UNIVERSIDAD
NACIONAL
DE COLOMBIA

Trabajo de Grado

"MÉTODO PARA EVALUAR EL SCORING DE CRÉDITO DE LA LÍNEA DE LIBRANZAS EN LAS COOPERATIVAS DE CRÉDITO DE MEDELLÍN"

Modalidad
"PROFUNDIZACIÓN"

Estudiante
VICTOR ALFONSO GONZÁLEZ MORA
Especialista en Ingeniería Financiera

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE MINAS
MAESTRÍA EN INGENIERÍA ADMINISTRATIVA
SEDE MEDELLÍN
2023

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Trabajo de Grado

"MÉTODO PARA EVALUAR EL SCORING DE CRÉDITO DE LA LÍNEA DE LIBRANZAS EN LAS COOPERATIVAS DE CRÉDITO DE MEDELLÍN"

Trabajo de grado presentado como requisito para optar al título de
Magister en Ingeniería Administrativa

Estudiante
VICTOR ALFONSO GONZÁLEZ MORA
Especialista en Ingeniería Financiera

Director Trabajo de Grado
D. Sc. SERGIO BOTERO BOTERO

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE MINAS
MAESTRÍA EN INGENIERÍA ADMINISTRATIVA
SEDE MEDELLÍN
2022

(El esfuerzo de un buen análisis, es la recompensa de un buen triunfo)

Este material si bien es de uso académico, no tiene censura para su uso en actividades profesionales o empresariales, según se requiera, y para quien pueda ayudar

A mis padres

"Que siempre han estado cuando me he sentido flaquecer, y por quienes me he vuelto a levantar".

En memoria de Rigoberto González Rivera y Myriam Teresa Mora Betancur, por ser mis formandos en la universidad de la vida

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Victor Alfonso Gonzalez Mora

_____

Nombre

Fecha 31/01/2023
Fecha Treinta y uno de Enero de 2023

Tabla de contenido

Resumen

En el presente proyecto de grado, se desarrolla una investigación empírica a modo de profundización sobre algunos modelos de clasificación de riesgo, sustentados en la metodología de la curva "ROC", a partir de la cual se selecciona el modelo que mejor área bajo la curva "AUC" tenga, y así, predecir de una mejor forma el comportamiento del buen y mal hábito de pago de los clientes de las Cooperativas Financieras de Crédito de Medellín, en su línea de libranzas. Dicha investigación se hace con el fin de reducir los re procesos manuales en los procesos de crédito, mitigar al máximo todos los riesgos de crédito que se presenten en cada operación de libranza y argumentar de mejor forma la toma de decisiones. Conforme lo expuesto, partiremos de la premisa que combinando algunos clasificadores de riesgo como Generalized Linear Models (GLMs)-Logit, Deep Learning (Neural Networks)-Redes Neuronales y Árboles Binomiales, con el propósito de implementar modelos más sólidos y precisos, que permitan predecir oportunamente el incumplimiento de los clientes en las operaciones de crédito.

Adicionalmente, se tuvo en cuenta la regulación prudencial del Comité de Supervisión Bancaria de Basilea, el cual ha generado a lo largo de los años 3 acuerdos, los cuáles han ido respondiendo a las necesidades humanas según la evolución y cambios presentados en la sociedad; sin embargo, adentrándonos en el territorio Colombiano, encontramos que de tal comité se derivó una normativa propia en Colombia, la cual adopta las recomendaciones presentadas por el Comité de Basilea sobre el riesgo de crédito que las entidades financieras deben adoptar y a su vez existe una serie de entes reguladores que se encargan de supervisarlas, controlarlas y vigilarlas.

Palabras Clave: Clasificación de riesgo, scoring de crédito, comité de regulación Basilea, ROC, AUC.

Abstract

"Method to evaluate the credit scoring of the libranzas line in the credit cooperatives of Medellin"

In the present grade project, an empirical research is developed as a way of deepening on some models of risk classification, based on the methodology of the "ROC" curve, from which the model that best area under the curve "AUC" is selected, And so, predict in a better way the behavior of the good and bad habit of payment of the customers of the credit cooperatives of Medellín, in their line of libranzas. This research is carried out with the aim of reducing manual rework in credit processes, minimising all credit risks involved in each credit resolution and providing a better basis for decision-making. As stated, we will start from the premise that combining some risk classifiers such as Generalized Linear Models (GLMs)-Logit, Deep Learning (Neural Networks)-Neuronal Networks and Binomial Trees, with the purpose of implementing more robust and accurate models, which allow timely prediction of customer default in credit operations.

In addition, account was taken of the prudential regulation of the Basel Committee on Banking Supervision, which has generated 3 agreements over the years, which have been responding to human needs according to developments and changes in society; however, going deep into the Colombian territory, we find that from such a committee came its own regulations in Colombia, which adopts the recommendations submitted by the Basel Committee on credit risk to be adopted by financial institutions, and there are a number of regulatory bodies responsible for monitoring, controlling and supervising them.

Keywords: Risk classification, credit scoring, basel regulatory committee, ROC, AUC.

1. Introducción

"El riesgo se ha convertido en un factor estratégico, no sólo para las entidades financieras, sino para cualquier organización, con independencia del tamaño que posea y del sector en que realice su actividad; un factor que puede marcar el futuro de cualquier entidad." (MEDINA, 2008)

1.1 Aproximación a la identificación del problema

A lo largo de los años, los gobiernos han tenido la tarea de regular y supervisar la actividad financiera con el fin de permitir un sistema confiable y "sano". Así mismo, cada institución financiera ha tenido que enfrentar el problema de cómo abordar el nivel de riesgo en sus operaciones de crédito, y para ello, cada uno ha tenido que decidir si adoptar modelos de riesgos existentes, personalizarlos o construir modelos propios desde cero, ésto con el fin de poder calificar a sus clientes de la forma más acertada posible.

Adicionalmente, se ha tenido que utilizar diversos métodos de minería, incluyendo regresiones logísticas, modelos no paramétricos, árboles de decisión, redes neuronales, entre otros, para poder estimar la probabilidad de cumplimiento de pago de los usuarios de una forma más oportuna. Los modelos de medición del riesgo permiten clasificar los clientes, y su vez, los modelos también pueden ser clasificados. Una forma de validar estos modelos es realizar una curva "ROC" (Receiver Operating Characteristic), la cual permite determinar la exactitud diagnóstica de éstos.

En esta investigación se presenta la necesidad de un grupo específico de instituciones (las Cooperativas de Crédito de Medellín), para construir un modelo de clasificación de datos que ayude a predecir el perfil de los clientes en sus operaciones de crédito (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

1.2 Justificación de la Investigación

A menudo, las empresas de servicios financieros del país, se enfrentan al reto sobre "como tomar sus pérdidas financieras por no pago, como oportunidades de mejora para su sistema financiero interno", sin embargo, y a pesar que esto lo hacen con el fin de poder seguir mejorando sus modelos de riesgo crediticio, les está costando cada vez más poder abordar mejor esta situación, debido a la fluctuación tan constante de la economía actual. Basado en lo anterior, se plantea la presente investigación con el fin de brindar algunas herramientas adicionales a las Cooperativas de Crédito de Medellín, las cuales les permitan gestionar los riesgos de una forma más adecuada, durante la toma de decisiones en el proceso de asignación de cupos de crédito.

Para esta investigación, se parte de que las compañías financieras del sector cooperativo han podido recopilar grandes volúmenes de datos con atributos propios de cada cliente, lo cual les puede servir para implementar mejoras sobre sus análisis predictivos, y desarrollar nuevos modelos de riesgo más acordes a sus necesidades.

De igual forma, la calidad de los datos es un factor determinante para el mantenimiento, rentabilidad y costo eficiencia de las Cooperativas de Crédito, ya que con ésto pueden desarrollar nuevos modelos de clasificación de clientes y ofrecerles productos mucho más diversificados.

1.3 Objetivos de la Investigación

1.3.1 General

Implementar un método para evaluar los modelos de medición de riesgo en la línea de libranzas de las Cooperativas de Crédito en Medellín, con el cual se pueda validar la gestión de riesgo de este tipo de productos, y así determinar cuál es el más conveniente, pertinente y eficaz.

1.3.2 Específicos

- Analizar las diversas técnicas estadísticas sobre las cuáles se desarrollan los modelos de riesgo "scoring".
- Desarrollar un método para evaluar el scoring de crédito con el cual se pueda medir la exposición al riesgo de crédito en el proceso de otorgamiento de solicitudes de financiación con el que operan las Cooperativas Financieras en Medellín en su línea de libranza.
- Aplicar un modelo estadístico para validar el método desarrollado, y determinar si se ajusta a las características del perfil de riesgo de los clientes bajo la modalidad de libranzas de las Cooperativas de Crédito de Medellín, tal que se identifiquen las variables cualitativas y cuantitativas disponibles en la población de estudio.

Para el desarrollo del primer objetivo, se tomó como base la investigación del autor Reyes Samaniego Medina, con su libro "El Riesgo de Crédito en el Marco del Acuerdo Basilea II" (MEDINA, 2008). Allí, el riesgo se determina como un factor estratégico en general, en todas las entidades financieras del mundo. Partiendo de esto, así como del nivel de concientización respecto a la importancia de una gestión adecuada del riesgo de crédito, existen directrices a nivel mundial en el sector financiero como los acuerdos del Comité de Basilea los cuales pueden o no ser adoptados en los países, con el fin de evitar y prevenir riesgos sistémicos. Cabe resaltar que si bien no es obligatorio que las entidades financieras de cada país adopten los acuerdos de Basilea, éstas lo hacen genéricamente para poder mantener una cartera cada vez más sana. En Colombia, a parte de adoptar dichos acuerdos, se crearon las entidades "Superintendencia Financiera" y "Superintendencia de Economía Solidaria", las cuales velan porque las entidades que ejecutan operaciones de crédito en general, hagan una gestión adecuada del riesgo, y brinden una mayor confianza a los clientes, usuarios, accionistas, inversionistas y demás consumidores del sistema financiero del país.

Para el segundo objetivo, se planteó una descripción demográfica sobre el sector cooperativo de Antioquia, el cual se encuentra distribuido territorialmente conforme las 39 entidades financieras solidarias en el Valle de Aburrá, y sobre las cuáles se aplicó una metodología de tipo cualitativa descriptiva, teniendo en cuenta el tipo de clientes de la región antioqueña, así como el proceso de otorgamiento de créditos en la línea de libranzas regulado por las entidades gubernamentales arriba descritas.

Para el tercer objetivo, se realizó un análisis comparativo de diferentes modelos de evaluación de scoring planteados, aplicados a las bases de datos de las Cooperativas de Crédito de Medellín.

2. Marco Teórico

En el presente marco teórico se pretende realizar un análisis de teorías, investigaciones y antecedentes generales, sobre la gestión del riesgo en la línea de libranzas de las Cooperativas de Crédito de Medellín.

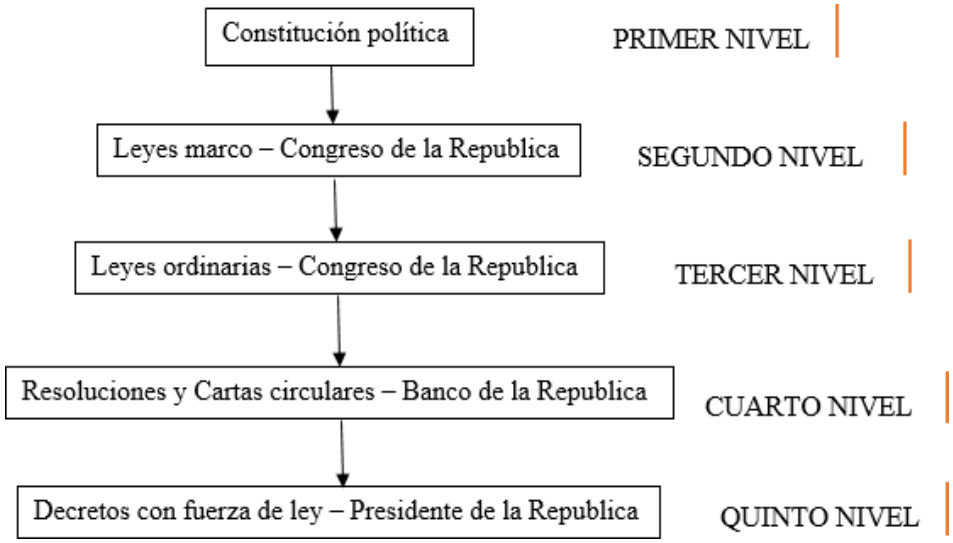
Para poder plantear la hipótesis de la presente investigación, así como su respectiva solución, es de vital importancia partir del hecho de que la economía ha sufrido grandes cambios a medida que el hombre ha evolucionado, y es por esta razón, que las entidades financieras han tenido cada vez más competencia entre sí y las fronteras geográficas y culturales han ido desapareciendo.

Es por ello que para plantear un adecuado análisis, se analizaron diversas definiciones técnicas de riesgo de crédito, dentro de las cuáles tenemos a Reyes Samaniego Medina (MEDINA, 2008)., Banco BBVA (Grupo BBVA, 2015), el Decreto 2555 de 2010 artículo 2.1.1.3.1. (Pública, 2010) y finalmente, la de la Superintendencia Financiera de Colombia (SFC) (Superintendencia Financiera de Colombia, 2008). Para efectos del presente trabajo, se toma la definición de riesgo de crédito de la Superintendencia Financiera de Colombia "Es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones".

2.1 Normatividades y entes de control

La normatividad Colombiana, es aplicada a todas las entidades del sector financiero conforme su orden jerárquico. En 1er lugar encontramos la Constitución Política de Colombia, partiendo del artículo 395 donde se estipula que la actividad financiera, es una actividad de interés público y por ende deberá ser autorizada por el Gobierno Nacional; en 2do lugar, encontramos las leyes marco expedidas por el Congreso de la República; en 3er lugar encontramos las Leyes Ordinarias emitidas por el Poder Legislativo Ordinario o por el Congreso de la República respectivamente; en 4to lugar encontramos las Resoluciones y Cartas Circulares expedidas por el Banco de la República.
Finalmente, en el 5to lugar encontramos los Decretos con Fuerza de Ley, cuyo autor es el Presidente de la República haciendo uso de sus facultades extraordinarias. (ASOBANCARIA, 2021)

Figura 1. Niveles Organismos Control del Crédito en Colombia



Fuente: Asobancaria (2021)

En Colombia, la entidad que se encarga de realizar, el control, vigilancia e inspección de las personas que desempeñan actividades financieras, bursátiles y de aseguramiento, es la Superintendencia Financiera de Colombia, el objetivo de esta es mantener la seguridad, confianza y estabilidad del sistema financiero (STEVENS, 2020)

2.1.1 Comité Basilea

"BCBS" – Es la sigla del Comité de Supervisión Bancaria de Basilea. Dicho comité es el organismo encargado a nivel mundial de la regulación prudencial de los bancos, y su solvencia.

Basilea expide recomendaciones que si bien no son de obligatorio cumplimiento, ya que no son legalmente vinculantes, su implementación requiere del compromiso de los países miembros, en pro de mantener buenas prácticas financieras y propender por una economía estable. Conforme a lo anterior, el principal objetivo de Basilea es regular y supervisar prácticas bancarias a nivel internacional, buscando siempre la mejora de su liquidez, solvencia, gobernanza y gestión del riesgo, esto con el fin de contribuir de forma positiva a la estabilidad financiera mundial.

Las decisiones trascendentales que ha tomado el comité siempre se han tomado por el grupo de Gobernadores y Jefes de Supervisión, quienes dentro del comité, son el máximo órgano de control; adentrándonos un poco en la historia del Comité de Supervisión Bancaria de Basilea, encontramos que el primer acuerdo de Basilea (Basilea I) se firma en 1988 este con el fin de establecer estándares internacionales de regulación bancaria internacional, posteriormente, 16 años después, en el año 2004 se crea el segundo acuerdo de Basilea (Basilea II) este modifico algunos aspectos del acuerdo anterior, finalmente, para el año 2017 se da un nuevo acuerdo, el tercer acuerdo de Basilea (Basilea III) donde se estipula un nuevo marco de regulación prudencial. (BANCO DE ESPAÑA, 2022)

2.1.1.1 Acuerdo Basilea I

Este primer acuerdo, se firma el 15 de julio de 1988, por el Grupo de los 10, que para la época, era conformado por Bélgica, Francia, Alemania, Japón, Italia, Suecia, Holanda, Reino Unido, Estados Unidos y Suiza; este acuerdo también es conocido como el Acuerdo de Capital, se da gracias a la necesidad que había de unificar criterios en la administración de riesgos de las entidades financieras a nivel mundial; dando esto el nacimiento a los parámetros regulatorios de la gestión del riesgo crediticio de los bancos comerciales. (Torres Avendaño, 2005)

Este acuerdo, define, el riesgo de crédito, como aquella probabilidad de que un deudor, incumpla con sus obligaciones de pago, obstaculizando la actividad financiera del banco o la entidad prestadora del servicio, poniendo en riesgo el capital de sus acreedores, en vista de esto, el Comité de Basilea establece la metodología para cubrir el riesgo, a través de, mantener una reserva, es decir un capital mínimo, cuyo objetivo es proteger posibles pérdidas. (Torres Avendaño, 2005)

Adicionalmente, se puede utilizar el siguiente procedimiento para realizar el cálculo de los requerimientos de capital, de cada entidad:

1. *Tomar el listado con valores de tipo "elementos de capital", suministrado por el acuerdo de Basilea I.*
2. *Clasificar en 5 categorías los riesgoss de los activos, donde cada categoría va porcentualmente desde 0% para los valores sin riesgo,*

hasta 100% para valores con riesgo. Estos últimos son considerados como normal.

3. *Categorizar los activos teniendo en cuenta:*
    - *Sector institucional al cual pertenece el emisor.*
    - *Emisor u Obligacionista.*
    - *Todas las garantías que se generen para la operación.*
4. *Ponderar el capital regulatorio con un valor mínimo del 8%, de los activos ponderados por riesgos, tal que a mayor riesgo, mayor capital.*

A continuación se presenta la distribución de porcentajes conforme las 5 categorías descritas en el acuerdo I de Basilea, de forma cualitativa:

Tabla 1. Distribución de Porcentajes – Acuerdo Basilea I

Pesos	Valores
0%	Emitidos por Estados o Bancos Centrales de los países de la OCDE.
10%	Emitidos por Administraciones Públicas distintas al Estado. En el caso de la Unión Europea, activos emitidos por entidades crediticias especializadas en el descuento de papel público.
20%	Operaciones interbancarias o bien con países no pertenecientes a la OCDE con duraciones menores al año.
50%	Préstamos con garantías hipotecarias de viviendas.
100%	El resto de las operaciones.

(BANCO DE ESPAÑA, 2022)

Donde $Z_{(n_1+n_2)}$ es el resultado de combinar la muestra $X_{n_1} = x_1, \ldots, x_{n_1}$ y la muestra $Y_{n_2} = y_1, \ldots, y_{n_2}$, en donde las observaciones de cada muestra son ordenadas de mayor a menor. N_i representa en número de observaciones en $X_{(n_1)}$ que son menores o iguales que la i-ésima observación de $Z_{(n_1+n_2)}$.

Conforme a esto, la hipótesis nula H0 de la expresión X(n1) y Y(n2) de la misma distribución continua es rechazada, siempre y cuando el estadístico de la prueba de AD, sea mayor que el valor crítico ADα, para un nivel de significancia α.

Ahora, para poder revalidar la certeza de la prueba como tal y comprobar aún más su homogeneidad, se pueden realizar múltiples pruebas para k muestras.

Se deberá tener en cuenta, que el capital que se obtiene aplicando estos pasos, es un requerimiento mínimo para subsanar el riesgo de crédito, sin embargo, si la entidad financiera considera que se requiere aumentar el capital mínimo, debido a diversas causales, como puede ser la escasez de provisiones o un incremento en otros riesgos que no son de crédito, podrá hacerlo.

Basados en lo expuesto hasta ahora, a continuación se listan algunos tópicos que fueron reevaluados por causa de la evolución tecnológica

y financiera mundial, y que debido a esto mismo, el acuerdo ya no cubría con todas las necesidades por las cuáles fue creado:

- No diferenciar las características de cada uno de los activos, por causa de establecer pesos iguales en las categorías de riesgos.
- No evaluar la concentración de riesgo, el vencimiento, el tipo de negocio, controles internos de la entidad financiera o la calidad de su gestión, a causa de la no implementación de modelos internos para la medición de los riesgos de crédito.
- Categorías de riesgos poco específicas, las cuáles no distinguían claramente los niveles de riesgo de la banca, causando así, que las entidades financieras asuman riesgos sin que estos se reflejen en los requerimientos mínimos de capital.
- Técnicas de reducción de riesgo de crédito que no se integraron al modelo de ponderación de riesgos.
- Esquema de ponderación frecuente, que si bien era periódico, no recolectaba los cambios en las situaciones crediticias.

De acuerdo a lo descrito y a la falta de estímulos bancarios para este 1er acuerdo, el Comité Basilea ejecutó un análisis exhaustivo del acuerdo, y luego de 16 años, generó el Acuerdo de Basilea II.

2.1.1.2 Acuerdo Basilea II

Creado en el 2004. Se conoce como "Convergencia Internacional de medidas y normas de capital: marco revisado" tuvo como objetivo principal, *"construir una base sólida para la regulación prudente del capital, la supervisión y la disciplina de mercado, así como perfeccionar la gestión del riesgo y la estabilidad financiera".* Este se encuentra diseñado para ofrecer a las entidades financieras alrededor del mundo diferentes posibilidades para la regulación prudente del capital. (Comite de Supervisión Bancaria de Basilea, 2004)

Este acuerdo se divide en dos partes, su aplicabilidad y sus tres pilares "información de mercado, el requerimiento mínimo de capital, y la supervisión del acuerdo como tal".

La aplicabilidad del acuerdo en general, parte de su misma normativa, en donde el capital bancario debe ser el suficiente para poder cubrir todos los riesgos de las entidades, sin que ninguna de las partes quede expuesta. Allí, todas las entidades deben cumplir con un mínimo de capital en los siguientes niveles:

- Consolidado: Grupos Financieros o Consolidaciones Globales.
- Sub consolidado: Sub Grupos Financieros o Entidades Individuales, donde cada banco es internacionalmente activo

Los anteriores grupos se caracterizan por:

- Tener o participar en Entidades Aseguradoras.
- Tener participaciones pequeñas en entidades bancarias, de valores u otras entidades financieras.
- Tener participaciones pequeñas en empresas comerciales.
- Tener grandes participaciones en otras entidades.

Ahora bien, el acuerdo, se encuentra dirigido principalmente a grandes bancos internacionales, sin embargo sus principios y metodologías podrán aplicarse a entidades de menor grado de complejidad y el comité espera que sea así

A continuación veremos los **3 pilares** arriba descritos, para la regulación del capital:

1. **Requerimiento mínimo de capital:** Realiza un cambio al acuerdo de Basilea I, en el denominador de los activos ponderados por riesgo, es decir, tanto la definición de capital y el porcentaje de capital se mantienen. (Gonzales Cervantes & Zornoza Batiz, 2006)

   En cambio en el denominador radica en, la metodología de medición, midiendo tres tipos de riesgos, riesgo de crédito, riesgo de mercado y riesgo operacional.

   Además, este pilar contempla, que si bien la clasificación de los riesgos de crédito de cada cliente debe ser anual, sin embargo, para aquellos que generan mayor riesgo o que se encuentran en mora, deberán tener una clasificación mucho más frecuente, con el fin de ir, actualizando la información obtenida de cada cliente. (Gonzales Cervantes & Zornoza Batiz, 2006)

2. **Proceso del examen supervisor:** Alienta a los bancos y entidades financieras a desarrollar sus propias técnicas de gestión de riesgo y ponerlas en práctica.

   La función principal de los supervisores es examinar la eficiencia en la cuantificación de las necesidades de capital de cada banco e intervenir cuando sea necesario. (Gonzales Cervantes & Zornoza Batiz, 2006)

   Se crearon 4 principios básicos del proceso de supervisión:

   1. Las entidades financieras deberían contar con un proceso de evaluación para la suficiencia del capital total de acuerdo al perfil de riesgo y la estrategia de mantenimiento del nivel de capital.
   2. Evaluación periódica del proceso utilizado por las entidades bancarias para la determinación de la suficiencia del capital, la calidad de la gestión y garantizar el cumplimiento de manera continua de estas condiciones.
   3. La meta que se deben poner los supervisores, es que los bancos operen por encima de los coeficientes mínimos del capital requerido.
   4. Los supervisores deben intervenir con prontitud antes de que el capital descienda por debajo de los niveles mínimos requeridos. (Gonzales Cervantes & Zornoza Batiz, 2006)

3. **Disciplina del mercado:** Complementa el proceso de supervisión, en el sentido de que, se desarrollen una serie de principios con los cuáles se pueda divulgar la información, permitiendo así, que los participantes del mercado evalúen el perfil de riesgo de una entidad

bancaria y cuál es su nivel de capitalización. (Gonzales Cervantes & Zornoza Batiz, 2006)

Esta pilar se aplica, únicamente a un nivel consolidado del grupo, por lo que para los bancos individuales no es de carácter obligatorio divulgar la información, a excepción de la exigencia e integración del capital mínimo de cada entidad; el acuerdo de Basilea II no requiere que esta información sea divulgada posterior a una auditoría externa, sin embargo, si una autoridad local lo exige, esta exigencia deberá seguirse. (Gonzales Cervantes & Zornoza Batiz, 2006)

La puesta en marcha del acuerdo de Basilea II, busca una mejora continua en los procesos de gestión del riesgo.

2.1.1.3 Acuerdo Basilea III

Este nuevo acuerdo, trae consigo una serie de reformas que tienen como objetivo reforzar las normas internacionales sobre el capital y la liquidez, la finalidad de esto es crear un sector bancario resistente; mejorando la capacidad de absorción de perturbaciones financieras o económicas reduciendo así el riesgo. (Comite de Supervisión Bancaria de Basilea , 2010)

Este nuevo acuerdo se da como respuesta a la crisis económica y financiera del 2007, en la que se evidencio que el sector financiero de un gran número de países acumulo un apalancamiento excesivo superando el balance; creando que los bancos no tuvieran la capacidad de absorber las pérdidas sufridas en las negociaciones y los créditos.

Con el fin de afrontar estos fallos en el mercado, los cuáles salieron a relucir durante la crisis, el comité presenta este nuevo acuerdo, el cual introduce una serie de mejoras en la gestión del riesgo, reforzando las buenas prácticas en las entidades y estimular la transparencia y la adecuada divulgación de la información, esto con el fin de que el sistema financiero se encuentre en la capacidad de aguantar los periodos de tensión. (Comite de Supervisión Bancaria de Basilea , 2010)

Las reformas que trae consigo el acuerdo de Basilea III, son:

- El requerimiento mínimo del capital del 8% establecido desde acuerdo de Basilea I, sigue siendo igual, sin embargo la composición de ese capital se modifica y se exige que este sea de alta calidad.
- Con el fin de que se dé un aumento de capital en los momentos de crecimiento económico y que este pueda ser utilizado en los casos de pérdidas, se crea un colchón de conservación de capital y se imponen ciertos límites a la distribución de beneficios.
- Al requerir más capital en los momentos de crecimientos excesivos del crédito, se establece un colchón en contra del capital cíclico, con el fin de evitar la formación de los que se llaman “burbujas”.
- Finalmente a las grandes entidades financieras, se les exige un capital adicional, esto como medida para evitar un riesgo sistémico. (Banco Bilbao Vizcaya Argentaria, 2017)

El acuerdo de Basilea III; es el primero, en la aplicación de requerimientos mínimos de liquidez a través de la ratio de cobertura de liquidez (LCR) y la ratio de financiación neta estable (NSFR); la función de estas ratios, es evaluar la supervivencia de las entidades bancarias en un eventual problema de liquidez bien sea a corto o largo plazo. (Banco Bilbao Vizcaya Argentaria, 2017)

2.2 Marco Conceptual del Crédito

Si bien hasta ahora hemos hablado del marco de acción global de los procesos de crédito en general, y como están organizados y estandarizados de forma transversal en la economía mundial, se plantean dos fases relevantes para realizar el proceso de crédito: la primera es el filtrado de datos, y la segunda es la aplicación de un método de evaluación de crédito.

2.2.1 Filtrado de Datos

Para lograr una mayor asertividad en la implementación de una técnica estadística, es necesario filtrar la data insumo y separarla de los datos que no generen valor en la investigación. Luego, es necesario testear el set de datos filtrado, con esquemas de pruebas que ayuden a disminuir el nivel de incertidumbre de las operaciones de crédito, y permitan tomar decisiones de si se aprueba o no cada solicitud de crédito.

Conforme lo anterior, se plantea el uso de los siguientes esquemas de pruebas:

- **Kolomogorov-Smirnov:** Aquí se establece la función de distribución empírica acumulada, y se caracteriza la prueba KS en 2 escenarios:

"*Variable aleatoria X donde,* $x_1, x_2, \ldots, x_n$ *son observaciones para un test de tamaño n y F* (x) *la función de distribución acumulada teórica subyacente de los datos. Definimos la distribución acumulada empírica de X como:*

Fórmula 1. Variable Aleatoria - KS

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(x_i \leq x)$$

(Fernández Castaño & Pérez Ramírez, 2005)

Con la prueba anterior, determinamos la cuantificación de la distancia vertical máxima entre 2 funciones de distribución acumulada empíricas, respecto de 2 muestras aleatorias disponibles independientes. Conforme a esto, optamos por aplicar la prueba de bondad de ajuste no paramétrica, teniendo en cuenta que no es sensible a las diferencias de escala de la función de distribución acumulada.

Luego, se tomarán las muestras x1, xn1 con el tamaño n1 de la variable aleatoria continua X y y1, yn2 con el tamaño n2 de la variable aleatoria continua Y, F1 y F2 funciones de una distribución acumulada teórica de las variables X e Y. Una vez obtenidos los resultados, con el modelo de pruebas de KS, analizaremos si las 2 muestras obtenidas, corresponden

originariamente de la misma distribución contínua hipotética, en donde no se especifica una distribución común:

Fórmula 2. Análisis de dos muestras - KS

$$\begin{cases} H_0 : F_1(x) = F_2(x) \\ H_1 : F_1(x) \neq F_2(x) \end{cases}$$

(Fernández Castaño & Pérez Ramírez, 2005)

En la anterior ecuación tenemos una hipótesis nula H_0 donde las 2 muestras se derivan de una distribución normal y la hipótesis alternativa H_1 describe que las 2 muestras no esta1n derivadas de la misma distribución. Ahora se presenta la siguiente ecuación estadística para demostrar la hipótesis nula *Ho:*

Fórmula 3. Hipótesis Nula - KS

$$KS = \max_x |\hat{F}_{n_1}(x) - \hat{F}_{n_2}(x)|$$

(Fernández Castaño & Pérez Ramírez, 2005)

De acuerdo a esto, $\hat{F}_{n_1}(x)$ corresponde al valor de la función de distribución empírica de X en la observación x y F^n2 (x) corresponde al valor de la función empírica de Y en la observación x. Por consiguiente, si las muestras $x_1, \ldots, x_n1$ y $y_1, \ldots, y_n2$ resultan originarias del mismo segmento de clientes en una población determinada, su contraste siempre va a ser una cola, teniendo en cuenta que su acumulación empírica $\hat{F}n1$ y $\hat{F}n2$ no es tan diferente. Por último, si tomamos un nivel de significancia α la hipótesis nula *H0* de igual distribución, ésta es rechazada, ya que Kolmogorov es superior a Kolmogorov α. Para corroborar esto, el valor de Kolmogorov -Smirnov α, KS lo toma como crítico de 2 muestras de la tabla de valores/datos utilizados.

- **Prueba de Anderson Darling:** Esta prueba tiene el mismo enfoque de la de Kolmogorov-Smirnov, y presenta los siguientes escenarios de aplicabilidad:

    - Con esta prueba es posible aplicar sensibilidad a la forma y escala de la distribución acumulada y a las colas de las distribuciones
    - Es posible aplicar esta prueba a muestras pequeñas o con un set de datos acotado.
    - Es posible aplicar esta prueba a muestras grandes para poder identificar diferencias en las muestras pequeñas.

La siguiente fórmula es acotada al ejercicio:

Fórmula 4. Prueba Anderson Darling - AD

$$AD = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1+n_2} (N_i Z_{(n_1+n_2-n_1 i)})^2 \frac{1}{iZ_{(n_1+n2-i)}}$$

(Fernández Castaño & Pérez Ramírez, 2005)

Donde $Z_{(n1+n2)}$ es el resultado de combinar la muestra $X_{n1} = x_1, \ldots, x_{n1}$ y la muestra $Y_{n2} = y_1, \ldots, y_{n2}$, en donde las observaciones de cada muestra son ordenadas de mayor a menor. N_i representa en número de observaciones en $X_{(n1)}$ queson menores o iguales que la i-ésima observación de $Z_{(n1+n2)}$.

Conforme a esto, la hipótesis nula H_0 de la expresión $X_{(n1)}$ y $Y_{(n2)}$ de la misma distribución continua es rechazada, siempre y cuando el estadístico de la prueba de *AD,* sea mayor que el valor crítico AD_α, para un nivel de significancia α. Ahora, para poder revalidar la certeza de la prueba como tal y comprobar aún más su homogeneidad, se pueden realizar múltiples pruebas para *k* muestras.

2.2.2 Modelos de evaluación de crédito

El objetivo del sistema de medición de riesgo de crédito es identificar cuáles son aquellos elementos determinantes en el riesgo de crédito en las carteras de las diferentes instituciones financieras; en la medición del crédito debe tenerse en cuenta los criterios de clasificación de las carteras crediticias; la estructura y composición de los portafolios de los productos crediticios; el impacto de las variables macroeconómicas y sectoriales y los antecedentes históricos de las carteras de crédito de cada entidad financiera. (Saavedra Garcia & Saavedra Garcia, 2010).

Existen diversos modelos de para evaluar el riesgo del crédito, sin embargo, en el presente trabajo de investigación, a continuación se describen los más representativos:

2.2.2.1 Modelo Scoring / De Calificación

Estos modelos, también reciben el nombre de score-cards o classifiers; traducido al español hace referencia tarjeta de puntuación, puntaje o clasificadores; fueron creados en los años de 1970 pero sólo hasta los años 1990, tomo fuerza y fueron ampliamente utilizados, con la entrada en vigencia del Acuerdo de Basilea II.

Estos modelos, consisten básicamente en evaluar de manera automática el riesgo de crédito de un solicitante de financiamiento o de una persona bien sea natural o jurídica que ya es cliente de la entidad financiera; estos modelos se enfocan principalmente en el riesgo de incumplimiento, es decir, tiene una dimensión individual, en la cual, su principal enfoque es clasificar a los solicitantes o clientes de la entidad financiera entre las clases de riesgo, que bien, pueden ser buenos o malos, por medio del uso de estadísticas evaluadoras, técnicas matemáticas, econométricas y de inteligencia artificial. Una vez realizada la evaluación a cada solicitante o cliente, se le es asignado una puntuación o clasificación, estos permiten entonces ordenar y clasificar a cada persona natural o jurídica, según el riesgo de crédito; posterior a esto, según el puntaje obtenido, se asignan los grupos de las personas con los perfiles de riesgo similares, para así finalmente obtener como resultado una aproximación a la probabilidad de incumplimiento del deudor.

Las técnicas que se emplean al momento de evaluar un crédito, depende del caso en particular, estas técnicas son muy variadas, dentro de las cuáles podemos encontrar: *"análisis discriminante, regresión lineal, regresión logística, modelos probit, modelos logit, métodos no paramétricos de suavizado, métodos de programación matemática, modelos basados en cadenas de Markov, algoritmos de particionamiento recursivo, sistemas expertos, algoritmos genéticos, redes neuronales y, finalmente, el juicio humano"* por lo que el uso de una y otra, dependerá de cuál sea la más eficiente y eficaz según el caso a evaluar. (Gutierrez Girault, 2007)

Ahora bien, cada técnica tiene sus ventajas y desventajas, respecto de esto las técnicas de enfoque econométrica, entre estos los de probabilidad lineal, se han ido quedando en el olvido, por sus desventajas, en vista de que este sólo clasifica a los deudores en grupos según el perfil del riesgo de crédito; mientras que las técnicas del probit, logit y la regresión logística, atribuyen a cada deudor la probabilidad de incumplimiento. Por otro lado nos encontramos los modelos no paramétricos y los de inteligencia artificial, los cuáles son, los árboles de clasificación o decisión, las redes neuronales

y los algoritmos genéticos, estos son superiores a los lineales, siempre que no requieren supuestos estadísticos sobre las distribuciones estadísticas.

Si bien la finalidad de los modelos de crédito es lograr una calificación lo más asertiva posible de los clientes/usuarios, es fundamental partir de la compración de diversas variables de perfiles de clientes tanto nuevos como antiguos, en el segmento de negocio donde se aplicará el scoring de crédito. Conforme se vaya aplicando la clasificación de clientes, los atributos de cada uno de ellos son los que determinan finalmente si son objeto de aprobarles la operación de crédito que estén solicitando, para este caso, Créditos de Libranza (pago directo a la obligación, desde su salario) y sucesivamente, son catalogados como ”Solicitantes Solventes” siempre y cuando no comprometan o arriesguen la estabilidad financiera de la empresa/entidad otorgadora del crédito.

Estas técnicas de calificación crediticia, son practicadas en Colombia por todas las entidades financieras en general, dentro de las cuáles tenemos las Cooperativas de Crédito, y específicamente las de Medellín para el presente trabajo; adicional, éstas técnicas son operadas por las Centrales de Riesgo existentes en Colombia, tales como CIFIN y Datacredito, las cuáles publican dichas calificaciones en sus respectivos portales y bases de datos.

En general, el modelo de calificación de clientes “Score de Crédito”, permite las entidades financieras afinar sus productos y servicios de cara al riesgo y la costo eficiencia, y a su vez, tener clientes con las mejores condiciones de pago, sin embargo, para el caso de las Cooperativas de Crédito de Medellín, en su proceso de crédito de libranzas, le han dado más peso a la calificación crediticia de sus clientes que a otros factores/modelos propios internos en sus empresas, lo cual no es del todo malo ya que eso ayuda a acelerar los procesos de aprobación/rechazo de solicitudes, disminuir la carga operativa y a filtrar clientes buenos y malos, pero, ponen a depender la operación de crédito más de una calificación hecha por terceros como las centrales de riesgo, los cuáles en caso de fallar, causan moras en los pagos, una cartera insana en colocación y arriesgan el capital/patrimonio de sus compañías como tal, (lo cual ya ha pasado tiempo atrás causando liquidación de compañías enteras) ya que si sus clientes no pagan completa y oportunamente sus obligaciones, y son las propias Cooperativas de Crédito de Medellín las que deben responder ante sus financiadores, “Los Bancos”.

La generalidad en Colombia a la hora de implementar calificaciones o modelos de crédito, ha sido adaptar buenas prácticas de la industria de Servicios Financieros de otros países, así como normativas de entidades regulatorias internacionales como Basilea, esto nos ha permitido trabajar e implementar más rápidamente técnicas reconocidas como modelos de regresión logística como LOGIT, Random Forest con Árboles de Decisión y Redes Neuronales, como se explica en el presente trabajo, las cuáles son mayormente utilizadas por analistas de crédito, prestamistas, inversionistas, investigadores, desarrolladores de software y profesionales en datos.

2.2.2.2 Modelo Logístico (LOGIT)

Este modelo se usa en el momento en que se desea predecir un resultado binario, por ejemplo, quiebra o no quiera; este es un modelo de regresión binaria, que se basa principalmente en la denominada las variables dependiente se relacionan con las variables independiente, teniendo en cuenta que, las variables dependientes, son

variables dummy, en las que el código 0 es un buen cliente y el código 1 es un mal cliente, se representa con la siguiente ecuación: (Fernández Castaño & Pérez Ramírez, 2005)

Fórmula 5. Prueba Modelo LOGIT

$$Y_i = \frac{1}{1+\exp(-z)} + u_i$$

(Fernández Castaño & Pérez Ramírez, 2005)

En esta ecuación:

Yi: es una variable dependiente, que podría tomar valor de 0 o 1.
Z: es un scoring logístico
U: es una variable aleatoria que se distribuye. (Fernández Castaño & Pérez Ramírez, 2005)

Este modelo se usa para aplicarla a una base de datos, donde permitan mejorar el control, la toma de decisiones en la administración financiera de la entidad y la gestión del riesgo, cuando se necesita realizar un análisis del comportamiento de las variables y cuál es la correlación entre ellas, esto se hace para determinar las relaciones entre grupos determinados de la población.

Dentro de este tipo de modelos, contemplamos el Modelo LOGIT para evaluar el caso particular de modelo lineal generalizado, en donde la distribución sea binomial y respectivamente, la función de enlace corresponde al logaritmo de las razones de probabilidad. De igual forma, LOGIT señala al coeficiente β1 como la pendiente que determina la variación propia de L, siempre que haya cambios de unidades al ingreso. Con este modelo se analizará, valorará y comprobará el nivel de asertividad del modelo score perse, demostrando si se incluye la variable de calificación score o no, y esto que tanto puede desvirtuar el modelo actual con el que operan las Cooperativas de Crédito en Medellín, en su línea de libranzas.

2.2.2.3 Redes Neuronales (DEEP LEARNING)

El modelo de redes neuronales artificiales “ANN”, responde a las analogías neuronales biológicas del cerebro, las cuáles cuentan con las siguientes características:

- **Aprender:** Aquí se realizan muchos ejercicios de entradas y salidas , y las redes neuronales se ponen en modo escucha, hasta que encuentren un modelo que relacione entre sí, las entradas y salidas.

- **Generalizar:** Previo se delimita un márgen de acción y se configuran entradas y salidas al modelo.

- **Abstraer:** A partir de la configuración previa hecha y las variables de entrada y salida, el modelo comienza a identificar aspectos y/o cualidades en común entre las variables de entrada.

A nivel general, las redes neuronales artificiales operan de la siguiente forma, reciben variables de entrada, cada variable se multiplica por su ponderación, se suman las ponderaciones obtenidas, y finalmente se generan respuestas por cada variable de entrada.

La clave para obtener un modelo de redes neuronales eficiente, consiste en definir un esquema de entrenamiento de cada red neuronal, conforme las necesidades perse que se tengan del modelo de crédito. Para ello, es necesario entrenar las redes antes que se implemente el modelo como tal, con las respectivas entradas y salidas deseadas. Dicho esto, existen diversos tipos de entrenamiento, dentro de los cuals para el presente trabajo, abordaremos 2:

- **Entrenamiento Supervisado:** Se acondiciona un agente externo el cual garantiza que la salida real VS la esperada, sea la misma. En caso que hayan diferencias, el modelo se debe seguir ajustando, conforme el peso de sus conexiones.

- **Entrenamiento No Supervisado:** No se cuenta con un agente como tal. La red se configura para que solamente reciba los datos e información de entrada, y luego abstraiga esta información y la separa o clasifique conforme sus caraterísticas en común, recibidas en sus datos de entrada.

Para el presente trabajo nos enfocaremos en el último esquema "No Supervisado", en donde las redes neuronales separarán sus datos de entrada y tipificarán el comportamiento del modelo conforme las variables de salida esperadas.

2.2.2.4 Bosque Aleatorio - Árboles de Decisión (RANDOM FOREST)

Random forest es un calsificador que colecta clasificadores estructuradamente en forma de árbol. Allí cada árbol se construye respecto a un vector aleatorio Θk, *donde* Θk , $k = 1, \ldots, L$ de forma independiente y con una distribución uniforme, en donde cada árbol contribuye o expide un resultado unitario respecto a la clase más frecuente.

El objetivo de este modelo es utilizar en un bosque aleatorio, los árboles de decisión con más bajo sesgo y alta volatilidad, los cuáles sirvan de base para a partir de ellos, construir modelos mucho más predictivos, y poderlos comprobar con la metodología de la curva ROC, la cual comprobaremos con AUC. Estos árboles si bien clasifican, no son paramétricos, por lo que no es necesario aplicar distribuciones específicas. Estos modelos de árboles de decisión permiten combinar muestras características, datos y modificar parámetros del árbol de clasificación.

Este modelo opera extrayendo múltiples sub muestras y reemplazando los conjuntos originales y luego, cada sub muestra de entrenamiento crea un árbol mediante selecciones aleatorias características. La forma tradicional de construir este modelo, es seleccionando aleatoriamente un subconjunto de variables para cada nodo del árbol y así, obtener resultados no tan relacionados entre sí con las muestras empleadas para entrenar el árbol creado. Este modelo es utilizado comunmente ya que permite hacer

uso de una cantidad considerable de árboles, así como una cantidad considerable de variables del subconjunto aleatorio en cada nodo. Sin embargo, para obtener un modelo eficiente con Random, el número de árboles requeridos debe aumentar conforme su cantidad de predictores. Para ello, es clave comparar las predicciones "Forest" con las predicciones de un subconjunto "Forest". Luego, se debe validar el número de variables seleccionadas con cada nodo de cada árbol, teniendo en cuenta el número de predictores.

A nivel general, esta técnica permite administrar:

- **Entradas:** Con muestras que reemplazan el conjunto de datos *T=T1,T2, ….,TN* y el subconjunto de variables *X=X1,X2, ….,XN*

- **Operaciones:** Para construir árboles de decisión empleando subconjuntos de variables en cada muestra, y luego, clasificar las variables de cada árbol de decisión. Adicional, permite asignar a X la clase con mayor cantidad de votos asignados por cada clasificador.

- **Salidas:** Para que conforme las entradas y operaciones realizadas, retornar nuevas etiquetas a los nuevos objetos definidos.

Sin embargo, para poder obtener un modelo eficiente con Random, el número de árboles requeridos debe aumentar conforme su cantidad de predictores. Para ello, es clave comparar las predicciones "Forest" con las predicciones de un subconjunto "Forest". Luego, se debe validar el número de variables seleccionadas con cada nodo del árbol, teniendo en cuenta el número de predictores. Adicional, es aconsejable no optar siempre por el voto con puntuación mayor, ya que este corre el riesgo de desvirtuarse conforme la cantidad de corridas que se hagan con el modelo.

Ahora bien, aunque existan variables que no cuentan con categóricas nominales, esto no es restricción para construir modelos estadísticos y probar su veracidad. En la presente investigación trabajamos con una distribución condicional de la variable respuesta y la forma de cambio ajustada a una variable explicativa, la cual aplicamos por juicio de experto, solamente al modelo de Bosque Aleatorio. Para ello, utilizamos una tabla cruzada que explica la relación entre variables categóricas, a la cual denominamos "Tabla de Contingencia" (Ver tabla 3), donde con la variable dependiente binaria ***y*** (con la cual determinamos si el Modelo Estadístico es eficiente y clasifica a los clientes como Bueno/Malo), y la variable polinómica cualitativa arbitraria ***x*** con ***k*** categorías, analizamos que cada observación pertenezca a una sola categoría, y determinamos que sean exhaustivas y/o excluyentes entre sí:

Tabla 2. Tabla de contingencia 2 × *k*

	X						
Y	C_1	C_2	...	C_i	...	C_k	***Total***
Bueno	b_1	b_2	...	b_i	...	b_k	*B*
Malo	m_1	m_2	...	m_i	...	m_k	*M*
Total	n_1	n_2	...	n_i	...	n_k	*n*

(Fernández Castaño & Pérez Ramírez, 2005)

Aquellas personas objeto de crédito que se clasifican en alguna categoría ***Y, b_i*** *y* ***m_i*** marcan las frecuencias de Bueno/Malo, y sucesivamente en una ***C_j*** categoría de ***X,*** *j* = 1 . . . *k,* en donde denota con ***B*** el número total de individuos Buenos, y con ***M*** el

número total de sujetos etiquetados como Malos, $n_i = b_i + m_i$ y $n = B + M$. Conforme lo expuesto anteriormente, hicimos las respectivas pruebas al modelo Bosque Aleatorio, de las cuáles se dejan evidencia en los anexos del presente trabajo.

2.3 Metodología para Evaluar Modelos de Riesgos

En la actualidad, existen diversos métodos para evaluar los modelos de riesgos de crédito, los cuáles se analizan, parametrizan y aplican conforme el riesgo a tratar. Para el presente trabajo, y teniendo en cuenta que validaremos 3 modelos de diversa índole como regresión "Logit", Bosque Aleatorio - Árboles de Decisión "Random Forest" y Redes Neuronales "Deep Learning", aplicaremos el método de la curva ROC, el cual se ajusta tanto al testeo de métodos binarios en los procesos de calificación de riesgos, como en la ejecución del proceso de aprobación y desembolso de los créditos.

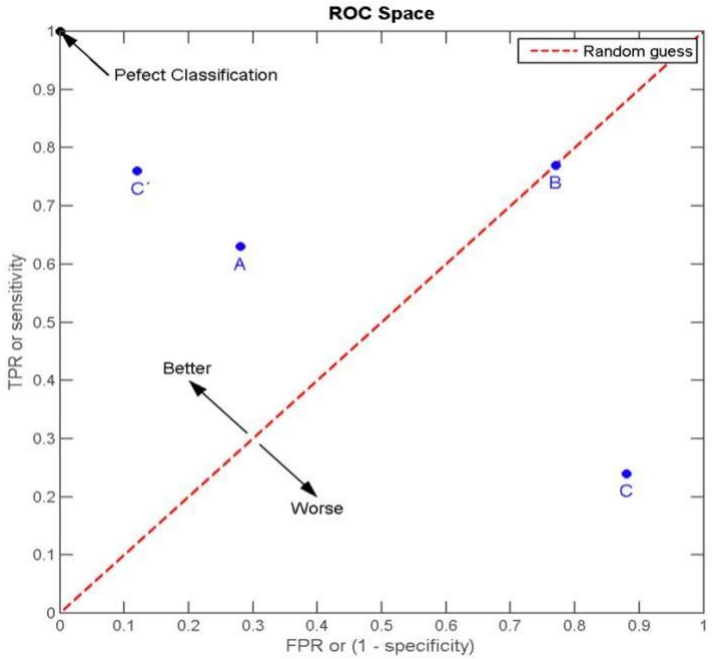
2.3.1 **M**etodología "ROC"

La metodología o curva ROC, es la demostración gráfica del nivel de sensibilidad respecto al nivel de especificidad, de un sistema de clasificación binario. Éste puede tener diversas variaciones según sus umbrales de discriminación. ROC también tiene la capacidad de interpretar las siguientes proporciones VPR "Razón de Verdaderos Positivos" y FPR "Razón de Falsos Positivos", igual según sus umbrales de discriminación, conforme los cuáles, se determina si los casos son positivos o falsos. La técnica de ROC, facilita herramientas para determinar que tan óptimo es o no, un modelo, ya que tiene una relación directa con la costo eficiencia de los tipos de riesgos que evalúa. Si bien ROC es una metodología que históricamente ha apoyado procesos relacionados con la salud, se ha descubierto que también sirve para aplicar técnicas de Inteligencia Artificial y Machine Learning a Modelos Matemáticos y/o de Minería de Datos.

Complementario a la curva ROC, tenemos la tabla de contingencia, la cual suministra diversas medidas de evaluación. Para ello, sólo basta con tener las razones de Verdaderos Positivos (VPR) para medir la cantidad de casos positivos correctos de una prueba diagnóstica, y los Falsos Positivos (FPR) para medir los resultados positivos incorrectos de ésta. Con estos valores, ROC es capaz de graficar la relación de los puntos positivos correctos VPR respecto de los puntos positivos incorrectos FPR. Dicho esto, ROC grafica VPR como puntos sensibles y FPR como puntos especificos, en donde cada resultado, representa un punto en la matriz de confusión.

A continuación, en la tabla 4 un ejemplo de la curva ROC, en donde el punto X=0 y Y=1 muestra todos los resultados positivos correctos, es decir resultados fiables, y el punto X=0 y Y=1 muestra todos los resultados positivos incorrectos, es decir resultados no fiables debido a falsos positivos:

Grafica 1. Curva ROC



(Fernández Castaño & Pérez Ramírez, 2005)

2.4 Estado del Arte de la Investigación

En el planteamiento e inicio del presente trabajo se realizó una búsqueda bibliográfica sobre los artículos y trabajos relacionados con métodos de evaluación de crédito con el fin de caracterizar el estado del arte. En esta revisión se encontró que ya existían trabajos consistentes en revisiones sistemáticas de literatura, con un alcance muy superior al que se podría lograr en este trabajo. Es por eso que se decidió referenciar de dichas revisiones, los artículos más relevantes.

Varios autores han hecho revisiones de los métodos de evaluación de crédito, en ese sentido se destacan las siguientes revisiones:

Louzada, Ara y Fernández (2016): Realizaron una revisión sistemática y una comparación general de los métodos de clasificación aplicados al "scoring" de crédito. En dicha revisión, se encontraron 189 artículos publicados entre enero de 1992 y diciembre de 2015.

Este intervalo de tiempo fue dividido en cuatro periodos. Los periodos fueron numerados así, I, II, III y IV con los intervalos respectivos [1992, 2005], [2006, 2009], [2010, 2012], y [2013,2015], para dichos intervalos se encontraron la siguiente cantidad de artículos: 45, 51, 39 y 54 artículos, respectivamente.

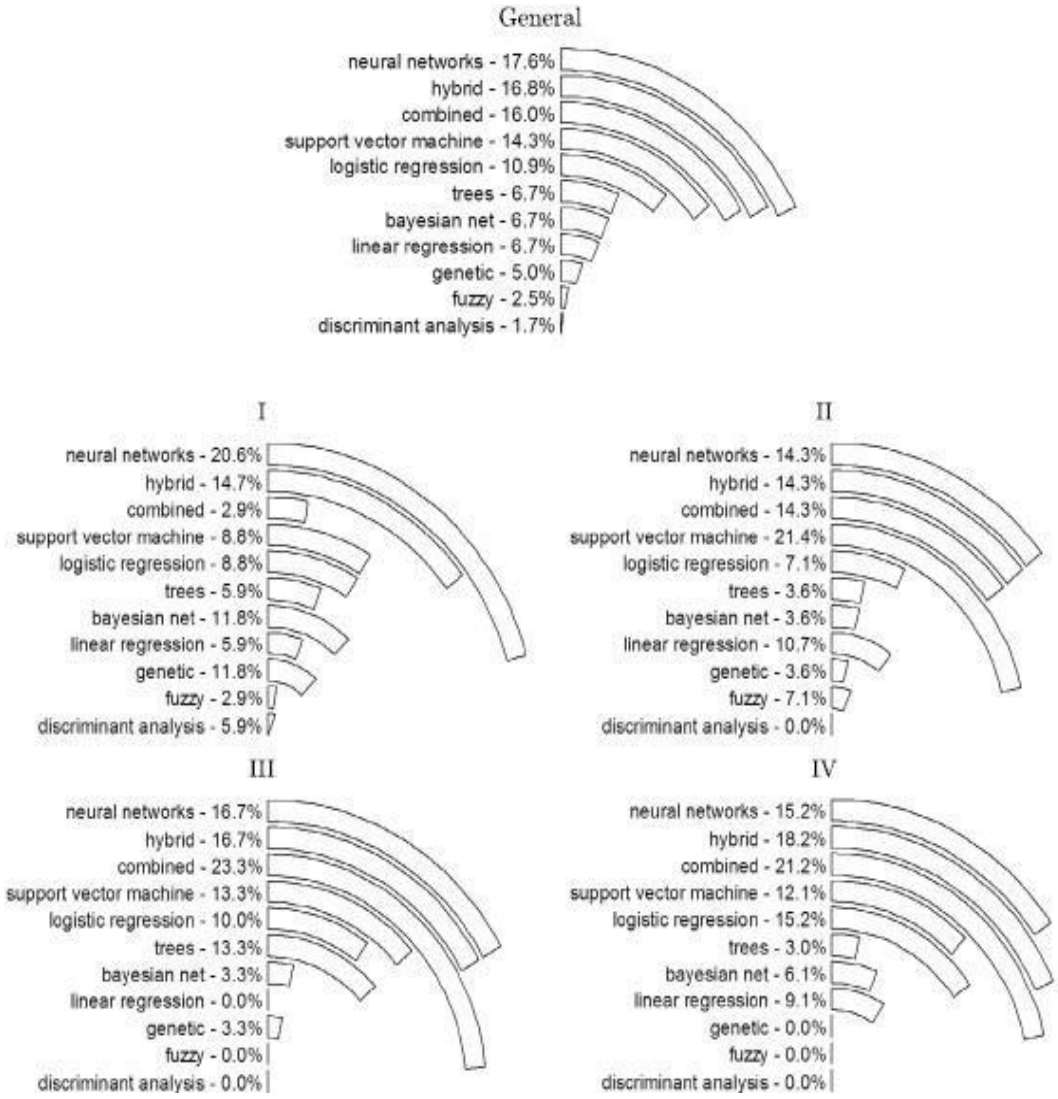
Markov, Seleznyova, Z y Lapshin (2022): Retomaron el trabajo de Louzada, Ara y Fernandez,(2016), siguiendo la misma metodología y cuestionario propuestos en ese estudio, y añadieron el periodo V, para el intervalo [2016, junio de 2021].

Para este periodo se analizaron 110 artículos (partiendo de los 150 más relevantes antes de ser filtrados).

De ambos trabajos, se encuentra que desde 1992 se ha presentado un crecimiento exponencial de artículos relacionados con métodos de evaluación de crédito.

La figura 2. muestra los principales métodos usados durante los cuatro primeros periodos (Louzada et al, 2016), y el resultado general para ese intervalo de tiempo.

Figura 2. Principales técnicas utilizadas para scoring (1992-2016)

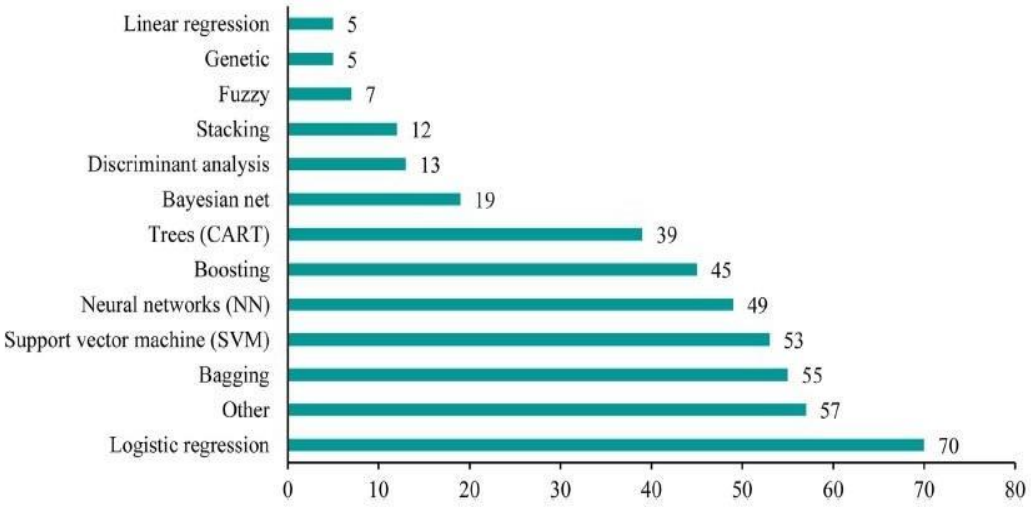


(Louzada F, 2026)

La figura 3 muestra el número de artículos que referenciaron los diferentes métodos, incluyendo el periodo V (Markov et al, 2022). Se puede ver que el método más popular es el Logit. Se podría prestar a confusiones puesto que en que en la figura 2 (hasta 2016) se muestra que el método más usado son las redes neuronales, sin embargo se puede inferir que los modelos logit son ampliamente usados en aplicaciones híbridas y combinadas, que

representan el segundo y tercer lugar en el análisis de Louzada et al. Además el modelo Logit también es usado con las nuevas tendencias computacionales.

Figura 3. Principales técnicas según número de artículos que las usan (1992-2022)



(Markov, Seleznyova, & Lapshin, 2022)

Adicionalmente, con respecto a las técnicas de medición de desempeño de los métodos, Markov et al (2022) verifican que las más usadas son:

- ✓ Medidas basadas en matriz de confusión
- ✓ AUC – ROC
- ✓ Puntaje Brier

De hecho la tendencia que se está presentando es a hacer una combinación de las anteriores.

3. Marco Referencial

3.1 Riesgo de Crédito en Colombia

El proceso más importante y fundamental en el sistema financiero colombio, fue la liberación financiera, llevada a cabo a finales de los años 80 y principios de la década de los 90, donde se realizó la reforma financiera y la apertura de capitales, en la que la expedición de normas de supervisión y control dictadas por la superintendencia bancaria, las cuáles se orientaron a la apertura de los capitales y a controlar el efecto de los riesgos, tomando como base las recomendaciones dadas por el acuerdo de Basilea, teniendo como resultado la emisión de resoluciones que imponen los parámetros para la clasificación de los crédito y las carteras; entre los años 1989 y 1994, se instaura la ruta para calcular el capital, las normas sobre las deudas y la concentración de riesgos crediticios.

3.2 Regulación del crédito en Colombia

En el momento en que la economía colombiana entra en el proceso de reformas estructurales, en los años 90, el sistema financiero se da una transformación en su organización, pasando de ser una banca especializada la cual estaba conformada por entidades separadas para la prestación de múltiples servicios financieros a ser una banca constituida por matrices bancarias que a su vez se componen de filiales separadas. Sin embargo, para comprender este proceso de transformación, es importante conocer los antecedentes. Como se mencionó anteriormente, en esta época Colombia introduce reformas estructurales a su sistema financiero, este tipo de reformas también se dieron en diferentes países latinoamericanos, el pilar fundamental de las mismas es la liberación de los mercados y la internacionalización de la economía.

La fuente de esta reforma en Colombia es la ley 45 expedida en diciembre de 1990; en ella se da una reorganización de las instituciones financieras y la actividad aseguradora; esta ley estableció la relación directa que existe entre el sistema financiero y el mercado de cesantías y la relación que hay entre el sistemas financiero y los fondos pensionales y de cesantías a pesar de estos apenas se encontraban en sus fases iniciales. Posterior a esta ley se crea la ley 100 de 1993; la cual trae consigo la reforma al Sistema de Seguridad Social Colombiano a como lo conocemos hoy, consecuentemente se reforman los fondos pensionales y de cesantías. Finalmente en el año 1996 comienza a regir el decreto 410 de 1971 Código de Comercio de Colombia, a pesar de haber trascurrido 25 años desde su creación, sólo comienza a regir en el año de 1996, y con él las normas sobre los grupos relacionados con las empresas financieras y no financieras.

Estas normas y sus diferentes reformas al pasar de los años, traen consigo las normas de la supervisión en Colombia, país en el cual hasta el momento la regulación prudencial, es la que domina la supervisión, con sus instrumentos principales, que son "*"capital adecuado" y régimen de exposición máxima, la existencia de un seguro de depósitos, la valoración a precios de mercado y las normas sobre cobertura de riesgos."* (Zuleta, 1997)

Las normas de supervisión en Colombia, se aplican inicialmente a entidades individuales y estas están encaminadas a evitar riesgos sistémicos, sin embargo, estas

leyes no se aplican a conglomerados financieros, en donde las matrices y las filiales, si bien existen, como las entidades de supervisión se encuentran separadas para entidades de créditos y servicios financieros y para las entidades del mercado de capitales, los filtros no son fácilmente detectable y la vigilancia es desacertada.

Dentro del proceso de regulación y supervisión del crédito en Colombia, existen una serie de entidades públicas, las cuáles se encuentran a continuación.

3.3 Ministerio de hacienda y crédito público

El 4 de julio de 1866, se crea la ley 68, por medio de la cual se da origen a la Secretaría de Hacienda y del Tesoro, la cual 20 años después con la expedición de la Constitución Política de 1886 se transforma en el Ministerio de Hacienda, consagrado en el título XII. En ese mismo año, pero en el mes de agosto, el presidente Rafael Núñez reglamentó la conformación del ministerio; a lo largo del tiempo este ha ido cambiado de nombre hasta el año de 1923 que con la ley 31 del 18 de julio, se unifica en uno solo los Ministerios de hacienda y Tesoro, en la figura del Ministerio de Hacienda y Crédito Público; nombre que conserva hasta la actualidad. (Ministerio de Hacienda y Crédito Público, 2022)

Las funciones del Ministerio de Hacienda y Crédito Público se encuentran en el artículo 3 del decreto 4712 de 2008; el cual se compone de 36 numerales, dentro de las cuáles las principales funciones son: *"el control de los mercados de capitales, la política cambiaria, el control de la balanza de pagos, el desarrollo de la política fiscal, el arancel y el presupuesto nacional"* (Ministerio de Hacienda y Crédito Público, 2022)

3.4 Superintendencia financiera de Colombia

Otra de las entidades que se vincula, en el proceso de supervisión en Colombia, es la Superintendencia Financiera, esta entidad se encuentra adscrita al Ministerio de Hacienda y crédito Público, esta se crea a partir del decreto 4327 de 2005, con el cual se unifican las superintendencias bancaria y de valores. Esta cuenta con personería jurídica, autonomía administrativa, financiera y un patrimonio propio. (MisAbogados.com.co, 2016)

Sus principales funciones radican en inspeccionar, vigilar, y controlar a las personas naturales o jurídicas que *"realicen actividades financiera, bursátil, aseguradora y cualquier otra relacionada con el manejo, aprovechamiento o inversión de recursos captados del público"* es decir, en Colombia toda entidad que se encuentre autorizada para captar, manejar y aprovechar dinero de particulares, se encuentran vigilados y controlados por la Superintendencia Financiera. (MisAbogados.com.co, 2016)

Los objetivos de la Superintendencia Financiera, son:

"1-Fortalecer la gestión funcional de la SFC: Dada la importancia de la Superintendencia Financiera de Colombia (SFC), esta institución desarrollará acciones orientadas a implementar mejores prácticas aceptadas a nivel internacional en temas de gobierno corporativo, actuando en forma coordinada con las demás entidades del Estado que velan por la estabilidad del sistema financiero".

"2- Fortalecer la supervisión consolidada: La Superintendencia Financiera avanzará en la práctica de las mejores prácticas internacionales para realizar una supervisión comprensiva y consolidada, partiendo del tema de conglomerado financiero, contando con facultades de supervisión a las holding, identificando la estructura de los conglomerados, jurisdicciones, accionistas y partes vinculadas y estableciendo un sistema de administración de riesgos financieros a nivel consolidado, para que las entidades vigiladas administren en forma conjunta sus riesgos, evitando su manejo en forma segregada o independiente".

"3- Contribuir con mecanismos de inclusión y educación financiera: Durante el período 2015 - 2018, la Superintendencia Financiera de Colombia seguirá acompañando las iniciativas del Gobierno Nacional tendientes a fomentar la inclusión financiera, procurando que todas las personas, especialmente las más pobres, tengan acceso a productos y servicios financieros formales acordes a sus necesidades". (MisAbogados.com.co, 2016)

3.5 SARC (Sistema de Administración de Riesgo Crediticio)

Este es el sistema que deben adoptar y complementar todas aquellas entidades solidarias vigiladas, cuyo objetivo deberá ser el de identificar, medir, controlar y monitorear el riesgo de crédito en el cual se encuentran las entidades en el desarrollo de los créditos, dándoles así la oportunidad de tomar decisión en el momento preciso para la mitigación del riesgo; la implementación del Sistema de Administración del Riesgo Crediticio en cada organización, deberá ir acompañada de una serie de políticas y procedimientos, que al ser claro y precisos, permitan la definición de criterios y la guía por medio de la cual la entidad identificara, evaluará, asumirá, calificara, controlara y cubrirá su riesgo crediticio. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Como se ha mencionado anteriormente, pero para enfatizar en ello, el Sistema de Administración de Riesgo Crediticio, aplica a todas aquellas organizaciones solidarias vigiladas que además posean cartera de crédito. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

El SARC al ser un procedimiento, contempla unas etapas básicas que deberán ser complementadas por cada entidad según sus necesidades, estas son: (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

- Identificación: se trata de identificar el riesgo de crédito al cual la organización se encuentra expuesta en el desarrollo de las operaciones autorizadas.

- Medición: Una vez se encuentra identificado el riesgo de crédito de la entidad, por lo que se deberán adoptar metodologías o criterios que permitan evaluar el perfil del solicitante del crédito y del deudor en la etapa de otorgamiento del crédito.

- Control: El SARC debe permitirle a la entidad la toma de medidas pertinentes y eficaces para controla y mitigar el riesgo de crédito.

- Monitoreo: las organizaciones solidarias vigiladas, de manera periódica y permanente deberán hacerle seguimiento a la evolución de su exposición al riesgo

de crédito. (Superintendencia de la Economia Solidaria"SUPERSOLIDARIA", 2021)

3.5.1 Definiciones SARC

La circular básica contable y financiera, Circulas Externa #35 del año 2021 de la Superintendencia de la Economía Solidaria, en su Título IV – Capitulo II numeral 3 contempla las siguientes definiciones:

3.5.1.1. Riesgo de crédito (RC)

El riesgo crediticio es la probabilidad de que una organización solidaria incurra en pérdidas y disminuya el valor de sus activos como consecuencia del incumplimiento del pago de las obligaciones contractuales por parte de sus deudores o contraparte.

Para propósitos de información, evaluación del RC, aplicación de normas contables y deterioros, entre otras, la cartera de créditos se debe clasificar en las siguientes modalidades:

3.5.1.2. Crédito de consumo

Se entiende por créditos de consumo, independientemente de su monto, los otorgados a personas naturales para financiar la adquisición de bienes de consumo o el pago de servicios para fines no comerciales o empresariales, distintos a los otorgados bajo la modalidad de microcrédito.

3.5.1.3. Crédito comercial

Se define como crédito comercial el otorgado a personas naturales o jurídicas para el desarrollo de actividades económicas organizadas, distintos a los otorgados bajo la modalidad de microcrédito.

3.5.1.4. Créditos de vivienda

Se entiende por créditos de vivienda, independientemente del monto, los otorgados a personas naturales para la adquisición de vivienda nueva o usada, o para la construcción de vivienda individual.

Estas operaciones deben cumplir con las características y criterios señalados en el artículo 17 de la Ley 546 de 1999 y las reglas previstas en los literales b) y c) del artículo 1º del Decreto 145 de 2000 y demás normas que los modifiquen, complementen o deroguen.

3.5.1.5. Microcrédito

Para efectos del presente capítulo, microcrédito es el constituido por las operaciones activas de crédito a las cuáles se refiere el Decreto 2555 de 2010, o las normas que la modifiquen, sustituyan o adicionen, así como las realizadas con microempresas en las cuáles la principal fuente de pago de la obligación provenga de los ingresos derivados de su actividad.

Para los efectos previstos en este capítulo, el saldo de endeudamiento del deudor no podrá exceder de ciento veinte (120) salarios mínimos mensuales legales vigentes al momento de la aprobación de la respectiva operación activa de crédito(40). Se entiende por saldo de endeudamiento el monto de las obligaciones vigentes a cargo de la correspondiente microempresa con el sector financiero y otros sectores, que se encuentren en los registros de los operadores de bancos de datos consultados por el respectivo acreedor, excluyendo los créditos hipotecarios para financiación de vivienda y adicionando el valor de la nueva obligación.

Se tendrá por definición de microempresa aquella consagrada en las disposiciones normativas vigentes.

La cartera de créditos comercial, de consumo y microcréditos, deben clasificarse además teniendo en cuenta la naturaleza de las garantías que las amparan (garantía admisible y otras garantías), acogiéndose a lo dispuesto sobre el particular en el Decreto 2555 de 2010 y las normas que lo adicionen, modifiquen o sustituyan.

Dentro de la metodología interna de cada organización solidaria, las anteriores modalidades pueden subdividirse en líneas de crédito (portafolios).

3.5.1.6. Créditos a asociados, administradores, miembros de juntas de vigilancia y sus parientes

Operaciones con asociados, administradores, miembros de las juntas de vigilancia y sus parientes, a que se refiere el artículo 61, de la Ley 454 de 1998, modificado por el artículo 109, de la Ley 795 de 2003.

3.5.1.7. Vinculados y partes relacionadas

Una parte relacionada o vinculada es una persona o entidad que está relacionada con la organización que prepara sus estados financieros.

En el caso de personas: miembro del personal clave de la gerencia y aquellas que ejercen control o control conjunto o influencia significativa sobre la organización solidaria.

Para el caso de una entidad: subsidiarias, asociadas, o un negocio conjunto, controladora, o cuando la entidad es un plan de

beneficios post-empleo para los trabajadores de la organización que informa.

3.5.2 Tipos de Operaciones de Crédito

El Decreto 4327 de 2005 CAPÍTULO 11.4.9 GESTIÓN DEL RIESGO CREDITICIO estipula que en Colombia existen cuatro tipos de crédito, los cuáles son:

- **Crédito Comercial**

  Otorgado a personas naturales o jurídicas, para el desarrollo de actividades económicas, industriales, comerciales o empresariales organizadas; atendiendo las necesidades de capital para el trabajo.

- **Crédito de Consumo**

  Otorgado a las personas naturales, cuya destinación sea la adquisición de bienes de consumo o pago de servicios, que no se encuentren relacionados con el desarrollo de actividades económicas o empresariales.

- **Crédito de Vivienda**

  Como su nombre lo indica, es el crédito que se otorga únicamente a personas naturales, para la adquisición de vivienda bien sea nueva o usada, la ley 546 de 1999 establece cuáles son las características de este tipo de créditos: **a.** La denominación debe estar en UVR o en la moneda legal. **b.** debe estar amparada en hipoteca en primer grado. **c.** el plazo de amortización debe estar comprendido entre 5 y 30 años. **d.** debe tener tasa de interés remuneratoria que se deberá pactar durante la vigencia del crédito. **e.** el monto del crédito puede ser hasta el 70% del valor del inmueble. f. el valor de la primera cuota no puede ser superior al 30% del valor de los ingresos familiares. **g.** el valor total del crédito podrá ser pagado en su totalidad o parcialmente sin que esto genere penalidades; cuando se realiza un pago parcial será el deudor quien decida si este disminuye el valor de la cuota o el plazo de la obligación. **h.** los inmuebles financiados deben estar asegurados contra incendios y terremotos.

- **Microcrédito**

  Es aquel crédito que se entrega únicamente a personas jurídicas constituidas bajo la figura de microempresa, en la cual, la fuente de ingresos con los cuáles responden por sus obligaciones, es la deriva de su actividad económica; la ley estipula que al momento de otorgamiento del crédito, este no debe superar los ciento veinte salarios mínimos mensuales legales vigentes.

3.5.3 Proceso para la solicitud y desembolso de los créditos

La misma circular 35 de la Superintendencia de la Economía Solidaria, que hemos venido estudiando, trae consigo el procedimiento para el otorgamiento de un crédito, en lo primero que hace hincapié es que este

otorgamiento debe iniciarse con el pleno conocimiento y consentimiento del deudor. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Los pasos establecidos para el otorgamiento del crédito son: Información previa al otorgamiento de un crédito, selección de variables y segmentación de líneas de crédito, perfil del deudor y los criterios mínimos para el otorgamiento de créditos (capacidad de pago, solvencia, consulta a las centrales de riesgo y demás fuentes que disponga la organización solidaria vigilada y la garantía) (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.1 Análisis de Información

La entidad tiene el deber de facilitar el entendimiento de los términos y condiciones del contrato de crédito por parte del futuro deudor, esta información deberá ser veraz, clara suficiente, comprensible y legible, esta deberá ser como mínimo la siguiente: (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

> *"• Monto del crédito.*
> *• Tasa de interés remuneratoria y moratoria expresada en efectiva anual.*
> *• Sistema de amortización*
> *• Plazo de amortización, incluyendo períodos muertos, periodos de gracia, etc.*
> *• Modalidad de la cuota (fija, variable, otras); si la tasa es variable, se debe informar el índice al cual quedará atada su variación y el margen, de igual forma, se deberá informar las implicaciones que tiene la variación de estas tasas en el mercado frente al valor de su cuota y la tabla de amortización del crédito.*
> *• Forma de pago (descuento por nómina, pago por caja, otras).*
> *• Periodicidad en el pago de capital y de intereses.*
> *• Tipo y cobertura de la garantía solicitada.*
> *• Información sobre las condiciones para prepagar la obligación o para realizar pagos anticipados.*
> *• Comisiones, recargos y demás conceptos que se aplicarán en la estimación de la cuota.*
> *• Entregar al asociado el plan de amortización del crédito y poner en su conocimiento el reglamento de crédito.*
> *Al momento del desembolso se deberán indicar los descuentos.*
> *• En caso de créditos reestructurados, se deberá mencionar el número de veces y condiciones propias de la reestructuración. Igualmente deben suministrar al deudor la información necesaria que le permita comprender las implicaciones de estas reestructuraciones en términos de costos, calificación*

crediticia, y los efectos de incumplir en el pago de la obligación.

• En caso de otros tipos de modificaciones de un crédito, se debe suministrar al deudor la información necesaria que le permita comprender las implicaciones de dicha modificación en términos de costos y calificación crediticia, así como un comparativo entre las condiciones actuales y las del crédito una vez sea modificado. Para el efecto deben suministrar como mínimo información respecto de las nuevas condiciones establecidas, los efectos de incumplir en el pago de la obligación bajo las nuevas condiciones, así como el costo total de la operación. Los derechos de la organización solidaria en caso de incumplimiento por parte del deudor.

• Los derechos del deudor, en particular, los que se refieren al acceso a la información sobre la calificación de riesgo de sus obligaciones con la organización solidaria.

• En el caso de los descuentos por libranza, se deberá tener en cuenta el tope máximo señalado en la normatividad vigente, lo cual podrá limitar el monto a otorgar." (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Adicional a esta información la organización solidaria vigilada, deberá entregar toda aquella información que considere sea relevante y necesario para que el futuro deudor tenga una comprensión clara, del alcance de sus derecho y las obligaciones a las que se está comprometiendo; de igual manera, para todo crédito otorgado a partir de agosto del año 2020, se deberá entregar al futuro deudor, información clara, verás, amplia y suficiente, sobre la posibilidad del pago total o parcial de las cuotas o saldos, de cualquier crédito que sea en moneda nacional, sin que se imponga ningún tipo de penalización. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Con respecto de la entrega de información previa al otorgamiento del crédito, es menester de la organización solidaria vigilada, dejar evidencia de la entrega de la misma a través de los formatos o medios que la entidad considere necesario, para lo cual también se podrá hacer uso de las TIC Tecnologías de la Información y la Comunicación; con el debido cumplimiento de la política de manejo de datos y ciberseguridad. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.2 Determinación de factores del crédito y segmentación de líneas de crédito

Cada entidad deberá estipular para cada tipo de crédito que maneje, las variables que le permitan identificar el tipo de crédito que más se ajuste al perfil del futuro deudor y al perfil del riesgo de la organización solidaria. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

La selección e importancia que se le dé a cada una de las variables de discriminación, debe ser un elemento de gran valor al momento de otorgar un crédito, en su seguimiento y como base para su clasificación. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.3 Perfil del deudor

Una vez realizada la selección de variables, según las líneas de crédito, la entidad deberá evaluar las características que debe tener el futuro deudor para acceder al crédito, el cual deberá ser conforme al mercado objetivo. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.4 Criterios mínimos para el otorgamiento de créditos

Por regulación expresa de la Superintendencia de la Economía Solidaria, las organizaciones solidarias vigiladas, deberán tener como mínimo los siguientes criterios de otorgamiento de los créditos. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.5 Capacidad de pago

Esta evaluación es fundamental en el proceso de otorgamiento de un crédito y deberá aplicarse a toda persona natural o jurídica que pueda estar directa o indirectamente obligada a pago de la obligación; esto es, codeudores, avalistas, deudores solidarios, entre otros. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021).

3.5.3.6 Solvencia

Es un análisis realizado por medio del uso de diversas variable, que permitan determinar el nivel de endeudamiento y la calidad de cumplimiento con la obligación, teniendo en cuenta los activos, pasivos, patrimonio y contingencias del deudor. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.7 Consulta Centrales de Riesgo y demás fuentes que disponga la Organización Solidaria Vigilada

En este punto es importante tener conocimiento sobre que son las centrales de riesgo: *"son entidades que almacenan, procesan y suministran información sobre cómo las personas naturales y jurídicas han cumplido con sus obligaciones en entidades como; financieras, cooperativas, almacenes y empresas del sector real."* (datacrédito experian, 2020) Esta consulta se basa en obtener información del deudor sobre el cumplimiento actual y pasado de sus obligaciones-, esta consulta deberá ser autorizada de manera previa por futuro deudo; en ella es de suma importancia evaluar las veces que este ha sido reestructurado y las características de las respectivas reestructuraciones, pues a mayor cantidad de operaciones reestructuradas es mayor el riesgo de no pago de la obligación. (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.5.3.8 Garantías

Son un componente esencial en el proceso de otorgamiento de un crédito, pues son las garantías las que respaldan la obligación en caso del no pago de la misma; estas garantías para que sean idóneas estas deberán contar con un respaldo jurídico eficaz para el pago de la obligación, además de evaluar su naturaleza, idoneidad, liquidez, valor y cobertura (Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

3.6 Cooperativas de Crédito

Este tipo de compañías son diferentes de otras entidades financieras, puesto que son denominadas de economía solidaria, en donde su financiamiento depende de los aportes de sus asociados. Debido a su naturaleza, este tipo de entidades son supervisadas por un ente regulador distinto al de la banca tradicional, denominado "Superintendencia de Economía Solidaria."

Según Vesga, Rafael y Lora, Eduardo una cooperativa de ahorro y crédito, es una entidad sin ánimo de lucro, la cual ofrece servicios financieros en cuanto a ahorro y préstamos; el capital con el cual prestan los servicios financieros, es captado por los aportes voluntarios que realizan sus asociados, que este a su vez se incrementa a medida que se van realizando préstamos con la tasa de interés permitida por el Gobierno Nacional. (Vesga & Lora, 1992)

Los requisitos para la constitución de una cooperativa en Colombia son:

- *"Mínimo 20 asociados.*
- *Constancia del representante legal frente al cumplimiento de las normas especiales del cooperativismo.*

- *El documento de constitución debe estar suscrito por todos los otorgantes o constituyentes.*
- *Su vigencia es indefinida (Circular Externa nro. 8 de 2012 de la Superintendencia de Economía Solidaria).*
- *Certificado de acreditación sobre educación solidaria, expedido por la Unidad Administrativa Especial de Organizaciones Solidarias (Circular Externa nro. 8 de 2012 de la Superintendencia de Economía Solidaria, Decreto 019 de 2012).”* (Camará de Comercio de Bogotá , 2022)

3.7 Superintendencia de la Economía Solidaria

En el año de 1986 se crea el decreto 2536 del 4 de agosto, con este nace el Consejo Nacional de la Economía Solidaria en Colombia; pero solo hasta el año 1988 se crea la ley 79 y con ella determina las formas solidarias del sistema financiero, que son lo que conocemos, como cooperativas, asociaciones mutuales y fondos de empleados. (Superintencia de la Economía Solidaria , 2022)

En el año de 1997, 9 años después de la creación del sistema solidario en Colombia, debido a la informalidad en que venía funcionando el sistema solidario, se evidencia la necesidad de supervisar y es aquí donde se la da vida a DANCOOP el Departamento Nacional de Cooperativas; este se encargaba de definir las políticas, la ejecución de los programas y el control de las cooperativas, las asociaciones mutuales y los fondos de empleados (Superintencia de la Economía Solidaria , 2022)

Con la crisis económica vivida en los años 90 y como el sistema solidario fue clave en ella, se creó la ley 454 de 1998, sancionada por el Presidente Ernesto Samper Pizano; con la cual el DANCOOP se transforma en DANSOCIAL Departamento Administrativo de la Economía Solidaria y este crea la Superintendencia de la Economía Solidaria y el FOGACOOP el Fondo de Garantías del Sector Cooperativo. (Superintencia de la Economía Solidaria , 2022)

La Superintendencia de la Economía Solidaria tiene como naturaleza, ser un organismo descentralizado, técnico, que se encuentra adscrito al Ministerio de Hacienda y Crédito Público, cuenta con personería jurídica y autonomía administrativa y patrimonial; su principal función es ser un ente inspector que vigila y controla las actividades financieras de las cooperativas, las asociaciones mutuales y los fondos de empleados. (Superintencia de la Economía Solidaria , 2022)

3.8 Scoring de Crédito - Cooperativas de Crédito de Medellín

Las Cooperativas de Crédito de Medellín implementan sus score de crédito para la línea de libranzas, a partir de modelos de riesgo predictivos; ésto, lo hacen teniendo en cuenta el bajo nivel de riesgo de crédito que tiene dicha línea, y la cantidad de garantías solicitadas a cada cliente. Adicionalmente, tienen la práctica de depositar la responsabilidad final de la aprobación de dicho crédito a sus analistas de crédito, ya que parten del supuesto que suministran las herramientas e información suficiente a expertos financieros calificados, para que aprueben o no tales operaciones.

Las Cooperativas de Crédito de Medellín fijan sus score de crédito para la línea de libranzas, a partir de los siguientes valores porcentuales:

- El historial de pagos del cliente 35%
- Cantidad adeudada por el cliente al momento del crédito 30%
- Antigüedad del historial crediticio del cliente 15%
- Tipos de cuentas que posee el cliente 10%
- Actividad del crédito reciente del cliente 10%

Las Cooperativas de Crédito en general, presentan dos tipos de riesgos, los empresariales y los financieros. Los primeros corresponden a problemas del sistema económico donde operan, y los segundos son asumidos por los socios de la cooperativa y se clasifican en riesgos de crédito, de liquidez, de mercado, políticos, de inflación, legales y operacionales.

Teniendo en cuenta que el "SARC" es un Sistema de Administración de Riesgo de Crédito que deben implementar las compañías solidarias que son vigiladas en cada país, todas las Cooperativas de Crédito de Medellín han implementado dicho sistema con el fin de identificar, medir, controlar, regular y monitorear el riesgo de crédito de sus operaciones financieras; sin embargo, en el proceso de identificar el perfil de riesgo de sus clientes, éstas han cometido errores como el de apalancar fuertemente su modelo de score de crédito en su línea de libranzas en las consultas a las centrales de riesgo en lugar de implementar modelos de riesgo especializados en la línea de libranzas, han aplicado modelos de riesgo más genéricos de la banca en lugar de potenciar su scoring de crédito en el nicho de mercado cooperativo, han afianzado su modelo de retorno de inversión con negocios con entidades gubernamentales en lugar de abrir nuevas oportunidades de negocio con clientes del sector privado o personas naturales, entre otros, esto ha llevado al sector cooperativo a un par crisis financieras, dentro de las cuales esta la de 1998, originada por el decreto 798 de Enero del 97 en el que el gobierno prohibió a sus entidades los depositos en cooperativas y con el cual se terminaron liquidando más de 28 cooperativas en el país en menos de 2 años.

4. Metodología

Éste es un trabajo final de profundización de la Maestría en Ingeniería Administrativa, en el cual se parte de la experiencia del estudiante, vivida en el sector financiero (bancario y cooperativo). Cabe resaltar que la presente investigacion mantendrá la reserva de información sobre los funcionarios de las Cooperativas de Crédito y Bancos analizados, como cumplimiento a los estándares, lineamientos, políticas y regulaciones con las que operan las entidades financieras en Colombia.

La metodología utilizada en el presente trabajo de investigación es de carácter cualitativo y cuantitativo. Se planteó hacer una revisión sobre como se estàn analizando y aprobando las operaciones de crédito en la línea de libranzas de las Cooperativas de Crédito de Medellin, y sucesivamente desarrollar un modelo para evaluar el scoring de crédito con el que operan dichas cooperativas. Para ello, se validarán 3 modelos de riesgos como son el Logit, Árboles Binomiales y Redes Neuronales, y se implementará el que más se ajuste a un modelo de clasificación binaria basado en la metodología ROC, esto con el fin de predecir cuáles clientes de las Cooperativas de Crédito de Medellín, cumplirán o no con sus obligaciones financieras. Adicional, se identificarán los perfiles de clientes buenos y malos de la línea de libranzas, se describirán las variables que

caractericen el segmento de clientes de las Cooperativas de Crédito de Medellín, se plantearán técnicas estadísticas de clasificación adecuadas para la técnica ROC, se realizará un análisis de la eficacia del método desarrollado y se harán comparaciones generales con las técnicas de medición de riesgos.

4.1 Levantamiento de la información

Se hizo la gestion de recolección y levantamiento de información de las Cooperativas de Crédito de Medellín, así como la de los pagos a sus entidades financieras apalancadores, los bancos que las financian.

Para comenzar cabe mencionar, que para la presentación de esta base de datos se partirá de los datos demográficos de las cooperativas tanto financieras como de ahorro y crédito y multiactivas con sección de ahorro y crédito, que operan en el territorio del valle del aburrá. Si bien tales datos fueron extraídos de una de las entidades financieras analizadas y personalizados para efectos prácticos de la presente investigación, no se revelarán en el presente documento por temas de confidencialidad y reserva bancaria. Para dar mayor veracidad a tales datos, se hizo una revisión sobre los mismos, y se contrasto contra los datos estadísticos que se encuentran en los sitios públicos de los entes financieros reguladores como la Superintendencia de la Economía Solidaria. En este sentido, algunos de los datos son de disponibilidad pública y otros son de carácter confidencial.

4.2 Descripción de la Base de Datos

Como se mencionó anteriormente, este trabajo de investigación parte de los datos obtenidos de una de las entidades analizadas, los cuales corresponden del periodo 2017 al 2019, y contrastados con información de la Superintendencia de Economía Solidaria, de donde se filtró la información de las Cooperativas Financieras, Entidades Especializadas en Ahorro y Crédito y las Multiactivas con Sección de Ahorro y Crédito, que para el mes de mayo de 2022, tenían como domicilio principal algún municipio del Valle de Aburrá:

Variables generales:

1. **Geografía:** Entidades que se encuentren domiciliadas en el Valle de Aburrá
2. **Locación:** Lugar donde se encuentra operando la entidad

Tabla 3: Relación de la muestra

CONCEPTO	CANTIDAD DE COOPERATIVAS POR GEOGRAFÍA Y LOCACIÓN
Cooperativas de Ahorro y Crédito y Multiactivas	39
Municipios donde están ubicadas las cooperativas	5

Fuente: (Superintendencia de la Economía Solidaria, 2022)

Si bien estas cooperativas se encuentran ubicadas en 5 de los 10 municipios del Valle del Aburrá, en la gran mayoría de casos, cuentan con sedes que cubren los demás municipios del Valle de Aburrá e incluso el resto de subregiones del departamento de Antioquia.

De esta muestra general, se puede concluir que estas cooperativas tienen domicilio principal en el 50% de los 10 municipios del Valle de Aburrá; así:

Tabla 4: Cooperativas domiciliadas en el Valle de Aburrá

No	RAZÓN SOCIAL	SIGLA	MUNICIPIO
1	Cooperativa independiente de empleados de Antioquia	CIDESA	Medellín
2	Cooperen, cooperativa de ahorro y crédito		Medellín
3	Cooperativa de ahorro y crédito cootramed	COOTRAMED	Medellín
4	Cooperativa nacional de trabajadores	COOPETRABAN	Medellín
5	Cooperativa especializada de ahorro y crédito orbiscoop	ORBISCOOP	Medellín
6	Cooperativa de trabajadores de las empresas departamentales de Antioquia	COEDA	Medellín
7	Cooperativa antioqueña de trabajadores grupo cafetero	COOAGRUPO	Medellín
8	Cooperativa de trabajadores del SENA	COOTRASENA	Medellín
9	Cooperativa medica de Antioquia Ltda.	COMEDAL	Medellín
10	Cooperativa de ahorro y crédito SERVUNAL	COOSERVUNAL	Medellín
11	Cooperativa especializada de ahorro y crédito universidad de Medellín	COMUDEM	Medellín
12	Cooperativa de ahorro y crédito soycoop	SOYCOOP	Medellín
13	Cooperativa de trabajadores departamentales de Antioquia	COOTRADEPTALES LTDA.	Medellín
14	Cooperativa telepostal Ltda.	TELEPOSTAL	Medellín
15	Cooperativa de profesores de la universidad de Antioquia	COOPRUDEA	Medellín
16	Cooperativa de ahorro y crédito cooyamor	COYAMOR	Medellín
17	Comfamigos cooperativa de ahorro y crédito	COMFAMIGOS	Medellín
18	Cooperativa de ahorro y crédito de empleados del sector financiero	COOEBAN	Medellín
19	AVANCOP cooperativa de ahorro y crédito	AVANCOP	Medellín
20	Cooperativa de empleados suramericana	COOPEMSURA	Medellín
21	Cooperativa San Vicente de Paul Ltda.	COOSVICENTE	Medellín
22	Cooperativa de ahorro y crédito crear Ltda. CREARCOP	CREARCOOP	Medellín
23	Forjar cooperativa de ahorro y crédito	FORJAR	Medellín
24	Cooperativa de ahorro y crédito universitaria bolivariana		Medellín
25	Cooperativa fraternidad sacerdotal Ltda.	COOFRASA	Medellín
26	Cooperativa de ahorro y crédito San Luis	COOSANLUIS	Medellín
27	Cooperativa Belén ahorro y crédito	COBELÉN	Medellín
28	Cooperativa de ahorro y crédito unión colombiana	COMUNIÓN	Medellín
29	Cooperativa de ahorro y crédito COLANTA	AYC COLANTA	Medellín
30	Microempresas de Colombia cooperativa de ahorro y crédito	MICROEMPRESAS DE COLOMBIA A.C.	Medellín
31	Confiar cooperativa financiera	CONFIAR	Medellín
32	Cooperativa financiera JFK	JFK	Medellín
33	Cooperativa financiera de Antioquia	CFA	Medellín
34	Cooperativa financiera empresas publicas	COOFINEP	Medellín
35	Cooperativa financiera COTRAFA	COTRAFA	Bello
36	COOPANTEX cooperativa de ahorro y crédito	COOPANTEX	Bello
37	Cooperativa de TT del colombiano Ltda.	CODELCO	Envigado
38	Cooperativa especializada de ahorro y crédito cooperenka	COOPERENKA	Girardota
39	COOCERVUNION cooperativa de ahorro y crédito	COOCERVUNION	Itagüí

(Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Además de esta información de la ubicación geográfica, es importante conocer la actividad económica de cada uno, se deberá comenzar por mencionar que en el valle de aburrá, existen: 5 Cooperativas Financieras, 31 Cooperativas Especializadas de

Ahorro y Crédito y 3 Cooperativas Multiactivas con Sección de Ahorro y Crédito. Información corroborada con la Supersolidaria.

Las Cooperativas Financieras, representan entonces, un 12.8% de la actividad financiera solidaria, con domicilio en el Valle de Aburrá; estas cooperativas son:

Tabla 5: Cooperativas del Valle de Aburrá que representan 12.8% de actividad

RAZÓN SOCIAL	SIGLA	MUNICIPIO
COOPERATIVA FINANCIERA EMPRESAS PUBLICAS	COOFINEP	Medellín
COOPERATIVA FINANCIERA DE ANTIOQUIA	CFA	Medellín
CONFIAR COOPERATIVA FINANCIERA	CONFIAR	Medellín
COOPERATIVA FINANCIERA JFK	JFK	Medellín
COOPERATIVA FINANCIERA COOTRAFA	COTRAFA	Bello

(Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Para el mes de mayo de 2022, las Cooperativas Especializadas en Ahorro y Crédito, con domicilio en el Valle de Aburrá, son 31, las cuáles representan el 79.4%, estas son:

Tabla 6: Cooperativas del Valle de Aburrá Especializadas en Ahorro y Crédito

RAZÓN SOCIAL	SIGLA	MUNCIPIO
COPERATIVA INDEPENDIENTE DE EMPLEADOS DE ANTIOQUIA	CIDESA	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO COTRAMED	COOTRAMED	Medellín
COOPERATIVA NACIONAL DE TRABAJADORES	COOPETRABAN	Medellín
COOPERATIVA ESPECIALIZADA DE AHORRO Y CRÉDITO ORBISCOOP	ORBISCOOP	Medellín
COOPERATIVA ANTIOQUEÑA DE TRABAJADOR ES GRUPO CAFETERO	COOAGRUPO	Medellín
COOPERATIVA DE TRABAJADORES DEL SENA	COOTRASENA	Medellín
COOPERATIVA MEDICA DE ANTIOQUIA LTDA	COMEDAL	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO SERVUNAL	COOSERVUNAL	Medellín
COOPERATIVA MULTIACTIVA UNIVERSIDAD DE MEDELLÍN	COMUDEM	Medellín
COOPERATIVA DE TRABAJADORES PANAMCO COLOMBIA S.A.	SOYCOOP	Medellín
COOPERATIVA DE TRABAJADORES DEPARTAMENTALES DE ANTIOQUIA	COOTRADEPTALES	Medellín
COOPERATIVA TELEPOSTAL LTDA	TELEPOSTAL	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO COOYAMOR	COMAYOR	Medellín
COMFAMIGOS COOPERATIVA MULTIACTIVA	COMFAMIGOS	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO DE EMPLEADOS DEL SECTOR FINANCIERO	COOEBAN	Medellín
AVANCOP COOPERATIVA DE AHORRO Y CRÉDITO	AVANCOP	Medellín

COOPERATIVA DE EMPLEADOS SURAMERICANA	COOPEMSURA	Medellín
COOPERATIVA SAN VICENTE DE PAUL LTDA	COOSVICENTE	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO CREAR LTDA CREARCOP	CREARCOOP	Medellín
FORJAR COOPERATIVA DE AHORRO Y CRÉDITO	FORJAR	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO UNIVERSITARIA BOLIVARIANA	COOBOLIVARIANA	Medellín
COOPERATIVA FRATERNIDAD SACERDOTAL	COOFRASA	Medellín
COOPERATIVA DE PILOTOS CIVILES DE COLOMBIA	COOPICOL	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO UNION COLOMBIANA	COMUNIÓN	Medellín
COOPERATIVA BELÉN AHORRO Y CRÉDITO	COOBELÉN	Medellín
COOPERATIVA DE AHORRO Y CRÉDITO COLANTA	AYC COLANTA	Medellín
MICROEMPRESAS DE COLOMBIA COOPERATIVA DE AHORRO Y CRÉDITO	MICROEMPRESAS DE COLOMBIA A.C.	Medellín
COOPANTEX COOPERATIVA DE AHORRO Y CRÉDITO	COOPANTEX	Bello
COOPERATIVA DE TT DE EL COLOMBIANO	CODELCO	Envigado
COOCERVUNION COOPERATIVA DE AHORRO Y CRÉDITO	COOSERVUNIÓN	Itagüí
COOPERATIVA DE TRABAJADORES DE ENKA	COOPERENKA	Girardota

(Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Finalmente las Cooperativas Multiactivas con Sección de Ahorro y Crédito, en total son 3 domiciliadas en el Valle de Aburrá, las cuáles representan un 7.8%, en ellas tenemos:

Tabla 7: Cooperativas MultiActivas del Valle de Aburrá

RAZÓN SOCIAL	SIGLA	MUNICIPIO
COOPERATIVA DE EMPLEADOS DE LA REGISTRADURÍA NACIONAL	COOPEREN	Medellín
COOPERATIVA DE TRABAJADORES DE LAS EMPRESAS DEPARTAMENTALES DE ANTIOQUIA	COEDA	Medellín
COOPERATIVA DE PROFESORES DE LA UNIVERSIDAD DE ANTIOQUIA	COOPRUDEA	Medellín

(Superintendencia de la Economia Solidaria "SUPERSOLIDARIA", 2021)

Esta información demográfica, es importante conocerla, ya que con ella se puede dilucidar, cuáles son las entidades que concentran los consumidores solidarios en el Valle de Aburrá, cuáles son sus actividades económicas y que clase de asociaciones son.

Partiendo de los datos expuestos sobre las entidades financieras solidarias del Valle de Aburrá, ahora iremos a lo particular, teniendo como referencia la base de datos simulada a partir de datos históricos de una Cooperativa de Crédito de la Ciudad de Medellín – Antioquia, sin suministrar nombres ni números de documentos, en la cual se encontraron 34 variables entre las cuáles habían cualitativas y cuantitativas.

Tabla 8: Variables de la base de datos - Cooperativas del Valle de Aburrá

Tipo de variable	Descripción de la Variable	Nombre de la Variable
Del cliente	Identificación del cliente.	ID
Del cliente	Valor adeudado por el cliente, calculado teniendo en cuenta el tiempo pactado con la entidad financiera y las fechas de corte. (Para el presente ejercicio oscila entre $19 y $56.549.783)	Plazo
Del cliente	Indicador de moras del cliente	Incumplimiento
Del cliente	Ahorro hecho por los clientes, a parte de sus obligaciones financieras adquiridas.	Aportes
Del cliente	Estado civíl actual del cliente. (Para el presente ejercicio se tomarán los siguientes estados: Casado, Eclesiástico, Separado, Soltero, Unión Libre y Viudo. Conforme el estado de cada cliente, ésta variable tomará el valor de 1 en el modelo y los demás estados se tomarán en 0)	Estado civíl
Del cliente	Género del deudor	Sexo
Del cliente	Cantidad de personas que dependen económicamente del cliente. (Para el presente ejercicio oscila entre 0 y 25)	Personas a cargo
Del cliente	Tipo de vivienda en la que habita el cliente: Propia, Familiar, Arrendada.	Tipo de vivienda
Del cliente	Tipo de contrato laboral que tiene el cleinte conforme su actividad económica: A término fijo, Por prestación de servicios, Pensionado, A término indefinido.	Tipo de contrato
Del cliente	Cantidad de días de retraso en el pago de obligaciones por parte del cliente (Para el presente ejercicio oscila entre 0 y 1.784 días calendario)	Días de mora
Del cliente	Edad del cliente en años (Para el presente ejercicio oscial entre 19 y 83)	Edad
Del cliente	Actividad a la que se dedica el cliente. (Para el presente ejercicio serán: Ama de casa, Desempleado, Empleado, Estudiante, Independiente, Pensiondo y Jubilado)	Ocupación
Del cliente	Nivel de estudios del cliente (Para el presente ejercicio será: Ninguno, Primaria, Bachillerato, Técnico, Tecnólogo, Universitario, Posgrado)	Nivel educativo
Del cliente	Cantidad de ingresos totales mensuales del cliente. (Para el presente ejercicio oscilan entre $0 y $15.100.000)	Ingresos
Del cliente	Gastos o salidas de dinero totales mensuales que tiene el cliente. (Para el	Egresos

	presente ejercicio oscila entre $0 y $11.800.000)	
Del cliente	Estrato socioeconómico del cliente. (Para el presente ejercicio oscila entre 1 “bajo-bajo”, 2 “bajo”, 3 “medio-bajo”, 4 “medio”, 5 “medio-alto” y 6 “alto”	Estrato
Del cliente	Cantidad de tiempo en años del cliente en su trabajo. (Para el presente ejercicio oscila entre 0 y 46)	Antigüedad laboral
Del modelo	Categoría de riesgo de cada cliente según su historial de pagos, las cuales están descritas en la circular externa 029/05/2007 de la Superfinanciera, y se citan 5 categorías: - **A “Riesgo Normal”:** Créditos atendidos oportunamente. Deudor con capacidad de pago. - **B “Riesgo Aceptable”:** Créditos atendidos apropiadamente. Sin embargo, se evidencian debilidades que pueden afectar la capacidad de pago del deudor, y suesivamente, se pueden causar pagos extemporáneos fuera de lo pactado con las entidades prestantes. **- C “Riesgo Apreciable”:** Créditos atendidos inapropiadamente. Deudor con problemas en su capacidad de pago. **- D “Riesgo Significativo”:** Créditos atendidos inapropiadamente. Deudor con problemas mayores en su capacidad de pago. No se asegura que se recauden oportunamente ni el capital ni los intereses del crédito. **- E “Riesgo de Incobrabilidad”:** Créditos atendidos inapropiadamente. Deudor sin capacidad de pago. El crédito es incobrable.	Categoría
Del modelo	Valor de la cuota pactada entre el deudor y la entidad financiera	Cuota
Del modelo	Tasa de interés mensual vencida que se aplica al crédito. (Para el presente ejercicio oscila entre el 0,16% y el 3,30%)	Tasa de interés
Del modelo	Valor de dinero asignado al cliente. (Para el presente ejercicio oscila entre 101.000 y $61.800.000)	Monto

Del modelo	Tiempo en meses durante el cual está vigente el crédito. (Para el presente ejercicio este tiempo puede aumentar si el cliente entra en mora)	Tiempo de maduración
Del modelo	Cantidad de créditos que ha tenido el cliente con las entidades financeras, tanto vigentes como cancelados.	Cantidad de créditos por cliente
Del modelo	Tiempo en años de antigüedad del cliente en la entidad financiera (Para el presente ejercicio oscila entre 0 y 49)	Antigüedad
Del modelo	Respaldo dado por el cliente antes la entidad financiera para avalar su obligación, el cual puede ser de tipo personal (avalista o codeudor), o real (inmueble, propiedad, bien)	Garantía
Del modelo	Modalidad de pago de las cuotas de la obligación por parte del cliente: Por ventanilla, Descuento por nómina.	Forma de pago
Del modelo	Modificación de las condiciones iniciales del crédito con el fin de asegurar los pagos completos y de forma oportuna por parte del cliente (Para el presente ejercicio será Credito Reestructurado o Crédito Sin Reestructuración)	Reestructurado
De la entidad financiera	Lugar donde se gestionó el crédito. (La entidad analizada tiene un total de 9 oficinas en el valle de Aburrá. La oficina donde se gestionó el crédito tomará el valor de 1 en el modelo para cada caso. Las demás tomarán el valor de 0)	Sucursal
De la entidad financiera	Tipo de sucursal donde se gestionó el crédito. (Para el presente ejercicio será: Grande, Mediana, Pequeña)	Tipo de sucursal
De la entidad financiera	Cantidad de clientes que tiene la sucursal donde se gestionó el crédito. (Para el presente ejercicio oscila entre 100 y 10000)	Cantidad de clientes
De la entidad financiera	Cantidad máxima que puede aprobar la sucursal donde se gestionó el crédito. (Para el presente ejercicio oscila entre $100.000 y $70.000.000)	Topes de crédito
De la entidad financiera	Cantidad de presupuesto asignado por sucursal para colocarlos en créditos de libranza.	Cupos de crédito para libranzas
De la entidad financiera	Nivel de riesgo externo al que se encuentra expuesta la sucursal donde se gestionó el crédito	Riesgo por ubicación de la sucursal
De la entidad financiera	Valor de la póliza que asegura la cantidad de efectivo disponible en la sucursal donde se gestionó el crédito.	Deducible de asegurabilidad

4.3. Depuración y transformación de la base de datos:

Para realizar la depuración de la base de datos, se analizaron las primeras bases de datos recolectadas, en donde se depuró la data histórica del 2018 menos (-1), es decir, se descartó trabajar con data del 2016 o años anteriores, con el fin de determinar si el fenomeno se presentó explicitamente en el año 2018, o por el contrario, comenzó en el 2017 y se agudizó en el 2018.

Para realizar la transformación , se validaron todas las variables y registros/clientes, de las bases de datos levantadas, quedando con una gran base de datos con 34 variables y 200.000 registros/clientes aproximadamente, de las cuales, se filtraron y se determinaron las variables objeto del modelo, quedando finalmente 6 variables y 148.670 clientes.

4.4. Construcción de modelos de evaluación de crédito:

Para poder plantear modelos óptimos que evalúen los scoring de crédito en las Cooperativas de Crédito de Medellín, es necesario entender cuales son los modelos de riesgo que ellas más utilizan. De acuerdo a la información levantada, las entidades cooperativas de la muestra utilizan varios modelos estadísticos y econométricos similares, con los que identifican los riesgos de sus procesos de crédito; entre los más utilizados están los de tipo logístico "Logit", que permiten predecir de mejor manera la probabilidad del no pago de sus clientes, y los "Probit", que permiten hacer estimaciones de valores efectivos para diferentes tasas de respuesta, ambos, en lugar de utilizar muchas regresiones logísticas, las cuales se enfocan más en mostrar las razones de probabilidad para variables independientes.

De acuerdo a lo anterior, todas las entidades analizadas presentaron ciertos atrasos en el pago a sus financiadores, debido a que sus libranzas entraron en mora; conforme a esto, se procedió a construir tres modelos score, con los cuales se ejecutaron las corridas de pruebas en diversas herramientas estadísticas como R y Rstudio. Luego, se filtraron los datos levantados con las siguientes métricas y variables dependientes, donde se estableció un puntaje mínimo de 600 puntos para las personas solicitantes de un crédito en la línea de libranzas, que su edad no sea mayor a 65 años para hombres y 62 para mujeres, que cuenten con 18 años, que haya tenido vida creditica previamente y que su empleador tenga convenio de nómina con la entidad bancaria donde se solicite el crédito.

A continuación los 3 modelos utilizados en la presente investigación. Como paso previo mediante la siguiente sentencia, se hizo el cargue de la base de datos depurada en la plataforma **R**, la cual contiene 148.670 regristos/clientes con 34 variables:

```
# https://www.kaggle.com/datasets/yasserh/loan-default-dataset

rm( list = ls())

setwd("~/Downloads")

d <- rio::import("Loan_Default.csv") table(d$Status)

nn <-
c("ID","year","open_credit","construction_type","Secured_by","Security_Type"
)
```

```
fil <- names(d)[!(names(d) %in% nn)]

dd <- d[,fil]
m1 <- glm(Status~Credit_Score,data = dd, family = binomial(link= "logit"))
m2 <- glm(Status~Credit_Score+LTV+income,data = dd, family = binomial(link=
"logit"))

library(pROC)
```

Esto arrojó una tabla de contingencia, la cual se utilizará en los tres modelos descritos y relaciona con (1) los clientes objeto de crédito de libranza, y con (0) los que no:

0	1
112031	36639

Para determinar que variables incluir en los ejercicios de simulación de la presente investigación, se aplicó el modelo Credimetrics a los datos levantados, a fin de poder medir el riesgo del portafolio de los créditos de libranza:

- Se calificó la data obtenida
- Se estableció un periodo de tiempo a medir
- Se desarrolló un modelo de valoración de la información
- Se analizaron los cambios en el valor de la cartera de libranzas
- Se definió como incumplimiento, al momento en el que el valor de los activos están por debajo del valor nominal, en los créditos de libranza.

4.4.1. Aplicación de Modelos

4.4.1.1. Modelo Logit

Se creo un modelo de elección binaria en R Studio, el cual hace una distribución acumulada estándar, y predice como falso o verdadero el cumplimiento de las obligaciones contraídas por los clientes de la base de datos de las Cooperativas de Crédito de Medellín, también cargada en R Studio, previamente:

```
# LOGIT

logit_P1 <- predict(m1 , newdata = dd ,type = 'response' ) roc_score1 <-
roc(dd$Status, logit_P1) #AUC score roc_score1$auc #AUC score
plot(roc_score1)

roc_score1
corte1 <- coords(roc_score1,"best")[[1]] tt1 <-
table((logit_P1>corte1)*1,dd$Status) caret::confusionMatrix(tt1)

ks.test(logit_P1[dd$Status==1],logit_P1[dd$Status==0])

logit_P2 <- predict(m2 , newdata = dd ,type = 'response' ) roc_score2 <-
roc(dd$Status, logit_P2) #AUC score plot(roc_score2)
logit_auc <- roc_score2$auc
```

```
corte2 <- coords(roc_score2,"best")[[1]] tt2 <-
table((logit_P2>corte2)*1,dd$Status) caret::confusionMatrix(tt2)

logit_ks <- ks.test(logit_P2[dd$Status==1],logit_P2[dd$Status==0])$statistic
```

Se puso a prueba el set de datos levantado. Con este modelo se determina que tanto se ajustan la operación de crédito como tal, con los score de crédito genéricos con los que operan las Cooperativas de Crédito de Medellín, y como resultado se obtuvo:

Area under the curve: 0.5027: Teniendo en cuenta que resultado es superio a 0.5, esta predicción demuestra que el modelo esta dentro de la curva AUC, con lo que se corrobora que el modelo esta afinado y operando correctamente.

Call:
roc.default(response = dd$Status, predictor = logit_P1)

Data: logit_P1 in 112031 controls (dd$Status 0) < 36639 cases (dd$Status 1). Area under the curve: 0.5027

Aquí se demuestra que demuestra que los clientes seleccionados son fiables de hacerles un estudio de crédito.

Confusion Matrix and Statistics

MATRIZ	0	1
0	88204	28545
1	23827	8094

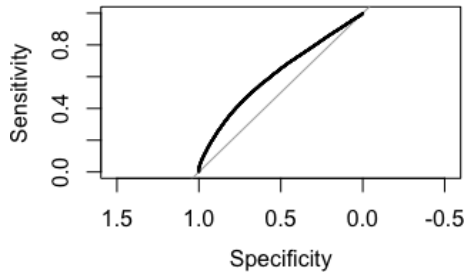
Aquí se demuestra que de los clientes seleccionados, el modelo predice que 88.204 son los cumplirán con sus obligaciones oprortunamente, que 28.545 y 23.827 se les debe hacer validaciones adicionales al scoring normal y que 8.094 deben ser descartados.

Accuracy	**0.6477**
95% CI	**(0.6453, 0.6502)**
No Information Rate	**0.7536**
P-Value [Acc > NIR]	**1**
Kappa	**0.0086**
Mcnemar's Test P-Value	**<2e-16**
Sensitivity	**0.7873**
Specificity	**0.2209**
Pos Pred Value	**0.7555**
Neg Pred Value	**0.2536**
Prevalence	**0.7536**
Detection Rate	**0.5933**
Detection Prevalence	**0.7853**
Balanced Accuracy	**0.5041**
'Positive' Class	**0**

Aquí se demuestra que el modelo construido, luego de las corridas de prueba, es confiable un 64%.

Asymptotic two-sample Kolmogorov-Smirnov test
data: logit_P1[dd$Status == 1] and logit_P1[dd$Status == 0]
D = 0.0082299, p-value = 0.0475
alternative hypothesis: two-sided

Aqui se aplica el test de KS y se comprueba el nivel de sensibilidad del modelo ante fallos.



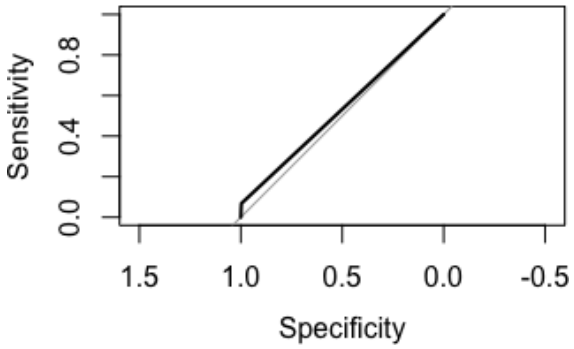
4.4.1.2. Modelo Bosque Aleatorio – Árboles de decisión

A continuación, se describe el Modelo de Bosque Aleatorio planteado para el presente trabaajo, a partir de la data cargada previamente. Este se construyó con algoritmos de machine learning con un mecanismo de aprendizaje supervisado, el cual se sometió dicho set de datos, para validar la cantidad de árboles de decisión y variables a las que se ajusta la data proporcionada, y así identificar si los score de credito de las cooperativas de credito de medellin, le pueden estar dando mas peso a las variables cualitativas que cuantitativas.

A continuación se muestra el test del Bosque Aleatorio – Arboles de Decisión, a partir de los resultados obtenidos en el modelo anterior. Los árboles de decisión también se ajustan a los datos cargados al principio y se ajustan correctamente al análisis que se esta realizando:

MATRIZ	**0**	**1**
0	**73642**	**10703**
1	**30476**	**9616**

Accuracy	**0.6691**
95% CI	**(0.6665, 0.6717)**
No Information Rate	**0.8367**
P-Value [Acc > NIR]	**1**



4.4.1.3. Modelo Redes Neuronales

A continuación se demuestra el nivel de predicción del modelo de redes neuronales:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3057	**0.6957**	**0.7047**	**0.6975**	**0.7072**	**0.7075**

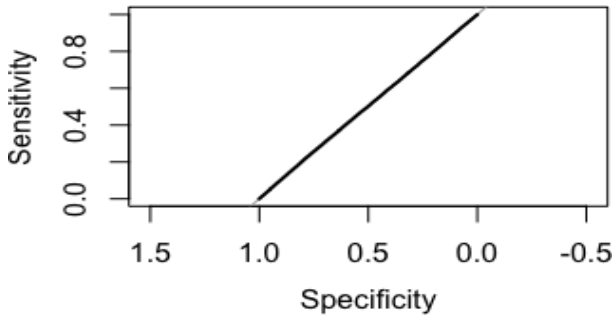
A continuación se describen las variables con las que opera de una forma más ajustada las redes neuronales para el presente ejercicio:

[1] "l0 ~ Credit_Score + LTV + income"

A continuación se detalla a matriz donde a partir de la tabla de contingencia, se especifican que 104.118 clientes son objeto de crédito, y 20.319 no lo son:

0	1
104118	**20319**

Donde se comprueba que aunque si bien los valores resultantes estan por muy por encima de 0.5 como es la media evaluativa de estos modelos de scoring, no sobre pasa el umbral de 1, y se comprueba que el modelo opera correctamente con los valores cargados al principio.



Finalmente, se sometió el set de datos levantado, ante el modelo de redes neuronales construído, para determinar si las opreraciones de crédito en la línea de libranzas de las cooperativas de medellin, está más enfocado en aprender de ciertos esquemas o proceso de crédito, que a evaluar cada solicitud de cre1dito de forma independiente y sin asumisiones. Esto con el fin de determinar que tan personalizadas son las solicitudes de créditos de libranza

de las Cooperativas de Crédito de Medellín, y que nivel de riesgo realmente están asumiendo o no.

4.4.1. Medición de desempeño de los modelos

Para medir el desempeño de los 3 modelos planteados, fue necesario desarrollar un método de clasificación binaria basado en la metodología ROC, con el cual se comparon tales modelos y se obtuvieron los siguientes resultados:

- Primero se aplicó el modelo de **Redes Neuronales** al método de evaluación planteado, y se identificó que las operaciones de crédito en la línea de libranzas de las Cooperativas de Medellín no le dan un peso adecuado a las variables "edad", "estado civil" y "antigüedad" de los clientes, por lo cual pueden poner en riesgo el pago de las obligaciones, debido a que tales variables son altamente variables en periodos de tiempo muy cortos.
- Luego, se aplicó el modelo de **Bosque Aleatorio**, con el cual se establecieron varios árboles de decisión y se evidenció que éste da más peso a las variables cualitativas que cuantitativas, y por lo tanto, no se ajusta mucho al método que se busca implementar en el proceso de asignación de créditos de libranza.
- Por último, se aplicó el modelo **Logit**, con el cual se obtuvo una distribución binomial mas ajustada a 1, es decir, predice de mejor forma el comportamiento de pago de los clientes.

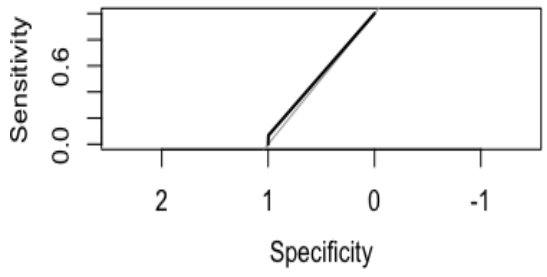
4.4.1.1. Validación de los 3 modelos

Conforme los tres modelos probados anteriormente, a continuación se muestran los resultados obtenidos por una prueba ejecutados entre estos:

```
> AUC <- c(logit_auc,tree_auc,nn_auc)
> KS <- c(logit_ks,tree_ks,nn_ks)
> data.frame(AUC,KS)
```

MATRIZ	AUC	KS
1	0.6138071	0.18054528
2	0.5325425	0.06505784
3	0.6250494	0.19222607

Aquí se observa, que si bien los tres modelos se ajustan al proceso para evaluar el riesgo de las solicitudes de solicitudes de las Cooperativas de Crédito de Medellín, en la línea de libranzas, finalmente el modelo más predictivo de los 3 analizados, es el LOGIT, debido a su nivel de sensibilidad mostrada en las pruebas.



5. Análisis de resultados

A nivel general, y luego de las corridas de pruebas hechas junto con los análisis realizados se estableció que:

- Las Cooperativas de Medellín fijan sus score de crédito en modelos de riesgo predictivos los cuales les permiten acelerar sus procesos de desembolso. Sin embargo, esta práctica los está poniendo a riesgos latentes de liquidez, debido a que están asumiendo que los clientes del segmento de libranza siempre seguirán teniendo el mismo comportamiento de pago, lo cual realmente es incierto.

- El Sector Cooperativo de Medellín en sus procesos de análisis de crédito, da un poco más de peso al juicio de experto que puedan tener sus responsables del proceso de desembolso, esto se hace con el fin de alivianar un poco más (en tiempo) los procesos que componen las operaciones de crédito en la línea de libranzas. Este esquema puede exponer a las Cooperativas de Crédito a riesgos de tipo interno relacionados con el conflicto de intereses, por asumisiones de parte de los ejecutores de los desembolsos.

- Con las corridas de pruebas hechas sobre los modelos testeados (ver resultados en la parte inferior del presente documento), se logró identificar que las operaciones de crédito de las Cooperativas de Medellín, basan en gran medida su nivel de riesgo para los desembolsos, en modelos predictivos binarios como el Logit.

- Los créditos de libranza que más entraron en mora en los periodos analizados, tienen la particularidad que fueron solicitados por clientes de edades cortas, por montos altos, y que no contaban con una buena antigüedad en su historial crediticio. Con esto se podría deducir que a pesar de que tales clientes presentaron las respectivas garantías cuando solicitaron sus créditos, estos no mitigaron el riesgo de tener una cartera sana.

5.1. Elección Del Mejor Modelo

Luego de analizar la situación financiera del Sector Cooperativo a nivel general, de validar la hipótesis planteada sobre el porqué de los incumplimientos de pago por parte de las Cooperativas de Crédito de Medellín ante sus entidades apalancadoras (los bancos), de analizar la información obtenida y de evaluar los score de credito con los que operan las Cooperativas de Crédito de Medellín, se estable que el mejor modelo para evaluar las solicitudes de créditos de libranza en dichas cooperativas, es el LOGIT, puesto que muestra el mejor desempeño AUC (Area bajo la curva).

6. Conclusiones

La primera conclusión que podemos tener del presente trabajo, es que el riesgo de crédito, debemos entenderlo como aquella probabilidad de que una de las partes de un contrato incumpla su obligación, para que así, podamos identificar posibles problemas de insolvencia o incapacidad de pago.

Para el caso en concreto de Colombia, tenemos que la normativa que se aplica a las entidades financieras, posee un orden jerárquico, partiendo desde la Constitución Política de 1991, como el primer nivel, hasta los decretos con fuerza de ley que crea el Presidente de la República, que los encontramos en el quinto nivel; sin embargo a nivel mundial encontramos el Comité de Supervisión Bancaria de Basilea, el cual se encarga de la regulación prudencial de los bancos y genera recomendaciones que pueden ser o no, de obligatorio cumplimiento; a través del tiempo el Comité de Basilea ha creado una serie de acuerdos, los cuáles han ido atendiendo y cambiando según las necesidades humanas del momento; el acuerdo de Basilea I se crea en el año de 1988, su objetivo fue unificar criterios sobre la administración de riesgos de las entidades financieras a nivel mundial; el acuerdo de Basilea II se aprobó en 2004 pero se aplicó en todos los países en el año 2013 el presenta tres pilares para el buen funcionamiento de la regulación de capital y finalmente el acuerdo de Basilea III trajo varias reformas con el fin de reforzar las normas sobre el capital y la liquidez.

Los modelos scoring o también llamados score-cards o classifiers, fueron creados en la década de los 70`s pero sólo hasta la década de los 90`s tomaron fuerza, debido al Acuerdo de Basilea II; estos modelos son herramientas estadísticas muy útiles para predecir la probabilidad de incumpliendo de un futuro deudor, teniendo en cuenta sus características cualitativas y cuantitativas, se analizaron tres modelos, estos fueron: análisis discriminante; modelo probabilístico; modelo logístico.

El proceso de transformación financiera en Colombia tuvo lugar a finales de la década de los 80`s y principios de los 90`s; pasando de ser una banca especializada a ser una banca constituida por matrices bancarias, la transformación financiera trajo consigo una serie de reformas estructurales, cuyo pilar fundamental fue la liberación de los mercados; la respuesta que tuvo Colombia ante esta transformación financiera, fue la creación de la ley 45 de 1990; en la que se da una reorganización a las instituciones financieras y los entes que los vigilan y controlan las diferentes entidades; la primera de estas instituciones es el Ministerio De Hacienda Y Crédito Publico; el cual fue creado desde 1866, pero sólo hasta el año de 1923 recibió el nombre con el cual se conoce hoy y su principal función es el control de mercados capitales, políticas bancarias y control de la balanza de pagos, entre otros; seguida del Ministerio tenemos la Superintendencia Financiera De Colombia ella al estar adscrita al Ministerio se encarga de inspeccionar, vigilar y controlar las entidades financiera; finalmente encontramos el SARC o Sistema De Administración De Riesgo Crediticio, que es aquel sistema que deben adoptar y complementar todas aquellas entidades financieras solidarias, como las Cooperativas Financieras, las Cooperativas Especializadas en Ahorro y Crédito y las Cooperativas Multiactivas.

Encontramos que en Colombia, el ciclo que debe cumplirse para el otorgamiento de un crédito se compone de cuatro pasos, los cuáles se encuentran establecidos por la Superintendencia de la Economía Solidaria; estos son: información previa al otorgamiento de un crédito, selección de variables y segmentación de líneas de crédito, perfil del deudor, criterios mínimos para el otorgamiento de créditos (capacidad de pago, solvencia, consulta a las centrales de riesgo y demás fuentes que disponga la organización solidaria vigilada, garantías).

Con respecto de la metodología, lo primero que se realizó una base de datos demográfica de aquellas entidades vigiladas que reportan sus actividades a la Superintendencia de la Economía Solidaria, que tiene como domicilio principal el Valle de Aburrá, encontrando que existen en total 39 entidades,

entre las cuáles 5 son Cooperativas Financieras; 31 son Cooperativas Especializadas de Ahorro y Crédito y 3 son Cooperativas Multiactivas con Sección de Ahorro y Crédito, estas se encuentran con domicilio principal en 5 de los 10 de los municipios del Valle de Aburrá, estos son: Medellín, Bello, Envigado, Itagüí y Girardota.

Se tuvo como referencia la data histórica de las Cooperativas de Crédito de la ciudad de Medellín – Antioquia; en la cual se tuvo en cuenta una base de datos que ellos construyeron a partir del conocimiento previo de una base de datos que una entidad financiera de la ciudad de Medellín les suministro para la realización de su investigación; posterior a ellos y luego de realizar validación de datos con la prueba de Kolmogorov-Smirnov; se determinó que el mejor modelo para diseñar un modelo de scoring al momento de otorgar un crédito es la regresión logística del Modelo LOGIT, pues los resultados que hallaron fue que el modelo erro en solo 40 casos de 9.957 lo que representa un 0,81% lo que en realidad quiere decir que el modelo es asertivo en el 99,19% de los casos

Al basarnos en dicha data histórica, se procesaron modelos y técnicas estadísiticas que dan respuesta al presente trabajo de investigación; toda vez que este busca encontrar el método mas efectivo para la evaluacion del Scoring de crédito en la linea de libranza de las 39 cooperativas que se encuentran en los municipios del Valle de Aburrá y este es precisamente la regresión logística, pues este por medio de la base de datos con las variables idóneas, le permite a la cooperativa conocer al futuro deudor, posterior a esto, el método le asigna un valor binominal a cada una de esas variable, permitiendo así cuantificar el riesgo, es decir, determinar la probabilidad de incumplimiento del futuro deudor, teniendo en cuenta variables como el valor solicitado, el valor de la cuota mensual, la solvencia del futuro deudor, entre otras; que pueden aumentar o disminuir la probabilidad de incumpliento.

A continuación, se describen las citas bibliográficas utilizadas en el presente trabajo:

6.1 Citas

MEDINA, R. S. (2008). EL RIESGO DE CREDITO EN EL MARCO DEL ACUERDO BASILEA II . SEVILLA : DELTA PUBLICACIONES .

Grupo BBVA. (2015). *Informe con Relevancia Prudencial 2015*. España.

Decreto 2555 de 2010 - Gestor Normativo - Función Pública

Superintendencia Financiera de Colombia. (4 de Abril de 2008). Obtenido de chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://fasecolda.com/cms/wp-content/uploads/2019/08/ce100-1995-cap-ii.pdf

ATLAX 360 . (14 de Agosto de 2020). *Los tipos de riesgo de crédito y la importancia de su gestión*. Obtenido de https://www.atlax360.com/es-ES/blog/los-tipos-de-riesgo-de-credito-y-la-importancia-de-su-gestion/

ASOBANCARIA. (2021). *Normatividad, decretos, resoluciones y leyes que rigen el sector*. Obtenido de https://www.asobancaria.com/normatividad/

STEVENS, R. (11 de FEBRERO de 2020). *RANKIA*. Obtenido de https://www.rankia.co/blog/analisis-colcap/3556208-superintendencia-financiera-colombia-funciones-historia-certificados

BANCO DE ESPAÑA. (2022). *Erosistema* . Obtenido de El Comité de Supervisión Bancaria de Basilea (BCBS): https://www.bde.es/bde/es/areas/supervision/actividad/BCBS/El_Comite_de_Su_13e462eab2e4961.html#

Torres Avendaño, G. (2005). El Acuerdo de Basilea: Estado del Arte del SARC en Colombia . *AD-MINISTER Universidad EAFIT* , 114-134 .

PowerData. (12 de Julio de 2013). *PowerData*. Obtenido de https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/307125/qu-son-los-acuerdos-de-basilea-basilea-i-basilea-ii-y-basilea-iii

Comite de Supervisión Bancaria de Basilea. (2004). Aplicación de Basilea II: aspectos prácticos . *BANCO DE PAGOS INTERNACIONALES* , 1-40.

Gonzales Cervantes, F., & Zornoza Batiz, O. (2006). Basiela II: una herramienta y tres pilares para un reto. *Estrategia Financiera, n°226*, 64 - 68 .

Comite de Supervisión Bancaria de Basilea . (2010). Basilea III: Marco regulador global para reforzar los bancos y sistemas bancarios . *Comite de Supervisión Bancaria de Basilea* , 1 - 2.

Banco Bilbao Vizcaya Argentaria. (20 de Febrero de 2017). *BBVA* . Obtenido de https://www.bbva.com/es/economia-todos-basilea-iii/

Saavedra Garcia, M. L., & Saavedra Garcia, M. J. (2010). Modelos para Medir el riesgo de crédito de la banca. *Cuadernos de administracion* , 295 - 319.

Gutierrez Girault, M. A. (2007). Modelos de Credit Scoring. *Munich Personal RePEc Archive*.

Villegas, F. (s.f.). *Modelos Clasicos*. Obtenido de Modelo Probabilistico: https://sites.google.com/site/modelosclasicosri/probabilistico

Fernández Castaño, H., & Pérez Ramírez, F. O. (2005). El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito. *Revista Ingenierías Universidad de Medellín, vol. 4, núm. 6*, 55-75.

https://www.eltiempo.com/archivo/documento/MAM-818687

Arango, M. (2006). Evolución y crisis del sistema financiero colombiano. *Series Estudios y perspectivas*(11), 1 - 101.

Zuleta, L. (1997). REGULACIÓN Y SUPERVICIÓN DE CONGLOMERADOS FINANCIEROS EN COLOMBIA. *SERIE FINANCIAMIENTO DEL DESARROLLO* , 5, 13, 21.

Ministerio de Hacienda y Crédito Público. (2022). *Ministerio de Hacienda y Crédito Público*. Obtenido de https://www.minhacienda.gov.co/webcenter/portal/AcercadelMinisterio/pages_home

MisAbogados.com.co. (02 de Junio de 2016). *MisAbogados.com.co*. Obtenido de ¿Qué es la Superintendencia Financiera de Colombia?: https://www.misabogados.com.co/blog/que-es-la-superintendencia-financiera-de-colombia

datacrédito experian. (Septiembre de 2020). *¿Qué son las centrales de riesgo y cómo funcionan?* Obtenido de https://www.datacreditoempresas.com.co/blog-datacredito-empresas/que-son-las-centrales-de-riesgo-y-como-funcionan/

Vesga, R., & Lora, e. (1992). *LAS COOPERATIVAS DE AHORRO Y CREDITO EN COLOMBIA: INTERMEDIACION FINANCIERA PARA SECTORES POPULARES* . Bogotá D.C: FEDESARROLLO.

Camará de Comercio de Bogotá . (2022). Obtenido de Cooperativas, fondos de empleados y asociaciones mutuales: https://www.ccb.org.co/Inscripciones-y-renovaciones/Fundaciones-asociaciones-y-corporaciones/Cooperativas-fondos-de-empleados-y-asociaciones-mutuales

Superintencia de la Economía Solidaria . (2022). *Superintencia de la Economía Solidaria* . Obtenido de Nuestra entidad : https://www.supersolidaria.gov.co/es/nuestra-entidad/resena-historica#:~:text=En%201986%20se%20adopta%20el,en%20el%20entorno%20econ%C3%B3mico%20nacional.

Superintendencia de la Economía Solidaria. (Mayo de 2022). *Entidades Vigilidas que reportan informacion 2022.* Obtenido de Estados Financieros de Entidades Solidarias: https://www.supersolidaria.gov.co/es/content/entidades-vigiladas-que-reportan-informacion-2022

Arango, L., & Restrepo, D. (2017). DISEÑO DE UN MODELO DE SCORING PARA EL OTORGAMIENTO DE CREDITO DE CONSUMO DE UNA COMPAÑIA DE FINANCIAMIENTO COLOMBIANA. Medellín: UNIVERSIDAD EAFIT.

Saldaña, M. (2016). Pruebas de bondad de ajuste a una distribucion normal . *Enfermería del Trabajo* , 36-45 .

De La Fuente, S. (2011). ANÁLISIS DISCRIMINANTE. *Instrumentos Estadísticos avanzados Facultad Ciencias Económicas y Empresariales Departamento de Economia Aplicada*.

Superintendencia de la Economia Solidaria "SUPERSOLIDARIA". (29 de Diciembre de 2021). Circular Externa #35 de 2021. *Circular basica contable y fanaciera* . Bogota D,C, Colombia .

Louzada F, A. A. (1 de 12 de 2026). *sciencedirect.* Obtenido de Classification methods applied to credit scoring: systematic review and overall comparison. Surv. Oper. Res. Mana.g Sci. 2016;21(2):117–134.: https://doi.org/10.1016/j.sorms.2016.10.001.

Markov, A., Seleznyova, Z., & Lapshin, V. (1 de 08 de 2022). *sciencedirect.* Obtenido de Credit scoring methods: Latest trends and points to consider. J. Financ. Data Sci. 2022, 8, 180–201.: https://doi.org/10.1016/j.jfds.2022.07.002

Louzada F, Ara A, Fernandes GB. Classification methods applied to credit scoring: systematic review and overall comparison. Surv. Oper. Res. Mana.g Sci. 2016;21(2):117–134. https://doi.org/10.1016/j.sorms.2016.10.001.

Markov, A.; Seleznyova, Z.; Lapshin, V. Credit scoring methods: Latest trends and points to consider. J. Financ. Data Sci. 2022, 8, 180–201. https://doi.org/10.1016/j.jfds.2022.07.002

6.2 Anexos

6.2.1 Figuras

6.2.2 Tablas

6.2.3 Gráficos

6.2.4 Fórmulas

6.2.5 Soportes y evidencias

- **Modelo Estadístico – LOGIT**

    - **MODELACION**

```
# https://www.kaggle.com/datasets/yasserh/loan-default-dataset
rm( list = ls())
setwd("/cloud/project/Data")
d <- read.csv("/cloud/project/Data/Loan_Default.csv", header = TRUE, sep = ",")
table(d$Status)
d <- rio::import("Loan_Default.csv")
table(d$Status)
nn <- c("ID","year","open_credit","construction_type","Secured_by","Security_Type")
fil <- names(d)[!(names(d) %in% nn)]
dd <- d[,fil]
m1 <- glm(Status~Credit_Score,data = dd, family = binomial(link= "logit"))
m2 <- glm(Status~Credit_Score+LTV+income,data = dd, family = binomial(link= "logit"))
m3 <- glm(Status~LTV+income,data = dd, family = binomial(link= "logit"))
summary(m1)
summary(m2)
summary(m3)

library(pROC)

logit_P1 <- predict(m1 , newdata = dd ,type = 'response' )
roc_score1 <- roc(dd$Status, logit_P1) #AUC score
roc_score1
corte1 <- coords(roc_score1,"best")[[1]]
tt1 <- table((logit_P1>corte1)*1,dd$Status)
caret::confusionMatrix(tt1)

ks.test(logit_P1[dd$Status==1],logit_P1[dd$Status==0])
```

```
logit_P2 <- predict(m2 , newdata = dd ,type = 'response' )
roc_score2 <- roc(dd$Status, logit_P2) #AUC score
roc_score2
corte2 <- coords(roc_score2,"best")[[1]]
tt2 <- table((logit_P2>corte2)*1,dd$Status)
caret::confusionMatrix(tt2)

ks.test(logit_P2[dd$Status==1],logit_P2[dd$Status==0])

logit_P3 <- predict(m3 , newdata = dd ,type = 'response' )
roc_score3 <- roc(dd$Status, logit_P3) #AUC score
roc_score3
corte3 <- coords(roc_score3,"best")[[1]]
tt3 <- table((logit_P3>corte3)*1,dd$Status)
caret::confusionMatrix(tt3)

ks.test(logit_P3[dd$Status==1],logit_P3[dd$Status==0])
```

- **COMPROBACIÓN**

```
> # https://www.kaggle.com/datasets/yasserh/loan-default-dataset
> rm( list = ls())
> setwd("/cloud/project/Data")
> d <- read.csv("/cloud/project/Data/Loan_Default.csv", header = TRUE, sep = ",")
> table(d$Status)

    0     1
112031 36639
> d <- rio::import("Loan_Default.csv")
> table(d$Status)

    0     1
112031 36639
> nn <- c("ID","year","open_credit","construction_type","Secured_by","Security_Type")
> fil <- names(d)[!(names(d) %in% nn)]
> dd <- d[,fil]
> m1 <- glm(Status~Credit_Score,data = dd, family = binomial(link= "logit"))
> m2 <- glm(Status~Credit_Score+LTV+income,data = dd, family = binomial(link= "logit"))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
> m3 <- glm(Status~LTV+income,data = dd, family = binomial(link= "logit"))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(m1)

Call:
glm(formula = Status ~ Credit_Score, family = binomial(link = "logit"),
    data = dd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7575  -0.7540  -0.7505  -0.7471   1.6809
```

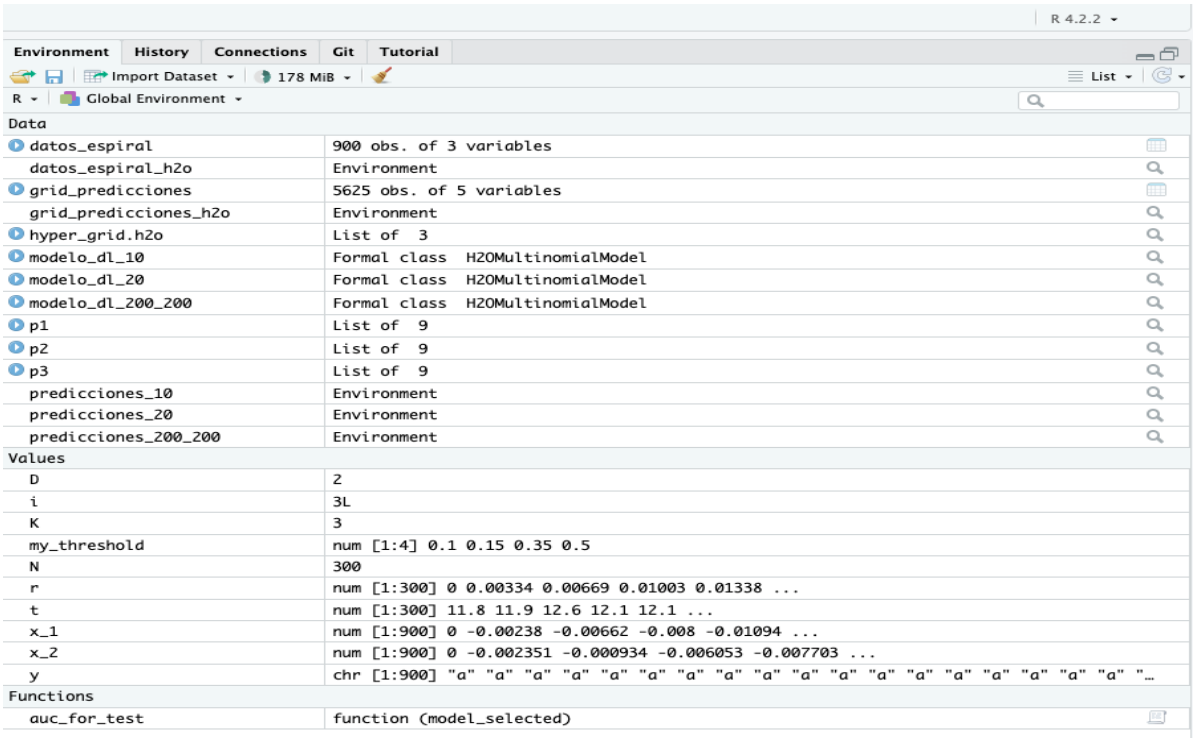
```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.174e+00 3.687e-02 -31.837 <2e-16
Credit_Score 8.018e-05 5.194e-05 1.544 0.123

(Intercept) ***
Credit_Score
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 166033 on 148669 degrees of freedom
Residual deviance: 166031 on 148668 degrees of freedom
AIC: 166035

Number of Fisher Scoring iterations: 4

> summary(m2)

Call:
glm(formula = Status ~ Credit_Score + LTV + income, family = binomial(link = "logit"),
    data = dd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8306 -0.6365 -0.5739 -0.4751 8.4904

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.310e+00 5.896e-02 -39.185 <2e-16
Credit_Score 9.793e-05 6.667e-05 1.469 0.142
LTV          1.303e-02 4.319e-04 30.158 <2e-16
income       -5.296e-05 2.030e-06 -26.085  <2e-16

(Intercept) ***
Credit_Score
LTV          ***
income       ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 110769 on 124436 degrees of freedom
Residual deviance: 108892 on 124433 degrees of freedom
  (24233 observations deleted due to missingness)
AIC: 108900

Number of Fisher Scoring iterations: 5

> summary(m3)

Call:
glm(formula = Status ~ LTV + income, family = binomial(link = "logit"),
    data = dd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8335 -0.6364 -0.5740 -0.4750 8.4904

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.242e+00 3.583e-02 -62.57 <2e-16
LTV          1.302e-02 4.319e-04 30.15 <2e-16
income      -5.296e-05 2.030e-06 -26.08  <2e-16
```

```
(Intercept) ***
LTV         ***
income      ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 110769 on 124436 degrees of freedom
Residual deviance: 108894 on 124434 degrees of freedom
  (24233 observations deleted due to missingness)
AIC: 108900

Number of Fisher Scoring iterations: 5

> library(pROC)
> logit_P1 <- predict(m1 , newdata = dd ,type = 'response' )
> roc_score1 <- roc(dd$Status, logit_P1) #AUC score
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> roc_score1

Call:
roc.default(response = dd$Status, predictor = logit_P1)

Data: logit_P1 in 112031 controls (dd$Status 0) < 36639 cases (dd$Status 1).
Area under the curve: 0.5027
> corte1 <- coords(roc_score1,"best")[[1]]
> tt1 <- table((logit_P1>corte1)*1,dd$Status)
> caret::confusionMatrix(tt1)
Confusion Matrix and Statistics


       0     1
 0 88204 28545
 1 23827  8094

               Accuracy : 0.6477
                 95% CI : (0.6453, 0.6502)
    No Information Rate : 0.7536
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0086

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.7873
            Specificity : 0.2209
         Pos Pred Value : 0.7555
         Neg Pred Value : 0.2536
             Prevalence : 0.7536
         Detection Rate : 0.5933
   Detection Prevalence : 0.7853
      Balanced Accuracy : 0.5041

       'Positive' Class : 0

> ks.test(logit_P1[dd$Status==1],logit_P1[dd$Status==0])

        Asymptotic two-sample Kolmogorov-Smirnov test

data: logit_P1[dd$Status == 1] and logit_P1[dd$Status == 0]
D = 0.0082299, p-value = 0.0475
alternative hypothesis: two-sided

Warning message:
In ks.test.default(logit_P1[dd$Status == 1], logit_P1[dd$Status == :
  p-value will be approximate in the presence of ties
```

```
> logit_P2 <- predict(m2 , newdata = dd ,type = 'response' )
> roc_score2 <- roc(dd$Status, logit_P2) #AUC score
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> roc_score2

Call:
roc.default(response = dd$Status, predictor = logit_P2)

Data: logit_P2 in 104118 controls (dd$Status 0) < 20319 cases (dd$Status 1).
Area under the curve: 0.6138
> corte2 <- coords(roc_score2,"best")[[1]]
> tt2 <- table((logit_P2>corte2)*1,dd$Status)
> caret::confusionMatrix(tt2)
Confusion Matrix and Statistics


      0     1
 0 73642 10703
 1 30476  9616

               Accuracy : 0.6691
                 95% CI : (0.6665, 0.6717)
    No Information Rate : 0.8367
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1297

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.7073
            Specificity : 0.4733
         Pos Pred Value : 0.8731
         Neg Pred Value : 0.2398
             Prevalence : 0.8367
         Detection Rate : 0.5918
   Detection Prevalence : 0.6778
      Balanced Accuracy : 0.5903

       'Positive' Class : 0

> ks.test(logit_P2[dd$Status==1],logit_P2[dd$Status==0])

        Asymptotic two-sample Kolmogorov-Smirnov test

data: logit_P2[dd$Status == 1] and logit_P2[dd$Status == 0]
D = 0.18055, p-value < 2.2e-16
alternative hypothesis: two-sided

Warning message:
In ks.test.default(logit_P2[dd$Status == 1], logit_P2[dd$Status == :
  p-value will be approximate in the presence of ties
> logit_P3 <- predict(m3 , newdata = dd ,type = 'response' )
> roc_score3 <- roc(dd$Status, logit_P3) #AUC score
Setting levels: control = 0, case = 1
Setting direction: controls < cases
> roc_score3

Call:
roc.default(response = dd$Status, predictor = logit_P3)

Data: logit_P3 in 104118 controls (dd$Status 0) < 20319 cases (dd$Status 1).
Area under the curve: 0.6138
> corte3 <- coords(roc_score3,"best")[[1]]
> tt3 <- table((logit_P3>corte3)*1,dd$Status)
> caret::confusionMatrix(tt3)
Confusion Matrix and Statistics
```

```
      0     1
0 74276 10827
1 29842  9492

               Accuracy : 0.6732
                 95% CI : (0.6706, 0.6758)
    No Information Rate : 0.8367
    P-Value [Acc > NIR] : 1

                  Kappa : 0.1311

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.7134
            Specificity : 0.4671
         Pos Pred Value : 0.8728
         Neg Pred Value : 0.2413
             Prevalence : 0.8367
         Detection Rate : 0.5969
   Detection Prevalence : 0.6839
      Balanced Accuracy : 0.5903

       'Positive' Class : 0

> ks.test(logit_P3[dd$Status==1],logit_P3[dd$Status==0])

        Asymptotic two-sample Kolmogorov-Smirnov test

data: logit_P3[dd$Status == 1] and logit_P3[dd$Status == 0]
D = 0.18053, p-value < 2.2e-16
alternative hypothesis: two-sided

Warning message:
In ks.test.default(logit_P3[dd$Status == 1], logit_P3[dd$Status == :
  p-value will be approximate in the presence of ties
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **Modelo Estadístico – Redes Neuronales**

  - **Flexibilidad de Redes Neuronales**
    - **MODELACIÓN**

```
# Número de observaciones por clase
N <- 300
# Número de dimensiones
D <- 2
# Número clases
K <- 3
# Matriz para almacenar las observaciones
x_1 <- vector(mode = "numeric")
x_2 <- vector(mode = "numeric")
y <- vector(mode = "numeric")

# Simulación de los datos
for(i in 1:K){
set.seed(123)
  r <- seq(from = 0, to = 1, length.out = N)
  t <- seq(from = i*4, to = (i+1)*4, length.out = N) + rnorm(n =
N) * 0.35
  x_1 <- c(x_1, r * sin(t))
  x_2 <- c(x_2, r*cos(t))
  y <- c(y, rep(letters[i], N))
}

datos_espiral <- data.frame(y = as.factor(y), x_1, x_2)
ggplot(data = datos_espiral, aes(x = x_1, y = x_2, color = y)) +
geom_point() +
  theme_bw() +
  theme(legend.position = "none",
        text = element_blank(),
        axis.ticks =  element_blank())
```

  - **COMPROBACIÓN**

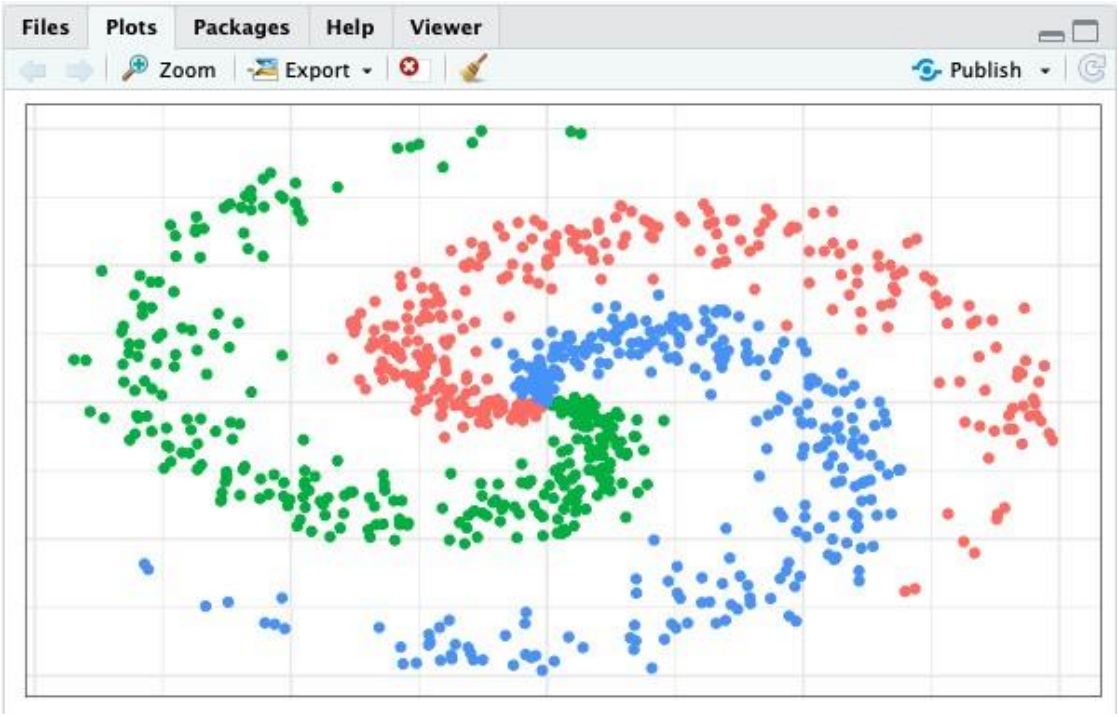
```
> # Número de observaciones por clase
> N <- 300
> # Número de dimensiones
> D <- 2
```

```
> # Número clases
> K <- 3
> # Matriz para almacenar las observaciones
> x_1 <- vector(mode = "numeric")
> x_2 <- vector(mode = "numeric")
> y <- vector(mode = "numeric")
> # Simulación de los datos
> for(i in 1:K){
+  set.seed(123)
+ r <- seq(from = 0, to = 1, length.out = N)
+    t <- seq(from = i*4, to = (i+1)*4, length.out = N) + rnorm(n = N) *
0.35
+ x_1 <- c(x_1, r * sin(t))
+ x_2 <- c(x_2, r*cos(t))
+ y <- c(y, rep(letters[i], N))
+ }
> datos_espiral <- data.frame(y = as.factor(y), x_1, x_2)
> ggplot(data = datos_espiral, aes(x = x_1, y = x_2, color = y)) +
+ geom_point() +
+ theme_bw() +
+ theme(legend.position = "none",
+         text = element_blank(),
+         axis.ticks = element_blank())
```

- **DATA UTILIZADA EN LA SIMULACIÓN**

Data	
datos_espiral	900 obs. of 3 variables
$ y : Factor w/ 3 levels "a","b","c": 1 1 1 1 1 1 1 1 1 1 ...	
$ x_1: num 0 -0.00238 -0.00662 -0.008 -0.01094 ...	
$ x_2: num 0 -0.002351 -0.000934 -0.006053 -0.007703 ...	
Values	
D	2
i	3L
K	3
N	300
r	num [1:300] 0 0.00334 0.00669 0.01003 0.01338 ...
t	num [1:300] 11.8 11.9 12.6 12.1 12.1 ...
x_1	num [1:900] 0 -0.00238 -0.00662 -0.008 -0.01094 ...
x_2	num [1:900] 0 -0.002351 -0.000934 -0.006053 -0.007703 …
y	chr [1:900] "a" "a" "a" "a" "a" "a" "a" "a" "a" "a" "a…

- **GRAFICACIÓN DE RESULTADOS DEL MODELO**



- **Flexibilidad de Redes Neuronales – Datos Espiral**
  - **MODELACIÓN**

```
datos_espiral_h2o <- as.h2o(datos_espiral)

modelo_dl_10 <- h2o.deeplearning(
  x = c("x_1", "x_2"),
  y = "y",
  distribution = "multinomial",
  training_frame = datos_espiral_h2o,
  standardize = TRUE,
  activation = "Rectifier",
  hidden = 10,
  stopping_rounds = 0,
  epochs = 50,
  seed = 123,
  model_id = "modelo_dl_10"
)

modelo_dl_20 <- h2o.deeplearning(
  x = c("x_1", "x_2"),
  y = "y",
```

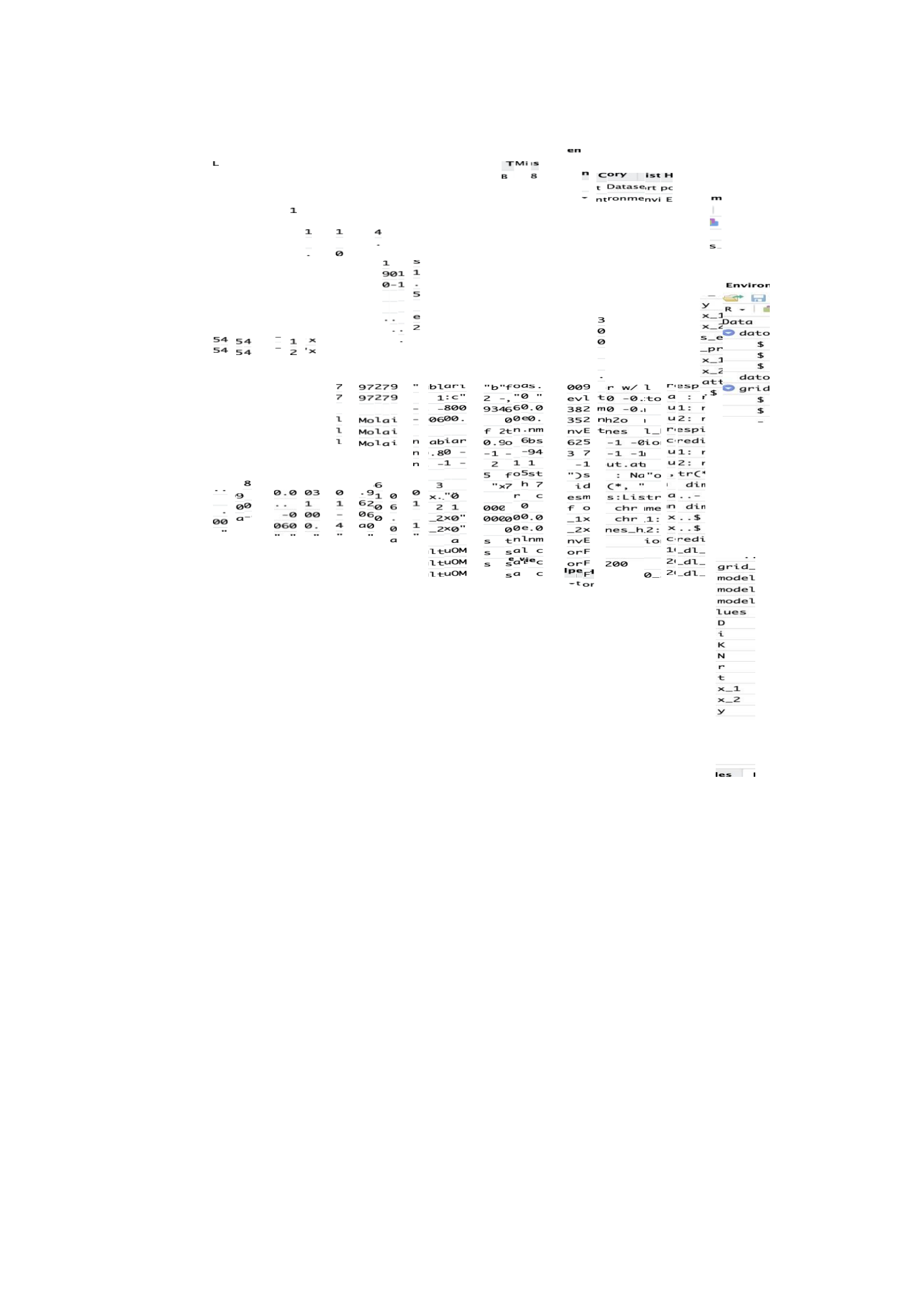
```
  distribution = "multinomial",
  training_frame = datos_espiral_h2o,
  standardize = TRUE,
  activation = "Rectifier",
  hidden = 20,
  stopping_rounds = 0,
  epochs = 1000,
  seed = 123,
  model_id = "modelo_dl_20"
)

modelo_dl_200_200 <- h2o.deeplearning(
  x = c("x_1", "x_2"),
  y = "y",
  distribution = "multinomial",
  training_frame = datos_espiral_h2o,
  standardize = TRUE,
  activation = "Rectifier",
  hidden = c(200, 200),
  stopping_rounds = 0,
  epochs = 1000,
  seed = 123,
  model_id = "modelo_dl_200_200"
)
```

- **COMPROBACIÓN**

```
> datos_espiral_h2o <- as.h2o(datos_espiral)

|=============================================
============================| 100%
> modelo_dl_10 <- h2o.deeplearning(
+ x = c("x_1", "x_2"),
+ y = "y",
+ distribution = "multinomial",
+ training_frame = datos_espiral_h2o,
+ standardize = TRUE,
+ activation = "Rectifier",
+ hidden = 10,
+ stopping_rounds = 0,
```

```
+ epochs = 50,
+ seed = 123,
+ model_id = "modelo_dl_10"
+ )

|=============================================
============================| 100%
> modelo_dl_20 <- h2o.deeplearning(
+ x = c("x_1", "x_2"),
+ y = "y",
+ distribution = "multinomial",
+ training_frame = datos_espiral_h2o,
+ standardize = TRUE,
+ activation = "Rectifier",
+ hidden = 20,
+ stopping_rounds = 0,
+ epochs = 1000,
+ seed = 123,
+ model_id = "modelo_dl_20"
+ )

|=============================================
============================| 100%
> modelo_dl_200_200 <- h2o.deeplearning(
+ x = c("x_1", "x_2"),
+ y = "y",
+ distribution = "multinomial",
+ training_frame = datos_espiral_h2o,
+ standardize = TRUE,
+ activation = "Rectifier",
+   hidden = c(200, 200),
+   stopping_rounds = 0,
+ epochs = 1000,
+ seed = 123,
+ model_id = "modelo_dl_200_200"
+ )

|=============================================
============================| 100%
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **Flexibilidad de Redes Neuronales – Fronteras de Clasificación**
  - **MODELACIÓN**

```
h2o.init()

grid_predicciones <- expand.grid(x_1 = seq(from = -1, to
= 1, length = 75),
                                 x_2 = seq(from = -1, to = 1, length =
75))
grid_predicciones_h2o <- as.h2o(grid_predicciones)

predicciones_10 <- h2o.predict(object = modelo_dl_10,
                               newdata = grid_predicciones_h2o)
grid_predicciones$y_10                                    <-
as.vector(predicciones_10$predict)

predicciones_20 <- h2o.predict(object = modelo_dl_20,
                               newdata = grid_predicciones_h2o)
grid_predicciones$y_20                                    <-
as.vector(predicciones_20$predict)

predicciones_200_200      <-      h2o.predict(object      =
modelo_dl_200_200,
                                  newdata = grid_predicciones_h2o)
grid_predicciones$y_200_200                               <-
as.vector(predicciones_200_200$predict)
```

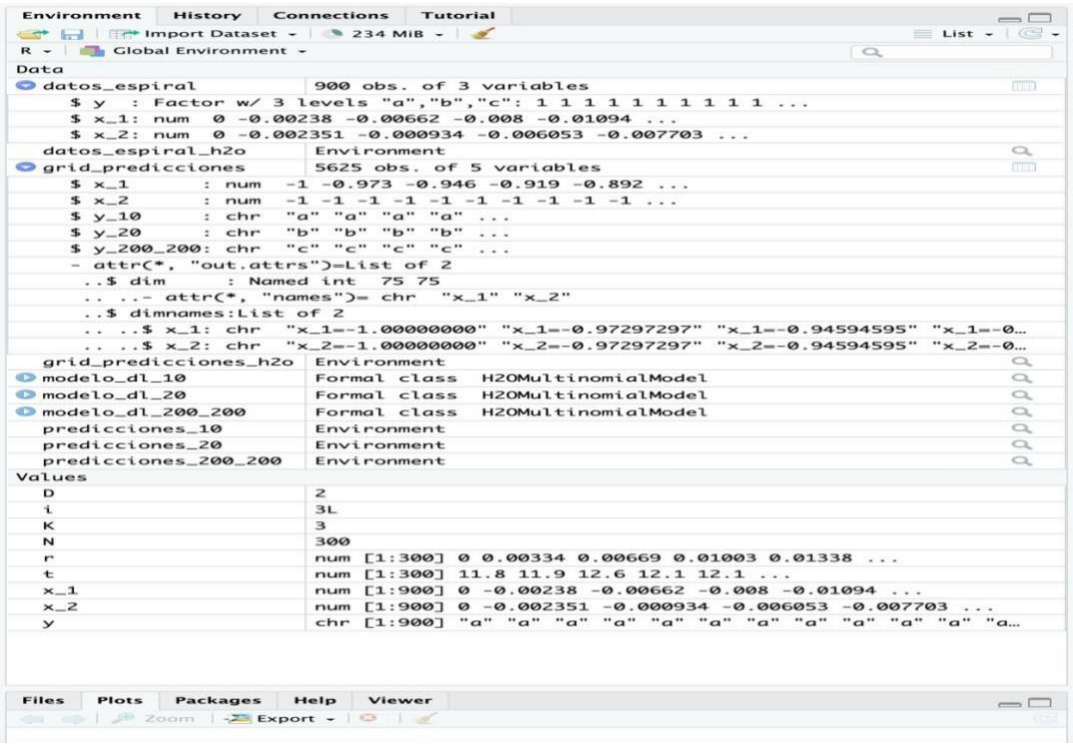
  - **COMPROBACIÓN**

```
> h2o.init()
 Connection successful!

R is connected to the H2O cluster:
   H2O cluster uptime:         1 hours 36 minutes
   H2O cluster timezone:       America/Guayaquil
   H2O data parsing timezone:  UTC
   H2O cluster version:        3.38.0.1
```

```
   H2O cluster version age:    4 months and 8 days !!!
   H2O cluster name:
H2O_started_from_R_viagonza_bxh486
   H2O cluster total nodes:    1
   H2O cluster total memory: 2.93 GB
   H2O cluster total cores:    16
   H2O cluster allowed cores: 16
   H2O cluster healthy:        TRUE
   H2O Connection ip:          localhost
   H2O Connection port:        54321
   H2O Connection proxy:       NA
   H2O Internal Security:      FALSE
   R Version:                  R version 3.6.2 (2019-12-12)

Warning message:
In h2o.clusterInfo() :
Your H2O cluster version is too old (4 months and 8 days)!
Please download and install the latest version from
http://h2o.ai/download/
> grid_predicciones <- expand.grid(x_1 = seq(from = -1,
to = 1, length = 75),
+                                  x_2 = seq(from = -1, to = 1, length
= 75))
> grid_predicciones_h2o <- as.h2o(grid_predicciones)

|=========================================
============| 100%
> predicciones_10 <- h2o.predict(object = modelo_dl_10,
+                                newdata = grid_predicciones_h2o)

|=========================================
============| 100%
>                  grid_predicciones$y_10                  <-
as.vector(predicciones_10$predict)
```

```
> predicciones_20 <- h2o.predict(object = modelo_dl_20,
+                                newdata = grid_predicciones_h2o)

|=========================================================| 100%
> grid_predicciones$y_20 <- as.vector(predicciones_20$predict)
> predicciones_200_200 <- h2o.predict(object = modelo_dl_200_200,
+                                     newdata = grid_predicciones_h2o)

|=========================================================| 100%
> grid_predicciones$y_200_200 <- as.vector(predicciones_200_200$predict)
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **Arboles Binomiales – RANDOM FOREST**
  - **MODELACIÓN**

```
#================================
#       Data Pre-processing
#================================

# Load some packages for data manipulation:
library(tidyverse)
library(magrittr)

# Clear workspace:
rm(list = ls())

# Import data:
hmeq                                                    <-
read.csv("http://www.creditriskanalytics.net/uploads/1/9/
5/1/19511601/hmeq.csv")

# Function replaces NA by mean:
replace_by_mean <- function(x) {
x[is.na(x)] <- mean(x, na.rm = TRUE)
return(x)
}

# A function imputes NA observations for categorical
variables:

replace_na_categorical <- function(x) {
  x %>%
    table() %>%
    as.data.frame() %>%
    arrange(-Freq) ->> my_df

  n_obs <- sum(my_df$Freq)
```

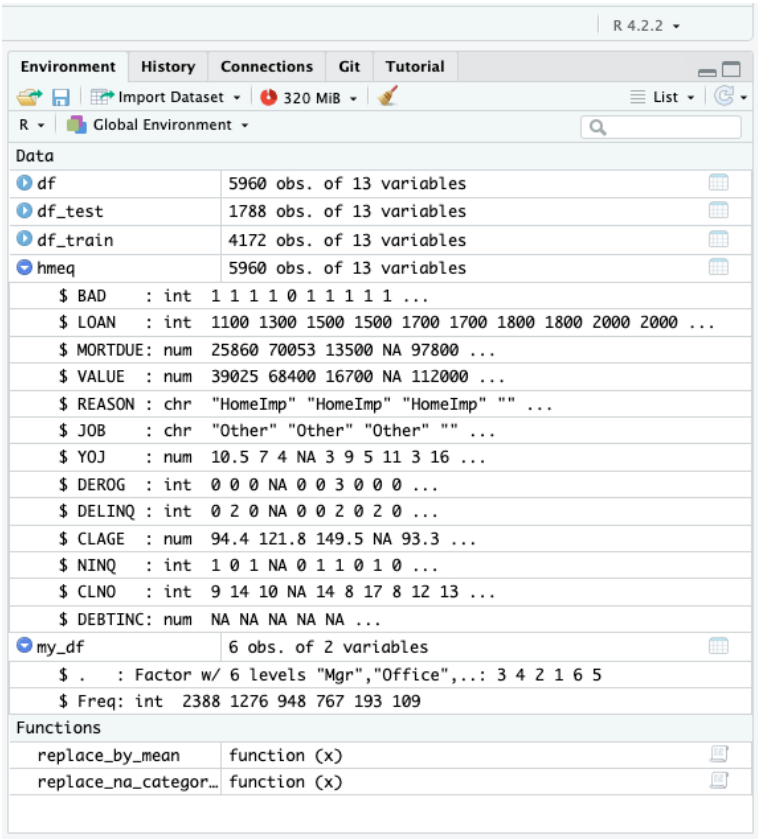
```
  pop <- my_df$. %>% as.character()
  set.seed(29)
  x[is.na(x)] <- sample(pop, sum(is.na(x)), replace =
TRUE, prob = my_df$Freq)
  return(x)
}

# Use the two functions:
df <- hmeq %>%
  mutate_if(is.factor, as.character) %>%
  mutate(REASON = case_when(REASON == "" ~
NA_character_, TRUE ~ REASON),
         JOB = case_when(JOB == "" ~ NA_character_,
TRUE ~ JOB),
         BAD = case_when(BAD == 1 ~ "BAD", TRUE ~
"GOOD")) %>%
  mutate_if(is_character, as.factor) %>%
  mutate_if(is.numeric, replace_by_mean) %>%
  mutate_if(is.factor, replace_na_categorical)

# Split our data:
set.seed(1)

df_train <- df %>%
  group_by(BAD) %>%
  sample_frac(0.7) %>%
  ungroup() # Use 70% data set for training model.

df_test <- dplyr::setdiff(df, df_train) # Use 30% data set for
validation.

# Activate h2o package for using:
library(h2o)
h2o.init(nthreads = 20, max_mem_size = "16g")
```

COMPROBACIÓN

```
> # Load some packages for data manipulation:
> library(tidyverse)
> library(magrittr)
> # Clear workspace:
> rm(list = ls())
> # Import data:
> hmeq <- read.csv("http://www.creditriskanalytics.net/uploads/1/9/5/1/19511601/hmeq.csv")
> # Function replaces NA by mean:
> replace_by_mean <- function(x) {
+ x[is.na(x)] <- mean(x, na.rm = TRUE)
+ return(x)
+ }
> replace_na_categorical <- function(x) {
+ x %>%
+     table() %>%
+     as.data.frame() %>%
+     arrange(-Freq) ->> my_df
+
+ n_obs <- sum(my_df$Freq)
+ pop <- my_df$. %>% as.character()
+ set.seed(29)
+ x[is.na(x)] <- sample(pop, sum(is.na(x)), replace = TRUE, prob = my_df$Freq)
+ return(x)
+ }
> # Use the two functions:
> df <- hmeq %>%
+ mutate_if(is.factor, as.character) %>%
+ mutate(REASON = case_when(REASON == "" ~ NA_character_, TRUE ~ REASON),
+         JOB = case_when(JOB == "" ~ NA_character_, TRUE ~ JOB),
+         BAD = case_when(BAD == 1 ~ "BAD", TRUE ~ "GOOD")) %>%
+ mutate_if(is_character, as.factor) %>%
+ mutate_if(is.numeric, replace_by_mean) %>%
+ mutate_if(is.factor, replace_na_categorical)
> # Split our data:
> set.seed(1)
> df_train <- df %>%
+   group_by(BAD) %>%
+   sample_frac(0.7) %>%
+ ungroup() # Use 70% data set for training model.
> df_test <- dplyr::setdiff(df, df_train) # Use 30% data set for validation.
> # Activate h2o package for using:
> library(h2o)
> h2o.init(nthreads = 20, max_mem_size = "16g")
 Connection successful!

R is connected to the H2O cluster:
    H2O cluster uptime:         57 minutes 53 seconds
    H2O cluster timezone:       UTC
```

- H2O data parsing timezone: UTC
- H2O cluster version: 3.38.0.1
- H2O cluster version age: 4 months and 8 days !!!
- H2O cluster name: H2O_started_from_R_r1866432_urh803
- H2O cluster total nodes: 1
- H2O cluster total memory: 0.18 GB
- H2O cluster total cores: 1
- H2O cluster allowed cores: 1
- H2O cluster healthy: TRUE
- H2O Connection ip: localhost
- H2O Connection port: 54321
- H2O Connection proxy: NA
- H2O Internal Security: FALSE
- R Version: R version 4.2.2 (2022-10-31)
-
- Warning message:
- In h2o.clusterInfo() :
- Your H2O cluster version is too old (4 months and 8 days)!
- Please download and install the latest version from http://h2o.ai/download/

▪ DATA UTILIZADA EN LA SIMULACIÓN



- **Arboles Binomiales – RANDOM FOREST – Conversión del Frame**
  - **MODELACIÓN**

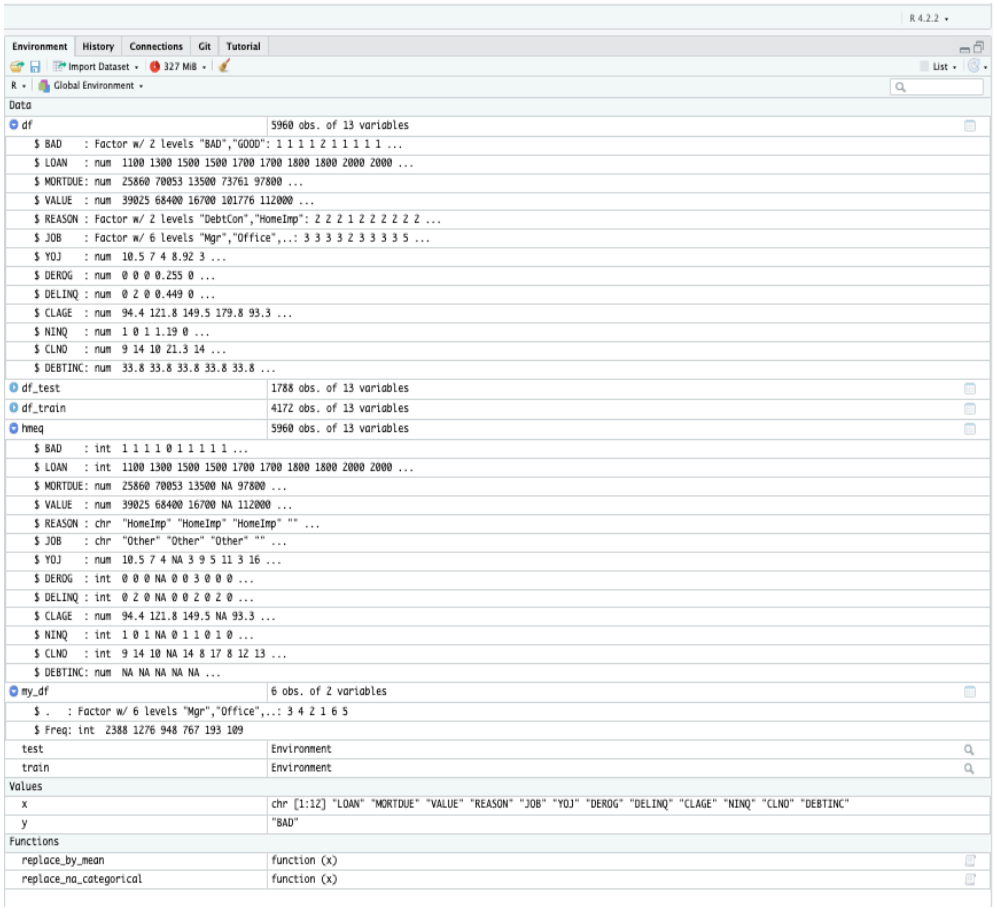
```
h2o.no_progress()

h2o.init()

# Convert to h2o Frame and identify inputs and output:
test <- as.h2o(df_test)
train <- as.h2o(df_train)
y <- "BAD"
x <- setdiff(names(train), y)
```

  - **COMPROBACIÓN**

```
>h2o.no_progress()
>h2o.init()

H2O is not running yet, starting it now...

 Note: In case of errors look at the following log files:
    /tmp/RtmpvfYTgv/filee56d8ce71d/h2o_r1866432_started_from_r.out
    /tmp/RtmpvfYTgv/filee54bbc03b3/h2o_r1866432_started_from_r.err

openjdk version "11.0.17" 2022-10-18
OpenJDK Runtime Environment (build 11.0.17+8-post-Ubuntu-1ubuntu220.04)
OpenJDK 64-Bit Server VM (build 11.0.17+8-post-Ubuntu-1ubuntu220.04, mixed mo
de, sharing)

Starting H2O JVM and connecting: ..... Connection successful!

R is connected to the H2O cluster:
    H2O cluster uptime:         8 seconds 98 milliseconds
    H2O cluster timezone:       UTC
    H2O data parsing timezone: UTC
    H2O cluster version:        3.38.0.1
    H2O cluster version age: 4 months and 8 days !!!
    H2O cluster name:           H2O_started_from_R_r1866432_qef711
    H2O cluster total nodes:    1
    H2O cluster total memory: 0.23 GB
    H2O cluster total cores:    1
    H2O cluster allowed cores: 1
    H2O cluster healthy:        TRUE
    H2O Connection ip:          localhost
    H2O Connection port:        54321
    H2O Connection proxy:       NA
    H2O Internal Security:      FALSE
    R Version:                  R version 4.2.2 (2022-10-31)

Warning message:
In h2o.clusterInfo() :
Your H2O cluster version is too old (4 months and 8 days)!
```

Please download and install the latest version from http://h2o.ai/download/
>\# Convert to h2o Frame and identify inputs and output:
>test <- as.h2o(df_test)
>train <- as.h2o(df_train)
>y <- "BAD"
>x <- setdiff(names(train), y)

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **Arboles Binomiales – RANDOM FOREST – Performance del Modelo con algunos criterios seleccionados**
  - **MODELACIÓN**

```
# Train default Random Forest:
default_rf <- h2o.randomForest(x = x, y = y,
                               training_frame = train,
```

```
                    stopping_rounds = 5,
                    stopping_tolerance = 0.001,
                    stopping_metric = "AUC",
                    seed = 29,
                    balance_classes = FALSE,
                    nfolds = 10)

# Function for collecting cross-validation results:

results_cross_validation <- function(h2o_model) {
  h2o_model@model$cross_validation_metrics_summary
%>%
    as.data.frame() %>%
    select(-mean, -sd) %>%
    t() %>%
    as.data.frame() %>%
    mutate_all(as.character) %>%
    mutate_all(as.numeric) %>%
    select(Accuracy = accuracy,
           AUC = auc,
           Precision = precision,
           Specificity = specificity,
           Recall = recall,
           Logloss = logloss) %>%
    return()
}

# Use function:
results_cross_validation(default_rf) -> ket_qua_default

# Model Performance by Graph:
theme_set(theme_minimal())

plot_results <- function(df_results) {
  df_results %>%
    gather(Metrics, Values) %>%
    ggplot(aes(Metrics, Values, fill = Metrics, color = Metrics))
+
    geom_boxplot(alpha = 0.3, show.legend = FALSE) +
    theme(plot.margin = unit(c(1, 1, 1, 1), "cm")) +
    scale_y_continuous(labels = scales::percent) +
    facet_wrap(~ Metrics, scales = "free") +
    labs(title = "Model Performance by Some Criteria Selected",
y = NULL)
```

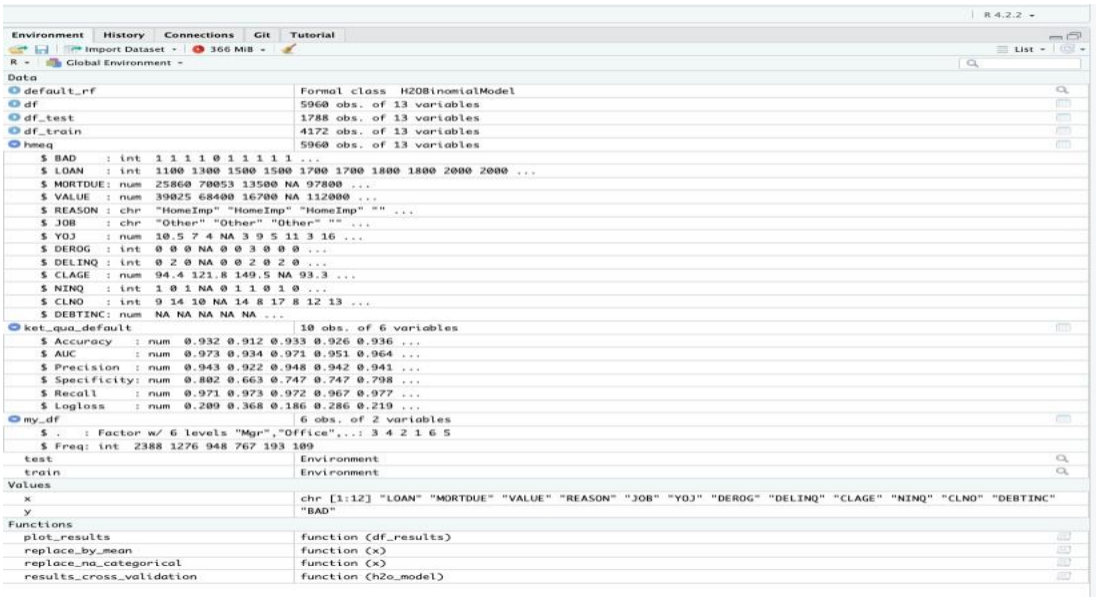
```
}

plot_results(ket_qua_default) +
  labs(subtitle = "Model: Random Forest (h2o package)")
```

- **COMPROBACIÓN**

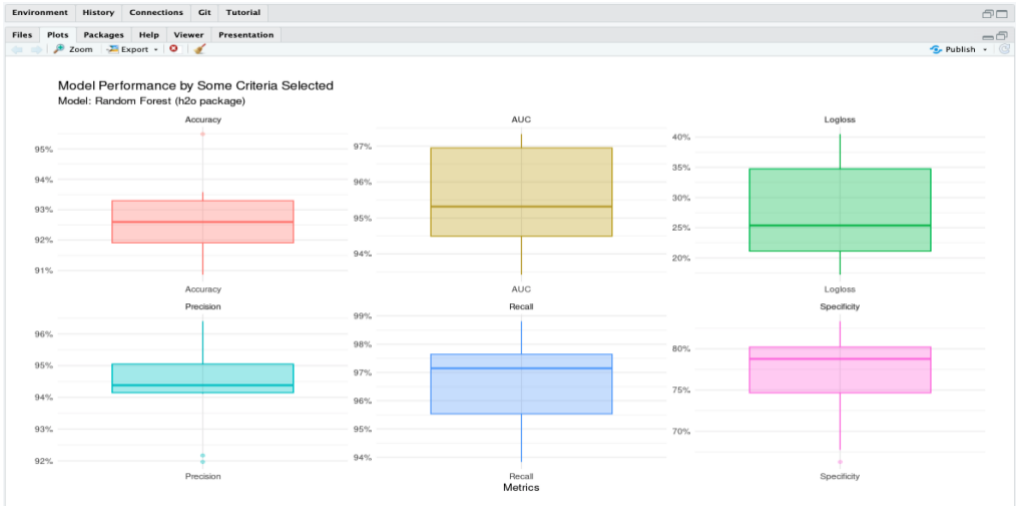
```
Warning message:
In .h2o.processResponseWarnings(res) :
  early stopping is enabled but neither score_tree_interval or score_each_iteration are d
efined. Early stopping will not be reproducible!.

> # Train default Random Forest:
> default_rf <- h2o.randomForest(x = x, y = y,
+                                training_frame = train,
+                                stopping_rounds = 5,
+                                stopping_tolerance = 0.001,
+                                stopping_metric = "AUC",
+                                seed = 29,
+                                balance_classes = FALSE,
+                                nfolds = 10)
Warning message:
In .h2o.processResponseWarnings(res) :
  early stopping is enabled but neither score_tree_interval or score_each_iteration are d
efined. Early stopping will not be reproducible!.

> results_cross_validation <- function(h2o_model) {
+ h2o_model@model$cross_validation_metrics_summary %>%
+     as.data.frame() %>%
+     select(-mean, -sd) %>%
+     t() %>%
+     as.data.frame() %>%
+     mutate_all(as.character) %>%
+     mutate_all(as.numeric) %>%
+     select(Accuracy = accuracy,
+            AUC = auc,
+            Precision = precision,
+            Specificity = specificity,
+            Recall = recall,
+            Logloss = logloss) %>%
+     return()
+ }
> # Use function:
> results_cross_validation(default_rf) -> ket_qua_default
> # Model Performance by Graph:
> theme_set(theme_minimal())
> plot_results <- function(df_results) {
+ df_results %>%
+     gather(Metrics, Values) %>%
+     ggplot(aes(Metrics, Values, fill = Metrics, color = Metrics)) +
+     geom_boxplot(alpha = 0.3, show.legend = FALSE) +
+     theme(plot.margin = unit(c(1, 1, 1, 1), "cm")) +
+     scale_y_continuous(labels = scales::percent) +
+     facet_wrap(~ Metrics, scales = "free") +
+     labs(title = "Model Performance by Some Criteria Selected", y = NULL)
+ }
> plot_results(ket_qua_default) +
```

+ labs(subtitle = "Model: Random Forest (h2o package)")

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **GRAFICACIÓN DEL MODELO**



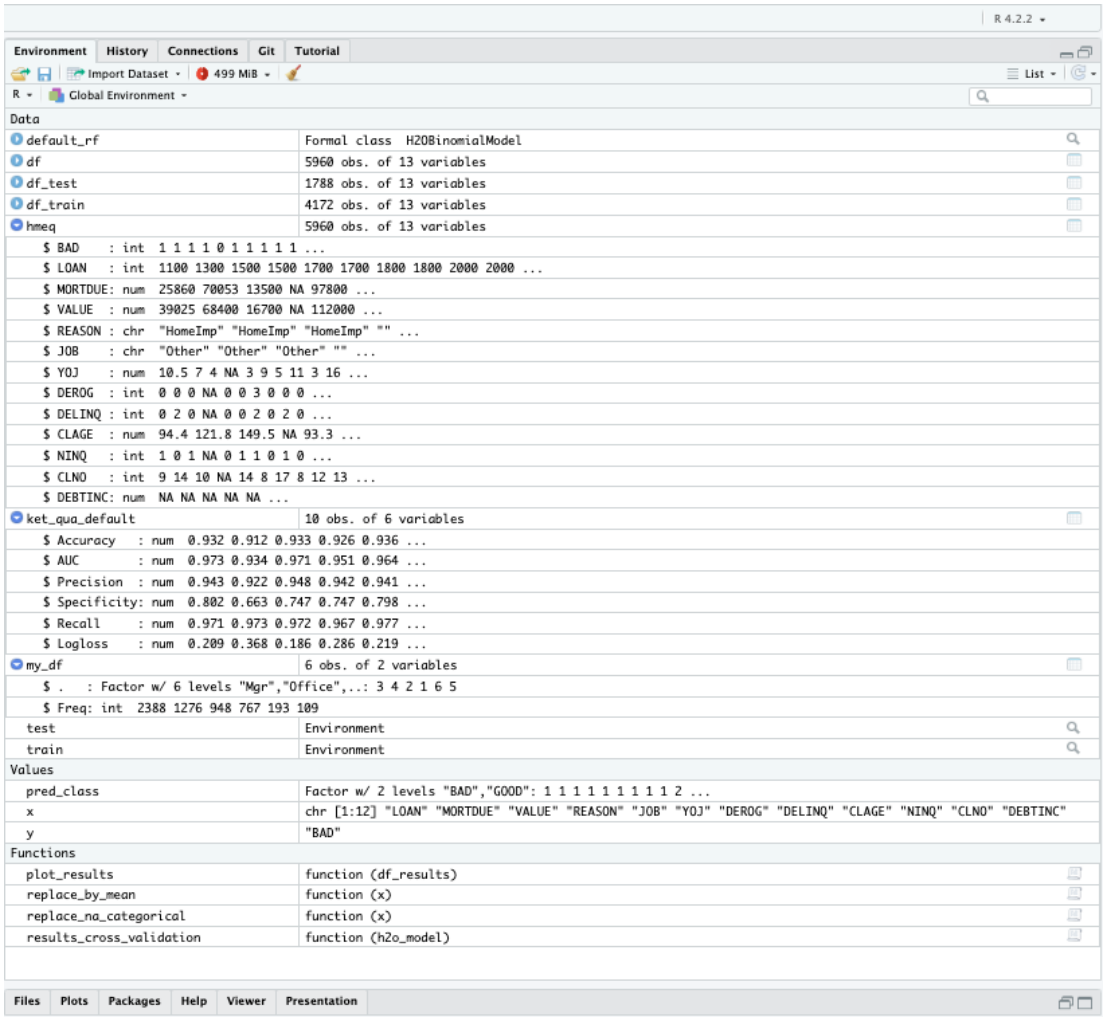
- **Arboles Binomiales – RANDOM FOREST – Modelo de Performance Basado en Data**
  - **MODELACIÓN**

```
# Model performance based on test data:
pred_class <- h2o.predict(default_rf, test) %>% as.data.frame() %>%
pull(predict)
library(caret)
confusionMatrix(pred_class, df_test$BAD, positive = "BAD")
```

- **COMPROBACIÓN**

```
># Model performance based on test data:
>pred_class <- h2o.predict(default_rf, test) %>% as.data.frame() %>% pull(predict)
>library(caret)
>confusionMatrix(pred_class, df_test$BAD, positive = "BAD")
Confusion Matrix and Statistics

          Reference
Prediction BAD GOOD
      BAD 276 44
      GOOD 81 1387

               Accuracy : 0.9301
                 95% CI : (0.9173, 0.9415)
    No Information Rate : 0.8003
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7724

 Mcnemar's Test P-Value : 0.001282

            Sensitivity : 0.7731
            Specificity : 0.9693
         Pos Pred Value : 0.8625
         Neg Pred Value : 0.9448
             Prevalence : 0.1997
         Detection Rate : 0.1544
   Detection Prevalence : 0.1790
      Balanced Accuracy : 0.8712

       'Positive' Class : BAD
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **Arboles Binomiales – RANDOM FOREST – Curva ROC & AUC**

  - **MODELACIÓN**

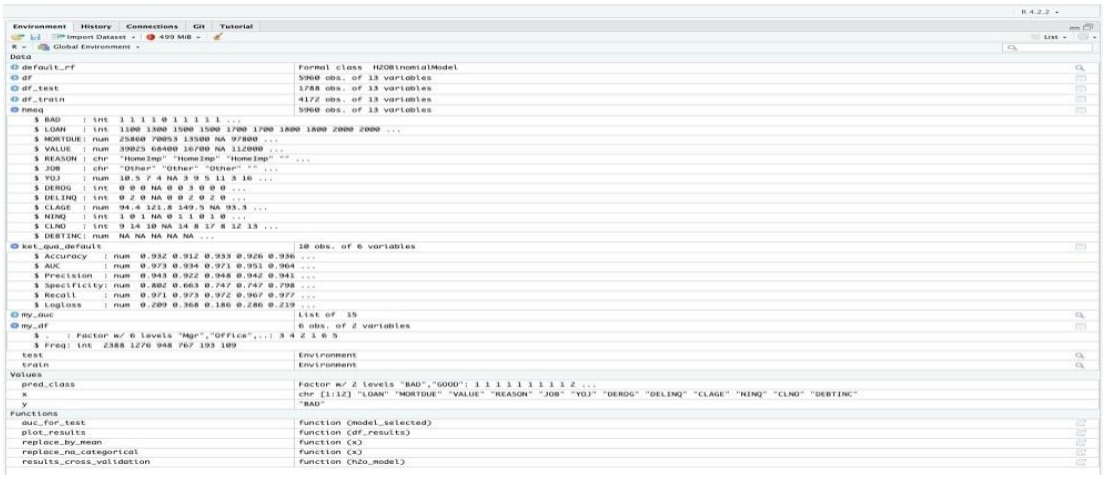
```
# ROC curve and AUC:
library(pROC)

# Function calculates AUC:
auc_for_test <- function(model_selected) {
  actual <- df_test$BAD
  pred_prob <- h2o.predict(model_selected, test) %>%
as.data.frame() %>% pull(BAD)
```

```
  return(roc(actual, pred_prob))
}

# Use this function:
my_auc <- auc_for_test(default_rf)
my_auc$auc
```

- **COMPROBACIÓN**

```
>
# ROC curve and AUC:
>library(pROC)
Type 'citation("pROC")' for a citation.

Attaching package: ‘pROC’

The following object is masked from ‘package:h2o’:

    var

The following objects are masked from ‘package:stats’:

    cov, smooth, var

>
# Function calculates AUC:
>auc_for_test <- function(model_selected) {
+ actual <- df_test$BAD
+ pred_prob <- h2o.predict(model_selected, test) %>% as.data.frame() %>% pull(BA
D)
+ return(roc(actual, pred_prob))
+ }
>
# Use this function:
>my_auc <- auc_for_test(default_rf)
Setting levels: control = BAD, case = GOOD
Setting direction: controls > cases
>my_auc$auc
Area under the curve: 0.9603
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **Arboles Binomiales – RANDOM FOREST – Modelo de Performance para clasificación RF basada en Data**

- **MODELACIÓN**

Graph ROC and AUC:

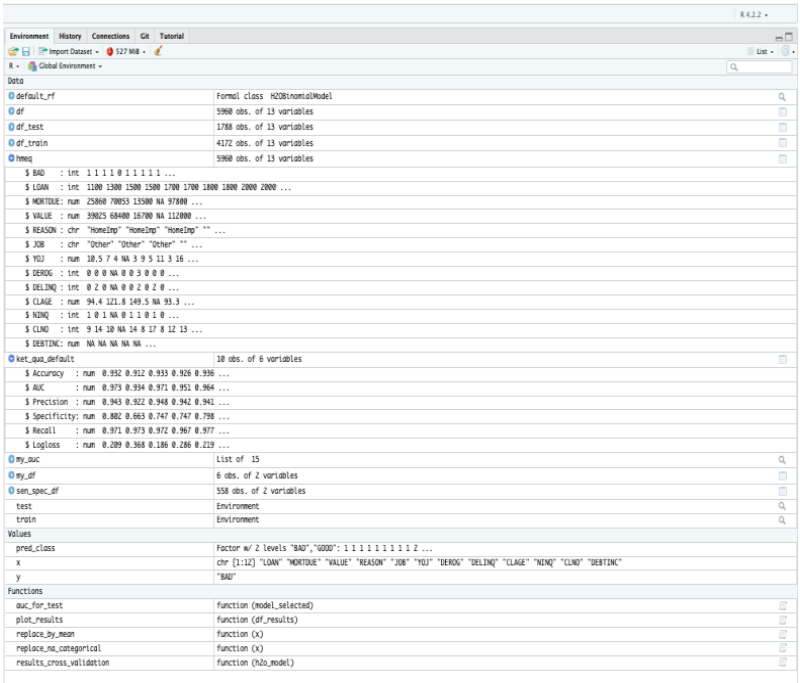
```
sen_spec_df <- data_frame(TPR =
my_auc$sensitivities, FPR = 1 - my_auc$specificities)

sen_spec_df %>%
  ggplot(aes(x = FPR, ymin = 0, ymax = TPR))+
  geom_polygon(aes(y = TPR), fill = "red", alpha = 0.3)+
  geom_path(aes(y = TPR), col = "firebrick", size = 1.2) +
          geom_abline(intercept = 0, slope = 1, color =
        "gray37", size = 1, linetype = "dashed") +
  theme_bw() +
  coord_equal() +
  labs(x = "FPR (1 - Specificity)",
      y = "TPR (Sensitivity)",
              title = "Model Performance for RF Classifier
        based on Test Data",
              subtitle = paste0("AUC Value: ", my_auc$auc
        %>% round(2)))
```

- **COMPROBACIÓN**

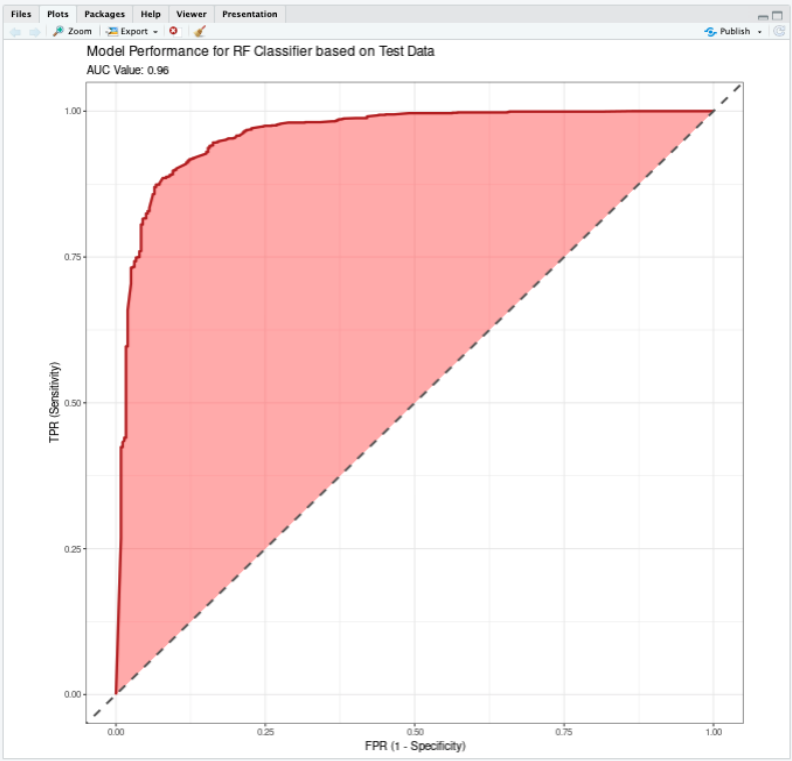
```
> sen_spec_df <- data_frame(TPR = my_auc$sensitivities, FPR = 1 - my
_auc$specificities)
Warning message:
`data_frame()` was deprecated in tibble 1.1.0.
ℹ Please use `tibble()` instead.
This warning is displayed once every 8 hours.
```

```
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
> sen_spec_df %>%
+ ggplot(aes(x = FPR, ymin = 0, ymax = TPR))+
+ geom_polygon(aes(y = TPR), fill = "red", alpha = 0.3)+
+ geom_path(aes(y = TPR), col = "firebrick", size = 1.2) +
+ geom_abline(intercept = 0, slope = 1, color = "gray37", size = 1, linet
ype = "dashed") +
+ theme_bw() +
+ coord_equal() +
+ labs(x = "FPR (1 - Specificity)",
+       y = "TPR (Sensitivity)",
+       title = "Model Performance for RF Classifier based on Test Data",
+       subtitle = paste0("AUC Value: ", my_auc$auc %>% round(2)))
Warning message:
Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
generated.
Connected to your session in progress, last started 2023-Jan-28 04:07:30
UTC (55 minutes ago)
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **GRAFICACIÓN**



- **FUNCIÓN PARA CALCULAR CM**

```
# Function for calculating CM:

my_cm_com_rf <- function(thre) {
  du_bao_prob <- h2o.predict(default_rf, test) %>%
as.data.frame() %>% pull(BAD)
  du_bao <- case_when(du_bao_prob >= thre ~ "BAD",
                      du_bao_prob < thre ~ "GOOD") %>%
as.factor()
  cm <- confusionMatrix(du_bao, df_test$BAD, positive =
"BAD")
  return(cm)

}
```

- **COMPROBACIÓN**

```
>my_cm_com_rf <- function(thre) {
+ du_bao_prob <- h2o.predict(default_rf, test) %>% as.data.frame() %>% pull(BA
D)
```

```
+ du_bao <- case_when(du_bao_prob >= thre ~ "BAD",
+                     du_bao_prob < thre ~ "GOOD") %>% as.factor()
+ cm <- confusionMatrix(du_bao, df_test$BAD, positive = "BAD")
+ return(cm)
+
+ }
```

- **Árboles De Decision – RANDOM FOREST – Tasa de Precisión para predecir aplicaciones buenas y malas, conforme el umbral definido**

  - **MODELACIÓN**

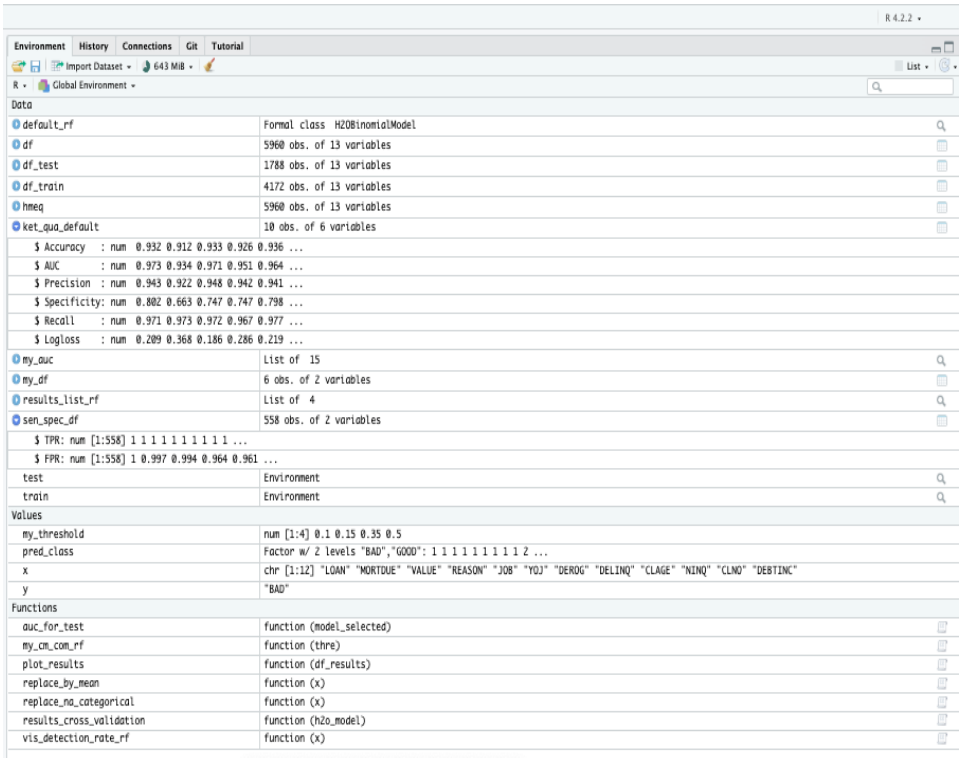
```
# Set a range of threshold for classification:
my_threshold <- c(0.10, 0.15, 0.35, 0.5)
results_list_rf <- lapply(my_threshold, my_cm_com_rf)

# Function for presenting prediction power by class:

vis_detection_rate_rf <- function(x) {

  results_list_rf[[x]]$table %>% as.data.frame() -> m
  rate <- round(100*m$Freq[1] / sum(m$Freq[c(1, 2)]), 2)
  acc <- round(100*sum(m$Freq[c(1, 4)]) / sum(m$Freq), 2)
  acc <- paste0(acc, "%")

  m %>%
    ggplot(aes(Reference, Freq, fill = Prediction)) +
    geom_col(position = "fill") +
    scale_fill_manual(values = c("#e41a1c", "#377eb8"), name
= "") +
    theme(panel.grid.minor.y = element_blank()) +
    theme(panel.grid.minor.x = element_blank()) +
    scale_y_continuous(labels = scales::percent) +
    labs(x = NULL, y = NULL,
         title = paste0("Detecting Default Cases when Threshold
= ", my_threshold[x]),
         subtitle = paste0("Detecting Rate for Default Cases: ",
rate, "%", ", ", "Accuracy: ", acc))
}
```

```
# Use this function:
gridExtra::grid.arrange(vis_detection_rate_rf(1),
                        vis_detection_rate_rf(2),
                        vis_detection_rate_rf(3),
                        vis_detection_rate_rf(4))
```

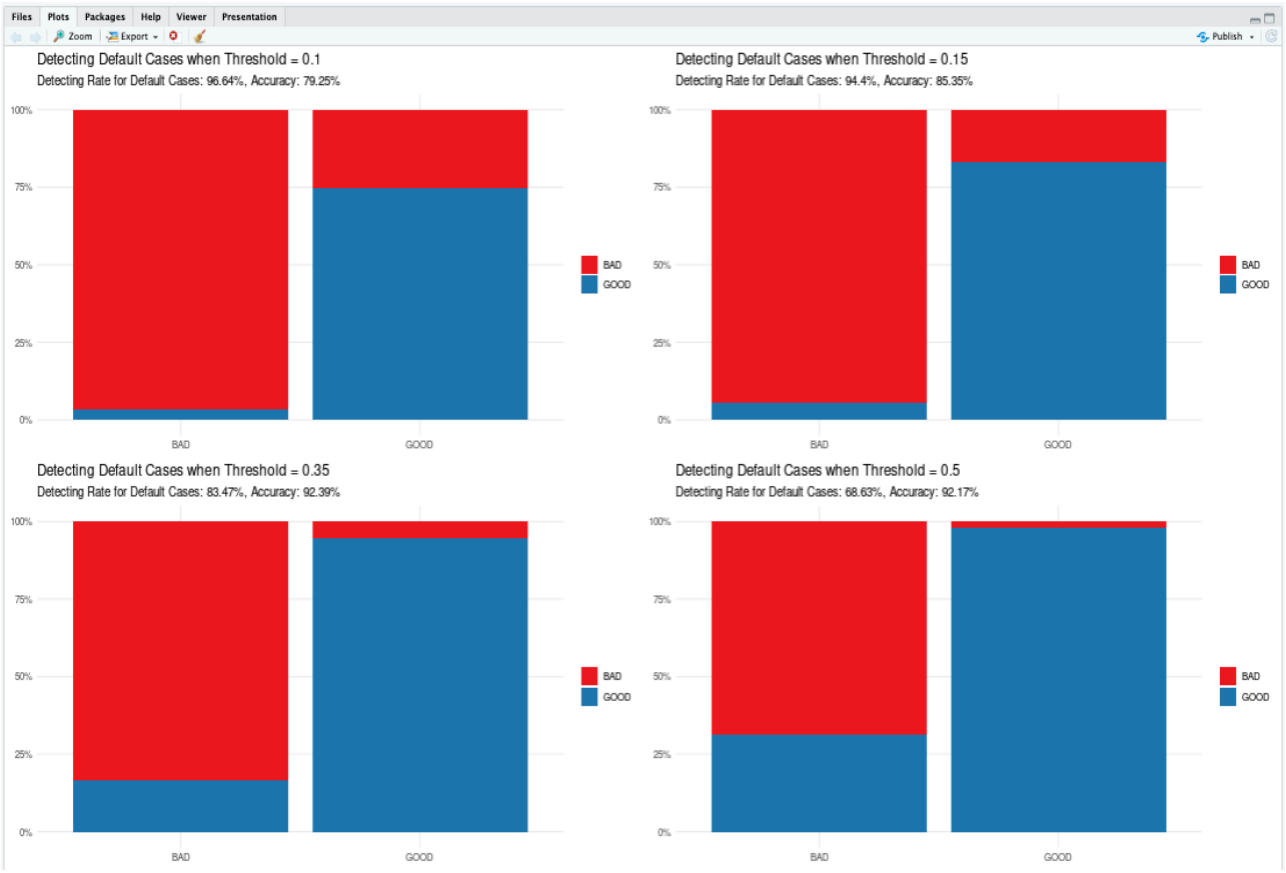
- **COMPROBACIÓN**

```
> # Set a range of threshold for classification:
> my_threshold <- c(0.10, 0.15, 0.35, 0.5)
> results_list_rf <- lapply(my_threshold, my_cm_com_rf)
> vis_detection_rate_rf <- function(x) {
+
+ results_list_rf[[x]]$table %>% as.data.frame() -> m
+ rate <- round(100*m$Freq[1] / sum(m$Freq[c(1, 2)]), 2)
+ acc <- round(100*sum(m$Freq[c(1, 4)]) / sum(m$Freq), 2)
+ acc <- paste0(acc, "%")
+
+ m %>%
+     ggplot(aes(Reference, Freq, fill = Prediction)) +
+     geom_col(position = "fill") +
+     scale_fill_manual(values = c("#e41a1c", "#377eb8"), name = "") +
+     theme(panel.grid.minor.y = element_blank()) +
+     theme(panel.grid.minor.x = element_blank()) +
+     scale_y_continuous(labels = scales::percent) +
+     labs(x = NULL, y = NULL,
+          title = paste0("Detecting Default Cases when Threshold = ", my_
threshold[x]),
+          subtitle = paste0("Detecting Rate for Default Cases: ", rate, "%",
", ", "Accuracy: ", acc))
+ }
> # Use this function:
> gridExtra::grid.arrange(vis_detection_rate_rf(1),
+                         vis_detection_rate_rf(2),
+                         vis_detection_rate_rf(3),
+                         vis_detection_rate_rf(4))
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **GRAFICACIÓN DEL MODELO**



- **CUADRICULA CARTESIANA COMPLETA**

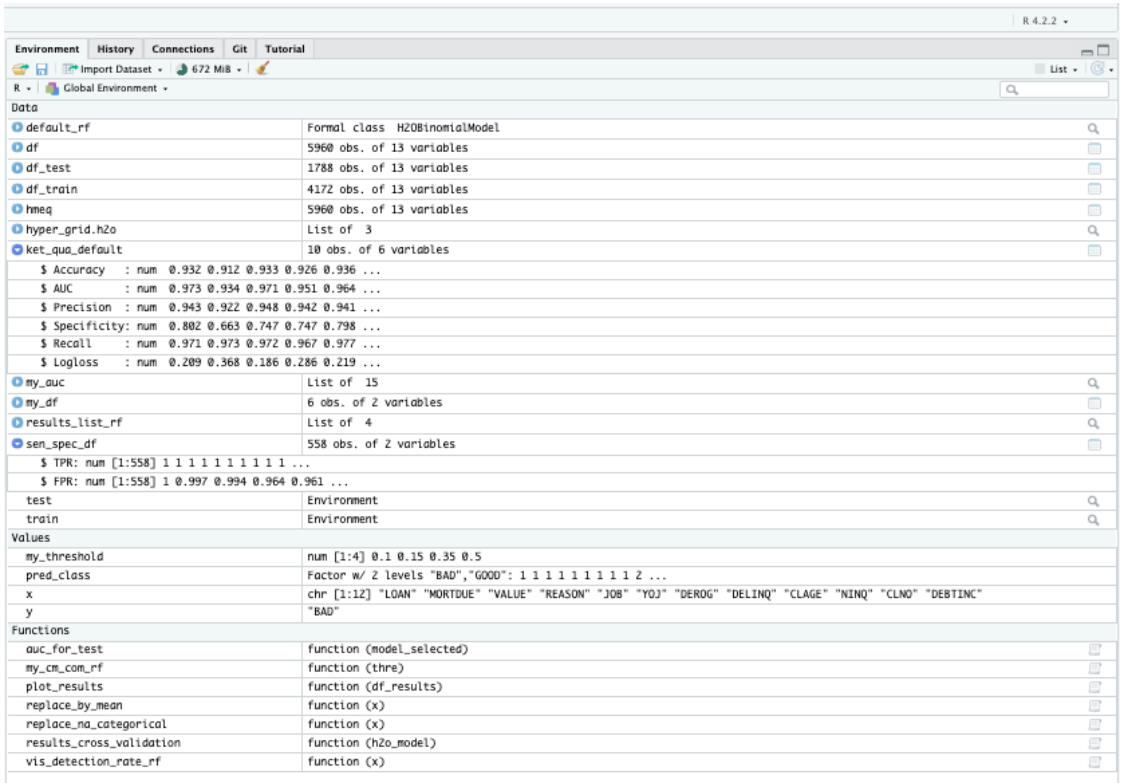
```
# Set hyperparameter grid:

hyper_grid.h2o <- list(ntrees = seq(50, 500, by = 50),
                       mtries = seq(3, 5, by = 1),
                       # max_depth = seq(10, 30, by = 10),
                       # min_rows = seq(1, 3, by = 1),
                       # nbins = seq(20, 30, by = 10),
                       sample_rate = c(0.55, 0.632, 0.75))

# The number of models is 90:
sapply(hyper_grid.h2o, length) %>% prod()
```

- **COMPROBACIÓN**

```
>hyper_grid.h2o <- list(ntrees = seq(50, 500, by = 50),
+                       mtries = seq(3, 5, by = 1),
+                       # max_depth = seq(10, 30, by = 10),
```

```
+                 # min_rows = seq(1, 3, by = 1),
+                 # nbins = seq(20, 30, by = 10),
+                 sample_rate = c(0.55, 0.632, 0.75))
>#The number of models is 90:
>sapply(hyper_grid.h2o, length) %>% prod()
[1] 90
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



- **MODELOS RANDOM ALEATORIOS**

  - **MODELACIÓN**

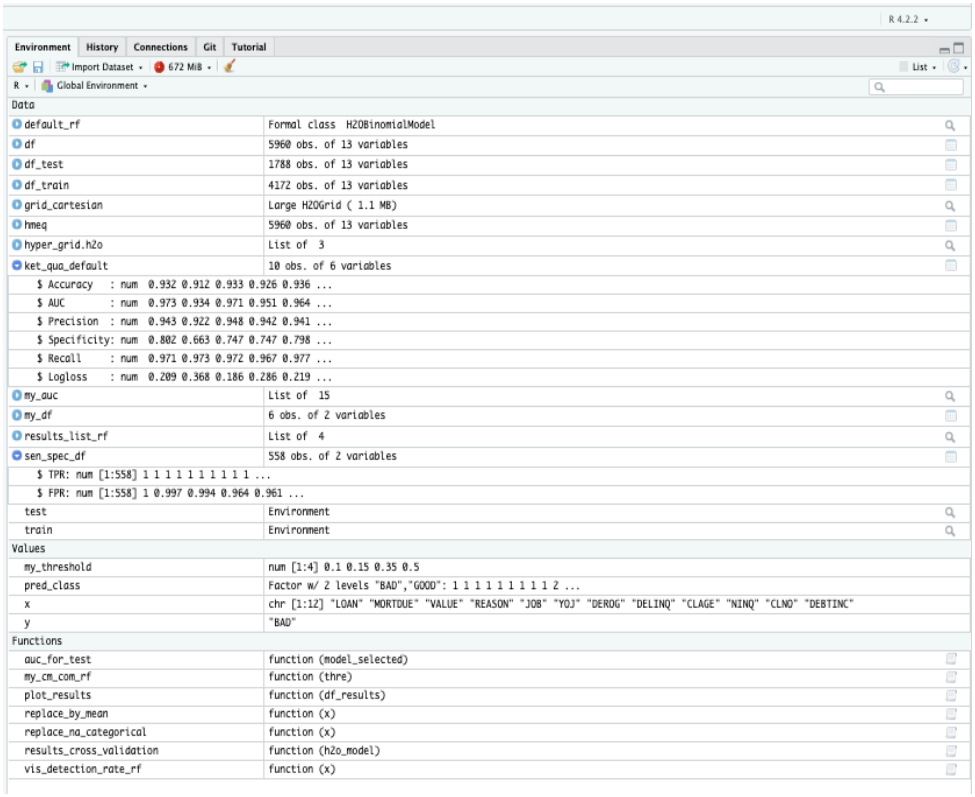
```
h2o.init()

# Train 6000 Random Forest Models:
system.time(grid_cartesian <- h2o.grid(algorithm = "randomForest",
                                       grid_id = "rf_grid1",
                                       x = x,
                                       y = y,
```

```
seed = 29,
nfolds = 10,
training_frame = train,
stopping_metric = "AUC",
hyper_params = hyper_grid.h2o,
search_criteria = list(strategy = "Cartesian")))
```

- **COMPROBACIÓN**

```
>h2o.init()
 Connection successful!

R is connected to the H2O cluster:
   H2O cluster uptime:         24 seconds 731 milliseconds
   H2O cluster timezone:       UTC
   H2O data parsing timezone: UTC
   H2O cluster version:        3.38.0.1
   H2O cluster version age: 4 months and 8 days !!!
   H2O cluster name:           H2O_started_from_R_r1866432_xen554
   H2O cluster total nodes:    1
   H2O cluster total memory: 0.18 GB
   H2O cluster total cores:    1
   H2O cluster allowed cores: 1
   H2O cluster healthy:        TRUE
   H2O Connection ip:          localhost
   H2O Connection port:        54321
   H2O Connection proxy:       NA
   H2O Internal Security:      FALSE
   R Version:                  R version 4.2.2 (2022-10-31)

Warning message:
In h2o.clusterInfo() :
Your H2O cluster version is too old (4 months and 8 days)!
Please download and install the latest version from http://h2o.ai/download/
>
># Train 6000 Random Forest Models:
>system.time(grid_cartesian <- h2o.grid(algorithm = "randomForest",
+                                       grid_id = "rf_grid1",
+                                       x = x,
+                                       y = y,
+                                       seed = 29,
+                                       nfolds = 10,
+                                       training_frame = train,
+                                       stopping_metric = "AUC",
+                                       hyper_params = hyper_grid.h2o,
+                                       search_criteria = list(strategy = "Cartesian")))
  |=======================================================
==================================================| 100%
   user system elapsed
  0.094   0.004   0.644
Warning message:
In h2o.getGrid(grid_id = grid_id) :
  Some models were not built due to a failure, for more details run `summary(grid_o
bject, show_stack_traces = TRUE)`
```

- **DATA UTILIZADA EN LA SIMULACIÓN**



R 4.2.2

Environment | History | Connections | Git | Tutorial

Import Dataset | 672 MiB | List

R | Global Environment

Data	
default_rf	Formal class H2OBinomialModel
df	5960 obs. of 13 variables
df_test	1788 obs. of 13 variables
df_train	4172 obs. of 13 variables
grid_cartesian	Large H2OGrid (1.1 MB)
hmeq	5960 obs. of 13 variables
hyper_grid.h2o	List of 3
ket_qua_default	10 obs. of 6 variables
$ Accuracy : num	0.932 0.912 0.933 0.926 0.936 ...
$ AUC : num	0.973 0.934 0.971 0.951 0.964 ...
$ Precision : num	0.943 0.922 0.948 0.942 0.941 ...
$ Specificity: num	0.802 0.663 0.747 0.747 0.798 ...
$ Recall : num	0.971 0.973 0.972 0.967 0.977 ...
$ Logloss : num	0.209 0.368 0.186 0.286 0.219 ...
my_auc	List of 15
my_df	6 obs. of 2 variables
results_list_rf	List of 4
sen_spec_df	558 obs. of 2 variables
$ TPR: num [1:558]	1 1 1 1 1 1 1 1 1 1 ...
$ FPR: num [1:558]	1 0.997 0.994 0.964 0.961 ...
test	Environment
train	Environment
Values	
my_threshold	num [1:4] 0.1 0.15 0.35 0.5
pred_class	Factor w/ 2 levels "BAD","GOOD": 1 1 1 1 1 1 1 1 1 2 ...
x	chr [1:12] "LOAN" "MORTDUE" "VALUE" "REASON" "JOB" "YOJ" "DEROG" "DELINQ" "CLAGE" "NINQ" "CLNO" "DEBTINC"
y	"BAD"
Functions	
auc_for_test	function (model_selected)
my_cm_com_rf	function (thre)
plot_results	function (df_results)
replace_by_mean	function (x)
replace_na_categorical	function (x)
results_cross_validation	function (h2o_model)
vis_detection_rate_rf	function (x)