





INFOTEC CENTRO DE INVESTIGACIÓN E  
INNOVACIÓN EN TECNOLOGÍAS DE LA  
INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y  
CONOCIMIENTO  
GERENCIA DE CAPITAL HUMANO  
POSGRADOS

# “Clasificación de Fondos de Aseguramiento Agropecuario y Rural mediante Inteligencia Artificial”

SOLUCIÓN ESTRATÉGICA EMPRESARIAL  
Que para obtener el grado de MAESTRO EN  
CIENCIA DE DATOS E INFORMACIÓN

Presenta:

**Nayeli Galván Palmerín**

Asesor:

**Dr. José Luis Jiménez Márquez**

Ciudad de México, julio 2025



**Biblioteca INFOTEC**

Ciudad de México, a 12 de agosto de 2025

## **VISTO BUENO DE TRABAJO TERMINAL**

**Maestría en Ciencia de Datos e Información  
(MCDI)**

**UNIDAD DE POSGRADOS  
PRESENTE**

Por medio de la presente se hace constar que el trabajo de titulación:

**“Clasificación de Fondos de Aseguramiento Agropecuario y Rural mediante Inteligencia Artificial”**

Desarrollado por la alumna: **Nayeli Galván Palmerín**, bajo la asesoría del **Dr. José Luis Jiménez Márquez**, cumple con el formato de Biblioteca, así mismo, se ha verificado la correcta citación para la prevención del plagio; por lo cual, se expide la presente autorización para entrega en digital del proyecto terminal al que se ha hecho mención. Se hace constar que la alumna no adeuda materiales de la biblioteca de INFOTEC.

No omito mencionar, que se deberá anexar la presente autorización al inicio de la versión digital del trabajo referido, con el fin de amparar la misma.

Sin más por el momento, aprovecho la ocasión para enviar un cordial saludo.

Dr. Juan Antonio Vega Garfias  
Subgerente de Innovación Gubernamental

JAVG/amr

C.c.p. Mtra. Anely Mendoza Rosales. – Encargada de la Gerencia de Capital Humano. - Para su conocimiento.  
Nayeli Galván Palmerín. – Alumna de la Maestría Ciencia de Datos e Información. – Para su conocimiento.



## AUTORIZACIÓN PARA PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL

Yo, **Nayeli Galván Palmerín**, alumna de la **Maestría en Ciencia de Datos e Información**, con matrícula **212DBD3587** por medio de la presente **autorizo a INFOTEC a publicar y difundir en el repositorio institucional** mi trabajo titulado:

**“Clasificación de Fondos de Aseguramiento Agropecuario y Rural mediante Inteligencia Artificial”**

En formato digital, con fines académicos y de investigación, de conformidad con la normatividad institucional.

Declaro que el contenido de la tesis es original y de mi autoría, y autorizo su consulta, descarga y reproducción parcial, siempre y cuando se cite la fuente.

Ciudad de México, a 29 de agosto de 2025.



---

Nayeli Galván Palmerín

# Tabla de contenido

Introducción .....	1
Contexto .....	2
Capítulo 1. Planteamiento del problema.....	5
1.1 Objetivos .....	12
1.1.1 Objetivo General.....	12
1.1.2 Objetivos Particulares.....	12
1.2 Hipótesis.....	12
1.3 Alcances y limitaciones.....	13
Capítulo 2. Conceptos previos.....	15
2.1 Aprendizaje automático y profundo .....	15
2.2 Aprendizaje automático.....	15
2.2.1 Modelo de aprendizaje no supervisado.....	16
2.2.1.1 K-medias.....	16
2.2.2 Modelos de aprendizaje supervisado .....	18
2.2.2.1 K vecinos más próximos (K-NN).....	18
2.2.2.2 Árboles de decisión.....	19
2.2.2.3 Máquina de vector de soporte (SVM).....	22
2.3 Aprendizaje profundo .....	26
2.3.1 Perceptrón multicapa .....	26
Capítulo 3. Objeto de estudio .....	33
3.1 Datos recolectados.....	33
Capítulo 4. Metodología.....	39
4.1 Comprensión del negocio.....	39
4.2 Comprensión de los datos.....	40
4.3 Preparación o preprocesamiento de datos.....	40
4.4 Modelado.....	41
4.5 Evaluación.....	41
Capítulo 5. Diseño de estrategias y resultados.....	44
5.1 Clasificación con k-medias .....	44
5.1.2 Métricas k-medias.....	49
5.2 Bosque aleatorio .....	55

5.2.1 Selección de parámetros .....	55
5.2.2 Resultados .....	56
5.3 Máquina de vector de soporte (SVM) .....	58
5.3.1 Kernel lineal.....	59
5.3.2 Kernel polinómico .....	59
5.3.3 Kernel radial o RBF.....	59
5.3.4 Kernel sigmoide.....	59
5.3.5 Selección de parámetros .....	60
5.3.6 Resultados.....	60
5.4 K vecinos más próximos (K-NN).....	62
5.4.1 Métricas .....	62
5.4.2 Selección de parámetros .....	62
5.4.3 Resultados.....	63
5.5. Perceptrón multicapa – Clasificador MLP .....	64
5.5.1 Selección de parámetros .....	65
5.5.2 Resultados.....	66
5.6 Comparación entre modelos .....	67
Conclusiones.....	69
<b>Clasificación de Fondos de Aseguramiento Agropecuario y Rural mediante Inteligencia Artificial: Caso Práctico 2024.....</b>	<b>71</b>
Bibliografía.....	75

## Índice de Ilustraciones

Ilustración 1 División de la Inteligencia Artificial .....	15
Ilustración 2 División de datos por árboles de decisión .....	21
Ilustración 3 Separación lineal por SVM.....	24
Ilustración 4 Neurona: Relación entre las señales de entrada recibidas por los nodos y la señal de salida.....	29
Ilustración 5 Separación lineal por perceptrón .....	30

## Índice de gráficos

Gráfico 1 Mapa de calor de los datos recolectados.....	36
Gráfico 2 Codo K-medias.....	44
Gráfico 3 Clasificación k-medias .....	45
Gráfico 4 Clasificación método actual .....	48
Gráfico 5 Comparativa metodología tradicional vs k-medias grupo 1.....	51
Gráfico 6 Comparativa metodología tradicional vs k-medias grupo 2.....	52
Gráfico 7 Comparativa metodología tradicional vs k-medias grupo 3.....	54
Gráfico 8 Exactitud vs profundidad bosque aleatorio .....	57
Gráfico 9 Clasificación bosque aleatorio .....	58
Gráfico 10 Clasificación SVM.....	61
Gráfico 11 Precisión vs cantidad de vecinos K-NN .....	63
Gráfico 12 Clasificación K-NN.....	64
Gráfico 13 Clasificación MLP .....	67
Gráfico 14 Caso práctico 2024 perceptrón multicapa.....	74
Gráfico 15 Caso práctico 2024 metodología tradicional .....	74

## Índice de cuadros

Tabla 1 Indicadores de riesgo financiero de Fondos de Aseguramiento .....	5
Tabla 2 División de indicadores de riesgo financiero.....	7
Tabla 3 Decrementos de impurezas .....	22
Tabla 4 Estadística datos recolectados.....	34
Tabla 5 Matriz de Confusión .....	42
Tabla 6 Contribución por variable a los componentes principales .....	46
Tabla 7 Matriz de confusión de riesgos k-medias .....	49
Tabla 8 Evaluación entre k-medias y el método actual .....	50
Tabla 9 Estadística grupo 1 .....	50
Tabla 10 Estadística grupo 2 .....	52
Tabla 11 Estadística grupo 3 .....	53
Tabla 12 Evaluaciones por método.....	68
Tabla 13 Comparativo clasificación con IA vs Excel .....	71

## Índice de diagramas

Diagrama 1 Modelos para la clasificación de la información financiera de Fondos de Aseguramiento.....	31
---	----

## Siglas y abreviaturas

**ARI:** Índice de Rand Ajustado.

**CNSF:** Comisión Nacional de Seguros y Fianzas.

**CRISP-DM:** Cross-Industry Standard Process for Data Mining (Proceso Estándar Interindustrial para Minería de Datos).

**CUSF:** Circular Única de Seguros y Fianzas.

**K-NN:** K vecinos más cercanos.

**LFAAR:** Ley de Fondos de Aseguramiento Agropecuario y Rural.

**MLP:** Multilayer Perceptron (Perceptrón Multicapa).

**NP-hard:** Nondeterministic Polynomial Time (Tiempo Polinomial No Determinista).

**PCA:** Principal Component Analysis (Análisis de Componentes Principales).

**SVM:** Support Vector Machine (Máquinas de Vector de Soporte)

**WCSS:** Within Cluster Sum of Squares (Suma de los Cuadrados Dentro de Cada Grupo).

## Glosario

**Cobertura de reservas técnicas:** Las reservas técnicas representan las provisiones necesarias para hacer frente a los riesgos asumidos con los asegurados. Dichas reservas deben ser respaldadas con inversiones que cumplen con condiciones adecuadas de seguridad, rentabilidad y liquidez. Es importante mencionar que estas inversiones deben mantenerse colocadas en todo momento conforme a la regulación aplicable (CNSF, s.f.-a).

**Indicadores financieros:** Relaciones entre cuentas de estados financieros de una empresa que permiten realizar análisis financieros para la toma de decisiones.

**Índice de precios al consumidor:** Mide la variación de los precios de una canasta de bienes y servicios representativos del consumo de los hogares mexicanos (INEGI, s.f.).

**Deudor por prima:** Importe de las primas, recargos y derechos pendientes de cobro.

**Disponibilidad:** Monto de efectivo, valor de documentos de cobro inmediato e importe de cuentas de cheque en moneda nacional o extranjera valuada al tipo de cambio del cierre de mes, publicado por el Banco de México (CNSF, 2019).

**Mutualidades:** Asociaciones de personas que, sin dar pólizas o contratos, conceden a sus miembros coberturas en circunstancias de muerte, beneficios en caso de accidentes y enfermedades o indemnización por daños, entre otros (CONDUSEF, s.f.).

**Prima:** Aportación económica que ha de satisfacer el contratante o asegurado a la entidad aseguradora en concepto de contraprestación por la cobertura de riesgo que este le ofrece (MAPFRE, s.f.).

**Prima cedida:** Proporción de la prima emitida que el asegurador transfiere al reasegurador por compartir el riesgo. Es decir, es la prima que el asegurador paga al reasegurador en contraprestación del riesgo asumido por este (Agroasemex, 2019).

**Prima emitida:** Monto total de ingresos que recibe una compañía de seguros por la venta de sus productos, es decir, el monto correspondiente a las pólizas suscritas por una aseguradora (Agroasemex, 2019).

**Prima en retención:** Proporción de la prima emitida, que no se transfiere a un reasegurador, es la parte del riesgo que la aseguradora asume con sus recursos (Agroasemex, 2019).

**Reaseguro:** Es un acuerdo mediante el cual un asegurador, denominado cedente, transfiere a otro asegurador, denominado reasegurador, la totalidad o una parte de sus riesgos y este recibe una parte de la prima (Fundación Mapfre, s.f.).

**Reserva de riesgos en curso:** Son las reservas para cubrir el valor esperado de las obligaciones futuras derivadas del pago de siniestros.

**Reserva especial de contingencia:** Los recursos de esta reserva podrán ser comprometidos para incrementar su retención de riesgos y negociar los términos y modalidades de la cobertura de reaseguro, así como realizar aportaciones al Fondo de Retención Común de Riesgos conforme a lo previsto en los artículos 75 y 84 de la LFAAR (Ley de Fondos de Aseguramiento Agropecuario y Rural, 2005).

**Reservas técnicas:** La constitución de provisión para hacer frente a contingencias futuras previstas en la Ley de Fondos de Aseguramiento Agropecuario y Rural.

**Resultado del ejercicio:** Está constituido por los ingresos del ejercicio menos los gastos de este, así como por los beneficios y quebrantos no relacionados con la actividad de la entidad. (Comisión Nacional de Mejora Regulatoria, 2024)

**Seguro:** Medio eficaz para la protección de los individuos frente a las consecuencias de los riesgos. Se basa en transferirlos a una aseguradora que se encarga de reparar o indemnizar todo o parte del perjuicio producido por la aparición de determinadas situaciones accidentales (Fundación Mapfre, s.f.).

**Seguro agrícola:** Los seguros agrícolas forman parte de la gestión de riesgos para la producción agropecuaria. En particular, contribuyen a gestionar los peligros inherentes a la cadena de valor alimentario agropecuaria, a estabilizar los ingresos agropecuarios y fomentan la inversión (Ramírez Aldana, 2017).

**Seguro directo:** Esta expresión se utiliza normalmente para hacer referencia a las primas obtenidas mediante contratación directa, y distinguiéndolas en este sentido de las recaudadas en concepto de reaseguro (MAPFRE, s.f.).

**Siniestros:** Cualquier materialización de un evento cubierto por el seguro o reaseguro, por ejemplo, una granizada (Agroasemex, 2019).

**Valores:** Importe de las inversiones efectuadas en valores emitidos por el Gobierno Federal y por empresas privadas.

## Introducción

“La Comisión Nacional de Seguros y Fianzas es un Órgano Desconcentrado de la Secretaría de Hacienda y Crédito Público, encargada de supervisar que la operación del sector asegurador y afianzador se apegue al marco normativo, preservando la solvencia y estabilidad financiera de estas instituciones” (CNSF, s.f.-b). Entre las entidades supervisadas se encuentran los Fondos de Aseguramiento Agropecuario y Rural, siendo sociedades con el objeto de ofrecer protección mutualista y solidaria a sus socios a través de operaciones de seguros del ramo agrícola y de animales entre otros (Ley de Fondos de Aseguramiento Agropecuario y Rural, 2005). Para procurar el apego a la normativa, la CNSF solicita los llamados Reportes Regulatorios, en específico para los Fondos de Aseguramiento, el RR-13, el cual consta de un conjunto de información financiera, técnica y operativa requerida a las instituciones en formato no estructurado.

Actualmente, se cuenta con un registro de más de 500 Fondos de Aseguramiento, aproximadamente el 50% de ellos presentan el reporte regulatorio; a pesar de ello, el volumen de datos es alto y ofrece la oportunidad para diversos análisis desde nuevos enfoques, como el modelado de las condiciones financieras.

El objetivo de este trabajo es proponer alternativas al proceso actual que lleva a cabo la CNSF para clasificar a los Fondos de Aseguramiento según su riesgo financiero, mediante técnicas de aprendizaje automático clásico y aprendizaje profundo, utilizando información al cierre del 2020 al 2023.

## Contexto

En 2006, Tapia señaló que entre las actividades productivas con mayor exposición a riesgos naturales y biológicos son las actividades agropecuarias con eventos como sequías, heladas, inundaciones, granizadas, plagas, entre otros; las pérdidas ocasionadas por este tipo de sucesos tienen un efecto directo en el ingreso y patrimonio de los productores e indirectamente tiene un costo para el país, teniendo un impacto en el índice de precios al consumidor.

A partir de la década de los cincuenta se concibió el concepto del seguro agrícola en México a consecuencia del incremento de la inversión pública en el área con la “Reforma Agraria Integral”, que entre sus diversos objetivos se encontraban que los productores tuvieran acceso a servicios financieros como el crédito y el seguro de producción agropecuaria, a su vez se estaban creando mutualidades como la Federación de Sociedades Mutualistas del Seguro Agropecuario.

Actualmente, en México, existen tres principales proveedores de seguros agrícolas y de animales: Agroasemex (institución nacional de seguros especializada en el sector rural), compañías aseguradoras privadas y los Fondos de Aseguramiento. Estos últimos, conocidos como sociedades, surgieron en la década de los setenta con el propósito de ofrecer protección mutualista y solidaria a sus socios mediante operaciones de seguros agrícolas y de animales, entre otros, tal como lo establece la definición de la Ley de Fondos de Aseguramiento Agropecuario y Rural.

Los Fondos de Aseguramiento fueron regulados hasta 1992 mediante la emisión de las reglas generales para su constitución, operación y funcionamiento, estas fueron emitidas por la Secretaría de Hacienda y Crédito Público, sin embargo, fue a finales de 2017 que el Diario Oficial de la Federación publicó las atribuciones de la Dirección General de Supervisión de Fondos de Aseguramiento, área perteneciente a la CNSF; entre las responsabilidades asignadas a esta dirección se encuentra la supervisión de diversos aspectos, como la contabilidad, la presentación y

aprobación de estados financieros, entre otros asuntos. El objetivo principal detrás de estas funciones es mantener la solvencia y estabilidad financiera de los Fondos, con miras a beneficiar a los socios que conforman estas sociedades.



**Capítulo 1**  
**Planteamiento del problema**

## Capítulo 1. Planteamiento del problema

Para el cumplimiento de sus funciones las entidades supervisoras y regulatorias requieren de una plantilla con expertos en el ámbito jurídico y técnico del ramo supervisado, enfocándose en el aspecto teórico-práctico de los seguros y fianzas, el supervisor utiliza métodos estadísticos, probabilísticos y computacionales que van de lo simple a lo complejo para replicar cálculos esenciales que en teoría realiza la institución supervisada para preservar la solvencia y reservas, asimismo la entidad regulatoria realiza y comparte análisis del mercado. En su mayoría, los métodos utilizados para la supervisión de entidades no son deterministas, requieren de análisis variados, calibración y comparación con otras metodologías.

La clasificación de riesgos financieros de los Fondos de Aseguramiento se basa en el resultado de los siguientes indicadores calculados con la información reportada de acuerdo con el catálogo contable definido para estas instituciones (el cual se encuentra detallado en el anexo 22.6.1 de la CUSF):

*Tabla 1 Indicadores de riesgo financiero de Fondos de Aseguramiento*  
Nota: Tabla de elaboración propia

Indicador	Definición	Interpretación
Cobertura de reservas técnicas	$\frac{\text{Valores} + \text{Inmobiliarias} + \text{Cheques en moneda extranjera} + \text{Deudor por prima} + \text{Participación de reaseguro en siniestros pendientes}}{\text{Reservas técnicas}}$	Las reservas técnicas representan las provisiones necesarias para hacer frente a los riesgos asumidos con los asegurados. Dichas reservas deben ser respaldadas con inversiones que cumplen con condiciones adecuadas de seguridad, rentabilidad y liquidez. Es importante mencionar que estas inversiones deben mantenerse colocadas en todo momento conforme a la regulación aplicable. Cuando este índice es mayor o igual a uno significa que las inversiones cubren las reservas técnicas y que la institución mantiene recursos suficientes para respaldar sus obligaciones (CNSF, s.f.-a).
Disponible e inversiones para cobertura de reservas técnicas	$\frac{\text{Valores} + \text{Inmobiliarias} + \text{Cheques en moneda extranjera} + \text{Deudor por prima} + \text{Participación de reaseguro en siniestros pendientes} + \text{Disponibilidad}}{\text{Reservas técnicas}}$	Bajo la premisa anterior, se agrega el rubro de "Disponibilidad" para los casos en que las inversiones no son suficientes para cubrir las reservas técnicas, el índice muestra si la institución cuenta con los recursos suficientes para respaldar sus obligaciones, el indicador debe ser mayor a uno.

<b>Indicador</b>	<b>Definición</b>	<b>Interpretación</b>
Liquidez	$\frac{\text{Activo circulante}}{\text{Pasivo circulante}}$	Mide la capacidad del Fondo para pagar obligaciones financieras a corto plazo, el indicador debe ser mayor a uno.
Estructura Financiera	$\frac{\text{Capital}}{\text{Pasivo}}$	Mide la suficiencia del capital para enfrentar obligaciones, se espera que el indicador sea mayor al 25%.
Costo de Adquisición	$\frac{\text{Costo Neto de Adquisición}}{\text{Primas de retención}}$	En el caso de los Fondos de Aseguramiento, los costos de adquisición es el desembolso realizado para la contratación de reaseguro y se espera que sean menor o igual al 25% respecto a la prima retenida.
Gastos de Operación	$\frac{\text{Gastos de Operación Netos}}{\text{Primas Emitidas}}$	Los gastos operativos del Fondo de Aseguramiento deben ser menor o igual al 20% respecto a la prima emitida.
Costo de siniestralidad	$\frac{\text{Costo Neto de Siniestralidad}}{\text{Primas de retención devengadas}}$	El costo neto de siniestralidad debe ser menor o igual al 25% respecto a la prima de retención devengada.
Porcentaje del resultado del ejercicio respecto a la prima	$\frac{\text{Resultado del ejercicio}}{\text{Primas emitidas}}$	Proporción del resultado del ejercicio respecto a la prima emitida, se espera que si las primas son positivas lo sea el resultado.
Porcentaje de reserva de riesgos en curso respecto a la prima	$\frac{\text{Reserva de riesgos en curso}}{\text{Primas de retención}}$	Proporción de la reserva de riesgos en curso respecto a la prima de retención, se espera que sea mayor o igual al 10%.

Para la clasificación de la información financiera, los indicadores se dividen en dos categorías: "Indicadores relevantes" e "Indicadores técnicos, financieros y de reaseguro", según se detalla a continuación:

*Tabla 2 División de indicadores de riesgo financiero*  
 Nota: Tabla de elaboración propia

<b>Indicadores relevantes</b>	<b>Indicadores técnicos, financieros y de reaseguro</b>
Cobertura de reservas	Costo de adquisición
Disponibilidad y cobertura	Costo de operación
Liquidez	Costo de siniestralidad
Estructura financiera	Porcentaje del resultado del ejercicio respecto a la prima
	Porcentaje de reserva de riesgos en curso respecto a la prima

Los "Indicadores relevantes" poseen gran importancia, ya que en su cálculo se incorporan las reservas técnicas, que son provisiones esenciales para enfrentar los riesgos contraídos con los socios, asimismo, evalúan la liquidez y la estructura financiera de los Fondos, proporcionando información crucial sobre la capacidad de cumplimiento de dichas instituciones. Por otro lado, los "Indicadores técnicos, financieros y de reaseguro", también suministran información valiosa en relación con los costos y los resultados de las operaciones, sin embargo, los "Indicadores relevantes" reciben un trato diferenciado debido a su mayor trascendencia.

A continuación, se describe el proceso de clasificación de las instituciones analizadas en categorías de "Riesgo Bajo", "Riesgo Medio" y "Riesgo Alto", utilizando ambos tipos de indicadores, cabe aclarar que este procedimiento contiene modificaciones respecto al utilizado por los supervisores de la CNSF, agregando y

descartando supuestos, con el fin de no revelar en su totalidad la metodología, así como los datos de las instituciones, ya que no son de dominio público:

- Para fines prácticos, se definen las categorías de riesgo con variables numéricas, siendo 1=Riesgo Bajo, 2= Riesgo Medio y 3=Riesgo Alto.
- Asignación de categoría para “Indicadores relevantes”:

- Cobertura de reservas

Si  $cobertura_{reservas} < 0.95$  la categoría es 3 (Riesgo Alto)

Si  $0.95 \leq cobertura_{reservas} < 0.99$  la categoría es 2 (Riesgo Medio)

Si  $cobertura_{reservas} \geq .99$  la categoría es 1 (Riesgo Bajo)

- Disponibilidad y cobertura

Si  $cobertura_{reservas} \geq .99$  (indicador anterior) la categoría es 1 (Riesgo Bajo)

Si  $cobertura_{reservas} < .99$  y  $disponibilidad_{cobertura} \geq 1$  la categoría es 2 (Riesgo Medio)

Si  $cobertura_{reservas} < .99$  y  $disponibilidad_{cobertura} < 1$  la categoría es 3 (Riesgo Alto)

- Liquidez

Si  $liquidez < .95$  la categoría es 3 (Riesgo Alto)

Si  $liquidez \geq .95$  y  $liquidez < .99$  la categoría es 2 (Riesgo Medio)

Si  $liquidez \geq .99$  la categoría es 1 (Riesgo Bajo)

- Estructura Financiera

Si  $Estructura_{financiera} < .25$  la categoría es 3 (Riesgo Alto)

Si  $Estructura_{financiera} \geq .25$  y  $Estructura_{financiera} \leq .50$  la categoría es 2 (Riesgo Medio)

Si  $Estructura_{financiera} > .50$  la categoría es 1 (Riesgo Bajo)

- Pasos para la asignación de categoría “Indicadores técnicos, financieros y de reaseguro”:

Paso 1: Se otorgan 0 o 2 puntos a los indicadores conforme los criterios descritos a continuación:

- Costo de adquisición

Si  $Costo_{adquisicion} \leq .25$  se otorgan 2 puntos al indicador, 0 en caso contrario

- Costo de siniestralidad

Si  $Gastos_{operacion} \leq .20$  se otorgan 2 puntos al indicador, 0 en caso contrario

- Porcentaje del resultado del ejercicio respecto a la prima

Si  $Resultado_{prima} \geq 0$  se otorgan 2 puntos al indicador, 0 en caso contrario

- Porcentaje de reserva de riesgos en curso respecto a la prima

Si  $RRC_{prima} \geq .10$  se otorgan 2 puntos al indicador, 0 en caso contrario

Paso 2: Sumar el total de puntos y dividir entre 10, cantidad que se denominará  $Promedio_{ITFR}$ .

Paso 3: Se aplican las siguientes condiciones:

Si  $Promedio_{ITFR} < .50$  la categoría es 2 (Riesgo Medio)

Si  $Promedio_{ITFR} \geq .50$  la categoría es 1 (Riesgo Bajo)

- **Asignación de categoría de riesgo definitiva** de la información financiera del Fondo de Aseguramiento:

$$Etapa = \max (Categoría_{IndicadoresRelevantes}, Categoría_{ITFR})$$

Donde:

- Etapa: Categoría de riesgo financiero
- $Categoría_{IndicadoresRelevantes}$ : Categoría de los “Indicadores Relevantes”
- $Categoría_{ITFR}$ : Categoría de los “Indicadores Técnicos, Financieros y de Reaseguro”

Dado lo expuesto, la evaluación del riesgo financiero de los Fondos de Aseguramiento implica realizar cálculos aparentemente sencillos, pero que

requieren la consideración de una variedad considerable de criterios y condiciones, asimismo, esta evaluación se ajusta de manera constante con base en las observaciones derivadas de la práctica.

En este contexto, este trabajo propone la implementación de enfoques metodológicos innovadores con el propósito de reducir la cantidad de supuestos, realizando la clasificación con métodos de aprendizaje automático y profundo, evaluando la viabilidad de continuar con el procedimiento actual mediante una comparativa entre los resultados obtenidos mediante las técnicas recién mencionadas y los resultados generados por el enfoque tradicional.

## **1.1 Objetivos**

### **1.1.1 Objetivo General**

Obtener alternativas al proceso actual que lleva a cabo la CNSF para clasificar a los Fondos de Aseguramiento según su riesgo financiero, mediante técnicas de aprendizaje automático clásico y aprendizaje profundo utilizando información al cierre del 2020 al 2023, así como realizar un comparativo para conocer el modelo que mejor se adapta a la clasificación.

### **1.1.2 Objetivos Particulares**

1. Concentrar y limpiar la información financiera al cierre del 2020 al 2023.
2. Calcular los indicadores financieros.
3. Analizar las variables más significativas.
4. Aplicar k-medias para la clasificación de la información.
5. Comparar la clasificación actual respecto a k-medias.
6. Aplicar métodos de aprendizaje supervisado y no supervisado con base en la clasificación obtenida con k-medias.
7. Aplicar modelo de red neuronal.
8. Comparar métodos y concluir cuál es el mejor para el caso de la clasificación de información financiera de Fondos de Aseguramiento.

## **1.2 Hipótesis**

Este trabajo parte del supuesto de que la agrupación de indicadores financieros puede ser estudiada con técnicas de aprendizaje automático clásico y aprendizaje profundo, siendo el principal objetivo el generar grupos bien delimitados para la toma de decisiones, para preservar la solvencia y estabilidad financiera de los Fondos de Aseguramiento.

Por lo tanto, se espera que los datos de cada agrupación cuenten con mayor homogeneidad, que los algoritmos de aprendizaje automático o profundo presenten mejores resultados que el método tradicional y las etiquetas obtenidas sean utilizadas para la clasificación de nuevos datos.

### **1.3 Alcances y limitaciones**

Las agrupaciones para clasificar indicadores financieros desarrolladas en este trabajo se aplican a las asociaciones mexicanas denominadas Fondos de Aseguramiento Agropecuario y Rural, así mismo se identifica la mejor metodología de aprendizaje automático clásico y/o profundo para dicha clasificación, así como la viabilidad del método actual.

A pesar de contar con información financiera de las instituciones analizadas desde el 2018, por los cambios en la estructura del reporte y por las características de los Fondos de Aseguramiento, la información con mejor calidad disponible es la de los periodos 2020 al 2023. Aunque existen mejoras en la presentación de la información, los Fondos de Aseguramiento cuentan con áreas de oportunidad al reportar, no apegándose a la normativa y manuales de manera involuntaria, lo cual podría conllevar alteraciones en las metodologías de agrupación y requerir un mayor tiempo de análisis.

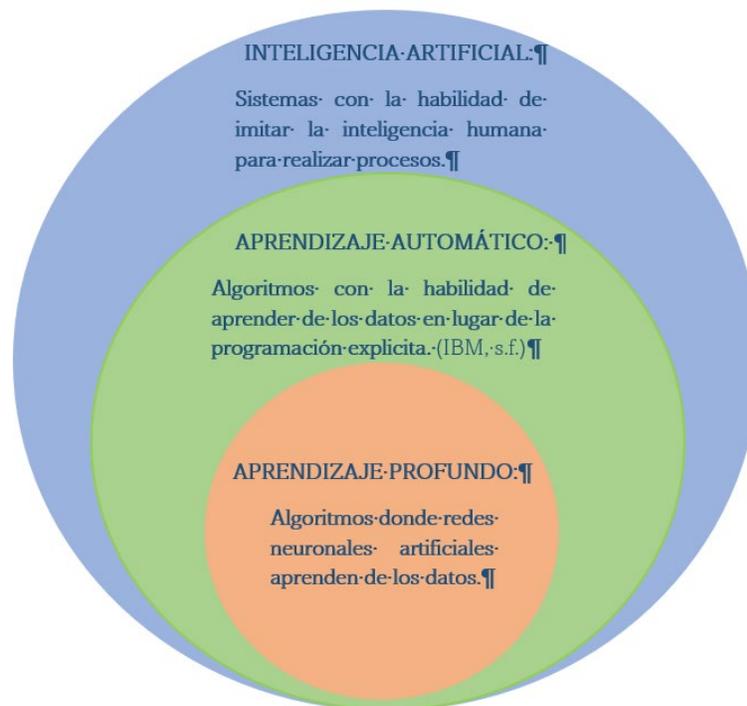


**Capítulo 2**  
**Conceptos previos**

## Capítulo 2. Conceptos previos

### 2.1 Aprendizaje automático y profundo

La inteligencia artificial es una disciplina que imita, a través de sistemas o máquinas, la inteligencia humana para realizar procesos, los cuales pueden ser mejorados iterativamente (Oracle, s.f.).



*Ilustración 1 División de la Inteligencia Artificial*

Nota: Ilustración de elaboración propia inspirada en Nuzzi et al. (2021)

### 2.2 Aprendizaje automático

El aprendizaje automático o aprendizaje computacional es una rama de la Inteligencia Artificial que estudia algoritmos capaces de aprender a partir de la identificación de patrones. Existen diferentes tipos de aprendizaje automático, los más comunes son: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo.

- **Aprendizaje supervisado:** Se utiliza para predecir un resultado específico a partir de una entrada dada. Específicamente, el modelo se construye a partir de pares de entrada y salida, conocidos como conjunto de entrenamiento. La entrada es un vector de características y la salida un conjunto de etiquetas. El objetivo es que el modelo realice predicciones precisas para nuevos datos. (Müller & Guido, 2017)
- **Aprendizaje no supervisado:** se tiene un conjunto de entradas y busca aprender la relación entre ellas, generando grupos o realizando transformaciones para localizar patrones.
- **Aprendizaje por refuerzo:** con este sistema de aprendizaje, llamado agente, se puede observar el entorno, seleccionar y realizar acciones, con la finalidad de obtener recompensas o castigos. Posteriormente, debe aprender por sí mismo cuál es la mejor estrategia, llamada política, para obtener la mayor recompensa. Con una política, se define qué acción debe hacer el agente cuando se encuentra en cierta situación (Valdemar Cuevas, Avalos, Primitivo, Valdivia, & Pérez, 2021).

## 2.2.1 Modelo de aprendizaje no supervisado

### 2.2.1.1 K-medias

El algoritmo k-medias pertenece a la categoría de agrupamiento basado en prototipos, lo que significa que cada grupo o clúster está representado por un prototipo, que puede ser tanto el centroide (promedio) de puntos similares con características continuas, como el medoide (el punto más representativo o que aparece con más frecuencia) en caso de las características categóricas (Rachka & Mirjalili, 2019).

El objetivo del algoritmo es agrupar los datos basándose en su semejanza de características y se puede resumir de la siguiente manera:

1. Seleccionar aleatoriamente  $k$  centroides.
2. Asignar cada dato al centroide más cercano.
3. Recalcular los centroides al centro de los datos asignados.
4. Repetir los pasos 2 y 3 hasta que las asignaciones de grupos no cambien o hasta conseguir una tolerancia definida por el usuario o llegar al número máximo de iteraciones.

De manera teórica, Shalev-Shwartz & Ben-David describen que los datos se dividen en conjuntos disjuntos  $C_1, \dots, C_k$  donde cada  $C_i$  es representado por un centroide  $\mu_i$ . Se asume que el conjunto de entrada  $X$  está embebido en un espacio métrico más grande  $(X', d)$  tal que  $(X \subseteq X')$  y los centroides pertenecen a  $X'$ . La función objetivo de  $k$ -medias mide la distancia al cuadrado entre cada punto  $x \in X$  y el centroide de su grupo.

El centroide de  $C_i$  se define como:

$$\mu_i(C_i) = \operatorname{argmin}_{\mu \in X'} \sum_{x \in C_i} d(x, \mu)^2$$

Entonces, el objetivo de  $k$ -medias es:

$$G_{k\text{-Medias}}((X, d), (C_1 \dots C_k)) = \min_{\mu_1, \dots, \mu_k \in X'} \sum_{i=1}^k \sum_{x \in C_i} d(x, \mu_i)^2$$

Sin embargo, resulta que encontrar la solución óptima de  $K$ -medias en ocasiones es computacionalmente inviable, debido a que se trata de un problema NP-hard. No hay algoritmo conocido que pueda encontrar la solución óptima en tiempo polinomial para todos los casos. Como alternativa se utiliza el siguiente algoritmo iterativo:

1. Inicialización: Se seleccionan los  $k$  centroides iniciales de manera aleatoria.
2. Asignación de clústeres: A cada punto  $x$  se le asigna el centroide  $\mu_i$  más cercano bajo la métrica seleccionada.
3. Recalculo de centroides: Para cada clúster se recalcula el centroide como la media de todos los puntos asignados a ese grupo. Si  $C_i$  es el conjunto de puntos asignados al clúster  $i$ , entonces el nuevo centroide  $\mu_i$  se calcula como:

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

## 2.2.2 Modelos de aprendizaje supervisado

### 2.2.2.1 K vecinos más próximos (K-NN)

El método de  $k$  vecinos más próximos (K-NN) es una técnica no paramétrica para clasificación, en donde se memorizan los grupos de datos que se clasificaron en la etapa de entrenamiento. Dada una nueva observación  $X_0$ , determinamos las distancias más cercanas de las  $N$  muestras  $X_1, \dots, X_N$  de entrenamiento con respecto a  $X_0$ ; entonces, la nueva observación será clasificada en el grupo  $K_i$  que tenga más puntos cercanos (Valdemar Cuevas, Avalos, Primitivo, Valdivia, & Pérez, 2021).

Resulta de gran importancia la elección de  $K$  para encontrar un buen equilibrio entre el ajuste de la muestra y la generalización o robustez de los resultados. Además, de seleccionar la métrica de distancia apropiada a las características del conjunto de datos. A menudo, la distancia euclidiana se utiliza para muestras de valores reales. Sin embargo, si se utiliza este tipo de distancia, es importante normalizar los datos

para que cada característica contribuya de forma equitativa en la distancia (Rachka & Mirjalili, 2019).

Entre las ventajas del algoritmo se encuentran la fácil implementación, adaptación, así como pocos hiperparámetros, por otro lado, sus desventajas son la necesidad de memoria de almacenamiento en comparación con otros clasificadores, tiende a ser propenso a sobreajustes, es decir a capturar el ruido y las fluctuaciones en los datos de entrenamiento, lo que resulta en un rendimiento pobre para los datos nuevos, así como a no funcionar bien con entradas de datos de alta dimensión.

### **2.2.2.2 Árboles de decisión**

Los árboles de decisión son una estructura de datos jerárquica y se encuentran conformados por nodos internos, donde se realizan operaciones y hojas, que indican la clase.

Valdemar Cuevas, Avalos, Primitivo, Valdivia, y Pérez indican que para clasificación son utilizados en problemas en los cuales los atributos de un objeto son usados para determinar la clase a la que pertenece. Para obtener un árbol de clasificación  $T$  es necesario definir un error  $E$  que cuantifique el desempeño de un nodo  $t$  en términos de su efectividad en la separación de los datos. Se refiere a este error como la función de impureza  $E(T)$ . Tal error produce un error mínimo de cero cuando los datos pertenecen a la misma clase y, contrariamente, arroja un error considerable cuando pertenecen a otra clase.

Una función de impureza realiza un mapeo de sus argumentos de entrada a un valor no negativo. La función de impureza se define considerando los resultados de cada nodo terminal del árbol, por lo que la impureza total está definida por:

$$E_E(T) = \sum_{t \in T} E(t)$$

Donde  $\tilde{T}$  es el conjunto de nodos hoja.

Los argumentos de entrada representan las probabilidades  $p_j$  de cada clase  $J$ . Las funciones más conocidas de impureza usadas por árboles de decisión son la función de entropía  $E_E(T)$  y el índice de Gini  $E_G(T)$ , definidas de la siguiente manera:

Entropía:

$$E_E(T) = - \sum_{j=1}^J p_j \ln(p_j)$$

Índice de Gini:

$$E_G(T) = \sum_{i \neq j} p_i p_j = 1 - \sum_{j=1}^J p_j^2$$

Para crear un árbol de decisión considerando las funciones de impureza, se debe calcular el coste de cada nodo. El objetivo radica en encontrar el valor óptimo que divide a las clases de los objetos, de tal modo que la reducción del costo sea el máximo. De esta forma, en un árbol binario, el cambio de impureza debido a la división se formula de la siguiente manera:

$$\Delta E(s, t) = E(t) - p_l E(t_l) - p_r E(t_r)$$

Donde

$t$  es el nodo actual en evaluación.

$E(t)$  es la impureza del nodo  $t$  antes de la división.

$t_l$  y  $t_r$  son los nodos hijo izquierdo y derecho, respectivamente.

$p_l$  y  $p_r$  es la probabilidad de la clase de cada uno de los lados.

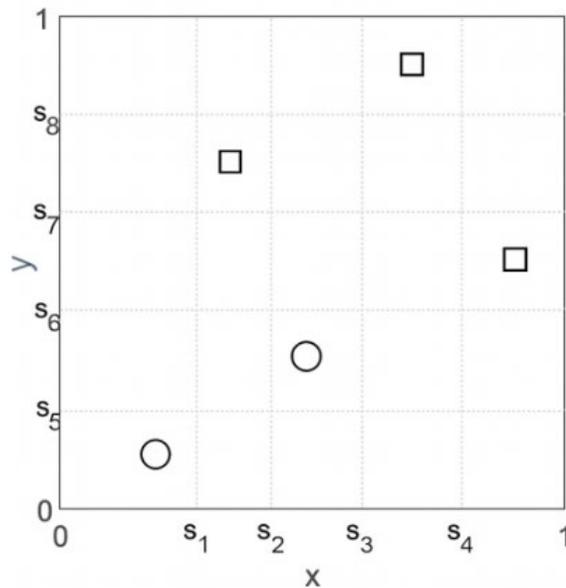
$s$  representa el punto de división dentro del nodo  $t$ .

Bajo este procedimiento, el objetivo es encontrar el punto óptimo que mejor divide a los datos para el nodo derecho e izquierdo. Tal valor corresponde al punto que produce el mayor decremento en términos de la impureza y se define como:

$$\Delta E(s^*, t_1) = \max_{s \in S} \Delta E(s, t_1)$$

Donde  $S$  representa todo el conjunto de valores en los cuales podrían dividirse los datos en el nodo  $t_1$ . Con el uso del valor óptimo  $s^*$ , se divide el conjunto de datos, de tal manera que los datos debajo del valor  $s^*$  sean bifurcados al nodo  $t_l$ , mientras que los datos por encima del valor  $s^*$  sean vinculados al valor  $t_r$ . Este procedimiento es repetido en cada uno de los nodos para poder encontrar la correcta división de los datos.

Por ejemplo, supongamos que se desea obtener el valor óptimo que permita una correcta división de los siguientes datos:



*Ilustración 2 División de datos por árboles de decisión*

*Obtenido de: Valdemar Cuevas, E., Avalos, O., Primitivo, E., Arturo, V., & Marco Antonio, P. (2021). Imagen tomada de: Introducción al Machine Learning con MATLAB. Marcombo.*

Además, se cuenta con los siguientes decrementos de impurezas para todos los puntos:

*Tabla 3 Decrementos de impurezas*  
*Obtenido de: Valdemar Cuevas, E., Avalos, O., Primitivo, E., Arturo, V., & Marco Antonio, P. (2021). Tabla tomada de: Introducción al Machine Learning con MATLAB. Marcombo.*

$\Delta E(s_1, t) = 0.2231$
$\Delta E(s_2, t) = 0.0138$
$\Delta E(s_3, t) = 0.2911$
$\Delta E(s_4, t) = 0.1185$
$\Delta E(s_5, t) = 0.2231$
$\Delta E(s_6, t) = 0.6730$
$\Delta E(s_7, t) = 0.2911$
$\Delta E(s_8, t) = 0.1185$

De estos valores resulta que la mejor opción para dividir los datos es  $s_6$ .

### 2.2.2.3 Máquina de vector de soporte (SVM)

La máquina de vector de soporte es un método de clasificación no paramétrico, cuya idea principal es construir un hiperplano que separe las clases de datos agrupados, maximizando el margen entre dichas clases.

Valdemar Cuevas, Avalos, Primitivo, Valdivia, y Pérez indican que hay que considerar a  $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$  como el conjunto de muestras pares de entrenamiento, donde  $x_i \in \mathbb{R}^d$  es el vector de características de la muestra  $i$ -ésima e  $y_i \in \{-1, +1\}$  es la etiqueta de clase correspondiente. Los datos se asumen linealmente separables, lo que significa que existe un hiperplano que puede separar las dos clases.

### Hiperplano y función discriminante

Un hiperplano es un espacio  $d$ -dimensional que se define mediante la ecuación:

$$w^T x + \alpha = 0$$

donde:

$w \in \mathbb{R}^d$  es el vector de pesos que define la orientación del hiperplano.

$\alpha \in \mathbb{R}$  es el término de sesgo o umbral que ajusta la posición del hiperplano.

La función discriminante para clasificar una muestra  $x$  se define como:

$$g(x) = w_0^T x + \alpha$$

El hiperplano debe cumplir con las siguientes condiciones:

$$g(x) \geq 1 \quad \forall x \in y_1 \text{ (muestras de la clase positiva)}$$

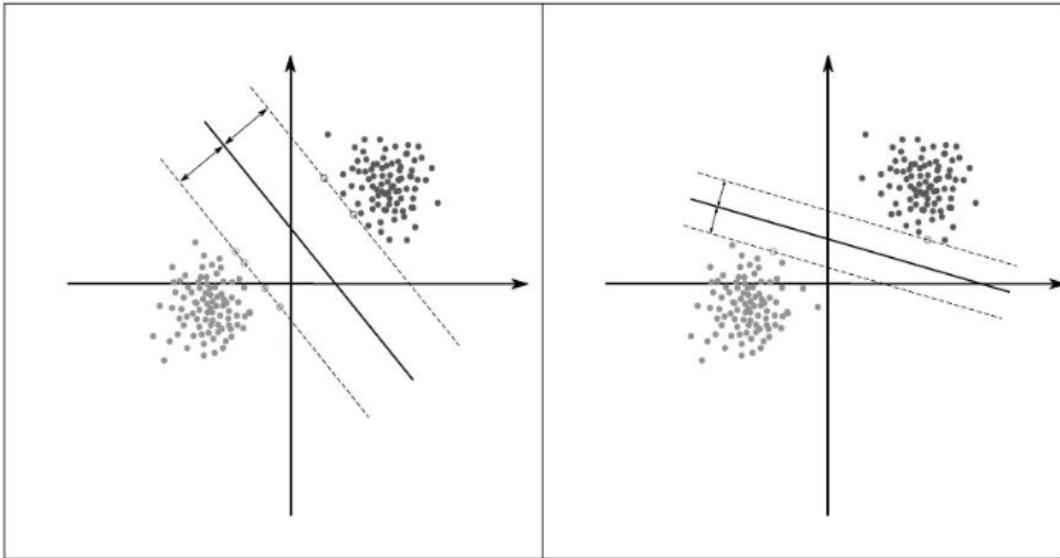
$$g(x) \leq -1 \quad \forall x \in y_2 \text{ (muestras de la clase negativa)}$$

Estas condiciones aseguran que los puntos de la clase  $y_1$  estén de un lado del hiperplano e  $y_2$  del otro.

### **Margen y vectores de soporte**

El objetivo es encontrar el hiperplano tal que el margen (distancia entre el hiperplano y las muestras más cercanas) sea lo más amplio posible. Las muestras más cercanas al hiperplano son las que definen este margen y se conocen como vectores de soporte.

A continuación, se muestra la separación lineal de dos clasificadores, la primera con el máximo margen de separación y la segunda con poco margen:



*Ilustración 3 Separación lineal por SVM*

*Obtenido de: Valdemar Cuevas, E., Avalos, O., Primitivo, E., Arturo, V., & Marco Antonio, P. (2021). Imagen tomada de: Introducción al Machine Learning con MATLAB. Marcombo.*

## Maximización del margen

El margen entre el hiperplano y las muestras de entrenamiento se define como la distancia desde  $x$  hasta el hiperplano. Para un punto  $x$ , se calcula como:

$$r = \frac{|g(x)|}{\|w\|}$$

donde  $\|w\|$  es la norma del vector de pesos  $w$ , para los vectores de soporte  $g(x) = \pm 1$ , por lo que la distancia al hiperplano es  $r = \frac{1}{\|w\|}$ , si se quiere maximizar el margen, se requiere minimizar  $\|w\|$ , lo que implica un problema de optimización.

El problema de optimización para la determinación de los parámetros se puede establecer como:

$$\min_{w, \alpha} \frac{1}{2} \|w\|^2$$

Sujeto a las siguientes restricciones:

$$y_i(w^T x_i + \alpha) \geq 1 \text{ para } i = 1, 2, \dots, N$$

Esto es un problema de optimización convexa con restricciones, el cual puede ser resuelto mediante multiplicadores de Lagrange, expresándose como:

$$L(w, \alpha, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^N \lambda_i [y_i(w^T x_i + \alpha) - 1] \dots \dots \dots \textbf{(1)}$$

donde  $\lambda_i$  son los multiplicadores de Lagrange.

Aplicando las condiciones de optimalidad, se deriva respecto  $w$  y  $\alpha$ .

$$\frac{\partial}{\partial w} L(w, \alpha, \lambda) = 0 \Rightarrow w = \sum_{i=1}^N \lambda_i y_i x_i$$

$$\frac{\partial}{\partial \alpha} L(w, \alpha, \lambda) = 0 \Rightarrow \sum_{i=1}^N \lambda_i y_i = 0$$

Sustituyendo estas expresiones en (1), se obtiene el problema dual:

$$\max_{\lambda} \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_i^N \sum_j^N y_i y_j \lambda_i \lambda_j x_i^T x_j$$

Sujeto a:

$$\lambda_i \geq 0 \quad i = 1, \dots, N$$

$$\sum_{i=1}^N \lambda_i y_i = 0$$

La solución óptima de este problema proporciona los multiplicadores  $\lambda_i$ , que luego se usan para obtener el vector de pesos  $w$  y el sesgo  $\alpha$  del hiperplano.

## 2.3 Aprendizaje profundo

El aprendizaje profundo es una evolución de las redes neuronales artificiales, que constituyen una de las técnicas más importantes de la inteligencia artificial, ya que trata de emular la manera como trabajan las neuronas del cerebro a través de redes formadas por unidades sencillas (neuronas artificiales) interconectadas entre sí, existen redes que se caracterizan por tener muchas capas o profundidad, este conjunto de redes neuronales artificiales y los mecanismos que hacen posible entrenarlas constituyen el denominado aprendizaje profundo (López Sotelo, 2021).

### 2.3.1 Perceptrón multicapa

Bobadilla explica que las neuronas artificiales cuentan con un funcionamiento muy similar a sus correspondientes en el modelo natural, siendo sus elementos:

*$u_i$  representa la neurona  $i$  –ésima*

*$y_i$  representa el resultado que genera la neurona  $i$*

*$w_{ji}$  representa el valor de inhibición entre las  $u_j$  y  $u_i$*

Es el equivalente al efecto de los neurotransmisores sobre la sinapsis que une a las neuronas. Cuando  $w_{ji} > 0$  se modela una sinapsis excitadora, cuando es menor a cero resulta una sinapsis inhibitoria y cuando es cero existe ausencia de conexión.

*Net<sub>i</sub> representa la suma de las señales que le llegan a la neurona u<sub>i</sub>*

$$Net_i = \sum_j y_j w_{ji}$$

*f(Net<sub>i</sub>) representa la función de salida o transferencia o activación.*

La función de activación determina cómo la neurona procesa la entrada  $Net_i$  para producir la salida  $y_i$ . Las funciones de activación más comunes son:

- Función lineal: Es una función simple, donde la salida es proporcional a la entrada. Se utiliza en problemas de regresión. La fórmula es:

$$f(x) = \alpha x + \beta$$

- Función escalón (Step function): Es utilizada para clasificaciones binarias, la salida es 0 si la entrada es menor que el umbral y 1 si la entrada es mayor o igual al umbral.

$$f(x) = \begin{cases} 1 & \text{si } x \geq \theta \\ 0 & \text{si } x < \theta \end{cases}$$

- Función sigmoide: Transforma la entrada en un valor entre 0 y 1, lo que es útil para modelar probabilidades.

$$f(x) = \frac{1}{1 + \exp(-x)}$$

- Función tangente hiperbólica (Tanh): Es muy utilizada en redes neuronales por su eficiencia computacional, su salida se encuentra entre -1 y 1 y a menudo necesita una activación centrada en cero.

$$f(x) = \tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$

- Función ReLU (Rectified Linear Unit): Esta función permite que las redes aprendan rápidamente y evita el problema de gradiente desaparecido.

$$f(x) = \max(0, x)$$

El proceso de aprendizaje consiste en variar los valores sinápticos  $w_{ji}$ . Los pesos  $w_{ji} = 0$  se podrían considerar como ausencia de conexión o conexión que no aporta información relevante. La red neuronal se considera entrenada cuando los pesos han sido ajustados tal que, para cada entrada que se le presenta, produce la salida esperada. Esta condición de finalización en el proceso de aprendizaje se puede expresar como:

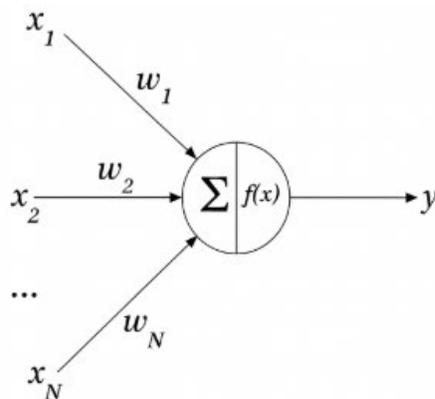
$$\frac{\partial w_{ij}}{\partial t} = 0$$

Lo cual indica que ya no es necesario seguir modificando los pesos  $w_{ij}$  en el tiempo  $t$ .

En el aprendizaje supervisado, Bobadilla (2020) describe los siguientes pasos para que la red neuronal ajuste sus parámetros:

1. Se aplica una muestra de entrada a la red neuronal.
2. Se obtiene la salida que la red neuronal calcula a partir de los datos de entrada.
3. Se compara la salida obtenida con la esperada. A las salidas esperadas se les denominan etiquetas. El aprendizaje supervisado requiere de data sets etiquetados.
4. Si existe error, significa que hay que variar los pesos de alguna manera para adaptar la red al objetivo deseado, como la clasificación.
5. El proceso se repite hasta que se considera aceptable la diferencia entre las salidas que se obtienen y las que se esperan.

Bobadilla (2020) también señala que el perceptrón es el caso más sencillo de red neuronal. En el perceptrón solo existe una neurona de cómputo y una salida y admite un número no restringido de datos de entrada a la neurona artificial. En su formulación, al existir una única neurona de cómputo, podemos eliminar el índice  $i$  de los pesos  $w$ ; asimismo, se añade el umbral  $\theta$  para añadir un grado de libertad en su respuesta.



*Ilustración 4 Neurona: Relación entre las señales de entrada recibidas por los nodos y la señal de salida  
Obtenido de: Bobadilla, J. (2020). Imagen tomada de: Machine Learning y Deep Learning. Bogotá, Colombia: Ra-Ma.*

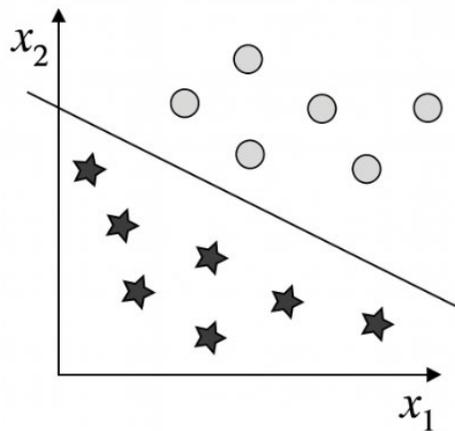
$$y = f \left[ \sum_{i=1}^N x_i w_i - \theta \right]$$

Para el caso de N=2

$$y = f(x_1 w_1 + x_2 w_2 - \theta)$$

$$y = 0 \rightarrow x_2 = \frac{w_1}{w_2} + \frac{\theta}{w_2}$$

Como se puede apreciar, el resultado es la ecuación de una recta. Variando los pesos  $w$ , modificamos la pendiente de la recta. De esta manera, un perceptrón es capaz de realizar separaciones lineales en las muestras de entrada.



*Ilustración 5 Separación lineal por perceptrón*

*Obtenido de: Bobadilla, J. (2020). Imagen tomada de: Machine Learning y Deep Learning. Bogotá, Colombia: Ra-Ma.*

Puesto que el perceptrón solamente permite realizar separaciones lineales, para muestras que no cumplen la condición de separabilidad lineal existe la opción de crear una red neuronal de varios perceptrones que realicen diferentes separaciones

lineales y posteriormente un último perceptrón que aplique funciones de tipo “Y” u “O”.

Bobadilla (2020) explica que, en 1986, Rumelhart, Hinton y Williams proponen un método para que una red neuronal aprendiera la asociación que existe entre los patrones de entrada a la misma y sus clases correspondientes utilizando más capas de neuronas, a este método se le denominó propagación del error hacia atrás o retropropagación de gradiente. La principal idea de la retropropagación se basa en la obtención de errores, empezando en la capa de salida y retrocediendo sucesivamente hasta la capa de entrada. Una vez obtenido el error de una neurona, se puede calcular la variación que hay que ir dando a los pesos  $w$  de entrada a esa neurona. Este proceso se repite con los sucesivos pares de *<muestra, etiqueta>*, actuando sobre todas las neuronas, hasta que se considera que la red ha aprendido y proporciona valores de salida adecuados.

Para la clasificación de la información financiera de Fondos de Aseguramiento, se implementaron y compararon los siguientes modelos:

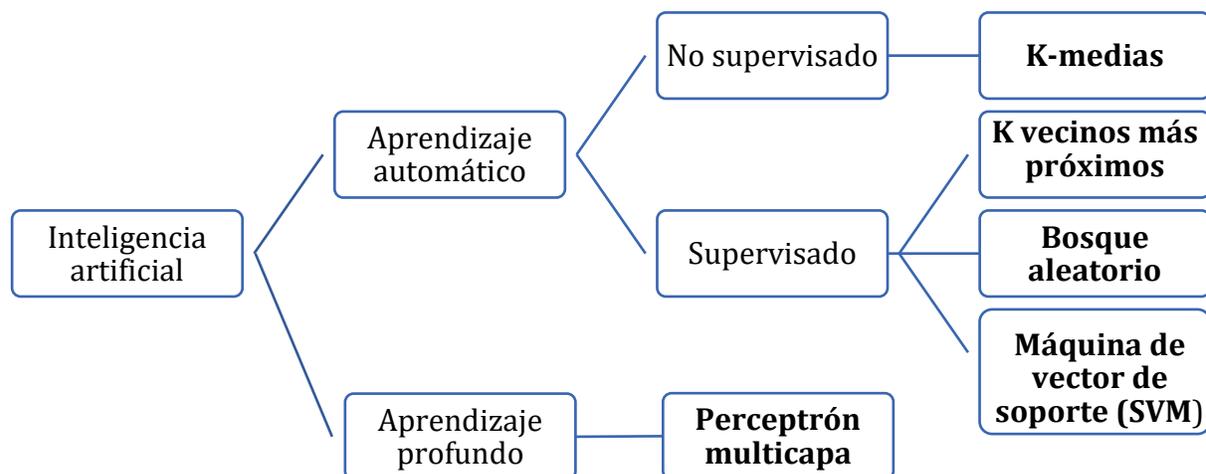


Diagrama 1 Modelos para la clasificación de la información financiera de Fondos de Aseguramiento  
Nota: Diagrama de elaboración propia



**Capítulo 3**  
**Objeto de estudio**

## Capítulo 3. Objeto de estudio

Con base en la hipótesis, el objeto de estudio son los indicadores financieros de los Fondos de Aseguramiento, esto permite un primer acercamiento para el alcance de los objetivos y el análisis de los datos.

### 3.1 Datos recolectados

La información proviene de los estados financieros al cierre de los ejercicios 2020 al 2023 de 311 Fondos de Aseguramiento, parte del Reporte Regulatorio sobre Fondos de Aseguramiento (RR-13) entregado a la CNSF. El análisis se realiza mediante los indicadores financieros expuestos en la Tabla 1 del Capítulo 1: Planteamiento del problema.

La exportación de información, limpieza y cálculo de indicadores se han realizado en Excel, para este trabajo únicamente se consideraron estados financieros que tuvieran un correcto cuadro entre activo y pasivo + capital, así como el cuadro de resultado del ejercicio entre Estado de Resultados y Balance General.

Se han identificado datos que se desvían de la norma, lo que ha llevado a la aplicación de consideraciones específicas para los siguientes indicadores:

**Cobertura de Reservas Técnicas y Disponibilidad de Inversiones para la Cobertura de Reservas Técnicas:** En situaciones donde tanto el dividendo como el divisor son positivos, se calculará el indicador mediante la simple división, sin embargo, cuando el divisor sea negativo y el dividendo positivo, el indicador se establecerá en 1, en el caso contrario, cuando el divisor sea positivo y el dividendo negativo, el indicador se fijará en 0, por último, si tanto el dividendo como el divisor resultan negativos, el indicador también se establecerá en 0.

El resto de los indicadores no cuentan con alguna consideración especial, solo en caso de indeterminación se realiza un análisis del valor más razonable para el

indicador, esto al criterio del supervisor, así mismo, el tratamiento de datos atípicos, para cada índice se toma el mínimo al percentil del 10% y el máximo al percentil del 90%.

La información estadística de los datos, considerando las correcciones realizadas a los datos atípicos descritos previamente, se presenta a continuación:

*Tabla 4 Estadística datos recolectados*  
Nota: Tabla de elaboración propia

<b>Indicador</b>	<b>Media</b>	<b>Desviación Estándar</b>	<b>Mín.</b>	<b>25%</b>	<b>50%</b>	<b>75%</b>	<b>Máx.</b>
Cobertura de Reservas Técnicas	1.13	0.66	0.00	0.77	1.12	1.51	2.23
Disponible e Inversiones para Cobertura de Reservas Técnicas	1.56	0.61	0.73	1.12	1.44	1.91	2.75
Liquidez	1.48	0.50	0.87	1.10	1.35	1.76	2.49
Estructura Financiera	0.52	0.52	-0.10	0.12	0.38	0.81	1.57
Costo de Adquisición	0.23	0.13	0.00	0.16	0.24	0.34	0.42
Gastos de Operación	0.48	0.30	0.15	0.23	0.38	0.67	1.04
Costo de Siniestralidad	0.29	0.30	-0.05	0.00	0.21	0.51	0.86
Porcentaje del Resultado del Ejercicio Respecto a la Prima	-0.14	0.27	-0.69	-0.30	-0.07	0.03	0.20
Porcentaje de Reserva de Riesgos en Curso Respecto a la Prima	0.36	0.31	0.00	0.05	0.36	0.53	1.00

En el contexto global, al examinar las estadísticas promedio, se identifican los siguientes indicadores con perspectiva positiva:

- Cobertura de Reservas Técnicas: Con una media de 1.13, cifra superior a uno, indicando que las inversiones respaldan adecuadamente las reservas.

- Disponibilidad e Inversiones para Cobertura de Reservas: Un promedio de 1.56 señala que existe suficiente respaldo financiero, tanto en cuentas de cheques como en efectivo, para cumplir con obligaciones.
- Liquidez: Con una media de 1.48, se evidencia un superávit en el activo circulante en comparación con el pasivo circulante.
- Estructura Financiera: Con un promedio del 52%, supera el umbral sugerido del 25%.
- Costo de Adquisición: Registrando una media del 23%, se aproxima al estándar del 25% sugerido.

Por el contrario, los siguientes indicadores muestran resultados menos favorables:

- Gastos de Operación: Con un promedio del 48%, excede significativamente el límite del 20% recomendado para instituciones de este tipo.
- Costo de Siniestralidad: Alcanzando una media del 29%, supera el umbral del 25% establecido.
- Resultados del ejercicio negativos.

## Mapa de calor

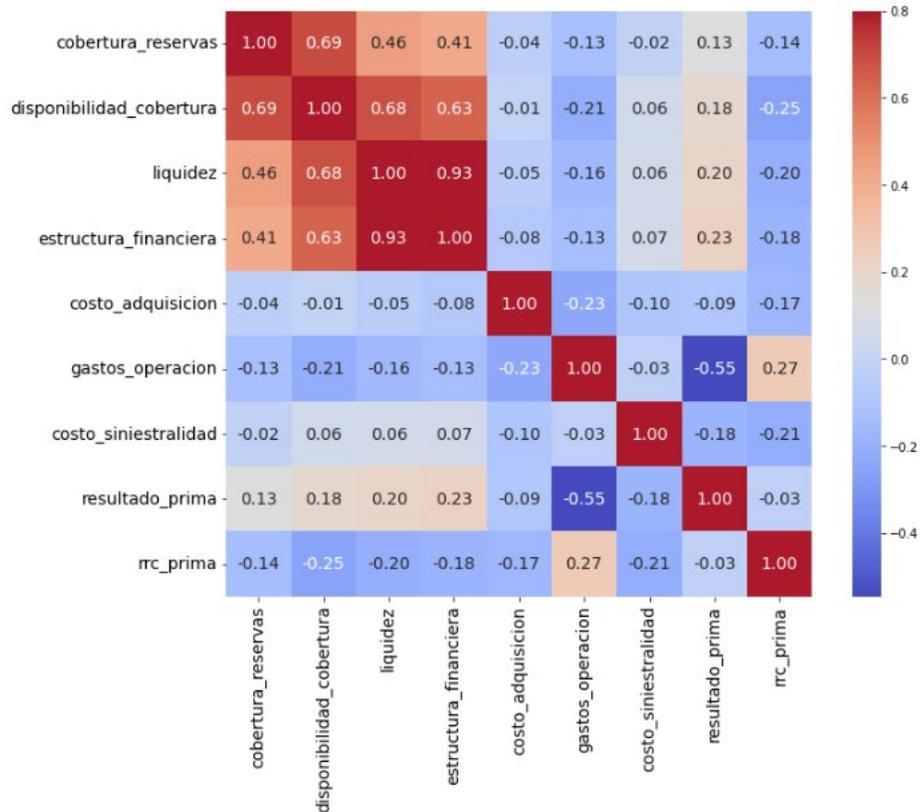


Gráfico 1 Mapa de calor de los datos recolectados  
Nota: Gráfico de elaboración propia

A través del mapa de calor, se observa una correlación positiva entre los indicadores de Cobertura de Reservas Técnicas y Disponible e Inversiones para Cobertura de Reservas Técnicas, lo cual es de esperarse, ya que utiliza las mismas cuentas contables para su cálculo con la diferencia de que el segundo indicador agrega el disponible, así mismo, en general se cuenta con correlaciones positivas entre los índices de Cobertura de Reservas Técnicas, Disponible e Inversiones para Cobertura de Reservas Técnicas, Liquidez y Estructura Financiera.

Es de destacarse la correlación negativa entre los Gastos de Operación y el índice de Resultado del Ejercicio, cómo se determinó en la estadística de datos, en promedio, los Fondos tienen altos gastos de operación y resultados

negativos, lo que permite suponer que estos gastos juegan un papel importante en la poca salud financiera de las asociaciones analizadas.



**Capítulo 4**  
**Metodología**

## Capítulo 4. Metodología

Con base en la metodología CRISP-DM, se desarrollaron los siguientes puntos para la elaboración de este trabajo:

### 4.1 Comprensión del negocio

Los Fondos de Aseguramiento son asociaciones de productores agrícolas y/o ganaderos mexicanos que operan bajo el principio de mutualidad; tienen personalidad jurídica y patrimonio propio y brindan protección a sus socios a través de operaciones activas de seguros y coaseguro sin fines de lucro (Agroasemex, 2016), lo anterior define su operación financiera, que se encuentra respaldada por sus propias reservas técnicas y por los contratos de reaseguro, a su vez solo podrán cobrar a sus socios por concepto de cuota lo indispensable para sufragar los gastos generales que ocasione su gestión y cuotas para sus Organismos Integradores; cubrir prima de reaseguro y accesorios a la misma; constituir o incrementar conforme la LFAAR las reservas técnicas necesarias para cumplir sus compromisos de aseguramiento con sus socios, así como crear e incrementar el Fondo Social (capital) (Ley de Fondos de Aseguramiento Agropecuario y Rural, 2005).

Por lo tanto, según la LFAAR, los Fondos de Aseguramiento solo podrán realizar las siguientes operaciones:

- i. Practicar operaciones de seguro conforme a la LFAAR.
- ii. Constituir e invertir las reservas previstas en la LFAAR.
- iii. Administrar los recursos retenidos a las instituciones de seguros correspondientes a las operaciones de reaseguro que hayan celebrado.
- iv. Constituir depósitos en instituciones de crédito.
- v. Operar con documentos mercantiles por cuenta propia para la realización de su objeto social.

- vi. Aquellas que sean necesarias para la realización de su objeto social.

Los puntos anteriores definen los indicadores financieros seleccionados para el análisis y clasificación de la situación financiera de los Fondos de Aseguramiento.

## **4.2 Comprensión de los datos**

Los datos que se utilizan para el cálculo de indicadores financieros se extraen del Reporte Regulatorio sobre Fondos de Aseguramiento (RR13), requerido por la CNSF, al cierre del cuarto trimestre de 2020 al 2023. El reporte debe incluir un archivo Excel denominado "IROT", el cual contiene múltiples hojas con el objeto de ser capturadas por los Fondos de Aseguramiento, entre las que se encuentran el estado de resultados y el balance general, para su llenado contienen la llave y el nombre de la cuenta contable.

Asimismo, el reporte contiene un PDF con los estados financieros firmados por los funcionarios del Fondo, para mejorar la calidad de la información, los supervisores realizan un comparativo, en un archivo concentrado, entre lo reportado en el Excel y el PDF, en caso de localizar inconsistencias el dato final es el reportado en los estados financieros firmados.

## **4.3 Preparación o preprocesamiento de datos**

Como parte de la concentración y limpieza de los datos y previo al modelado de la información se eliminaron los estados financieros que presentan descuadres entre estado de resultado y balance general, en Excel se manejó en un formato numérico, información no proporcionada o vacía se sustituyó por 0 y se eliminaron los valores atípicos, para el cálculo de indicadores. Este cálculo y consideraciones especiales se describen en el Capítulo 1: Planteamiento del problema y Capítulo 3: Objeto de estudio.

Los indicadores son almacenados en formato CSV para su importación en Python, en donde se aplica la siguiente normalización:

$$z_i = \frac{(x_i - \text{mínimo}(x))}{(\text{máximo}(x) - \text{mínimo}(x))}$$

#### 4.4 Modelado

Después de normalizar los datos, se procede a modelarlos utilizando los siguientes algoritmos empleando la biblioteca Scikit-Learn:

- K-medias utilizando la importación de Kmeans.
- Bosque aleatorio a través de la importación de RandomForestClassifier.
- Máquina de vector de soporte mediante la importación de SVC.
- K- vecinos más próximos a través de la importación de KNeighborsClassifier.
- Red neuronal: Se emplea el clasificador MLP importado desde MLPClassifier.

Para facilitar la visualización gráfica, se aplicará una reducción de dimensiones con análisis de componentes principales importando PCA también de Scikit-Learn.

#### 4.5 Evaluación

Utilizando nuevamente la librería Scikit-Learn, se evaluarán los modelos mediante las siguientes métricas:

Siendo NV= Negativos verdaderos, NF=Negativos falsos, PF=Positivos falsos y PV=Positivos verdaderos.

- Precisión (Precision) usando precision\_score

$$precision = \frac{PV}{PV + PF}$$

- Exhaustividad (Recall) usando recall\_score

$$recall = \frac{PV}{PV + NF}$$

- Valor F (F1-score) usando f1\_score

$$precision = 2 * \frac{precision * recall}{precision + recall}$$

- Exactitud (Accuracy) usando accuracy\_score

$$recall = \frac{PV + NV}{PV + NV + PF + NF}$$

- Matriz de Confusión (Confusion matrix) usando confusion\_matrix

Tabla 5 Matriz de Confusión  
Nota: Tabla de elaboración propia

		Predicción	
		0	1
Realidad	0	NV	PF
	1	NF	PV

The background features a complex technical illustration. On the left, there are several interlocking gears of different sizes, some with dashed outlines. To the right, there are various geometric shapes and lines, including a dashed hexagon, a solid hexagon, and several parallel lines with arrows pointing in different directions. The overall style is clean and technical, using a light gray color palette.

# Capítulo 5

## Diseño de estrategias y resultados

## Capítulo 5. Diseño de estrategias y resultados

En esta sección se presentan los resultados. Retomando el cuarto objetivo particular, realizamos el comparativo del método actual con el método de aprendizaje supervisado k-medias.

### 5.1 Clasificación con k-medias

Para comprobar que el número de clústeres que se proponen es óptimo para la clasificación, se aplica la metodología del codo, la cual utiliza la suma de los cuadrados de las distancias con su centroide correspondiente, obtenida al aplicar k-medias con distintas  $k$ 's.

$$WCSS = \sum_{i \in n} (X_i - Y_i)^2$$

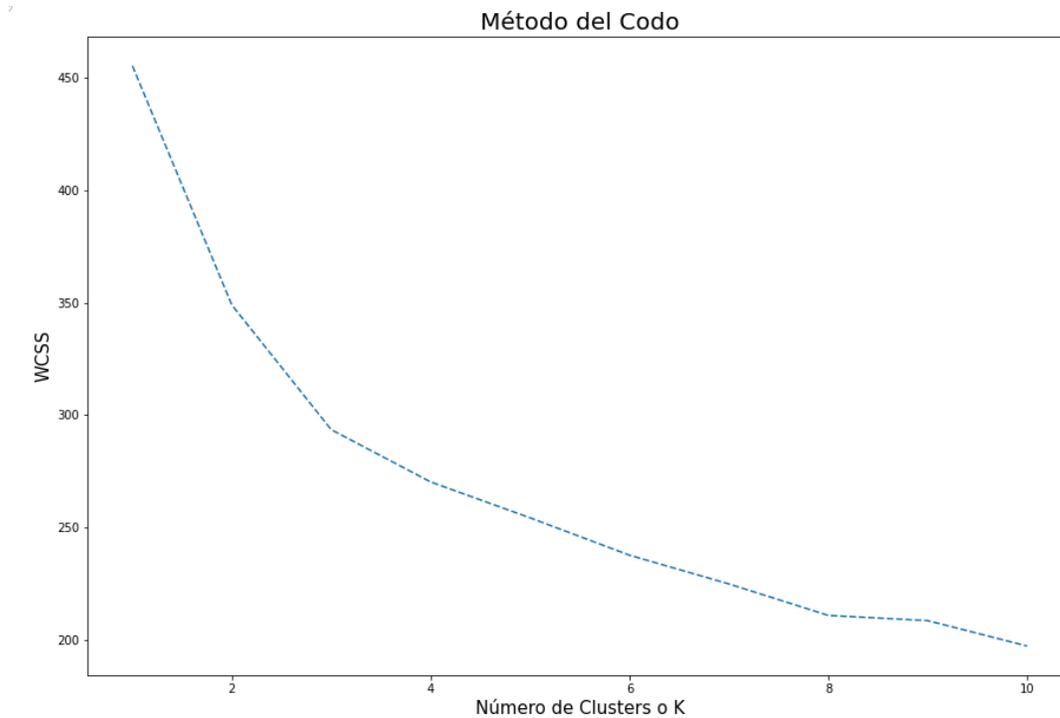


Gráfico 2 Codo K-medias  
Nota: Gráfico de elaboración propia

Gráficamente, no se observa un decrecimiento drástico de WCSS, como se espera al aplicar el método del codo, sin embargo, se continúa con el supuesto de realizar la clasificación en tres tipos de riesgo: 1=Bajo, 2=Medio y 3=Alto.

La información normalizada es clasificada con k-medias, para la visualización gráfica, se aplica reducción de dimensiones con PCA:

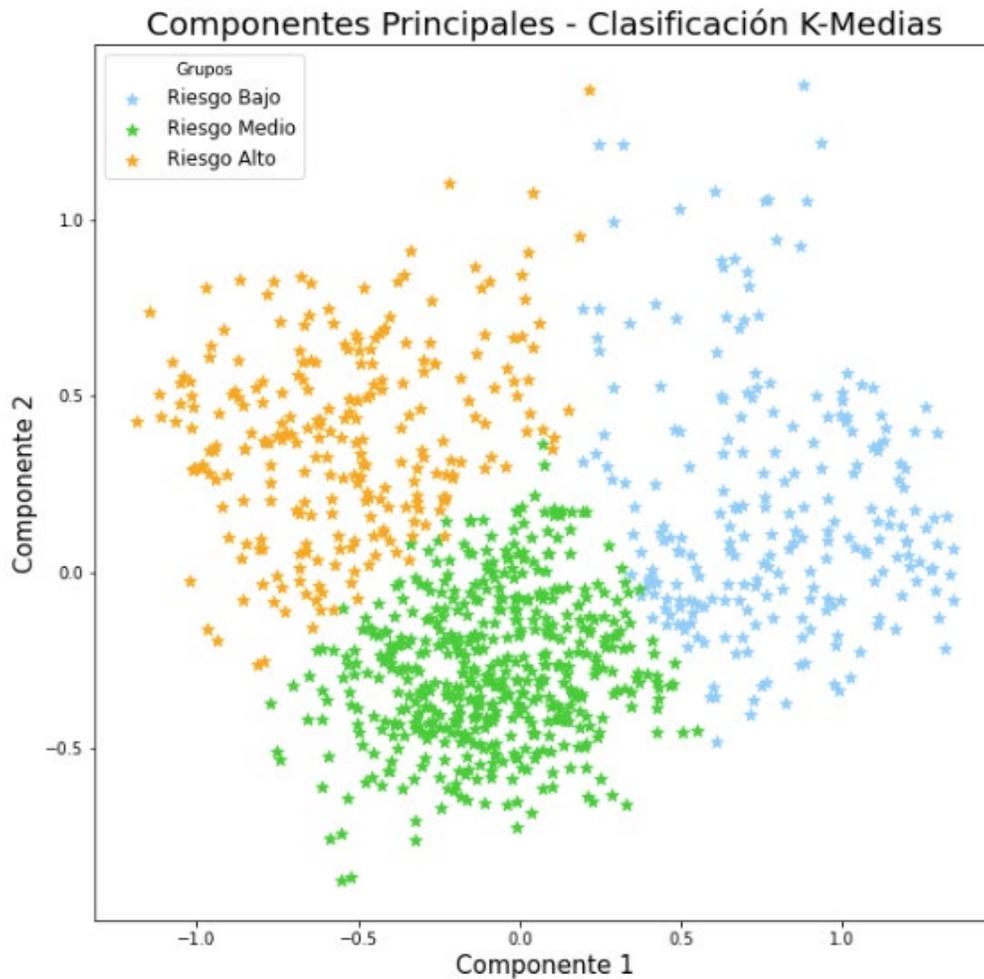


Gráfico 3 Clasificación k-medias  
Nota: Gráfico de elaboración propia

A través del modelo de k-medias se obtiene una gráfica con grupos delimitados, donde la mayor concentración se encuentra en el grupo de color verde.

La asignación del tipo de riesgo mediante k-medias se realizó con base en la información estadística de cada grupo, presentada en las Tablas 7, 8 y 9, las cuales se encuentran en la siguiente sección: 5.1.2 Métricas k-medias.

*Tabla 6 Contribución por variable a los componentes principales*  
 Nota: Tabla de elaboración propia

<b>Indicador</b>	<b>PC1</b>	<b>PC2</b>
<b>cobertura_reservas</b>	0.36	0.13
<b>disponibilidad_cobertura</b>	0.47	0.13
<b>liquidez</b>	0.50	0.20
<b>estructura_financiera</b>	0.49	0.21
<b>costo_adquisicion</b>	0.00	-0.35
<b>gastos_operacion</b>	-0.27	0.70
<b>costo_siniestralidad</b>	0.05	0.22
<b>resultado_prima</b>	0.22	-0.45
<b>rrc_prima</b>	-0.21	0.16

La asignación de tipo de riesgo se robustece considerando la contribución de cada variable a los componentes principales, se observa que las variables positivas con mayor peso en el primer componente (PC1) son: liquidez (0.50), disponibilidad\_cobertura (0.47), estructura\_financiera (0.49) y cobertura\_reservas (0.36), esto indica que valores altos en PC1 están asociados a un mejor desempeño y mayor solidez financiera.

En cuanto al segundo componente (PC2), el valor más alto es gastos de operación (0.70) y como valor negativo el resultado\_prima (-0.45), lo que implica que valores altos de PC2 pueden estar relacionados con una menor eficiencia técnica y un alto nivel de gastos operativos.

Al observar la gráfica de dispersión, se identifica que el grupo clasificado como riesgo bajo (color azul) está influenciado por los indicadores de liquidez,

estructura\_financiera, disponibilidad\_cobertura y cobertura\_reservas, lo que representa Fondos financieramente más sólidos.

Por su parte, los fondos clasificados como riesgo medio (color verde) presenta valores intermedios, a diferencia de los riesgos clasificados como altos (color naranja) se concentran en la zona con valores bajos con indicadores relacionados con liquidez y reservas, que reflejan menor capacidad financiera.

Como medida adicional para evaluar la estabilidad del modelo, se utilizó el Índice de Rand Ajustado (ARI), realizando 20 ejecuciones del algoritmo con diferentes condiciones de inicialización se calculó el índice, el análisis arrojó una media de 0.9847 con desviación estándar de 0.0106, lo que refleja consistencia en los agrupamientos obtenidos, incluso al modificar la semilla del modelo.

Para iniciar con el comparativo, se presentan los datos gráficamente con las etiquetas del método actual, obtenido por el área de supervisión de Fondos de Aseguramiento de la CNSF:

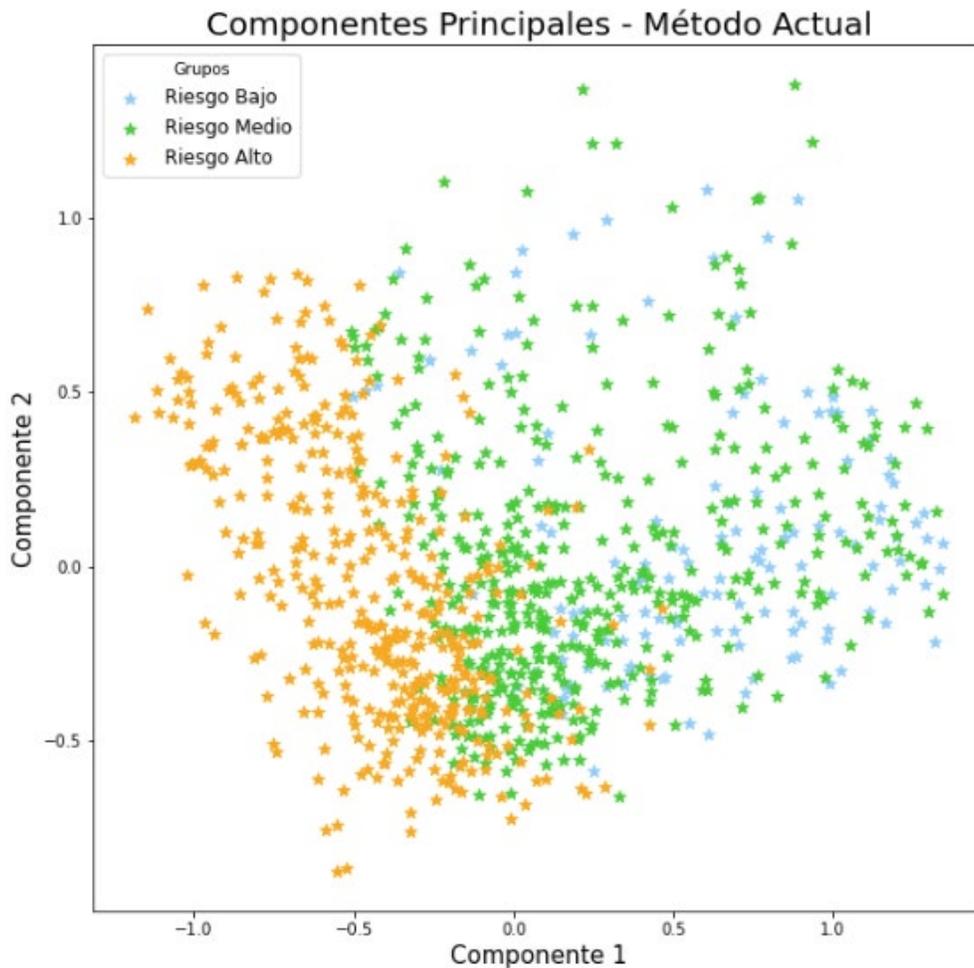


Gráfico 4 Clasificación método actual  
 Nota: Gráfico de elaboración propia

Gráficamente, se contempla que hay mezcla de los perfiles de riesgo en cada una de las agrupaciones, sin embargo, se observa una tendencia en el grupo central al color verde observado en el Gráfico 3 Clasificación k-medias, así como el color amarillo en el grupo izquierdo y del color azul en el grupo de la derecha, siendo el grupo con el menor número de elementos.

Para evaluar la concordancia entre el algoritmo k-medias y el método actual utilizado por la CNSF, se emplean diversas métricas, entre las cuales se incluyen la matriz de confusión, precisión, recall, F1-score y accuracy.

## 5.1.2 Métricas k-medias

### Matriz de confusión

Tabla 7 Matriz de confusión de riesgos k-medias  
Nota: Tabla de elaboración propia

		Método tradicional		
		Bajo	Medio	Alto
K-medias	Bajo	88	145	2
	Medio	33	252	227
	Alto	15	59	179

La matriz de confusión refleja que el clasificador k-medias agrupa un total de 145 elementos como riesgo bajo, cuando en el enfoque tradicional estos fueron considerados riesgo medio. En total se tiene una coincidencia de solo el 37.45% entre ambos métodos para este tipo de riesgo.

En cuanto a los Fondos clasificados como riesgo medio, se observa una coincidencia del 49.22% entre los dos métodos, mientras que un 44.34% de estos Fondos fueron categorizados como riesgo alto en el método actual.

Por último, los elementos clasificados como riesgo alto son los que muestran una mayor coincidencia, alcanzando un 70.75%.

Por lo tanto, el 71.24% de los Fondos bajaron de nivel de riesgo y el 20.38% subieron de categoría; ambos métodos tienen una concentración mayor en el grupo de riesgo medio, aproximadamente la mitad del total de datos, no obstante, el método tradicional clasificó el 40.80% de sus elementos como riesgo alto, a comparación de k-medias que disminuyó el mencionado grupo al 25.30%, teniendo un mayor equilibrio entre el número de elementos de riesgo bajo y alto.

Dado lo anterior, el método tradicional tiende a concentrar una mayor proporción en Fondos de riesgo alto, mientras que k-medias los dispersa con mayor equidad.

*Tabla 8 Evaluación entre k-medias y el método actual*

Nota: Tabla de elaboración propia

<b>K-medias</b>				
<b>Clase</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>Accuracy</b>
<b>Bajo</b>	0.65	0.37	0.44	0.52
<b>Medio</b>	0.55	0.49	0.54	
<b>Alto</b>	0.44	0.71	0.56	

El algoritmo k-medias muestra una coincidencia del 52% en las clasificaciones respecto al método actual. Derivado de los resultados anteriores, es crucial realizar un análisis más detallado de los resultados para evaluar la viabilidad de realizar un cambio en la metodología actual.

Para profundizar en el comparativo entre k-medias y el método tradicional, se presentan cuadros con la media y desviación estándar de ambos métodos por grupo, comenzando con el grupo 1 o perfil de riesgo bajo:

*Tabla 9 Estadística grupo 1*

Nota: Tabla de elaboración propia

	<b>Comparativo grupo 1 o perfil bajo</b>			
	<b>Media metodología tradicional</b>	<b>Media K-medias</b>	<b>Desviación estándar metodología tradicional</b>	<b>Desviación estándar K-medias</b>
<b>Cobertura de reservas técnicas</b>	1.72	1.72	0.46	0.65
<b>Disponibilidad y cobertura</b>	2.12	2.33	0.53	0.47
<b>Liquidez</b>	1.96	2.19	0.43	0.34
<b>Estructura financiera</b>	1.07	1.24	0.38	0.37
<b>Costo de adquisición</b>	0.18	0.19	0.12	0.14
<b>Gastos de operación</b>	0.41	0.41	0.30	0.26
<b>Costo de siniestralidad</b>	0.18	0.36	0.25	0.33

Comparativo grupo 1 o perfil bajo				
	Media metodología tradicional	Media K-medias	Desviación estándar metodología tradicional	Desviación estándar K-medias
<b>Prima emitida contra resultado del ejercicio</b>	0.03	-0.06	0.20	0.24
<b>Reserva de riesgos en curso contra prima emitida</b>	0.42	0.23	0.32	0.29



Gráfico 5 Comparativa metodología tradicional vs k-medias grupo 1  
Nota: Gráfico de elaboración propia

En el perfil de riesgo bajo, se observa una brecha significativa en los indicadores de Disponibilidad y Cobertura, Liquidez y Estructura Financiera, los indicadores obtenidos mediante el uso de k-medias muestran valores más elevados, lo que indica un perfil financiero más sólido, sin embargo, es importante destacar que el indicador de costo de siniestralidad es más alto, lo que podría resultar en que el índice de Prima Emitida contra Resultado del Ejercicio sea inferior en comparación con el método tradicional, debido a la disminución del Resultado del Ejercicio por la alta siniestralidad.

Tabla 10 Estadística grupo 2  
 Nota: Tabla de elaboración propia

	Comparativo grupo 2 o perfil medio			
	Media metodología tradicional	Media K-medias	Desviación estándar metodología tradicional	Desviación estándar K-medias
<b>Cobertura de reservas técnicas</b>	1.28	1.02	0.62	0.53
<b>Disponibilidad y cobertura</b>	1.78	1.40	0.52	0.40
<b>Liquidez</b>	1.69	1.32	0.43	0.27
<b>Estructura financiera</b>	0.72	0.35	0.44	0.30
<b>Costo de adquisición</b>	0.26	0.28	0.13	0.11
<b>Gastos de operación</b>	0.43	0.32	0.26	0.15
<b>Costo de siniestralidad</b>	0.32	0.25	0.31	0.26
<b>Prima emitida contra resultado del ejercicio</b>	-0.15	-0.06	0.24	0.17
<b>Reserva de riesgos en curso contra prima emitida</b>	0.30	0.34	0.27	0.25

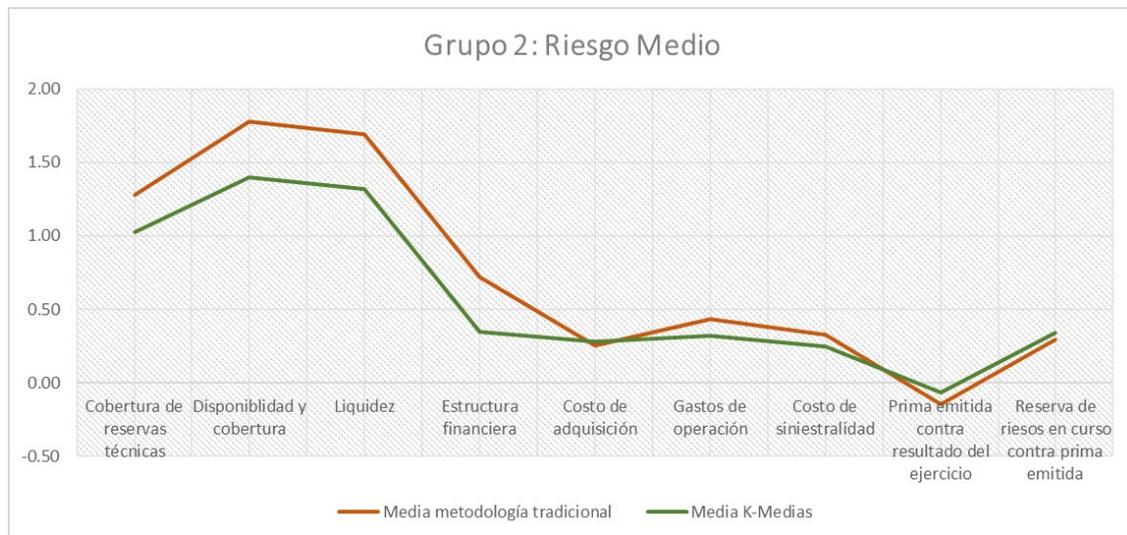


Gráfico 6 Comparativa metodología tradicional vs k-medias grupo 2  
 Nota: Gráfico de elaboración propia

En el grupo 2, que corresponde al perfil medio, se observa que, en comparación con la metodología tradicional, la mayoría de los indicadores son más altos, con excepción de los indicadores de Costo de Adquisición, Prima Emitida contra el Resultado del Ejercicio y Reserva de Riesgos en Curso contra la Prima Emitida, donde se presenta una situación inversa a la del grupo 1 o perfil de riesgo bajo.

*Tabla 11 Estadística grupo 3*  
 Nota: Tabla de elaboración propia

	<b>Comparativo grupo 3 o perfil alto</b>			
	<b>Media metodología tradicional</b>	<b>Media K-medias</b>	<b>Desviación estándar metodología tradicional</b>	<b>Desviación estándar K-medias</b>
<b>Cobertura de reservas técnicas</b>	0.75	0.78	0.54	0.54
<b>Disponibilidad y cobertura</b>	1.13	1.17	0.42	0.43
<b>Liquidez</b>	1.08	1.15	0.23	0.30
<b>Estructura financiera</b>	0.11	0.21	0.27	0.35
<b>Costo de adquisición</b>	0.23	0.18	0.13	0.13
<b>Gastos de operación</b>	0.56	0.87	0.33	0.21
<b>Costo de siniestralidad</b>	0.28	0.31	0.30	0.34
<b>Prima emitida contra resultado del ejercicio</b>	-0.20	-0.39	0.29	0.30
<b>Reserva de riesgos en curso contra prima emitida</b>	0.41	0.52	0.33	0.36

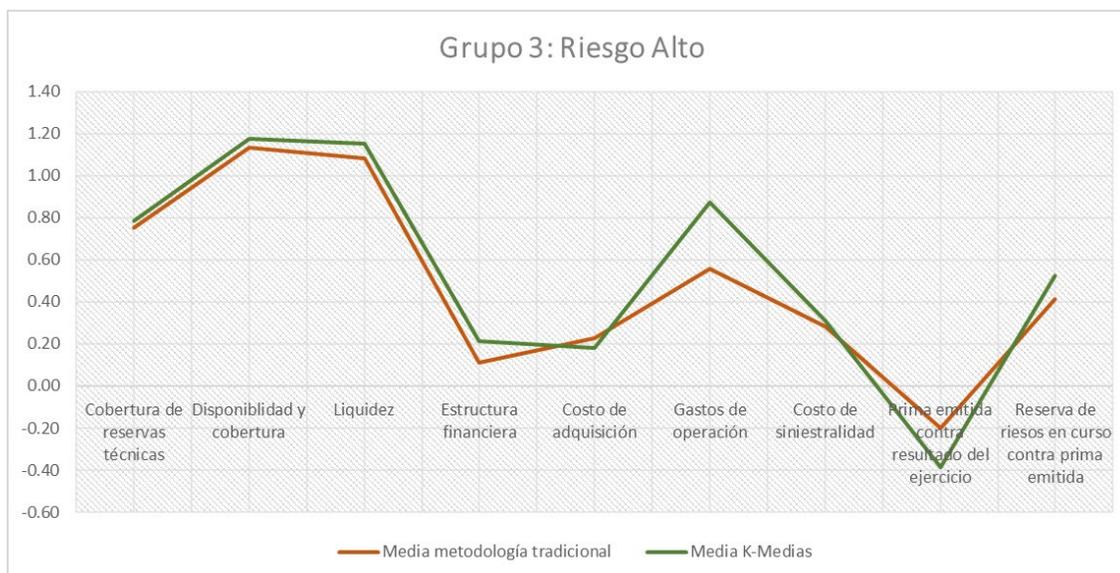


Gráfico 7 Comparativa metodología tradicional vs k-medias grupo 3  
 Nota: Gráfico de elaboración propia

En el grupo 3, que corresponde al perfil alto, se observa una brecha más estrecha entre la mayoría de los indicadores, no obstante, destaca una notable diferencia en los Gastos de Operación, los cuales ascienden en promedio al 87% con el método k-medias y al 56% con el enfoque tradicional, esta diferencia conduce a un posible Resultado del Ejercicio inferior en el caso de k-medias.

Dado lo anterior, el modelo de k-medias se encuentra fuertemente influenciado por el indicador de Gastos de Operación, en el análisis de datos se constató que existe una correlación negativa entre estos gastos y el Resultado del Ejercicio, sin embargo, este efecto no se refleja en el método tradicional, ya que no es considerado un indicador relevante.

Entre las variables más significativas para el análisis de la información financiera de los Fondos de Aseguramiento se encuentra la Cobertura de Reservas Técnicas, derivado de la importancia de tener respaldado el capital para hacer frente a las obligaciones con los socios del Fondo, así como los indicadores de Liquidez, Estructura Financiera y Costo Neto de Siniestralidad, según lo observado en los

resultados anteriores, además se destaca la importancia de los Gastos de Operación en los resultados de estas instituciones, dado lo anterior se considera como mejor metodología k-medias.

En los próximos apartados se exponen los resultados de la clasificación a través de los métodos de aprendizaje supervisado: bosque aleatorio y máquina de vector de soporte, así como el método de aprendizaje supervisado: k vecinos más próximos, con base en los grupos obtenidos con k-medias.

## **5.2 Bosque aleatorio**

El bosque aleatorio (random forest) es un método conjunto de aprendizaje automático, que combina los resultados de varios árboles de decisión, cada uno entrenado en un subconjunto de datos que posteriormente combina los resultados de todos los árboles, la clasificación final son los datos resultantes con mayor coincidencia entre ellos. Este tipo de técnica suele otorgar mejores resultados que los árboles de decisión por sí solos.

Para la clasificación mediante la técnica bosque aleatorio se utilizó la librería de Python: Scikit-Learn, en específico, las funciones GridSearchCV y RandomForestClassifier.

### **5.2.1 Selección de parámetros**

GridSearchCV permite evaluar de forma automática el rendimiento de diversos parámetros de un modelo, contando con las siguientes entradas:

- estimator: Modelo a evaluar.
- param\_grid: Selección de parámetros a evaluar.
- cv: Número de conjuntos en el que se dividen los datos para validación cruzada.

RandomForestClassifier es una clase en Python que utiliza diversos parámetros, para simplificar el modelo, únicamente se evaluaron diferentes valores de los siguientes:

- n\_estimators: Número de árboles del bosque.
- max\_depth: Profundidad máxima del árbol.

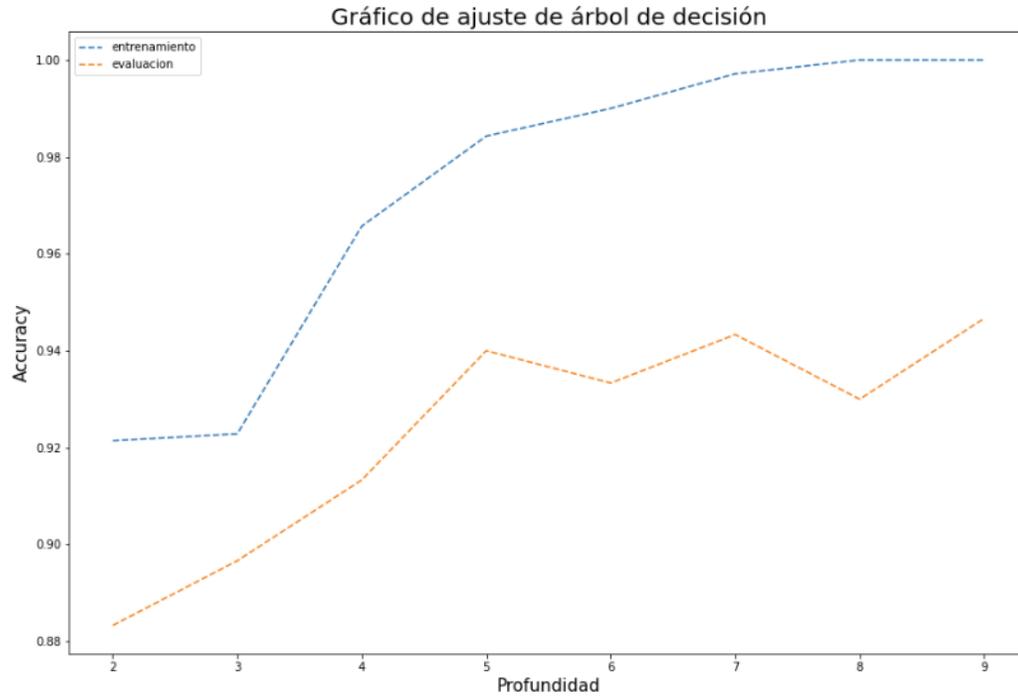
Se calculó, a través de GridSearchCV, la exactitud o accuracy con 100, 200 y 300 árboles y una profundidad de 2 a 10 nodos, ya que se observó que presentaba sobreajuste con una mayor cantidad de nodos.

### **5.2.2 Resultados**

Los parámetros seleccionados, a través de GridSearchCV, fueron los siguientes:

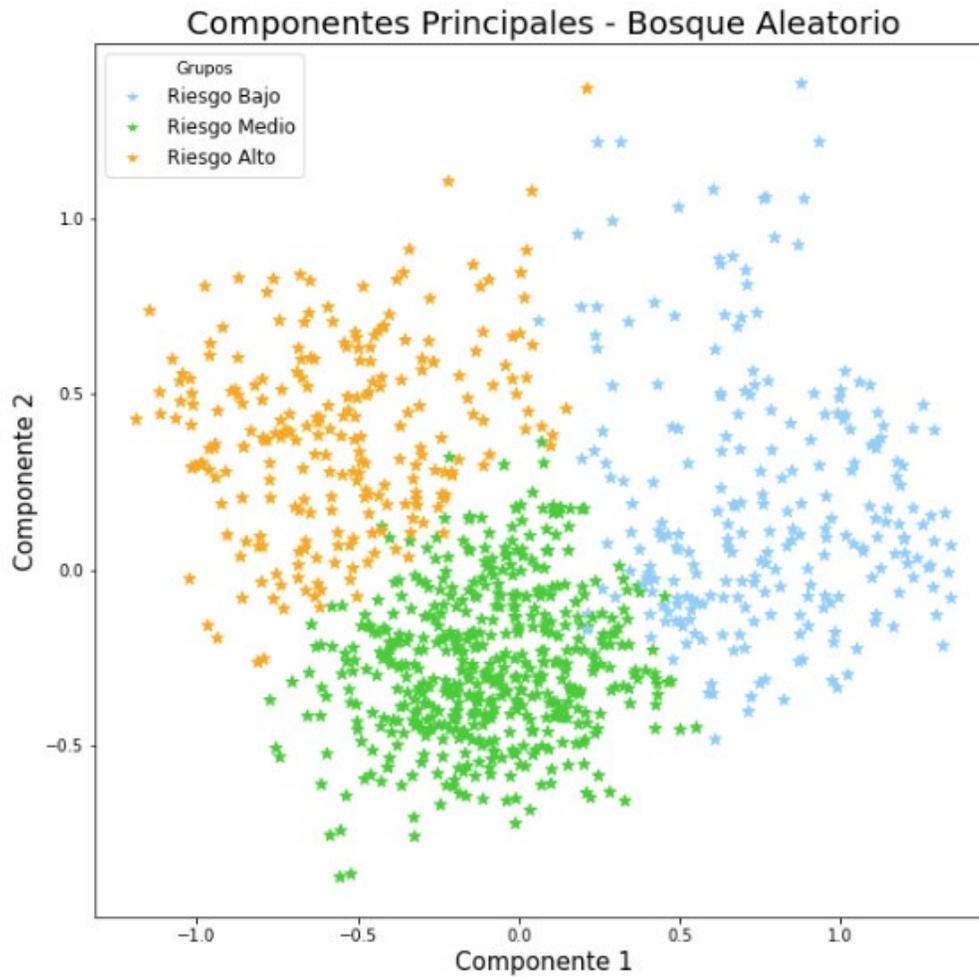
- n\_estimators=100
- max\_depth=10

Adicionalmente, se muestra el gráfico con diferentes profundidades:



*Gráfico 8 Exactitud vs profundidad bosque aleatorio*  
 Nota: Gráfico de elaboración propia

Como resultado, sobre el conjunto de prueba se obtuvo un **94% de exactitud o accuracy**; mediante el siguiente gráfico se pueden observar los puntos no coincidentes con la clasificación realizada a través de k-medias:



*Gráfico 9 Clasificación bosque aleatorio*  
 Nota: Gráfico de elaboración propia

### 5.3 Máquina de vector de soporte (SVM)

Máquina de vector de soporte (SVM) utiliza la función matemática llamada kernel para la transformación de características, esta asigna los datos a un espacio de otra dimensión con el objetivo de facilitar la clasificación, a continuación, se proporciona una breve explicación de las funciones más populares:

### 5.3.1 Kernel lineal

Gil Martínez (2018) explica que la función kernel lineal es recomendable si la separación lineal de los datos es sencilla. Este kernel cuantifica la similitud de un par de observaciones usando la correlación de Pearson.

$$K(x_i, x_j) = x_i x_j$$

### 5.3.2 Kernel polinómico

En Aprendeia.com se indica que el kernel polinómico no se limita a utilizar las características dadas por las muestras de entrada para determinar la similitud, sino también utiliza las combinaciones de estas.

$$K(x_i, x_j) = (1 + x_i x_j)^d$$

### 5.3.3 Kernel radial o RBF

Según Gil Martínez (2018), el RBF permite mapear la entrada a un espacio de dimensión infinita.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Donde  $\gamma$  es una constante positiva que al incrementarse aumenta la flexibilidad del SVM, el RBF tiene un comportamiento local, lo que hace que las observaciones de entrenamiento cercanas a una observación de prueba tendrán efecto sobre la clasificación.

### 5.3.4 Kernel sigmoide

Wood señala que las funciones sigmoideas tienen la propiedad de asignar valores reales a un rango específico que puede interpretarse como una probabilidad, por lo

anterior se han vuelto populares en el aprendizaje automático, existen diferentes tipos de funciones sigmoideas, entre las que se encuentran, la función logística, hiperbólica, arcotangente, entre otras. La función utilizada por la clase SVC de Scikit-Learn es la tangente hiperbólica, definiéndose como:

$$K(x,y) = \tanh(ax^T y + c)$$

SVC utiliza otros parámetros, que pueden ser consultados en la documentación de la clase, de los cuales para la clasificación de información financiera se consideró asignar diferentes valores a los siguientes parámetros mediante GridSearchCV.

- Gamma: es un hiperparámetro utilizado para los kernels RBF, polinómico y sigmoide. Los valores bajos de gamma proporcionan grandes radios de similitud, dando como resultado una mayor agrupación. Para valores altos de gamma, los puntos deben ser cercanos para considerarse de la misma clase.
- C: proporciona una penalización por cada dato mal clasificado. Si el parámetro C es pequeño, la penalización es baja; por el contrario, una C grande minimiza la cantidad de ejemplos mal clasificados dada la alta penalización.

### 5.3.5 Selección de parámetros

Los parámetros seleccionados para comparar su desempeño son:

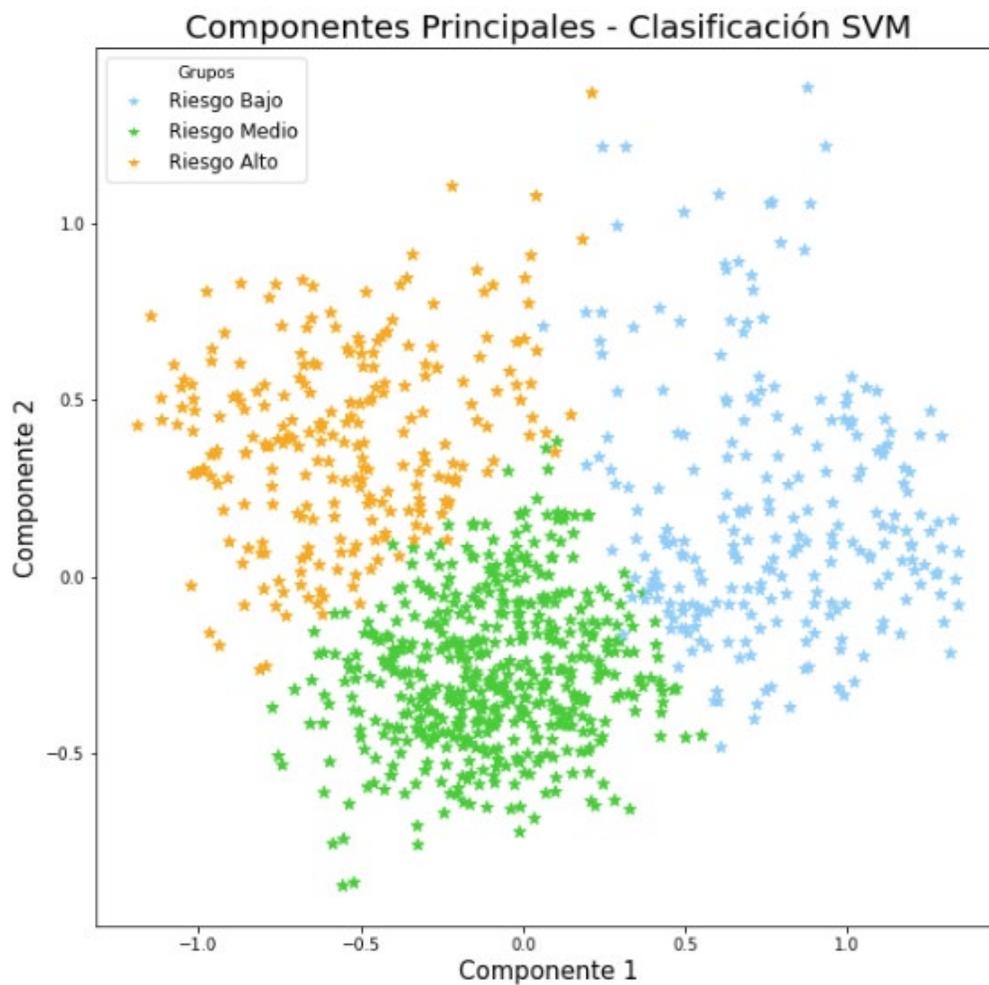
- Kernel: "rbf", "poly" y "sigmoid".
- Gamma: 1, 1e-2 y 1e-3.
- C: 1, 10 y 100.

### 5.3.6 Resultados

Los parámetros seleccionados, a través de GridSearchCV, fueron los siguientes:

- Kernel: "poly"
- Gamma: 1
- C: 10

Como resultado, sobre el conjunto de prueba se obtuvo un **98% de exactitud o accuracy**; mediante el siguiente gráfico se pueden observar los puntos no coincidentes con la clasificación realizada a través de k-medias:



*Gráfico 10 Clasificación SVM*  
Nota: Gráfico de elaboración propia

## 5.4 K vecinos más próximos (K-NN)

Según la documentación de Scikit-Learn el principio del algoritmo K-NN es encontrar un número predefinido de muestras de entrenamiento más cercanas en distancia a un nuevo punto, clasificándolo a partir de ellas, el número de muestras puede ser una constante definida por el usuario o ser variable, así mismo se pueden utilizar distintos tipos de métricas para medir la distancia, siendo la euclidiana estándar la más frecuente.

### 5.4.1 Métricas

$$Euclidiana = d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

$$Manhattan = d(x, y) = \left( \sum_{i=1}^n |x_i - y_i| \right)$$

$$Minkowski = d(x, y) = \left( \sum_{i=1}^n |x_i - y_i| \right)^{1/p}$$

La clase de Python utilizada en este apartado es `KNeighborsClassifier`, que cuenta con diversos parámetros como: `n_neighbors`, `weights`, `algorithm`, `leaf_size`, `p`, `metric`, `metric_params`, y `n_job`, para la implementación en la clasificación de la información financiera de Fondos de Aseguramiento únicamente se usaron diferentes valores para conocer los parámetros óptimos de `n_neighbors` y `metrics` con `GridSearchCV`.

### 5.4.2 Selección de parámetros

Los parámetros seleccionados para comparar su desempeño son:

- n\_neighbors: rango entre 1 y 15 vecinos.
- metrics: “euclidean”, “manhattan” y “minkowski”

### 5.4.3 Resultados

Los parámetros seleccionados, a través de GridSearchCV, fueron los siguientes:

- n\_neighbors: 5 vecinos.
- metrics: “euclidean”

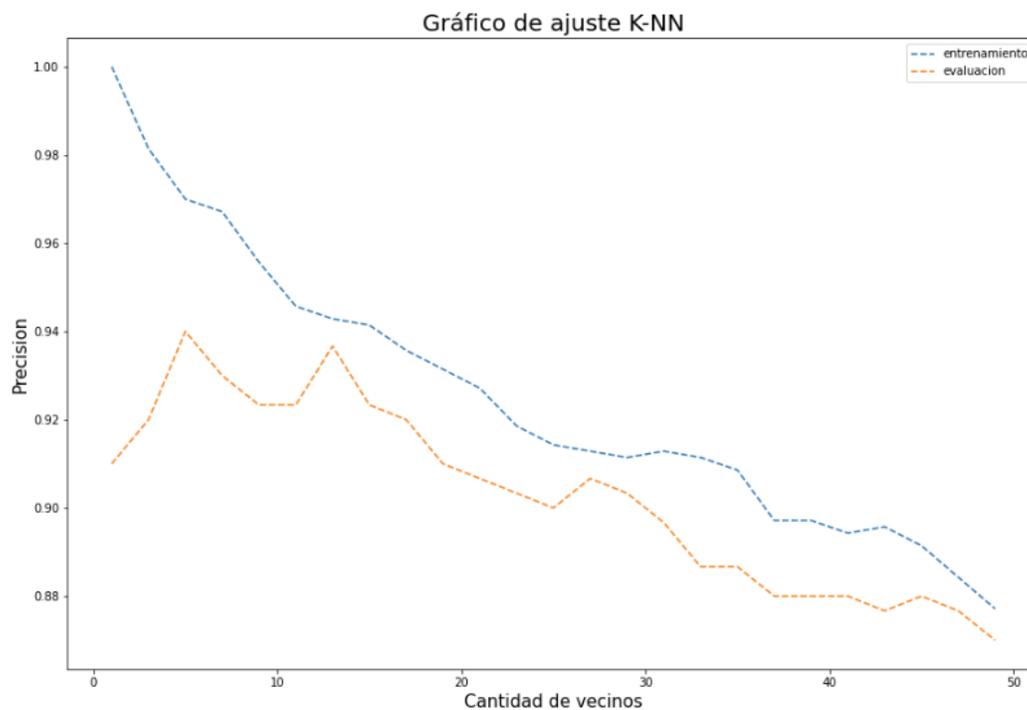


Gráfico 11 Precisión vs cantidad de vecinos K-NN  
Nota: Gráfico de elaboración propia

Como resultado, sobre el conjunto de prueba se obtuvo un **94% de exactitud o accuracy**; mediante el siguiente gráfico se pueden observar los puntos no coincidentes con la clasificación realizada a través de k-medias:

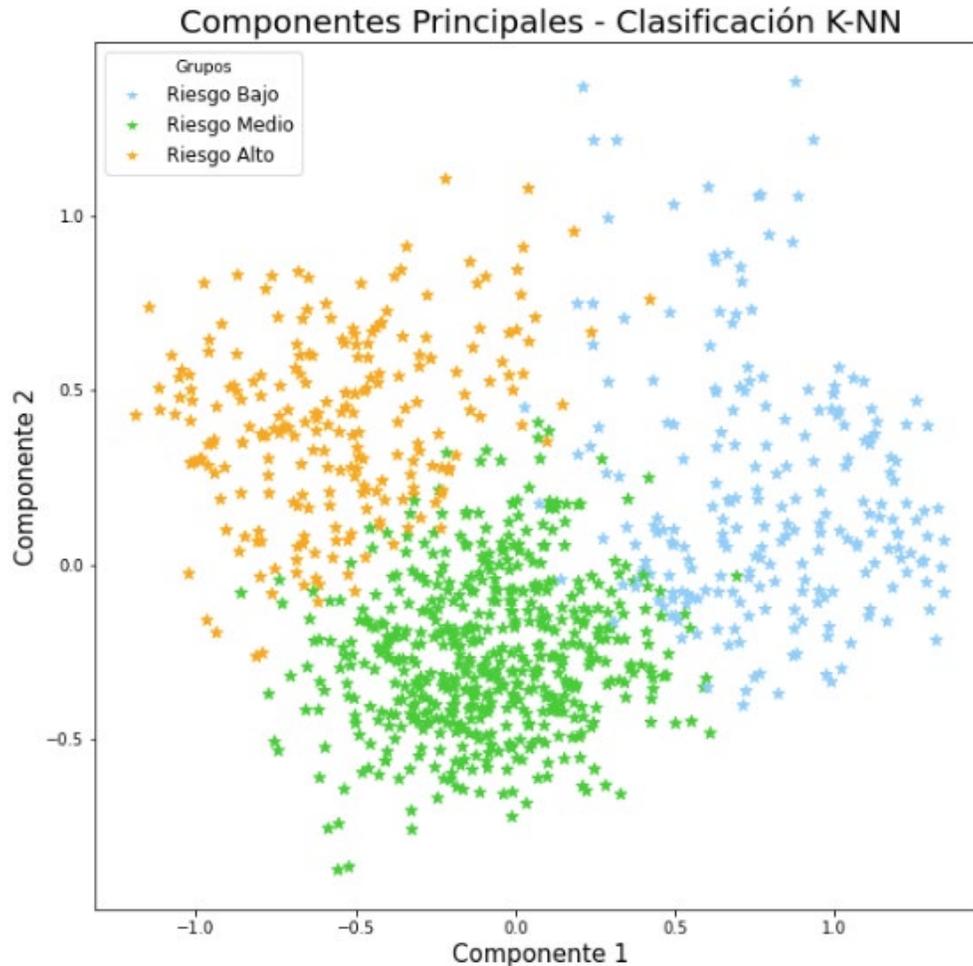


Gráfico 12 Clasificación K-NN  
 Nota: Gráfico de elaboración propia

Para finalizar, el último modelo utilizado para la clasificación de la información financiera de Fondos de Aseguramiento es el modelo de red neuronal: perceptrón multicapa (MLP), del cual se presentan sus resultados en la siguiente sección.

### 5.5. Perceptrón multicapa – Clasificador MLP

El algoritmo de perceptrón multicapa es aplicado mediante MLPClassifier de Scikit-Learn, el modelo es entrenado usando Backpropagation, método de dos fases:

adaptación-propagación, la red aprende la asociación entre los patrones de entrada y sus salidas propagándose desde la primera capa hacia capas superiores de la red, generando así una salida.

### 5.5.1 Selección de parámetros

La función que implementa el algoritmo permite diversos parámetros, los cuales se pueden consultar en la documentación de Scikit-Learn. En la implementación, únicamente se buscan los valores óptimos de los siguientes parámetros:

- `hidden_layer_sizes`: El *i*-ésimo elemento representa el número de neuronas en la *i*-ésima capa oculta. En el caso de la clasificación de Fondos de Aseguramiento, se realizó el comparativo entre 10, 50 y 100 capas ocultas.
- `activation`: Función de activación de la capa oculta:

- 'identity', función lineal

$$f(x) = x$$

- 'logistic', la función logística sigmoidea

$$f(x) = \frac{1}{(1 + \exp(-x))}$$

- 'tanh', la función tan hiperbólica

$$f(x) = \tanh(x)$$

- 'relu', la función de unidad lineal rectificadas

$$f(x) = \max(0, x)$$

- solver
  - 'lbfgs' es un optimizador de la familia de métodos cuasi-Newton.
  - 'sgd' se refiere al descenso del gradiente estocástico.
  - 'adam' optimizador estocástico basado en gradientes propuesto por Kingma, Diederik y Jimmy Ba.

### 5.5.2 Resultados

Los parámetros seleccionados, a través de GridSearchCV, fueron los siguientes:

- hidden\_layer\_sizes: 10.
- activation: "relu".
- solver: "lbfgs".

Como resultado, sobre el conjunto de prueba se obtuvo un **98.7% de exactitud o accuracy**; mediante el siguiente gráfico se pueden observar los puntos no coincidentes con la clasificación realizada a través de k-medias:

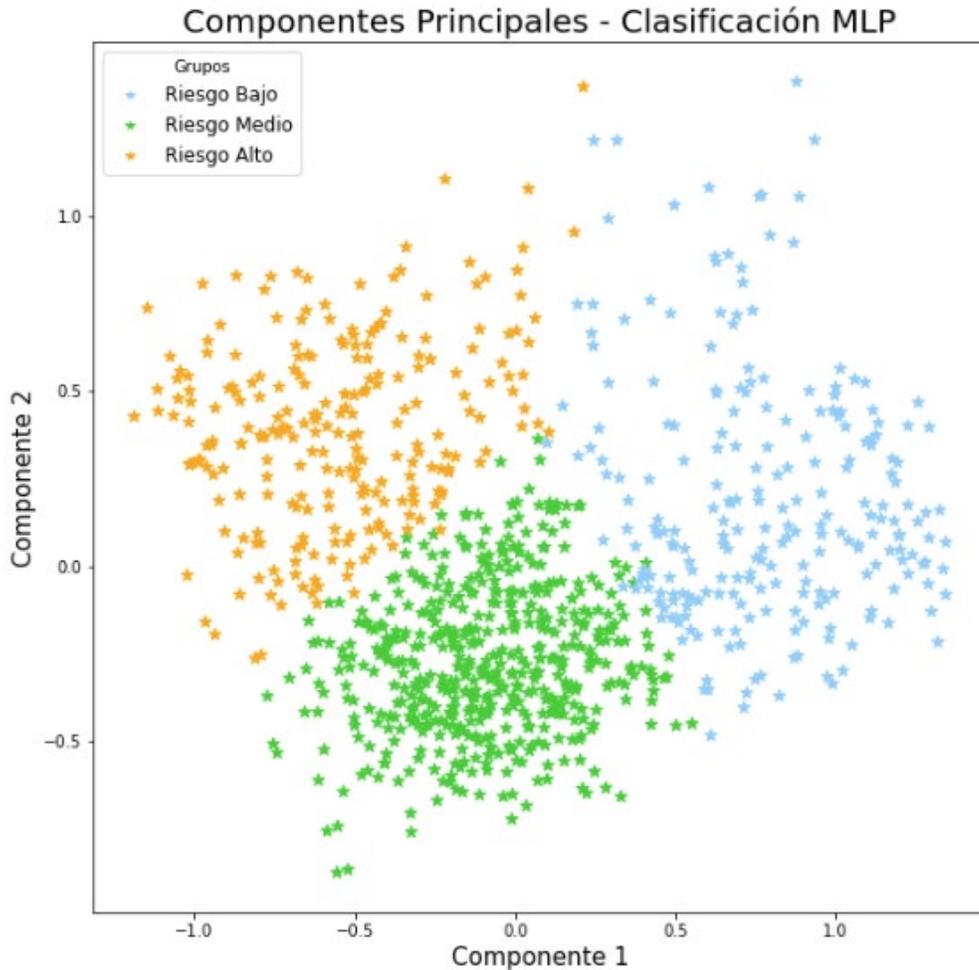


Gráfico 13 Clasificación MLP  
 Nota: Gráfico de elaboración propia

## 5.6 Comparación entre modelos

La siguiente tabla muestra la evaluación a través de las métricas “Precision”, “Recall”, “F1-score” y “Accuracy” de los diferentes modelos utilizados, los mejores resultados se obtienen a través de perceptrón multicapa, seguido de máquina de vector de soporte y bosque aleatorio, por último, el de menor desempeño es k vecinos más próximos, no obstante, en general se presentaron buenos resultados en cada uno de los modelos.

Tabla 12 Evaluaciones por método  
 Nota: Tabla de elaboración propia

<b>Método</b>	Clase	Precision	Recall	F1-score	Accuracy
<b>Bosque aleatorio</b>	Bajo	0.967	0.983	0.975	0.984
	Medio	0.988	0.986	0.987	
	Alto	0.992	0.980	0.986	
<b>Máquina de vector de soporte (SVM)</b>	Bajo	0.987	0.992	0.990	0.993
	Medio	0.994	0.996	0.995	
	Alto	0.996	0.988	0.995	
<b>K vecinos más próximos (K-NN)</b>	Bajo	0.978	0.924	0.950	0.961
	Medio	0.943	0.994	0.968	
	Alto	0.987	0.927	0.956	
<b>Perceptrón Multicapa</b>	Bajo	0.996	0.987	0.992	0.996
	Medio	0.994	1.000	0.997	
	Alto	1.000	0.996	0.998	

## Conclusiones

El objetivo principal de este trabajo fue encontrar un modelo de aprendizaje automático clásico o aprendizaje profundo que sirva como alternativa al proceso actual para la clasificación de la información financiera de Fondos de Aseguramiento. Se evaluaron diferentes modelos, utilizando el modelo k-medias como clasificador base para la exploración y etiquetado de datos. Los demás modelos cuentan con el objetivo de facilitar la implementación futura.

Los resultados obtenidos mostraron que el perceptrón multicapa tuvo un desempeño favorable, alcanzando una exactitud del 99.6% en las predicciones. Le siguen máquina de vector de soporte y el bosque aleatorio, con una exactitud, del 99.3% y 98.4%, respectivamente. Por último, el modelo de k vecinos más próximos obtuvo el menor desempeño con una exactitud del 96.1%. En general, se observa un buen desempeño en los modelos evaluados, con indicadores de precisión superiores al 90%.

Como principales aspectos revisados se han planteado algunas consideraciones en el cálculo de los indicadores financieros para evitar valores atípicos, así como la exclusión de información inconsistente, asimismo, se analizó la viabilidad de utilizar k-medias como clasificador base, observando que brinda un mayor peso a variables significativas, en especial a los gastos de operación de las instituciones analizadas que tienen una correlación negativa respecto al resultado del ejercicio.

Además, este modelo de clasificación financiera de fondos de aseguramiento también ofrece nuevas alternativas para los procesos de supervisión, brindando pautas para examinar aspectos previamente no analizados de la información. Por otro lado, se logró reducir la dependencia de supuestos en la clasificación por riesgo financiero, lo que conlleva una mayor objetividad en la evaluación. Además,

mediante la elaboración de manuales de procedimiento, la implementación de estos procesos se simplificaría considerablemente para los supervisores de la Comisión Nacional de Seguros y Fianzas (CNSF). En consecuencia, se promueve una supervisión más efectiva y eficiente, al tiempo que se asegura un análisis más completo y una toma de decisiones más fundamentada.

Se sugiere que futuros estudios consideren la evaluación de la clasificación existente, la cual muestra una coincidencia aproximada del 60% con los resultados obtenidos mediante el modelo k-medias. Esto permitiría comprender mejor las similitudes y diferencias entre ambos enfoques, brindando una perspectiva más completa sobre su desempeño y utilidad.

Además, se recomienda realizar análisis más exhaustivos para garantizar que no haya sobreajuste en los datos utilizados en el modelo de aprendizaje automático. Esto implica la aplicación de técnicas de validación cruzada y la exploración de diferentes conjuntos de datos para verificar la robustez y generalización de los resultados obtenidos.

Asimismo, sería beneficioso buscar alternativas para mejorar la calidad de la información reportada por los Fondos de Aseguramiento. Esto podría incluir la implementación de controles de calidad más estrictos, la estandarización de los formatos de presentación de los datos financieros y la promoción de buenas prácticas en la generación y reporte de la información.

# Clasificación de Fondos de Aseguramiento Agropecuario y Rural mediante Inteligencia Artificial: Caso Práctico 2024

El análisis de datos de gran volumen y complejidad requiere de herramientas que no solo procesen la información, sino que también encuentren relaciones y patrones no perceptibles para todos los usuarios, ya sea por cuestiones de experiencia de este o diferencias en sus criterios.

Para no depender de reglas predefinidas se seleccionó el algoritmo k-medias para la clasificación base de los indicadores financieros de los Fondos de Aseguramiento, este modelo se adapta y aprende de los datos, identificando los mencionados patrones ocultos, proporcionando una clasificación más precisa y eficiente, que puede ser útil cuando no se tienen etiquetas previas y se desea explorar la estructura de los datos.

Con la información del 2020 al 2023 etiquetada, para la clasificación de los datos de 2024 se usa el algoritmo perceptrón multicapa que cuenta con capacidad para adaptarse y aprender de datos previos.

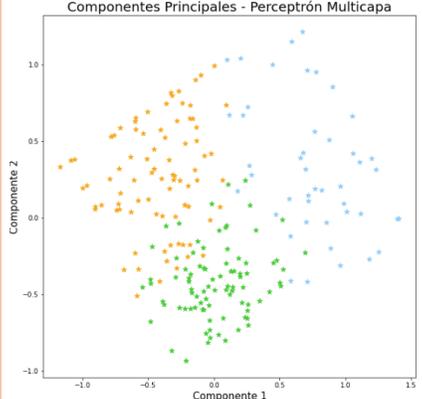
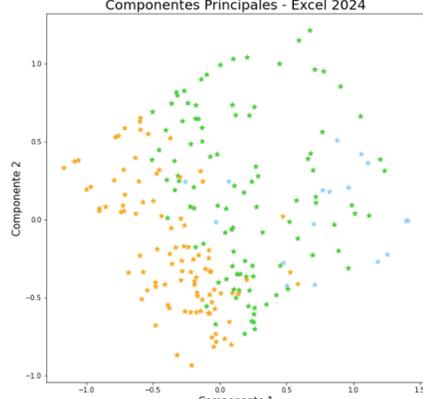
Para constatar la mejora del proceso, a continuación, se presenta el comparativo del proceso actual y el propuesto con base en la clasificación de los datos de 2024:

*Tabla 13 Comparativo clasificación con IA vs Excel*  
Nota: Tabla de elaboración propia

	<b>Clasificación con inteligencia artificial</b>	<b>Clasificación actual con Excel</b>
<b>Descripción del proceso</b>	1. Recolección y limpieza de la información del 2024 a través de macros (concentrado de archivos en formato Excel entregado a las CNSF).	1. Recolección y limpieza de la información del 2024 a través de macros (concentrado de archivos en formato Excel entregado a la CNSF).

	<b>Clasificación con inteligencia artificial</b>	<b>Clasificación actual con Excel</b>
	<ol style="list-style-type: none"> <li>2. Cálculo de indicadores financieros 2024 en Excel.</li> <li>3. Normalización de los datos.</li> <li>4. Aplicación del modelo de perceptrón multicapa.</li> <li>5. Análisis estadístico de grupos obtenidos.</li> <li>6. Representación gráfica de los datos.</li> </ol> <p><b>Número de pasos: 6</b></p>	<ol style="list-style-type: none"> <li>2. Cálculo de indicadores financieros 2024 en Excel.</li> <li>3. Clasificación a través de los indicadores denominados "Indicadores relevantes"</li> <li>4. Clasificación a través de los indicadores denominados "Indicadores técnicos financieros y de reaseguro"</li> <li>5. Obtención de la clasificación definitiva a través del máximo entre los indicadores: "Indicadores relevantes" e "Indicadores técnicos financieros y de reaseguro".</li> <li>6. Comparativo de la clasificación con referencia al periodo anterior.</li> </ol> <p><b>Número de pasos: 6</b></p>
<b>Herramientas utilizadas</b>	Excel Python <b>Número de herramientas: 2</b>	Excel <b>Número de herramientas: 1</b>
<b>Tiempo de Ejecución</b>	<b>5 horas</b>	<b>7 horas</b>
<b>Nivel de automatización</b>	<ul style="list-style-type: none"> <li>• 2 procesos manuales (puntos 2 y 5 del proceso).</li> <li>• 1 proceso parcialmente automatizado (punto 1 del proceso).</li> <li>• 3 procesos automatizados (puntos 3, 4 y 6 del proceso).</li> </ul> <p><b>Porcentaje de automatización: 58.33%</b></p>	<ul style="list-style-type: none"> <li>• 2 procesos manuales (puntos 2 y 6 del proceso).</li> <li>• 4 procesos parcialmente automatizados (puntos 1, 3, 4 y 5) del proceso.</li> </ul> <p><b>Porcentaje de automatización: 33.33%</b></p>

	<b>Clasificación con inteligencia artificial</b>	<b>Clasificación actual con Excel</b>
<b>Métricas de Evaluación</b>	<p>El modelo se evaluó con:</p> <p>Accuracy Precision Recall F1- Score</p>	No se aplican métricas de evaluación.
<b>Ventajas</b>	<ol style="list-style-type: none"> <li>1. Se puede procesar a mayor velocidad.</li> <li>2. La inteligencia artificial permite manejar conjunto de datos más grandes.</li> <li>3. Mayor nivel de automatización.</li> <li>4. El modelo utilizado puede aprender y adaptarse a través del tiempo.</li> <li>5. Se permite el manejo de datos más complejos y no lineales, identificando patrones y relaciones.</li> <li>6. Elimina sesgos o subjetividades del criterio humano.</li> <li>7. Reducción de errores humanos.</li> </ol>	<ol style="list-style-type: none"> <li>1. Mejor manejo de información que no está bien estructurada o es ambigua.</li> <li>2. Costo de implementación menor.</li> <li>3. No se requieren conocimientos avanzados en programación.</li> <li>4. Compatible con otras herramientas de Microsoft para informes y presentaciones.</li> </ol>
<b>Desventaja</b>	<ol style="list-style-type: none"> <li>1. Sesgo en los datos y modelos.</li> <li>2. La lógica interna del modelo resulta difícil de comprender.</li> <li>3. Mayor costo en recursos computacionales y personal especializado para el mantenimiento del modelo.</li> <li>4. Se requiere actualización continua.</li> <li>5. Poca adaptación en situaciones que difieren de las condiciones de</li> </ol>	<ol style="list-style-type: none"> <li>1. Requiere de mayor tiempo y esfuerzo.</li> <li>2. Susceptible a errores humanos.</li> <li>3. Subjetividad en la interpretación.</li> <li>4. Falta de capacidad predictiva.</li> <li>5. Este método da un mayor peso a ciertos indicadores, invisibilizando la relación entre los gastos operativos y los resultados de los Fondos.</li> </ol>

	<b>Clasificación con inteligencia artificial</b>	<b>Clasificación actual con Excel</b>
	entrenamiento del modelo.	
<b>Gráfica</b>	 <p><i>Gráfico 14 Caso práctico 2024 perceptrón multicapa (Elaboración propia)</i></p>	 <p><i>Gráfico 15 Caso práctico 2024 metodología tradicional (Elaboración propia)</i></p>

Por lo anterior, ambos procesos cuentan con el mismo número de pasos, sin embargo, el modelo que emplea inteligencia artificial tiene un porcentaje mayor de automatización, cuenta con métricas de evaluación, así como, con un mayor número de ventajas y gráficamente tiene una mejor delimitación de los grupos.

## Bibliografía

Agroasemex. (diciembre de 2019). *CONAMER*. Obtenido de Plan de Negocios 2019 - 2024  
Agroasemex:  
<https://cofemersimir.gob.mx/expediente/25848/mir/51539/anexo/6110920>

Agroasemex. (s.f.). *S265 "Programa de Aseguramiento Agropecuario"*. Obtenido de  
Transparencia presupuestaria:  
[https://www.transparenciapresupuestaria.gob.mx/work/models/PTP/Reingenieria\\_  
Gasto/imagenes/Ventanas/Ramo\\_6/06S265.pdf](https://www.transparenciapresupuestaria.gob.mx/work/models/PTP/Reingenieria_Gasto/imagenes/Ventanas/Ramo_6/06S265.pdf)

Agroasemex, S.A. (7 de enero de 2016). *Gobierno de México*. Obtenido de Presentación y  
actualización de Fondos de Aseguramiento 2016:  
[https://www.gob.mx/agroasemex/documentos/presentacion-y-actualizacion-de-  
fondos-de-aseguramiento-2016](https://www.gob.mx/agroasemex/documentos/presentacion-y-actualizacion-de-fondos-de-aseguramiento-2016)

Aprende IA. (s.f.). *Aprende IA*. Obtenido de Kernel: [https://aprendeia.com/kernel-maquinas-  
vectores-de-soporte-clasificacion-regresion/](https://aprendeia.com/kernel-maquinas-vectores-de-soporte-clasificacion-regresion/)

Bobadilla, J. (2020). *Machine Learning y Deep Learning*. Bogotá, Colombia: Ra-Ma.

Nuzzi, R., Boscia, G., Marolo, P., & Ricardi, F. (2021). *The Impact of Artificial Intelligence and Deep Learning in Eye Diseases: A Review*. *Frontiers in Medicine*, 8, artículo 710329: <https://doi.org/10.3389/fmed.2021.710329>

CNSF. (s.f.-a). Obtenido de *Índice de cobertura de reservas técnicas*:  
<https://www.cnsf.gob.mx/EntidadesSupervisadas/InstitucionesSociedadesMutualistas/Indicadores/Paginas/IndiceCoberturaReservasTecnicas.aspx>

CNSF. (s.f.-b). Obtenido de *¿Qué hacemos? Gobierno de México*:  
<https://www.gob.mx/cnsf/que-hacemos>

CNSF. (2019). Obtenido de *Anexo 22.6.1 de la Circular Única de Seguros y Fianzas: Catálogo de cuentas para los Fondos de Aseguramiento Agropecuario y Rural*:  
[https://www.gob.mx/cms/uploads/attachment/file/522694/ANEXO\\_22.6.1.pdf](https://www.gob.mx/cms/uploads/attachment/file/522694/ANEXO_22.6.1.pdf)

Comisión Federal de Mejora Regulatoria (COFEMER). (s. f.). *Contabilidad de las entidades aseguradoras* [PDF]. SMIR. Obtenido de:  
<https://www.cofemersimir.gob.mx/expediente/4129/mir/11329/archivo/416075>

CONDUSEF. (s.f.). Obtenido de <https://www.gob.mx/condusef/prensa/sabes-que-es-una-mutualista>

Fundación Mapfre. (s.f.). Obtenido de Educación Financiera *¿Qué es el seguro?*:  
<https://www.fundacionmapfre.mx/educacion-divulgacion/educacion-financiera/seguros/que-es-el-seguro/>

Fundación Mapfre. (s.f.). *Seguros y Pensiones para TODOS*. Obtenido de  
<https://segurosypensioneparatodos.fundacionmapfre.org/seguros/definicion-seguro-asegurar/funcionamiento-mercado-seguros-asegurar/reaseguro/>

Gil Martínez, C. (junio de 2018). *RPubs by RStudio*. Obtenido de Máquina de vector soporte:  
[https://rpubs.com/Cristina\\_Gil/SVM](https://rpubs.com/Cristina_Gil/SVM)

IBM. (s.f.). Obtenido de ¿Qué es Machine Learning?: <https://www.ibm.com/mx-es/analytics/machine-learning>

INEGI. (s.f.). Obtenido de <https://www.inegi.org.mx/programas/inpc/2018/>

*Ley de Fondos de Aseguramiento Agropecuario y Rural.* (13 de mayo de 2005). Obtenido de <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFAAR.pdf>

López Sotelo, J. A. (2021). *Deep Learning Teoría y Aplicaciones*. Bogotá: Alpha Editorial.

MAPFRE. (s.f.). *Fundación Mapfre*. Obtenido de <https://www.fundacionmapfre.org/publicaciones/diccionario-mapfre-seguros/>

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media

Oracle. (s.f.). Obtenido de ¿Qué es la inteligencia artificial? Obtenga más información sobre la inteligencia artificial: <https://www.oracle.com/mx/artificial-intelligence/what-is-ai/>

Rachka, S., & Mirjalili, V. (2019). *Python Machine Learning: Aprendizaje automático y aprendizaje profundo con Python, scikit-learn y TensorFlow*. Marcombo

Ramírez Aldana, L. (diciembre de 2017). *Comisión Nacional de Seguros y Fianzas*. Obtenido de El seguro Agrícola y de Animales en México: [https://www.gob.mx/cms/uploads/attachment/file/293683/165.\\_El\\_Seguro\\_Agr\\_col\\_a\\_y\\_de\\_Animales\\_en\\_M\\_xico.pdf](https://www.gob.mx/cms/uploads/attachment/file/293683/165._El_Seguro_Agr_col_a_y_de_Animales_en_M_xico.pdf)

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Nueva York: Cambridge University Press

Tapia, E. D. (2006). *El seguro agropecuario en México: experiencias recientes*. México: Naciones Unidas

Valdemar Cuevas, E., Avalos, O., Primitivo, E., Valdivia, A., & Pérez, M. A. (2021). *Introducción al Machine Learning con MATLAB*. Marcombo

Wood, T. (s.f.). *DeepAI*. Obtenido de Sigmoid Function: <https://deepai.org/machine-learning-glossary-and-terms/sigmoid-function>