
20	f	unión libre	0	familiar	comercio	1700000	2464072	2271839	no	
21	f	soltero	0	arrendada	comercio	2500000	3579373	3633108	no	
51	f	casado	3	propia	comercio	1500000	3328763	3006424	no	
63	f	casado	4	propia	comercio	6000000	8443839	7626085	no	
63	f	separado	2	familiar	comercio	5000000	6343333	6135000	no	
20	f	soltero	0	familiar	comercio	3000000	3809500	3447240	no	
53	f	casado	2	familiar	comercio	2500000	3926666	3822500	no	
41	f	separado	3	arrendada	comercio	2200000	4562277	4510312	no	
39	f	casado	2	arrendada	comercio	16700000	17241256	17434549	no	
18	f	soltero	0	familiar	comercio	6000000	6022666	6324480	no	
23	f	soltero	0	familiar	comercio	17013846	16928749	16478733	no	
33	f	casado	2	arrendada	comercio	1700000	3640753	3684668	no	
45	f	separado	3	propia	comercio	2500000	4344647	3975938	no	
32	f	soltero	2	arrendada	comercio	6000000	8204652	7294186	no	
36	f	casado	3	arrendada	comercio	17350000	20725035	18315242	no	
61	f	soltero	2	propia	comercio	2200000	3576245	3389195	no	
54	f	separado	2	familiar	producción	6000000	7437500	6970625	no	
47	f	casado	3	familiar	producción	5000000	6619166	6515000	no	
24	f	soltero	1	arrendada	producción	6000000	7036743	7213580	no	
18	f	soltero	0	familiar	producción	16700000	16784783	16176321	no	
36	f	casado	2	propia	producción	2200000	3831783	3283605	no	
31	f	soltero	1	arrendada	producción	2500000	3824995	4018495	no	
22	f	soltero	1	familiar	servicio	3000000	3849583	3954562	no	
48	f	casado	2	propia	servicio	5502082	6816981	6464221	no	
38	f	soltero	3	propia	servicio	1700000	3339719	3457121	no	
45	f	casado	4	arrendada	servicio	17350000	19866497	18760709	no	
34	f	soltero	2	familiar	servicio	16700000	17962283	16921121	no	
61	f	soltero	4	propia	servicio	2200000	4277306	4174033	no	
42	f	casado	2	arrendada	servicio	12924512	13968179	13966845	no	
57	f	separado	3	propia	servicio	5000000	6141615	6520444	no	
27	m	unión libre	2	propia	servicio	14360301	14906917	14737829	no	
48	m	unión libre	3	propia	servicio	17350000	17268609	17869348	no	
64	m	unión libre	4	propia	servicio	2200000	4239979	4262445	no	
42	m	unión libre	2	familiar	servicio	2500000	3735833	3904775	no	
38	f	soltero	3	arrendada	rentista capital	20640262	23582447	21168378	no	

28	f	soltero	2	propia	rentista capital	1500000	3062684	2643915	no	
46	f	casado	3	propia	rentista capital	6000000	7850985	7223259	no	
38	f	casado	0	familiar	agro	1700000	2335108	2329842	no	
37	f	soltero	0	propia	agro	3000000	3670931	3367465	no	
41	f	separado	2	propia	agro	16700000	18060601	16715331	no	