



FACULTAD DE POSTGRADO

TESIS DE POSTGRADO

**MODELO PREDICTIVO DE VENTA CRUZADA EN
PRODUCTOS DE AUTO Y VIDA EN UNA ASEGURADORA**

SUSTENTADO POR:

ROBERTO ENRIQUE MEZA VILLEDA

PREVIA INVESTIDURA AL TÍTULO DE

MÁSTER EN:

ANALÍTICA DE NEGOCIOS

TEGUCIGALPA, M.D.C, HONDURAS, C.A.,

SEPTIEMBRE, 2023

**UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA
UNITEC**

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

RECTORA

ROSALPINA RODRÍGUEZ GUEVARA

PRORECTOR/ SECRETARIO GENERAL

ROGER MARTÍNEZ MIRALDA

VICERRECTOR ACADÉMICO

JAVIER ABRAHAM SALGADO LEZAMA

DIRECTORA NACIONAL DE POSTGRADO

ANA DEL CARMEN RETALLY VARGAS

**MODELO PREDICTIVO DE VENTA CRUZADA EN
PRODUCTOS DE AUTO Y VIDA EN UNA ASEGURADORA**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE
MÁSTER EN**

ANALÍTICA DE NEGOCIOS

ASESOR METODOLÓGICO

Msc ALBA GABRIELA GARAY ROMERO

ASESOR TEMÁTICO

ING. ANTONIO SANDRES

MIEMBROS DE LA TERNA:

ING. CARLOS AMADOR

ING. JOSUÉ MEJÍA

ING. GERARDO LUJANO

DERECHOS DE AUTOR

© Copyright (2023)

(ROBERTO ENRIQUE MEZA VILLEDA)

Todos los derechos son reservados.



FACULTAD DE POSTGRADO

**MODELO PREDICTIVO DE VENTA CRUZADA EN
PRODUCTOS DE AUTO Y VIDA EN UNA ASEGURADORA**

NOMBRE DEL MAESTRANDO:

ROBERTO ENRIQUE MEZA VILLEDA

RESUMEN

El presente estudio pretendía determinar la factibilidad técnica de diseñar e implantar un modelo numérico de predicción de venta cruzada de clientes de seguros de autos y vida en una entidad aseguradora en Honduras, evidenciando el auge y valor agregado del análisis de datos como herramienta de impulso comercial para la organización, influyendo en el aumento de colocación de productos y permitiendo un incremento en la generación de utilidad por cliente. El estudio ha sujetado un enfoque cuantitativo y causal, ya que ha requerido de análisis y transformación de datos estadísticos asociados a las ventas de seguros sucedidas durante los años 2020 y 2023 en MAPFRE Seguros Honduras y la experimentación a través de distintos algoritmos de predicción, a fin de conocer el modelo idóneo para predecir los candidatos al seguro adicional de forma asertiva, y los recursos necesarios para implantar el modelo dentro de la organización.

Palabras claves: Aprendizaje automático, seguro de auto, seguro de vida, algoritmo, predicción.



GRADUATE SCHOOL

**PREDICTIVE MODEL OF CROSS-SELLING IN AUTO AND
LIFE PRODUCTS IN AN INSURANCE COMPANY**

STUDENT NAME:

ROBERTO ENRIQUE MEZA VILLEDA

ABSTRACT

The present study aimed to determine the technical feasibility of designing and implementing a numerical model for predicting cross-selling of auto and life insurance clients in an insurance entity in Honduras, evidencing the rise and added value of data analysis as a commercial promotion tool. for the organization, influencing the increase in product placement and allowing an increase in the generation of profit per customer. The study has taken a quantitative and causal approach, since it has required analysis and transformation of statistical data associated with insurance sales that occurred during the years 2020 and 2023 at MAPFRE Seguros Honduras and experimentation through different prediction algorithms, to in order to know the ideal model to predict the candidates for additional insurance assertively, and the resources necessary to implement the model within the organization.

Keywords: Machine learning, auto insurance, life insurance, algorithm, prediction.

DEDICATORIA

Dedico este trabajo a YHWH, quien con su bendición llena siempre mi vida, por ser mi guía y acompañarme en el transcurso de este bonito camino, brindándome paciencia y sabiduría para culminar con éxito mis metas propuestas. A mis padres, Mario Meza y Daisy Villeda quien con sus consejos han sabido guiarme siempre por el camino correcto y han sido claves para todos mis logros. A mi amada esposa, Michell Cabrera, que con su apoyo incondicional me ha animado a continuar con mis estudios y ha sido un pilar importante inspirándome a siempre continuar superándome. A mis hermanos, Mario Meza, Alejandra Meza, Josue Meza y Andrea Meza por ser un ejemplo y motivación para mí durante todos mis años de formación, les agradezco todas sus enseñanzas y comprensión. Finalmente, infinitas gracias a mis familiares y amigos por el apoyo moral, que me brindaron a lo largo de esta etapa de mi vida.

Roberto Enrique Meza Villeda

AGRADECIMIENTO

De manera especial agradezco a mi familia Meza Villeda, por su constante e incondicional apoyo en mi formación profesional, por ser los principales promotores de mis sueños, por confiar y creer en mis expectativas, por los consejos, valores y principios que me han inculcado. A mis amistades, que me brindaron su apoyo, gracias infinitas por toda su ayuda y buena voluntad.

A la empresa MAPFRE Honduras, por abrirme las puertas y haber permitido realizar todo el proceso analítico, agradezco la confianza y accesibilidad a la información, así como la apertura a proponer un modelo predictivo de venta cruzada para cumplir con el requisito exigido por la Universidad Tecnológica Centroamericana para optar al título de Máster en Analítica de Negocios.

A la Universidad Tecnológica Centroamericana, por su valioso aporte para nuestra formación académica, por ser la sede de todo el conocimiento adquirido en este camino. Al personal académico quienes supieron impartir sus conocimientos con mucho esfuerzo, para hacer posible nuestra preparación profesional. Finalmente expresamos nuestro más grande y sincero agradecimiento a la Ingeniera Alba Gabriela Garay, a quien consideramos una persona muy profesional, de quien admiramos su inteligencia y gran calidad humana. También por su orientación metodológica, quien con su dirección, conocimiento, enseñanza y colaboración permitió el desarrollo de este trabajo.

INDICE DE CONTENIDO

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN	1
1.1INTRODUCCIÓN	1
1.2ANTECEDENTES DEL PROBLEMA	2
1.3DEFINICIÓN DEL PROBLEMA	4
1.3.1. ENUNCIADO DEL PROBLEMA	4
1.3.2. FORMULACIÓN DEL PROBLEMA	6
1.3.3. PREGUNTAS DE INVESTIGACIÓN	6
1.4OBJETIVOS DEL PROYECTO	6
1.4.1. OBJETIVO GENERAL.....	6
1.4.2. OBJETIVOS ESPECÍFICOS	7
1.5JUSTIFICACIÓN.....	7
CAPÍTULO II. MARCO TEÓRICO DE LA INVESTIGACIÓN.....	9
2.1ANÁLISIS DE LA SITUACIÓN ACTUAL.....	9
2.2CONCEPTUALIZACIÓN.....	14
2.2.1. RIESGO	14
2.2.2. PRIMA.....	15
2.2.3. ASEGURADO.....	15
2.2.4. VENTA DEL SEGURO	15
2.2.5. VENTA CRUZADA	15
2.2.6. TECNOLOGÍA.....	16
2.2.7. APRENDIZAJE AUTOMATICO.....	16
2.2.8. SEGURO DE AUTOMOVIL.....	16
2.2.9. SEGURO DE VIDA	16
2.2.10. SEGMENTACIÓN DE MERCADO	17
2.3TEORÍAS DE SUSTENTO	17
2.3.1. BASES TEORICAS	17
2.3.1.1. APRENDIZAJE AUTOMÁTICO SUPERVISADO	17
2.3.1.2. APRENDIZAJE AUTOMÁTICO NO SUPERVISADO	23

2.3.1.3. APRENDIZAJE AUTOMÁTICO SEMISUPERVISADO.....	25
2.3.1.4. APRENDIZAJE AUTOMÁTICO REFORZADO.....	25
2.3.2. METODOLOGÍAS DESARROLLADAS POR OTROS	
INVESTIGADORES O EXPERTOS.	25
2.3.2.1. MODELO PREDICTIVO DE VENTA CRUZADA EN PRODUCTOS DE	
VIDA Y SALUD:.....	25
2.3.2.2. DESARROLLO DE UN MODELO DE RECOMENDACIÓN DE	
COMPRA PARA CLIENTES DE UNA EMPRESA DE SEGUROS	26
2.3.3. INSTRUMENTOS UTILIZADOS.....	27
2.4 MARCO LEGAL.....	29
CAPÍTULO III. METODOLOGÍA	30
3.1 CONGRUENCIA METODOLÓGICA	30
3.1.1. MATRIZ METODOLÓGICA.....	30
3.1.2. ESQUEMA DE VARIABLES DE ESTUDIO.....	32
3.1.4. HIPÓTESIS	34
3.2 ENFOQUE Y MÉTODOS	34
3.3 DISEÑO DE LA INVESTIGACIÓN	35
3.3.1. POBLACIÓN	35
3.3.2. MUESTRA	35
3.3.3. TÉCNICAS DE MUESTREO.....	35
3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS	36
3.5 FUENTES DE INFORMACIÓN	39
3.5.1. FUENTES PRIMARIAS.....	39
3.5.2. FUENTES SECUNDARIAS.....	¡Error! Marcador no definido.
CAPÍTULO IV. RESULTADOS Y ANÁLISIS	41
4.1 INFORME DE PROCESO DE RECOLECCIÓN DE DATOS.....	41
4.2 RESULTADO Y ANALISIS DE LAS TECNICAS APLICADAS	45
4.2.1. EXPLORACIÓN DE DATOS	45
4.2.2. PROCESO DE TRANSFORMACIÓN.....	51
4.2.2.1. BBDD_Segmentación_Demografica.....	51

4.2.2.2. BBDD_Siniestros_Autos	51
4.2.2.3. BBDD_Prima.....	52
4.2.2.4. BBDD_Suma asegurada	52
4.2.2.5. BBDD_Coberturas Adicionales.....	52
4.2.2.6. BBDD_Nivel de ingresos	52
4.2.2.7. BBDD_Venta Cruzada	53
4.2.3. ANÁLISIS DE CORRELACIÓN	55
4.2.4. APLICACIÓN DE ALGORITMO.....	57
4.2.5. COMPARACIÓN DE RESULTADOS	60
4.3 RESULTADOS Y ANÁLISIS DE LOS DATOS ENCONTRADOS CON OTRAS TÉCNICAS QUE LOS AUTORES ESTIMEN Y JUSTIFIQUEN QUE SON NECESARIAS.....	63
4.4 ESTUDIO DE ALGORITMOS CON BASE DE NUEVA EMISIÓN.....	65
4.4.1. PROYECCIÓN DE VENTAS.....	65
4.4.2. PROYECCIÓN DE VENTAS.....	66
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.....	67
5.1 CONCLUSIONES	67
5.2 RECOMENDACIONES	69
CAPITULO VI. APLICABILIDAD	70
6.1 NOMBRE DE LA PROPUESTA	70
6.2 JUSTIFICACION DE LA PROPUESTA	70
6.3 ALCANCE DE LA PROPUESTA	71
6.3.1. OBJETIVO GENERAL.....	71
6.3.2. OBJETIVOS ESPECÍFICOS	71
6.4 DESCRIPCIÓN Y DESARROLLO A DETALLE DE LA PROPUESTA	71
6.4.1. DESCRIPCIÓN DE LA PROPUESTA.....	71
6.4.2. DESARROLLO DE LA PROPUESTA	72
6.4.2.1. Querys de consulta a base de datos.....	72
6.4.2.2. Análisis y transformación de datos.....	73
6.4.2.3. Entrenamiento de algoritmo	76

6.4.2.4. Predicción de candidatos	78
6.4.2.5. Proyección de ventas	78
6.5 CRONOGRAMA DE IMPLEMENTACIÓN.....	79
6.6 CONCORDANCIA DE SEGMENTO DE LA TESIS CON LA PROPUESTA.....	80
BIBLIOGRAFIA	83
ANEXOS	88

INDICE DE TABLAS

TABLA 1 - MATRIZ METODOLÓGICA.....	31
TABLA 2 - ESQUEMA DE LAS VARIABLES.....	32
TABLA 3 - ESQUEMA DE LAS VARIABLES.....	33
TABLA 4 – METODOLOGÍA DE INVESTIGACIÓN IMPLEMENTADA.....	34
TABLA 5 – PREDICCIÓN DE BASES.....	38
TABLA 6 – MATRIZ DE CONFUSIÓN.....	39
TABLA 7 – DICCIONARIO DE VARIABLES	53
TABLA 8 – CARACTERÍSTICAS DE SET DE DATOS	57
TABLA 9 – COMPARACIÓN RESULTADOS DE ALGORITMOS	62
TABLA 10 – COMPARACIÓN RESULTADOS DE CURVA ROC-AUC.....	65
TABLA 11 –PREDICCIÓN DE CANDIDATOS	65
TABLA 12 –PROYECCIÓN DE VENTAS.....	66
TABLA 13 – FASES IMPLANTACION MODELO PREDICTIVO	71
TABLA 14 –PREDICCIÓN DE CANDIDATOS	78
TABLA 15 –PROYECCIÓN DE VENTAS.....	78
TABLA 16 – GANTT DE IMPLANTACIÓN DE MODELO.....	79
TABLA 17 – PERFILES DE PUESTOS Y PRESUPUESTO REQUERIDO	79

INDICE DE FIGURAS

Figura 1. Regresión lineal.....	18
Figura 2. Árbol de decisión.....	19
Figura 3. Bosque Aleatorio.....	20
Figura 4. Esquema de red neuronal.	21
Figura 5. Regresión logística.	22
Figura 6. Máquina de Vectores de Soporte.....	23
Figura 7. Clustering.	24
Figura 8. PCA.	24
Figura 9. Comparación resultados Matriz de confusión.	26
Figura 10. Comparación predicción Compra.....	26
Figura 11. Comparación Clasificación Total.....	27
Figura 12. Nodo Segmentación Demográfica.....	41
Figura 13. Nodo Siniestros Autos.....	42
Figura 14. Nodo Prima.....	42
Figura 15. Nodo Suma Asegurada.	43
Figura 16. Nodo Coberturas Adicionales.	43
Figura 17. Nodo Nivel de Ingresos.....	44
Figura 18. Nodo Venta Cruzada.	44
Figura 19. Comparación Distribución por Genero.	45
Figura 20. Comparación Distribución por Edad.....	46
Figura 21. Comparación Distribución por Siniestro.....	47
Figura 22. Comparación Distribución por Suma Asegurada.....	48
Figura 23. Comparación Distribución por Prima.....	49
Figura 24. Comparación Distribución por Producto.....	50
Figura 25. Matriz de Correlación.....	56
Figura 26. Partición de Datos.	57
Figura 27. Entrenamiento Algoritmo XGBoost.....	58

Figura 28. Entrenamiento de Algoritmo Bosques Aleatorios.....	59
Figura 29. Matriz de Confusión XGBoost.....	60
Figura 30. Matriz de Confusión Bosques Aleatorios.....	61
Figura 31. Diagrama ROC Curve.	63
Figura 32. Resultado ROC Curve XGBoost.....	64
Figura 33. Resultado ROC Curve Bosques Aleatorios.....	64
Figura 34. División de Datos.	76
Figura 35. Entrenamiento Algoritmo XGBoost.....	77

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

Satisfacer las necesidades del cliente dándole el producto o servicio adecuado eficientemente de manera que se cumplan o superen sus expectativas, es uno de los desafíos más importante dentro de las entidades aseguradoras en la actualidad, cuando uno de los procedimientos pertenecientes a la dirección comercial es concebido para lograr la diversificación del portafolio adquirido por los asegurados existentes mediante el ofrecimiento de productos adicionales, que respondan a las necesidades particulares de los mismos, nos referimos a venta cruzada, la técnica de venta se basa en la comprensión del comportamiento del cliente y en como satisfacerlo, lo cual permite maximizar la utilidad del usuario y de la empresa. Al escribir de venta cruzada asociamos estas palabras de forma inmediata al impacto económico (rentabilidad) que genera esta técnica de venta, sin embargo, el alcance va más allá y su materialización podría representar un impacto en la inversión de esfuerzos comerciales y la fidelización de clientes, estos impactos se vuelven más notables cuando el objeto de aplicabilidad es una entidad Aseguradora, con una clara misión inherente al rubro de identificar candidatos a seguros, anticipándose a sus necesidades y ofrecer productos para satisfacerlas. El cliente como centro en las organizaciones nos obliga a ser cada día más ágiles en la gestión del mismo, aprovechando las oportunidades de negocio de forma oportuna. Machine Learning es una rama de la inteligencia artificial que permite a un sistema aprender de los datos, patrones habituales y generar predicciones en base una gran cantidad de datos, por lo cual puede ser entrenado para resaltar los asegurados que puedan ser candidatos, para la colocación de productos adicionales al previamente contratado. Un modelo predictivo entrenado para identificar los militantes de una campaña de venta cruzada, brindara a la organización ese efecto diferenciador en la carrera por el crecimiento rentable y paliar la fuga de capital.

Se pretende a través del presente proyecto diseñar un modelo predictivo de venta cruzada en productos de auto y vida, entrenando un modelo numérico con variables correlacionadas, para que sea capaz de predecir los clientes potenciales para aceptar un seguro adicional al contratado, a fin de que sean pasados por un proceso posterior de venta a cargo de la entidad MAPFRE SEGUROS Honduras. Se realizará para la entidad el estudio de datos disponibles en sus sistemas de emisión

de seguros y se probaran algoritmos supervisados de clasificación, a fin de obtener un modelo con un grado de asertividad mayor al 70% en la predicción de potenciales clientes objetivo. Es igual de importante a la definición de un modelo predictivo, la determinación de la dirección comercial de la entidad de poner en producción la campaña de venta cruzada, estas aristas serán objeto de estudio de forma superficial durante el proyecto.

1.2 ANTECEDENTES DEL PROBLEMA

Parte fundamental del diseño de técnicas para mejorar la venta, es conocer el génesis del término y su aplicabilidad.

Jurídicamente hablando debe entenderse el concepto de “venta” según la ley de impuesto sobre venta de Honduras:

Todo acto que importe transferencia a título oneroso de una mercadería del dominio de una persona natural o jurídica al dominio de otra, o que tenga por fin último la transmisión de dicho dominio, independientemente de la designación que las partes den al contrato de origen o a la negociación en que se incluya o involucre y de la forma del pago del precio, sea éste en dinero o en especie. (Gobierno de Honduras, 1994)

De la misma forma La American Marketing Association define la venta:

El proceso personal o impersonal por el que el vendedor (consultor, asesor, empresario, Gerente, o cualquier persona que ofrezca un producto/servicio, con el ánimo de intercambiarlo por dinero), comprueba, activa y satisface las necesidades del comprador para el mutuo y continuo beneficio de ambos (del vendedor y el comprador) (De manera presencial o utilizando herramientas digitales). (American Marketing Association, s. f.)

Sobre estas definiciones jurídicas y etimológicas del concepto de venta se puede definir el concepto de venta cruzada.

Jones et al. (2012) define la venta cruzada “como la oferta a los clientes de productos o servicios adicionales que pueden aportarles un valor añadido” (p.27).

La regulación de comercio del país (Código de Comercio) habilita el aseguramiento de varios intereses a una misma persona, “Pueden asegurarse uno o varios intereses sobre una misma cosa o persona, y uno o varios intereses sobre conjuntos unitarios de cosas o personas” (Congreso Nacional de Honduras, 1950).

A pesar del uso generalizado de la técnica de venta cruzada en la práctica financiera, las entidades aseguradoras de Honduras carecen de estrategias impetuosas que establezcan mecanismos para catapultar este esfuerzo y lograr captar la fidelidad de sus clientes. Las técnicas de identificación de potenciales asegurados para integrar una campaña de venta cruzada en MAPFRE SEGUROS HONDURAS, se basan en técnicas muy tradicionales, básicamente en el análisis de la documentación disponible, la aparición de las nuevas tecnologías ha hecho evolucionar las técnicas para predecir candidatos a productos de seguro, particularmente las herramientas de aprendizaje automático o “machine learning” se han vuelto el recurso de los nuevos tiempos para impulsar las ventas en las compañías.

Machine learning es un componente importante del creciente campo de la ciencia de datos.

Mediante el uso de métodos estadísticos, los algoritmos se entrenan para hacer clasificaciones o predicciones, y descubrir información clave dentro de los proyectos de minería de datos. Esta información clave facilita posteriormente la toma de decisiones dentro de las aplicaciones y las empresas, lo que afecta idealmente a las métricas de crecimiento clave. (¿Qué es machine learning? | IBM, s. f.)

Parte fundamental en los antecedentes son los orígenes de la organización objeto de la investigación, la cual ya ha sido mencionada en algunos párrafos previos, el sitio web oficial de la compañía relata los hitos destacados en la evolución de la organización;

MAPFRE SEGUROS HONDURAS fue fundada en 1954 con el nombre de Aseguradora Hondureña, S.A. por un grupo de inversionistas que constituyeron una Sociedad Anónima para la comercialización de seguros y fianzas, abriendo sus puertas en la ciudad de Tegucigalpa para dar inicio a una larga trayectoria de servicio y constante innovación, extendiéndose posteriormente a las ciudades de San Pedro Sula, La Ceiba, Choluteca, Comayagua, Tocoa, Santa Rosa de Copan y Guanaja. En el 2006 la Aseguradora Hondureña, S.A. pasa a formar parte del Grupo Mundial, uno de los principales conglomerados financieros de Centroamérica, adquiriendo así una mayor solidez en el mercado asegurador a nivel regional. La compañía pasó a denominarse Aseguradora Hondureña Mundial S.A.

En el 2010, la Compañía Mundial Desarrollo de Negocios S. A., fue adquirida por Grupo MAPFRE, un grupo empresarial español independiente. En consecuencia, a partir del mes de mayo de 2011, la compañía cambió su denominación social a MAPFRE Seguros Honduras S.A., con un portafolio de productos que abarca todos los ámbitos de la actividad aseguradora. (Historia de MAPFRE Honduras - Seguros MAPFRE, 2022)

1.3 DEFINICIÓN DEL PROBLEMA

1.3.1. ENUNCIADO DEL PROBLEMA

Una cultura de seguros, es una cultura de previsión, es prepararse para los eventos inesperados, para las pérdidas o gastos urgentes que se tengan que enfrentar, sin embargo, en un país como Honduras, donde más del 59% de la población vive en pobreza, no permea este criterio, en su defecto se percibe el seguro como un lujo al alcance de un privilegiado sector o como un gasto con no retorno. La mala conceptualización del seguro en el país centroamericano enfrenta a las compañías dedicadas a este rubro a una batalla constante por atraer y mantener clientes y explotar al máximo la utilidad de ellos.

Desde sus comienzos en el país, MAPFRE SEGUROS HONDURAS se ha caracterizado por ser una empresa bastante tradicional en sus canales de venta y ha perseguido de forma exitosa la eficiencia en su estructura. En la actualidad la unidad comercial cuenta con 17 puestos de trabajo, los cuales son distribuidos en sus 4 sedes a nivel nacional, impulsando las ventas de pólizas de seguros a través de corredores externos, corredores internos, bancos y atención directa. Dentro de las estrategias de venta de la entidad, ha prevalecido la captación de nuevos asegurados como ruta para el logro de metas, hasta que en el año 2022 el nuevo CEO ha impulsado las campañas de fidelización de clientes a través de técnicas de venta cruzada.

Para mayo del 2022, la dirección comercial de MAPFRE SEGUROS HONDURAS lanzó su campaña de venta cruzada de pólizas de seguro a nivel nacional. El proyecto incluía etapas de generación de base de datos de cartera de asegurados, segmentación de cartera de asegurados entre asesores de venta y distribución de los prospectos de venta cruzada a través de los canales correspondientes. Una de las principales ventajas de la campaña, era que el costo de retención de clientes es significativamente más bajo en comparación con el costo de adquirir un nuevo asegurado, adicional refuerza la relación cliente-compañía. Actualmente se ha vuelto indispensable para la entidad la inclusión de campañas de venta cruzada dentro de sus ejercicios de planeación estratégica, llegando incluso a definirse metas de colocación de primas, apostando por el crecimiento rentable.

A pesar de la evidente ventaja que representa la venta cruzada de Seguros, la entidad ha señalado, que la identificación de prospectos a un seguro adicional está basada en un esfuerzo empírico de presunción y selección de candidatos por parte de la fuerza de ventas, ponderando por igual la probabilidad de que los candidatos acepten ampliar su porfolio de seguros adquiridos.

Con una cartera de más de 35,000 clientes activos, MAPFRE ha señalado la imperante necesidad de implantar herramientas que prioricen de forma automática los prospectos a venta cruzada, optimicen los esfuerzos al máximo y mitiguen la frustración por no alcanzar el umbral máximo de utilidad de cada cliente.

1.3.2. FORMULACIÓN DEL PROBLEMA

Determinar la factibilidad de implantación de un modelo de predicción de candidatos a venta cruzada de seguros de auto y vida, por lo cual se plantea la interrogante: ¿Cómo influye el histórico de datos de asegurados, en la obtención de un modelo predictivo confiable?

1.3.3. PREGUNTAS DE INVESTIGACIÓN

1. ¿Cuál es la data de segmentación registrada actualmente por la entidad aseguradora referente a los asegurados que disponen de seguros de auto y vida, vigente en el periodo 2020-2023?
2. ¿Qué variables incluidas en los expedientes del asegurado de auto y vida tienen mayor correlación con la potencial venta cruzada?
3. ¿Qué algoritmo supervisado de clasificación será matemáticamente confiable para predecir al asegurado candidato para adquirir un seguro adicional a su póliza de auto?

1.4 OBJETIVOS DEL PROYECTO

1.4.1. OBJETIVO GENERAL

Diseñar e implementar un modelo predictivo enfocado en la detección de asegurados candidatos a adquirir un seguro de vida, como producto adicional a la póliza de autos previamente obtenida, que influya en catapultar la técnica de venta cruzada, logrando la fidelización de clientes y explotación máxima de utilidad de cartera, en MAPFRE Seguros Honduras.

1.4.2. OBJETIVOS ESPECÍFICOS

1. Realizar el análisis de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2020-2023.
2. Identificar un marco referencial de variables a considerar para el análisis de detección de asegurados candidatos a venta cruzada de póliza de autos y vida.
3. Entrenar un modelo numérico de predicción de candidatos a adquirir un seguro adicional a la póliza de autos para MAPFRE Seguros Honduras.

1.5 JUSTIFICACIÓN

La razón más obvia por la cual es importante emprender una campaña de venta cruzada en una aseguradora es captar mayor dinero en primas. Una inversión adecuada en unidades, herramientas y planes de venta cruzada permite a las entidades alcanzar sus metas de colocación de pólizas a corto, mediano y largo plazo. De forma adicional evita pérdida de tiempo y frustración a los canales de comercialización, producto de un rechazo por parte del cliente durante el tanteo de venta de un producto adicional. Retener y fidelizar al cliente de una entidad es igual de importante, asegura la confianza, la credibilidad y estrecha relación, la captación de un cliente nuevo puede impactar de forma directa en el presupuesto de gastos de la organización. Una estrategia adecuada de venta cruzada, influye también en los colaboradores de la unidad comercial de la organización, incentiva un comportamiento animado y optimista durante la venta, y sirve como medida de atracción a mediadores externos para comercializar productos de MAPFRE Seguros Honduras, siendo disuadidos por ser apoyados en las tareas de captación de primas.

Analizando más a fondo el impacto económico que representa los clientes con seguros de auto y vida, los datos otorgados por la MAPFRE Seguros Honduras indican que desde el año 2020 y lo transcurrido en el 2023, el monto de prima que representa para la entidad el segmento de cliente

de auto asciende a más de L 1,000,000,000.00, con un promedio de L 1,193.00 por asegurado. Por otra parte, el segmento de vida representa más de L 55,000,000.00, con un promedio de L 2,642.00 por asegurado en el mismo periodo de estudio anterior.

Tomando como parámetros los puntos de acceso a la información, costo, mandos intermedios y tiempo se determina que la investigación para determinar la factibilidad técnica de implementar un modelo de predicción de candidatos a adquirir un seguro adicional a la póliza de autos en la entidad MAPFRE SEGUROS HONDURAS es viable en todos sus aspectos, porque se cuenta con los siguientes elementos necesarios para poder hacer el proceso investigativo:

1. Se cuenta con el acceso a la información de segmentación referente a los asegurados de auto y vida correspondiente a los años 2020 y 2023.
2. El costo es accesible, no se requiere de una suma de dinero para recopilar la información y modelar los resultados.
3. El tiempo para realizar la investigación, es de ocho semanas tiempo suficiente para recopilar y procesar la información para entregar resultados en el tiempo definido.

CAPÍTULO II. MARCO TEÓRICO DE LA INVESTIGACIÓN

2.1 ANÁLISIS DE LA SITUACIÓN ACTUAL

La rentabilización y fidelización de clientes, son uno de los principales desafíos que enfrentan las compañías de todo el mundo en la actualidad, la definición de venta cruzada en sector seguros es en general, “Ventas efectuadas a clientes que han comprado otros productos de una compañía, y que por su perfil son potenciales consumidores de otros productos o servicios de la misma. Con esta técnica también se persigue la fidelización del cliente” («venta cruzada - Fundación MAPFRE», 2020).

Durante la primera década de este siglo, los ejecutivos comerciales concentraron sus esfuerzos en retener su base de clientes, debido a que se extendió la idea de que retener un cliente es más rentable que adquirir uno nuevo (Kotler, 2003). Esta concepción ha sido probada alrededor del mundo en múltiples sectores, entre los que es posible resaltar, entre otros, al bancario, al de telecomunicaciones y al de seguros. Se estima que concretar una nueva venta a clientes existentes cuenta con un 60 y 70% de probabilidad, mientras que concretar una nueva venta a un no-cliente se encuentra entre un 5 y 20%. Un componente central e importante de la venta cruzada para la industria aseguradora, es que el cliente asegurado tiende a conservar su póliza por varios años.

La construcción de relaciones con los clientes tiene un impacto directo tanto en los márgenes de rentabilidad como en la forma en que compiten las aseguradoras, y hay un valor significativo en juego. Un análisis de referencia reciente de más de 20 aseguradoras, que ofrecen cobertura de vida, salud y pensiones en Europa y Estados Unidos encontró que una compañía de seguros promedio tenía una densidad de pólizas por asegurado de 1.2 a 1.5. En las aseguradoras y sus agentes, con relaciones más profundas con los asegurados, la densidad general de productos aumentó a dos productos por cliente, o incluso tres o más en algunos casos, lo que indica que esas aseguradoras, han hecho que sea mucho más conveniente para sus asegurados cubrir todo su conjunto de necesidades en un solo sitio (Unlocking the potential of deeper customer-agent-insurer relationships, 2021).

La envergadura de las estrategias de venta cruzada y su impacto en la industria siempre es difícil de determinar, sin embargo, a nivel macro, Honduras cuenta con un alto potencial de venta cruzada si lo comparamos contra otros países de Centroamérica. Según el informe sectorial de seguros lanzado en abril 2021, Honduras se encuentra 5ta en el ranking de volumen de primas – con un valor total de \$476,421.38 miles en 2020 (Garita Araya, 2021).

Un enfoque comercial de venta cruzada comienza con cambios estratégicos y estructurales los cuales sientan la base para construir relaciones confiables y duraderas con los clientes, estos cambios combinados con la mejora en la efectividad de las ventas y el aprovechamiento del poder de las herramientas de análisis avanzado. Es necesario activar varias palancas simultáneamente para mejorar la densidad del producto y las relaciones con los clientes (Unlocking the potential of deeper customer-agent-insurer relationships, 2021).

Las herramientas de análisis avanzado se están convirtiendo cada vez más, en apuestas en todas las facetas de la distribución de seguros. Las aseguradoras que combinan con éxito estas herramientas con una fuerza de ventas bien capacitada a menudo están mejor posicionadas para construir relaciones profundas con los clientes y maximizar las oportunidades de venta cruzada de productos y servicios. Combinadas, estas dos tácticas son poderosas, pero es probable que no sean óptimas si trabajan de forma aislada o a ritmos distintos. La COVID-19 tendrá un impacto duradero en los seguros, en gran parte porque la pandemia ha acelerado los procesos digitales, sin embargo, en la línea de los estudios realizados por EY y Aite-Novarica, el informe de FRISS indica que las aseguradoras deben continuar abordando su deuda tecnológica mediante la digitalización de los procesos principales, la migración a la nube y la adopción de modelos de abastecimiento flexibles (FRISS, 2022).

El futuro de la detección de candidatos a venta cruzada radica en el uso de tecnologías avanzadas, las herramientas y capacidades de análisis avanzado pueden agregar una visión tremenda al proceso de venta cruzada al ayudar a los agentes de ventas a identificar los productos que mejor se adaptan a sus clientes existentes en función de las necesidades de los clientes, la cartera de productos y el historial de compras. Un modelo de puntuación de propensión al cliente puede ayudar a priorizar eficazmente a qué clientes acercarse con productos y servicios adicionales, lo que es especialmente útil para los agentes con grandes carteras de clientes. Armados con datos

sobre la probabilidad de compra de un cliente, los agentes pueden dirigir sus esfuerzos e inversiones de venta cruzada de maneras más específicas, lo que mejorara su productividad, al tiempo que brindara mejores ofertas personalizadas a los clientes (Unlocking the potential of deeper customer-agent-insurer relationships, 2021).

El código de Comercio de Honduras contempla dentro de sus artículos un ítem que invita y habilita a las entidades aseguradoras a realizar acciones de venta cruzada, en la búsqueda de dinamitar la cartera de pólizas de un asegurado. “Pueden asegurarse uno o varios intereses sobre una misma cosa o persona, y uno o varios intereses sobre conjuntos unitarios de cosas o personas” (Congreso Nacional de Honduras, 1950).

Los datos publicados por el superintendente referente al crecimiento de las primas en el país hondureño, vislumbran un escenario fértil para el sector asegurador, de acuerdo al reporte de primas netas del mercado asegurador hondureño en 2020 éstas ascendieron a L11.569.0 millones (477 millones de USD), esta evolución significó un crecimiento interanual del 2.9%, superior al mostrado el año anterior 2019 (0.8%), es una de las razones por la cual existe una clara urgencia a nivel de gremio asegurador en impulsar campañas y herramientas que exploten la colocación de pólizas, que les permita garantizar la rentabilidad e integridad del negocio (CNBS, 2021).

La contratación de seguros para proteger las personas y los vehículos ha cobrado auge en los últimos años en el sistema financiero de Honduras, un análisis de inclusión financiera de la CNBS indica que a diciembre de 2019 el porcentaje de la población que se encuentra asegurada en los diferentes ramos es del 4.8%. Con relación a la población adulta mayor de 18, se estima que 78 personas de cada 1000 habitantes en el país poseen una póliza de seguros, ya sea de personas o general. El número de pólizas globales en el sistema asegurador de Honduras creció de 377,372 a 441,326 en el bienio 2018-2019 (Rodríguez, 2020).

Si bien el sector asegurador en el país mantienen una oferta de 38 tipos de seguros, se mantiene una alta concentración en pocos seguros; pues en los seguros de vida que representan el 35% del total de primas netas del sistema asegurador, el 85% se concentra en seguros de vida colectivo; respecto a los seguros generales (34%), el 60% se concentra sobre incendios y líneas aliadas; accidentes y enfermedades (22%), el 86% de las primas se generan en el tipo de salud y hospitalización; finalmente, en otros seguros generales (8%), el 60% se concentra sobre coberturas

de equipo electrónico, transporte terrestre, responsabilidad civil y robo y hurto; tal situación no ha experimentado un cambio importante en los últimos cinco años (Rodríguez, 2020).

En el país operan 13 compañías aseguradoras, así están aglutinadas en la Camara Hondureña de Aseguradoras (CAHDA) y que funcionan desde el 25 de junio de 1974, se enlistan a continuación:

1. Seguros Atlántida, S.A.
2. Seguros Bolívar Honduras, S.A. (Seguros Davivienda)
3. Pan American Life Insurance Company (PALIC)
4. Interamericana de Seguros, S.A. (FICOHSA | Seguros)
5. MAPFRE | Seguros
6. Seguros CREFISA, S.A.
7. Seguros del País, S.A.
8. Seguros Equidad, S.A.
9. ASSA Compañía de Seguros Honduras S.A.
10. Seguros Lafise Honduras, S.A.
11. Seguros Banrural Honduras, Sociedad Anónima
12. Seguros Continental, S.A.
13. Seguros Cuscatlán

Para diferentes compañías del sector asegurador del país sin importar su tamaño, han tenido que reinventar las estrategias para hacer frente a la débil cultura de seguro, la aplicación de técnicas estadísticas y econométricas dirigidas a cuantificar la probabilidad de que un cliente acepte un seguro basado en su histórico de datos está ganando terreno en los últimos años.

En el apartado siguiente se realiza un breve análisis interno de MAPFRE Seguros Honduras, organización objetivo de esta investigación, en el cual se describe la historia, visión, propósito y valores con el propósito de entender la forma en la que opera la empresa.

Breve reseña de MAPFRE

MAPFRE nace en 1933 como la Mutuality de Seguros de la Agrupación de Fincas Rústicas de España con el fin de asegurar a los trabajadores de las explotaciones agrícolas. A partir de 1955, se sientan las bases de la entidad tal como la conocemos en la actualidad, extendiéndose su actividad, de forma inicial, a otros ramos aseguradores como Vida, Accidentes o Transportes. Durante la década de los 70, se inició en América Latina la estrategia internacional con las actividades de Asistencia y Reaseguro como puntas de lanza del negocio de Seguros. En 1975 inicia sus actividades FUNDACIÓN MAPFRE y en los años 80, MAPFRE –que ya era la primera entidad aseguradora de España- consolidó su apuesta latinoamericana con un importante esfuerzo inversor, que culminaría en los 90 con la creación de una red de seguro que atendía a las particularidades locales y era similar al modelo de éxito español.

A partir del año 2000, se comenzó a percibir la aceleración del crecimiento de las compañías de MAPFRE en América Latina, un efecto que se combinó con la diversificación de los canales de comercialización en España. En 2007, se llevó a cabo una reorganización que proporcionó una estructura corporativa y una capacidad financiera que permite seguir ampliando las actividades y la expansión internacional. FUNDACIÓN MAPFRE se convirtió en el accionista mayoritario y en el garante de la independencia de la nueva sociedad MAPFRE S.A., entidad que integra todas las actividades del Grupo. (Historia de MAPFRE, s. f.)

A partir del mes de mayo de 2011, MAPFRE Seguros Honduras abrió sus puertas a la población hondureña con un portafolio de productos que abarca todos los ámbitos de la actividad aseguradora.

Visión

MAPFRE, tu aseguradora de confianza.

Propósito

Estamos a tu lado, acompañándote para que avances con tranquilidad, contribuyendo al desarrollo de una sociedad más sostenible y solidaria. Cuidamos lo que te importa.

Valores

- Solvencia
- Innovación
- Servicio
- Integridad
- Equipo multicultural y diversos

(Misión de MAPFRE Honduras - Seguros MAPFRE, 2022)

2.2 CONCEPTUALIZACIÓN

A continuación, se describen algunos conceptos básicos que es necesario conocer para comprender de una mejor manera la investigación.

2.2.1. RIESGO

Podemos decir que con ella se describe, desde el sentido común, la posibilidad de perder algo (o alguien) o de tener un resultado no deseado, negativo o peligroso. El riesgo de una actividad puede tener dos componentes: la posibilidad o probabilidad de que un resultado

negativo ocurra y el tamaño de ese resultado. Por lo tanto, mientras mayor sea la probabilidad y la pérdida potencial, mayor será el riesgo. (Tocabens, 2011)

2.2.2. PRIMA

Es el precio del seguro, la contraprestación que el tomador debe pagar a la aseguradora por el riesgo que ésta asume. Se refiere a la suma resultante de aplicar la tarifa calculada sobre las bases de la probabilidad de ocurrencia del siniestro y de su cuantía posible. (Correa Rodriguez, s. f.)

2.2.3. ASEGURADO

“Es la persona, titular del objeto, materia o aleas sobre cuyo riesgo recae el contrato de seguro. En otras palabras, es la persona sobre quién recae la cobertura del seguro” (Correa Rodriguez, s. f.).

2.2.4. VENTA DEL SEGURO

La venta, considerada en términos generales, consiste en la colocación de determinado producto o servicio entre el público consumidor; en la actividad aseguradora, se identifica con la obtención de nuevas pólizas y aumento de las garantías de las ya existentes. («venta del seguro - Fundación MAPFRE», 2020)

2.2.5. VENTA CRUZADA

“La venta cruzada, consiste en sugerir a los clientes, durante el proceso de compra de un producto ya seleccionado, otro u otros productos adicionales que puedan resultarles de interés por estar relacionados entre sí” (Soledad, 2014).

2.2.6. TECNOLOGÍA

Ellul, (1960, c.p. Carlos, 2002), define "la technique, como la totalidad de Los métodos a los que se ha llegado racionalmente y que tienen una eficacia absoluta (para una fase de desarrollo dada) en todos los campos de la actividad humana".

2.2.7. APRENDIZAJE AUTOMATICO

El aprendizaje automático es una rama de la inteligencia artificial que busca que un programa de computador aprenda de un conjunto de datos con los cuales se entrena, y buscare identificar un patrón con el que pueda realizar predicciones sobre nuevos datos. (PINEDA PERTUZ, 2021)

2.2.8. SEGURO DE AUTOMOVIL

El seguro del automóvil tiene por objeto, fundamentalmente, reparar o indemnizar los daños accidentales producidos en los vehículos de terceros y en el vehículo asegurado. La cobertura principal consiste en asegurar los daños causados a terceros en el uso del vehículo (responsabilidad civil del automóvil). En España, todo propietario de vehículos a motor está obligado a suscribir un seguro por cada vehículo del que sea titular. (Fundación MAPFRE, 2020)

2.2.9. SEGURO DE VIDA

El seguro de vida es un seguro que paga dinero ya sea por la muerte del asegurado o después de un período de tiempo establecido. Este dinero se paga a un beneficiario. Los beneficiarios son las personas que usted elige para que reciban el dinero de la póliza de

seguro de vida. (Seguro de Vida- Conceptos Básicos - Office of Public Insurance Counsel, 2022)

2.2.10. SEGMENTACIÓN DE MERCADO

El proceso de subdividir un mercado en subconjuntos distintos de clientes que se comportan de la misma manera o que presentan necesidades similares. Cada subconjunto se puede concebir como un objetivo que se alcanzará con una estrategia distinta de comercialización. (THOMPSON, 2007)

2.3 TEORÍAS DE SUSTENTO

2.3.1. BASES TEORICAS

Distintas metodologías de Aprendizaje Automático han sido aplicadas a lo largo de los años para atender distintos escenarios, la presente investigación busca ser capaz de predecir los asegurados de auto candidatos a venta cruzada de un seguro de vida mediante la clasificación automática de los asegurados. Se analizan distintos algoritmos de Aprendizaje Automático, se consideran únicamente modelos de clasificación de variables dicotómicas binarias (Si ha realizado venta cruzada vs no ha realizado venta cruzada). En los párrafos siguientes analizaremos los algoritmos de aprendizaje automático más utilizados en la actualidad. En primer lugar, examinaremos los algoritmos de aprendizaje automático supervisado para luego dar paso a los algoritmos de aprendizaje no supervisado.

Hastie et al. (2001) de cuatro categorías principales de aprendizaje automático:

2.3.1.1. APRENDIZAJE AUTOMÁTICO SUPERVISADO

En este tipo de aprendizaje todas las muestras de conjunto de datos son etiquetadas con un resultado (valor predefinido) o una categoría utilizando una variable llamada destino u objetivo

(Pertuz, 2022). Estos datos etiquetados ayudan a que el modelo internamente construya un patrón y sea capaz de predecir la variable destino de una muestra nueva. Buscando asegurar la calibración del modelo y evitar un sobreajuste, se realiza la separación del conjunto de datos en una base de entrenamiento y una base de datos de validación, finalmente se contrasta y mide el grado de precisión.

A su vez, se distinguen dos tipos de aprendizaje supervisado:

Aprendizaje Automático Supervisado de regresión: La regresión permite predecir un valor continuo, tomando en consideración varias variables de entrada, en otras palabras busca encontrar la relación entre una variable dependiente con algunas variables independientes. Algunos problemas por ejemplo que se pueden abordar con la técnica de regresión podrían ser: predecir el precio de una casa a partir del área en metros cuadrados, predecir la temperatura del próximo mes, predecir el tiempo de vida de un dispositivo electrónico. Según Hastie et al. (2001), los tipos más comunes de modelos de regresión son:

Regresión lineal: La regresión permite predecir un valor continuo, tomando en consideración varias variables de entrada. O en otras palabras busca encontrar la relación entre una variable dependiente con algunas variables independientes. Entre los tipos de regresiones lineales podemos encontrar regresiones lineales múltiples y la regresión polinómica (Pertuz, 2022).



Figura 1. Regresión lineal.

Fuente: (Hastie et al., 2009)

Árbol de decisión: Un árbol de decisión es un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Constituyen probablemente el modelo de clasificación más utilizado y popular. El conocimiento obtenido durante el proceso de aprendizaje inductivo se representa mediante un árbol. Un árbol gráficamente se representa por un conjunto de nodos, hojas y ramas. El nodo principal o raíz es el atributo a partir del cual se inicia el proceso de clasificación; los nodos internos corresponden a cada una de las preguntas acerca del atributo en particular del problema. Cada posible respuesta a los cuestionamientos se representa mediante un nodo hijo. Las ramas que salen de cada uno de estos nodos se encuentran etiquetadas con los posibles valores del atributo. Los nodos finales o nodos hoja corresponden a una decisión, la cual coincide con una de las variables clase del problema a resolver (MARTÍNEZ, 2009).

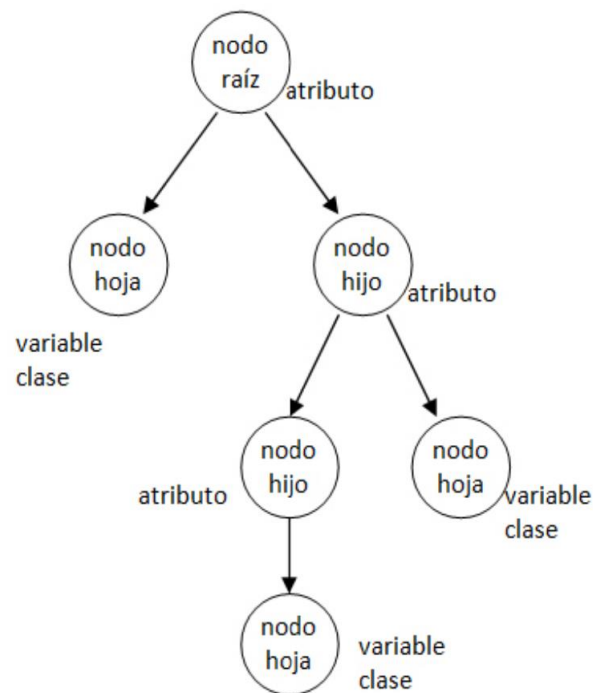


Figura 2. Árbol de decisión.

Fuente: (MARTÍNEZ, 2009)

Bosque aleatorio: El algoritmo Random Forest (Breiman, 2001) es una técnica de aprendizaje supervisado que genera múltiples árboles de decisión sobre un conjunto de datos de entrenamiento:

conjuntos de datos; suelen ser utilizadas como herramientas para la predicción de tendencias y como clasificadoras de conjuntos de datos. Las redes neuronales tienen tres capas: entrada, salida y oculta. Las capas de entrada y salida son visibles y como su nombre dice son los datos de entrada y de salida del modelo (Ocaris, s. f.).

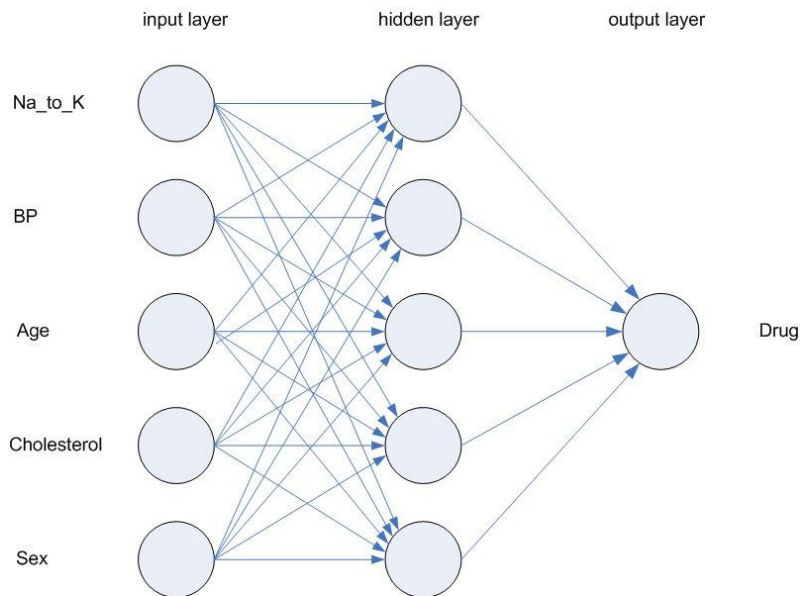


Figura 4. Esquema de red neuronal.

Fuente: (Ocaris, s. f.)

Aprendizaje Automático Supervisado de clasificación: Tipo de algoritmo de aprendizaje en los que el valor a predecir es de tipo discreto, lo que quiere decir que se asocia una instancia a una categoría o clase particular a la cual pertenece (Pertuz, 2022). A continuación, se presentan los tipos más comunes de modelos de clasificación:

Regresión logística: La regresión logística consiste en obtener una función logística de las variables independientes que permita clasificar a los individuos en una de las dos subpoblaciones o grupos establecidos por los dos valores de la variable dependiente. La función logística es aquella que halla, para cada individuo según los valores de una serie de variables (X_i), la probabilidad (p) de que presente el efecto estudiado. Una transformación logarítmica de dicha ecuación, a la que se le llama logit, consiste en convertir la probabilidad (p) en odds. De aquí surge la ecuación de la regresión logística, que es parecida a la ecuación de la regresión lineal múltiple (Fiuza Pérez, 2000).

Para poder interpretar el resultado del modelo de regresión logística debemos recurrir al concepto de ‘odds’, una de las medidas de las que se dispone para poder cuantificar el riesgo. De esta forma el o la ‘odds’ se define como el cociente de la probabilidad de presentar una característica y la probabilidad de no presentarla, o lo que es lo mismo el cociente del número de casos que presentan la característica entre el número de casos que no la presentan. $Odds = p / (1 - p)$. Cuando realizamos una regresión logística lo que pretendemos es estimar los parámetros de la ecuación ($\beta_0, \beta_1, \beta_2, \dots, \beta_k$) de la función que pretendemos evaluar: $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ (Moral Peláez, 2006).

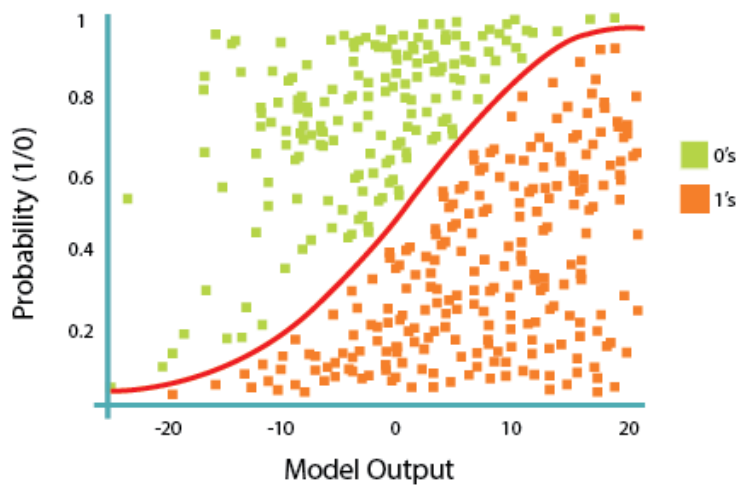


Figura 5. Regresión logística.

Fuente: (Hojas, 2021)

Máquina de vectores de soporte: las máquinas de vector soporte (Support Vector Machines, SVMs) son un conjunto de algoritmos de aprendizaje estadístico supervisado pertenecientes a la familia de los clasificadores lineales desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T en torno a 1995. Sin pérdida de generalidad, suponiendo que tenemos ejemplos de sólo dos categorías una SVM construye un hiperplano en un espacio de dimensionalidad muy alta o incluso infinita. Este hiperplano separa de forma óptima los puntos de una clase de la de otra. En el concepto de “separación óptima” es donde reside la característica fundamental de las SVM, se busca el hiperplano que tenga la máxima distancia (margen) con los puntos que estén más cerca de él mismo. Por eso también a veces se les conoce a las SVM como clasificadores de margen máximo [WSVM] (Campo Leon, 2017).

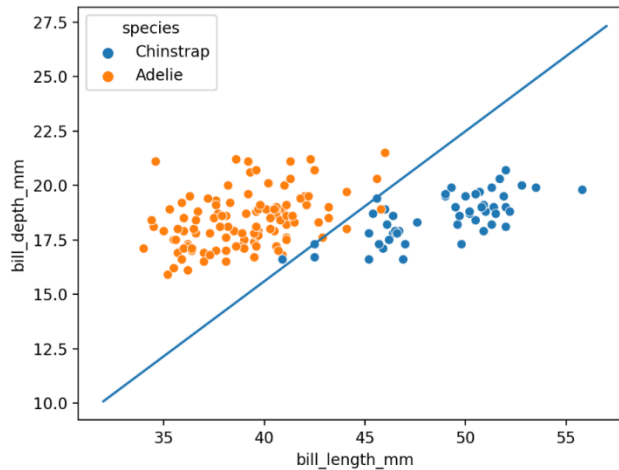


Figura 6. Máquina de Vectores de Soporte.

Fuente: (Campo Leon, 2017)

Árboles de decisión, Bosque aleatorio y redes neuronales: Estos modelos siguen la misma lógica definida en los párrafos anteriores, su diferencia está relacionada a es que su salida es discreta en lugar de continua.

2.3.1.2. APRENDIZAJE AUTOMÁTICO NO SUPERVISADO

En esta forma de aprendizaje no contamos con etiquetas para las muestras de entrenamiento del modelo. Es decir, el modelo actúa directamente sobre los datos de entrada y busca hallar relaciones entre ellos basándose en características común. Los dos problemas que el aprendizaje automático no supervisado tarta de resolver principalmente son el agrupamiento o clustering y la reducción de la dimensionalidad (Pertuz, 2022).

Dichos métodos se destacan a continuación:

Clustering: El proceso de clustering consiste en la división de los datos en grupos de objetos similares. Para medir la similitud entre objetos se suelen utilizar diferentes formas de distancia: distancia euclídea, de Manhattan, de Mahalanobis, etc. El representar los datos por una serie de clústeres, conlleva la pérdida de detalles, pero consigue la simplificación de los mismos. Las

técnicas de clustering más comunes son el uso de k-medias, el clustering jerárquico, el clustering de desplazamiento de medias y el clustering basado en la densidad (Garre, 2007).

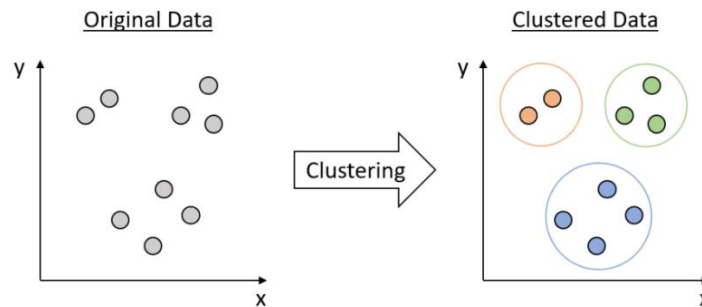


Figura 7. Clustering.

Fuente: (Garre, 2007)

Reducción de la dimensionalidad: Los métodos de reducción de dimensionalidad son algoritmos que mapean el conjunto de los datos a subespacios derivados del espacio original, de menor dimensión, que permiten hacer una descripción de los datos a un menor costo. Dentro de los métodos más populares de reducción de la dimensionalidad podemos encontrar descomposición en valores singulares (SVD) y, en especial el análisis de componentes principales (PCA).

El PCA o Principal Component Analysis, es un método espectral de análisis multivariado, basado en correlaciones entre variables, mediante un proceso estocástico que satisface la reducción de dimensiones, realizando una transformación lineal de las variables iniciales, para proyectar vectores propios ortonormales denominados componentes principales (Anaya Isaza, 2017).

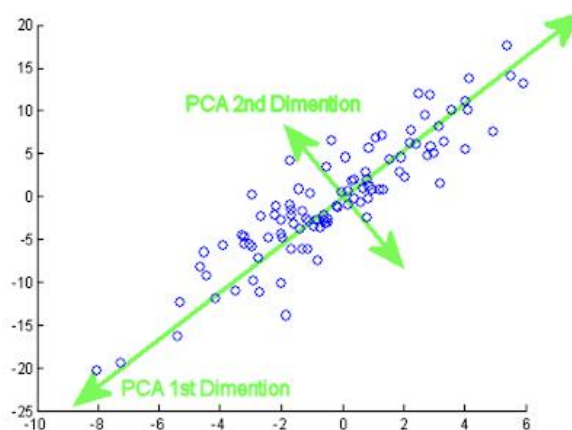


Figura 8. PCA.

Fuente: (Anaya Isaza, 2017)

2.3.1.3. APRENDIZAJE AUTOMÁTICO SEMISUPERVISADO

Utiliza datos de entrenamiento tanto etiquetados como no etiquetados, Si juntamos datos etiquetados con datos no etiquetados en ciertas situaciones se consiguen modelos más exactos (Pertuz, 2022).

2.3.1.4. APRENDIZAJE AUTOMÁTICO REFORZADO

En este tipo de modelos el aprendizaje se consigue a partir de ensayo y error, teniendo en cuenta que cada vez que se da acierto en una predicción se establece una recompensa y en caso contrario una penalización. Esto es en esencia, la aplicación de una retroalimentación (feedback) en el proceso de aprendizaje. Lo podemos comparar a la forma como, por ejemplo, aprende un niño en sus primeros años de vida, en cada vez que acierta en la identificación de un objeto de su entorno recibe una recompensa de parte de sus padres, mientras que por el contrario cuando falla es corregido por sus progenitores para que en el futuro no vuelva a equivocarse. Estos algoritmos de aprendizaje llamados agentes, en definitiva, buscan aprender estrategias para minimizar estas penalizaciones y maximizar las recompensas, con el objetivo de encontrar el mejor ajuste que potencie su capacidad predictiva (Pertuz, 2022).

2.3.2. METODOLOGÍAS DESARROLLADAS POR OTROS INVESTIGADORES O EXPERTOS.

2.3.2.1. MODELO PREDICTIVO DE VENTA CRUZADA EN PRODUCTOS DE VIDA Y SALUD:

Sanchez Sardaña (2022), presenta un estudio en el cual busca predecir los candidatos a venta cruzada en seguros de vida y salud de una Asegurado radicada en el país del Salvador, se han contrastado el algoritmo random forest y el XGBoost en su investigación. Dentro del análisis ha

sido utilizado la matriz de confusión como herramienta para presentar los resultados de las predicciones, según se detalla a continuación:

Bosque Aleatorio		XGBoost	
Exactitud	92,14 %	Exactitud	94,63 %
Precisión	87 %	Precisión	91,43 %
Sensibilidad	72,65 %	Sensibilidad	81,61 %
F-Score	79,18 %	F-Score	86,24 %

Figura 9. Comparación resultados Matriz de confusión.

Fuente: (Sanchez Sardaña, 2022)

Se hace evidente en la figura 9 que el mejor modelo para predecir el escenario de venta cruzada de salud y vida es el XGBoost ya que su porcentaje de acierto es ligeramente superior al bosque aleatorio, y en paralelo las métricas de exactitud, precisión, sensibilidad y F-score son mejores.

2.3.2.2. DESARROLLO DE UN MODELO DE RECOMENDACIÓN DE COMPRA PARA CLIENTES DE UNA EMPRESA DE SEGUROS

En otro ejercicio de modelado, Alberto (2013), se centra en desarrollar una metodología que, tomando como base data e información disponible de los clientes de la línea vida de la compañía Consorcio Nacional de Seguros Chile, dinamite la efectividad de las campañas de venta cruzada de los seguros de automóvil, a través de los modelos predictivos de árbol de decisión, Naive Bayes y Support Vector Machines:

	CRT	Naive Bayes	SVM
TPR	89.53%	91.07%	90.11%
FPR	17.70%	32.37%	25.97%

Figura 10. Comparación predicción Compra.

Fuente: (Alberto, 2013)

	CRT	Naive Bayes	SVM
Correctamente clasificadas	71.00%	62.32%	77.28%
Incorrectamente clasificadas	29.00%	37.68%	22.72%

Figura 11. Comparación Clasificación Total.

Fuente: (Alberto, 2013)

El algoritmo utilizado en el modelo con mejor resultado ha sido el árbol de decisión, evidenciando una clara superioridad en el porcentaje de clasificación total, respecto al modelo Naive Bayes. En otra arista es un 6% inferior a SVM. Para la clase a predecir “compra”, el True Positive Rate (TPR) es un 1% inferior al resto de los modelos utilizados, finalmente también se logra un False Positive Rate (FPR) 8% menor que el algoritmo más cercano (SVM).

2.3.3. INSTRUMENTOS UTILIZADOS

Todos los procedimientos de aprendizaje automático deben realizarse a partir de una base de datos, los descritos en la presente investigación consiste en datos reales pertenecientes a entidades aseguradoras que operan en diferentes mercados en el mundo con presencia en el ramo de auto y vida dentro de su portafolio comercial. La base de datos recopilada comprende información relativa a las particularidades inherentes de los productos en vigor, de los distintos canales de distribución, características propias de los asegurados, variables sociodemográficas e información referente a siniestros acaecidos en el pasado (Valero et al., 2019).

Las técnicas de transformación de los datos previo al modelado, coinciden en las investigaciones analizadas, se resumen a continuación:

Selección de variables: una adecuada selección de variables permitiría aminorar la dificultad del modelo, facilitando su correcta comprensión y acrecentando la precisión de los algoritmos si se distingue un conjunto apropiado.

División del conjunto de datos: Elegido el conjunto de variables que constituye parte del análisis, se realiza un proceso de adquisición de información de los datos mediante la división de los datos en dos subconjuntos: El primero de entrenamiento y el segundo de constatación o test (Valero et al., 2019).

Realizado el modelado de datos se vuelve inherente a la metodología de aprendizaje automático la medición de la precisión de los resultados, la matriz de confusión es apropiada en el aprendizaje automático para cotejar los valores reales con los pronósticos en un modelo basal con datos entrenados. De este modo, se pueden pergeñar cuatro clases diferenciadas entre sí: Verdaderos negativos, verdaderos positivos, falsos negativos y falsos positivos (Valero et al., 2019).

2.4 MARCO LEGAL

Jurídicamente hablando debe entenderse el concepto de “venta” según la ley de impuesto sobre venta de Honduras:

Todo acto que importe transferencia a título oneroso de una mercadería del dominio de una persona natural o jurídica al dominio de otra, o que tenga por fin último la transmisión de dicho dominio, independientemente de la designación que las partes den al contrato de origen o a la negociación en que se incluya o involucre y de la forma del pago del precio, sea éste en dinero o en especie. (Gobierno de Honduras, 1994)

El código de Comercio de Honduras contempla dentro de sus artículos un ítem que invita y habilita a las entidades aseguradoras a realizar venta cruzada, en búsqueda de dinamitar la cartera de pólizas de un asegurado. “Pueden asegurarse uno o varios intereses sobre una misma cosa o persona, y uno o varios intereses sobre conjuntos unitarios de cosas o personas” (Congreso Nacional de Honduras, 1950).

Todo acto de venta de seguro por alguna de las instituciones del país es regulado y supervisado por la CNBS, citado en el artículo 5 de la Ley de Instituciones de Seguro y Reaseguro:

“Las instituciones de seguro, afianzadores, reaseguradoras y reafianzadores y demás personas a que se refiere el Artículo 2 se regirán por la presente Ley” (República de Honduras, 2001).

CAPÍTULO III. METODOLOGÍA

En el presente capítulo se plantearán los métodos, procedimientos y técnicas que permitirán alcanzar los objetivos planteados en el estudio, se definirán las variables y también se planteará las hipótesis del problema.

3.1 CONGRUENCIA METODOLÓGICA

Se refiere de manera general, al problema, representado por medio de la matriz metodológica, seguidamente con la función de las variables, y las hipótesis posibles que requiera la investigación. La congruencia metodológica permite conocer el grado de coherencia que tienen los diferentes elementos definidos dentro de la investigación.

3.1.1. MATRIZ METODOLÓGICA

Consiste en la descripción de una secuencia lógica, en donde se presentan varios cuadros formados por filas y columnas, que permiten al investigador evaluar el grado de conexión lógica y coherencia entre el título, el problema, los objetivos, la hipótesis, las variables, el tipo, método, diseño e instrumentos de investigación.

“En consecuencia, la matriz facilita tener una visión general de estudio, puesto que permite al investigador ubicar las actividades que se plantean como necesarias para dar cumplimiento a los resultados” (Galindo, 2016).

TABLA 1 - MATRIZ METODOLÓGICA

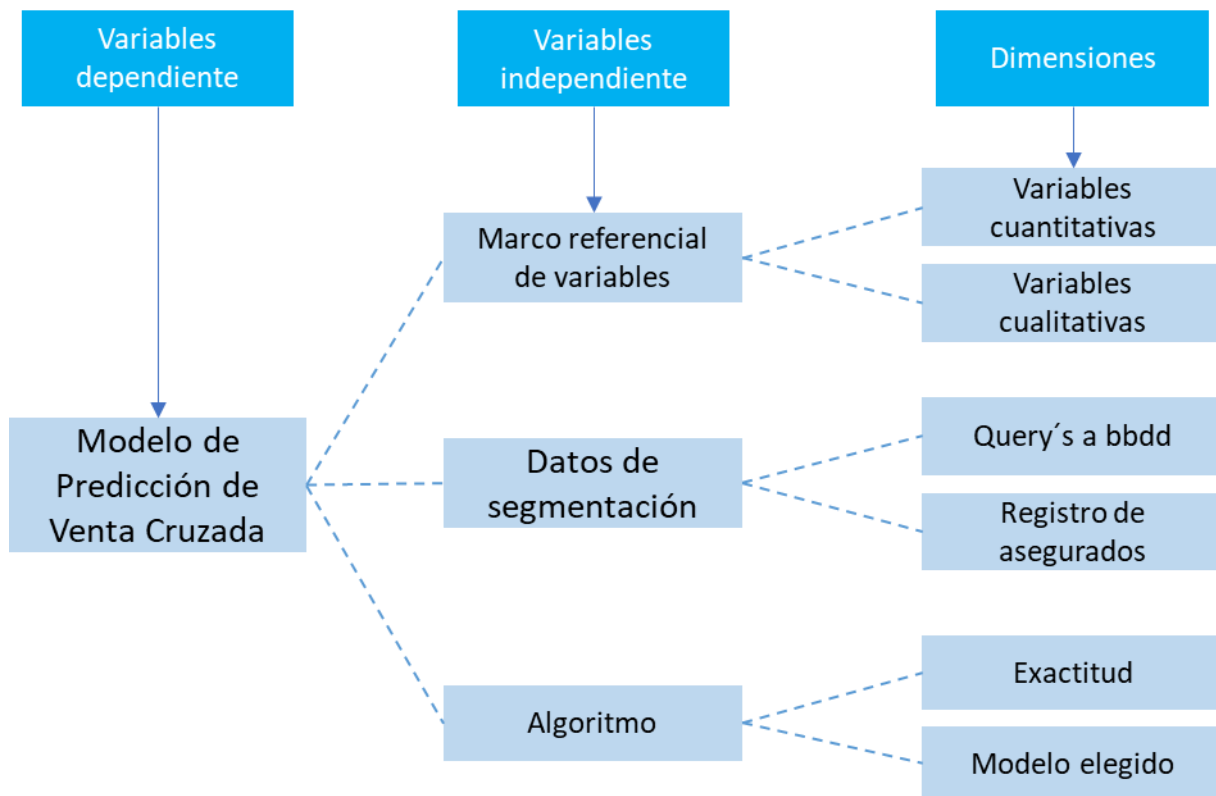
Formulación del Problema	Objetivo General	Objetivos específicos	Variable Independiente	Variables Dependientes
Determinar la factibilidad de implantación de un modelo de predicción de candidatos a venta cruzada de seguros de auto y vida, por lo cual se plantea la interrogante: ¿Cómo influye el histórico de datos de asegurados, en la obtención de un modelo predictivo confiable?	Diseñar e implementar un modelo predictivo enfocado en la detección de asegurados candidatos a adquirir un seguro de vida, como producto adicional a la póliza de autos previamente obtenida, que influya en catapultar la técnica de venta cruzada, logrando la fidelización de clientes y explotación máxima de utilidad de cartera, en MAPFRE Seguros Honduras.	Identificar un marco referencial de variables a considerar para el análisis de detección de asegurados candidatos a venta cruzada de póliza de autos y vida.	Marco referencial de variables	Modelo de Predicción de Candidatos a Venta Cruzada
		Realizar el análisis de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2022-2023.	Datos de segmentación de asegurados	
		Entrenar un modelo numérico de predicción de candidatos a adquirir un seguro adicional a la póliza de autos para MAPFRE Seguros Honduras.	Algoritmo con mayor exactitud	

Fuente: Elaboración propia, 2023

3.1.2. ESQUEMA DE VARIABLES DE ESTUDIO

En esta sección se presentará el esquema de las variables el cual está estrechamente vinculada al tipo de técnica o metodología empleada para la recolección y análisis de los datos de segmentación de asegurados en MAPFRE Seguros Honduras. Se mostrará la dimensión de cada variable con el fin de poder definir un instrumento con el cual medir las variables.

TABLA 2 - ESQUEMA DE LAS VARIABLES



Fuente: Elaboración propia, 2023

3.1.3. OPERACIONALIZACIÓN DE LAS VARIABLES

TABLA 3 - ESQUEMA DE LAS VARIABLES

Variable independiente	Definición		Dimensiones	Indicador
	Conceptual	Operacional		
Marco referencial de variables	El marco de referencia o marco referencial es un texto que identifica y expone los antecedentes, las teorías, las regulaciones y/o los lineamientos de un proyecto de investigación, de un programa de acción o de un proceso.	Identificar el tipo de variable, complejidad que enriquece los registros de asegurados de autos dentro de la entidad aseguradora.	VARIABLES CUANTITATIVAS	Cantidad de columnas y tipo de variable almacenada en sistema (core) de seguro.
			VARIABLES CUALITATIVAS	
Datos de segmentación de asegurados	Los datos de segmentación son características/variables de un público de una marca o negocio que permite dividirlo en grupos más pequeños, que se identifican por compartir ciertas características entre sí.	Recopilar los registros históricos de segmentación de asegurados que sirvan de entrenamiento para el modelo numérico.	Query's a bdd de core de seguros de MAPFRE Honduras.	Cantidad de registros de asegurados con componente de segmentación dentro de sistema (core) de seguro.
			Registro de asegurados de MAPFRE Honduras.	
			Promedio de pólizas por cliente	
Algoritmo con mayor exactitud	Los modelos predictivos son un grupo de técnicas que, mediante los campos del aprendizaje automático, la recolección de datos históricos, el Big Data y el reconocimiento de patrones, pretende dar una predicción de resultados futuros; con el objetivo de precisar la toma de decisiones mediante técnicas de análisis de datos.	Determinar la técnica correcta que de forma medible ofrezca confiabilidad al estudio.	Exactitud	Matriz de confusión
			Algoritmo elegido	

Fuente: Elaboración propia, 2023

3.1.4. HIPÓTESIS

Tomando en cuenta el alcance del trabajo de tesis, descriptivo, no se consideran hipótesis para esta fase de estudio.

3.2 ENFOQUE Y MÉTODOS

El enfoque usado en la investigación es cuantitativo, a causa de que en los objetivos fue definido un enfoque de investigación específico, el cual se concluye a través de análisis estadísticos y uso de modelos matemáticos. Se realizará un análisis de datos y variables, las cuales han sido el resultado de consultas (queries) a base de datos del sistema de MAPFRE Seguros Honduras, específicamente registros de asegurados de autos y vida, el análisis será realizado a través de herramientas de minería de datos y hojas de cálculo, incluyendo en el mismo una proyección/cuantificación del monto de prima que se puede colocar con la estrategia de venta.

TABLA 4 – METODOLOGÍA DE INVESTIGACIÓN IMPLEMENTADA



Fuente: Elaboración propia, 2023

3.3 DISEÑO DE LA INVESTIGACIÓN

La presente investigación habrá de seguir un tipo de investigación no experimental, bajo la premisa que las variables no serán manipuladas, los datos y variables de estudio serán obtenidos de forma natural, procesados y medidos a través de diferentes modelos estadísticos, modelos predictivos aplicados a los datos.

El diseño será transversal, porque los datos se analizarán en un periodo específico de tiempo, según se estipula en los objetivos específicos de la investigación, así mismo se trata de concluir o predecir la relación entre las variables, interpretando el resultando sin la existencia de una manipulación tendencial previa de las variables.

Finalmente, el alcance será descriptivo.

3.2.1. POBLACIÓN

La población considerada para llevar a cabo la investigación será la cartera de (14,000) asegurados que componen el negocio de auto y vida en la entidad aseguradora al cierre del año dos mil veinte y dos (2022).

3.2.2. MUESTRA

La muestra de la investigación se realizará de forma dirigida, utilizando una muestra de tres mil ochocientos veintiún (3,821) registros, que corresponden a la totalidad de asegurados, con registro completo de todas las variables de cliente y que disponen de pólizas de autos y/o vida, con vigencia en el periodo 2020-2023 en MAPFRE Seguros Honduras.

3.2.3. TÉCNICAS DE MUESTREO

La técnica de muestreo a utilizar será intencional y el tipo de muestra a utilizar en la investigación es no probabilístico, considerando que la elección de los elementos no depende de la probabilidad sino a causa de que cada muestra tiene la característica de ser un cliente que dispone de pólizas de autos y vida o ambas, así mismo la técnica de muestreo no se basa en fórmulas de probabilidad.

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

Este estudio se centra en un enfoque cuantitativo, utilizando las siguientes técnicas:

1. Revisión sistemática de Literatura.
2. Consultas/Querys a bbdd de core de la entidad.
3. Transformación de Datos.
4. Análisis numérico y Estadístico, específicamente: XGBoost y Bosques Aleatorios.
5. Algoritmos de Aprendizaje Automático Supervisado

Las técnicas antes descritas serán aplicadas de manera virtual a través de MS Office y el software de minería de datos Knime.

En la investigación se utilizarán distintos instrumentos para analizar, validar y comprobar la hipótesis, así mismo darán una radiografía fundamental para la toma de decisión de factibilidad de implantación del modelo numérico dentro de MAPFRE Seguros Honduras, se contemplaron los instrumentos siguientes:

1. Hojas de cálculo Excel.
2. Diccionario de variables.
3. Knime workflow: TESIS_CROSS_SELLING

Se dispone de la información resultante de las consultas (querys) a bases de datos de MAPFRE Seguros Honduras. Dentro del workflow denominado TESIS_CROSS_SELLING se realizarán la preparación de las mismas, las cuales están originalmente en formato archivo CSV compuesto por determinadas variables, entre las cuales destacan características inherentes a los productos en vigor, canales de distribución y variables sociodemográficas características de los asegurados, el nombre identificador de cada base se detalla a continuación:

1. BBDD_Segmentación_Demografica
2. BBDD_Siniestros_Autos
3. BBDD_Prima
4. BBDD_Suma asegurada

5. BBDD_Coberturas Adicionales
6. BBDD_Nivel de ingresos
7. BBDD_Venta Cruzada

Este estudio de investigación seguirá un procedimiento secuencial para la validación de las hipótesis planteadas, contribuyendo a responder las preguntas de investigación. El primero de los pasos es la aplicación de una depuración general de los datos, eliminando registros que no generan valor a la investigación, como ser datos ausentes, duplicados, erróneos así como valores en cero, de forma paralela se genera una modificación de los nombres de las variables, transformación de formatos de fechas (formato dd/mm/yyyy), para finalmente diseñar el diccionario de variables el cual se percibe como una fuente de definición y caracterización de las variables a incluir en el modelo, con un propósito de asegurar el tratamiento idóneo del modelo y datos en workflow, así como reducir el riesgo de resultados tendenciosos o erróneos.

El siguiente paso es el establecimiento de la variable dependiente dentro de los datos, para lo cual es necesaria una correcta clasificación a través de una misma variable a los clientes que tienen venta cruzada (póliza de autos y vida) y a los que no (póliza de autos). Dentro de la clasificación se hace evidente el desbalanceo de los datos, ya que del total de asegurados con venta cruzada corresponde únicamente el 28.57% del total de asegurados, en esta investigación se realiza un muestreo de los datos a través de la adición y eliminación de registros.

Una vez balanceada la data se prosigue con el proceso de selección de variables, a través de una adecuada selección de variables nos permite aminorar la dificultad del modelo y facilitar su comprensión, es usual que la utilización de determinadas variables sea mejor de cara a los resultados del modelo que la selección de todas las variables, para la selección de las variables se hará uso del coeficiente de correlación de Pearson pensado para variables cuantitativas y tomando como referencia el diccionario de variables previamente diseñado.

Acorde a Lalinde et al. (2017) se define como coeficiente de correlación:

“El coeficiente de correlación de Pearson, pensado para variables cuantitativas (escala mínima de intervalo), es un índice que mide el grado de covariación entre distintas variables relacionadas linealmente”.

En la etapa final de la transformación de los datos nos encontramos con la división de los mismos en dos subconjuntos, un subconjunto que servirá para entrenamiento del modelo y otro para test de la precisión del mismo.

La técnica usada para particionar los datos será aleatoria con una distribución 70/30, para la data de entrenamiento y testing respectivamente.

Concluido el proceso de transformación de datos, entramos a las técnicas de aprendizaje automático, se procederá a aplicar los algoritmos matemáticos que permitan clasificar de forma recurrente los asegurados con venta cruzada, tal cual lo describe el capítulo de bases metodológicas, se aplicará través de las librerías del software Knyme Analytics las funciones de xgboost y bosques aleatorios.

El método de comprobación/evaluación del modelo será a través de la matriz de confusión, en donde serán comparados sus índices de exactitud para determinar la idoneidad y factibilidad del o los modelos.

TABLA 5 – PREDICCIÓN DE BASES

Base de Entrenamiento	Clasificación predecida		
Clasificación observada	Con Venta Cruzada	Sin Venta Cruzada	Total
Con Venta Cruzada			
Sin Venta Cruzada			
Total			

Base de Prueba	Clasificación predecida		
Clasificación observada	Con Venta Cruzada	Sin Venta Cruzada	Total
Con Venta Cruzada			
Sin Venta Cruzada			
Total			

Fuente: Elaboración propia, 2023

TABLA 6 – MATRIZ DE CONFUSIÓN

		Estimado por el Modelo			
Matriz de confusión		Negativo (N)	Positivo (S)		
Real	Negativo	a:(TN)	b:(FP)	Precisión ("precisión") Porcentaje predicciones positivas correctas:	d/(b+d)
	Positivo	c:(FN)	d:(TP)		
		Sensibilidad, exhaustividad ("Recall") Porcentaje casos positivos detectados	Especificidad, (Specifity) Porcentaje casos negativos detectados	Exhaustividad, ("accuracy") Porcentaje de predicciones correctas	
		d/(d+c)	a/(a+b)	(a+d)/(a+b+c+d)	

Fuente: Elaboración propia, 2023

3.5 FUENTES DE INFORMACIÓN

Acorde a Hernández Sampieri, Fernández Collado, & Baptista Lucio (2014, p.61) se define como fuentes de información:

La revisión de la literatura implica detectar, consultar y obtener la bibliografía (referencias) y otros materiales que sean útiles para los propósitos del estudio, de donde se tiene que extraer y recopilar la información relevante y necesaria para enmarcar nuestro problema de investigación.

3.4.1. FUENTES PRIMARIAS

Se considera como fuente primaria la información obtenida de primera mano de MAPFRE Seguros Honduras, la cual fue analizada y fundamental para el correcto desarrollo de la investigación. A continuación, detallamos las fuentes primarias:

1. Datos y registros de asegurados de auto y vida de MAPFRE Seguros Honduras con pólizas vigentes durante el periodo 2022-2023.
2. Registros de prima promedio por póliza de auto y vidas.
3. Registro de cantidad promedio de pólizas por cliente al cierre 2022.

La técnica usada para la recolección de datos de reclamos fue la ejecución de consultas (queries) a la base de datos del core de MAPFRE Seguros Honduras.

Guillén (2022) define como query:

“Una query es una petición precisa para obtener información en una base de datos o sistema de información”.

A través del lenguaje SQL se invocó a los registros de asegurados que cumplían el filtro configurado en la consulta y que permitían la consolidación de una base única.

En lo referente a los registros de primas por póliza, estos de igual manera fueron recabados a través de Query considerando la variable prima total por cliente.

CAPÍTULO IV. RESULTADOS Y ANÁLISIS

4.1 INFORME DE PROCESO DE RECOLECCIÓN DE DATOS

Referente al desarrollo del proceso de recolección de datos, se utilizaron consultas o queries a bases de datos a través del lenguaje estándar SQL, que permitió extraer datos de diversas tablas del sistema CORE de la entidad aseguradora, las consultas se formularon esencialmente a través de la siguiente estructura:

SELECT <campos o columnas de una tabla>

FROM <tabla o tablas>

WHERE <condición que deben cumplir los registros>

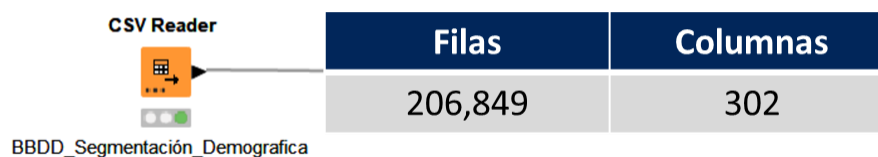
ORDER BY <campos> ASC/DESC

Como resultado se han obtenido archivos en formato CSV, los cuales han sido almacenados bajo nombres asociativos a las variables almacenadas en cada colección. Las características de cada base se amplían en el apartado siguiente:

1. BBDD_Segmentación_Demografica

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de autos, se incluyen datos demográficos y geográficos.

Tipo de variables: numéricos y de cadena.



CSV Reader	Filas	Columnas
BBDD_Segmentación_Demografica	206,849	302

Figura 12. Nodo Segmentación Demográfica.

Fuente: (Elaboración propia, 2023)

2. BBDD_Siniestros_Autos

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de reclamo por siniestro de póliza de auto, se incluyen datos conductuales.

Tipo de variables: numérica

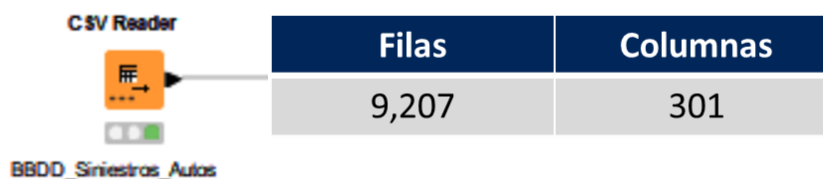


Figura 13. Nodo Siniestros Autos.

Fuente: (Elaboración propia, 2023)

3. BBDD_Prima

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos de la prima pagada por el cliente.

Tipo de variables: numérica

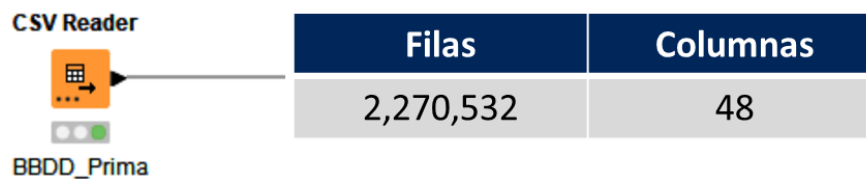


Figura 14. Nodo Prima.

Fuente: (Elaboración propia, 2023)

4. BBDD_Suma asegurada

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos de la cuantía de la suma asegurada.

Tipo de variables: numérica

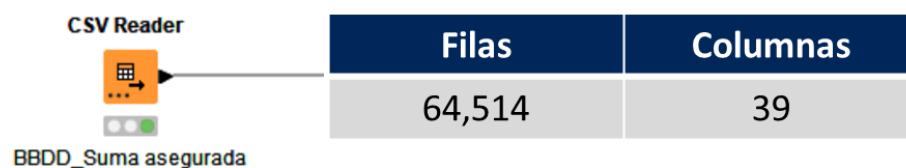


Figura 15. Nodo Suma Asegurada.

Fuente: (Elaboración propia, 2023)

5. BBDD_Coberturas Adicionales

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos de las coberturas adicionales asumidas por el cliente.

Tipo de variables: de cadena

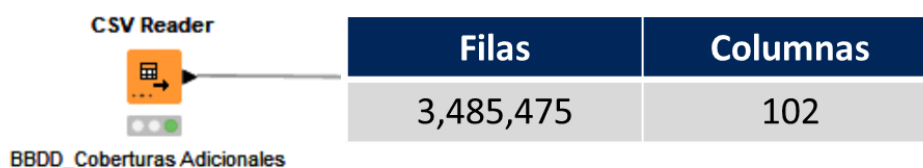


Figura 16. Nodo Coberturas Adicionales.

Fuente: (Elaboración propia, 2023)

6. BBDD_Nivel de ingresos

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de autos, se incluyen datos del ingreso económico del cliente.

Tipo de variables: numéricos y de cadena.

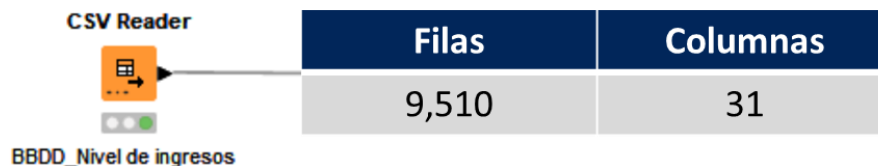


Figura 17. Nodo Nivel de Ingresos.

Fuente: (Elaboración propia, 2023)

7. BBDD_Venta Cruzada

Definición: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos que describen la densidad de la cartera de productos del asegurado.

Tipo de variables: numérico y de cadena

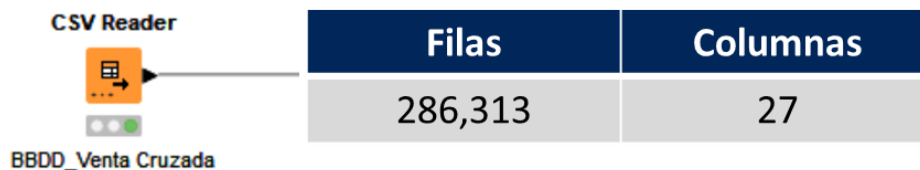


Figura 18. Nodo Venta Cruzada.

Fuente: (Elaboración propia, 2023)

4.2 RESULTADO Y ANALISIS DE LAS TECNICAS APLICADAS

4.2.1. EXPLORACIÓN DE DATOS

Uno de las etapas más importantes en la ciencia de datos y particularmente en la investigación realizada en MAPFRE Seguros Honduras, es el análisis exploratorio de datos conocida esta técnica comúnmente por sus siglas en Ingles EDA: Exploratory Data Analysis, Megan (2022) define como EDA:

“Es un enfoque estadístico y visual que se utiliza para obtener información sobre un conjunto de datos.”. A continuación, los hallazgos más relevantes de este proceso:

1. HALLAZGO: Distribución por Género

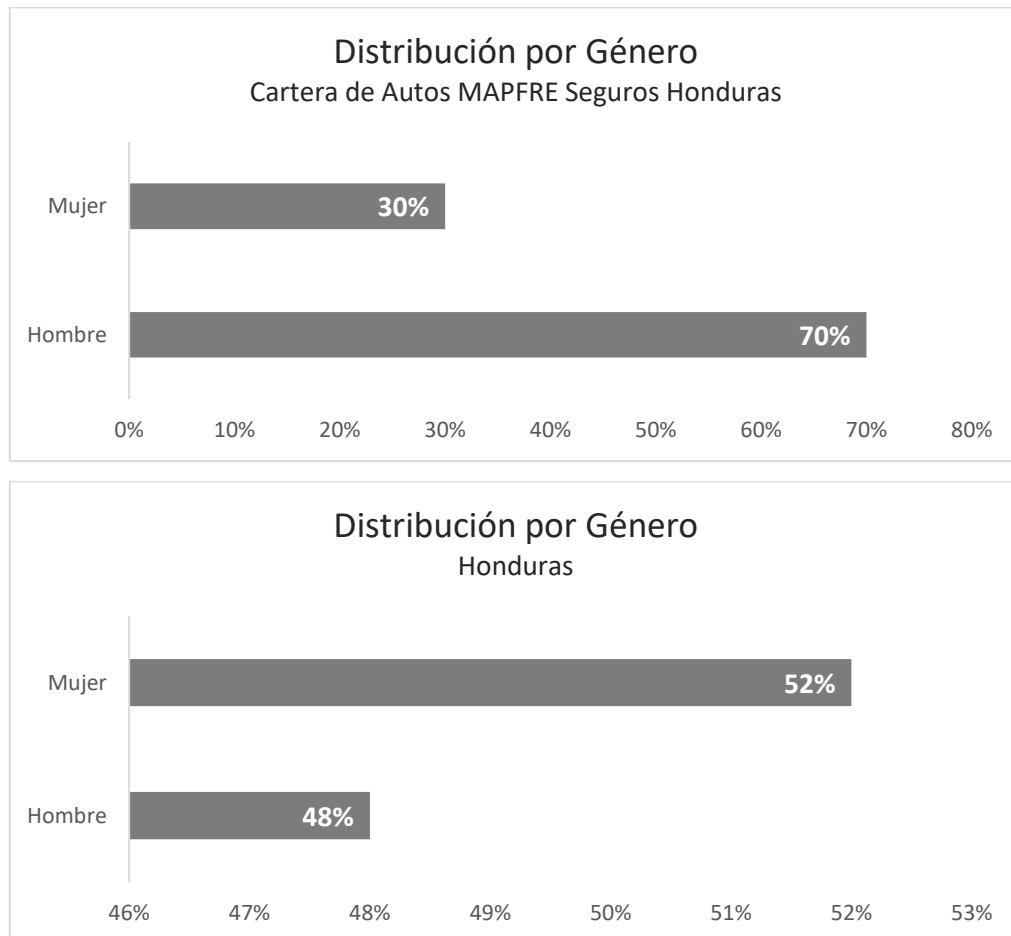


Figura 19. Comparación Distribución por Genero.

Fuente: (Elaboración propia, 2023)

Es reseñable la identificación de que la población de la cartera de autos de MAPFRE Seguros Honduras se comporta de manera dispar a la población del país hondureño, lo anterior puede estar asociado a la mayor participación del sexo masculino en el mercado laboral, lo cual resulta en un mayor ingreso y capacidad de adquisición de vehículo junto a la correspondiente póliza de auto.

2. HALLAZGO: Distribución por Edad

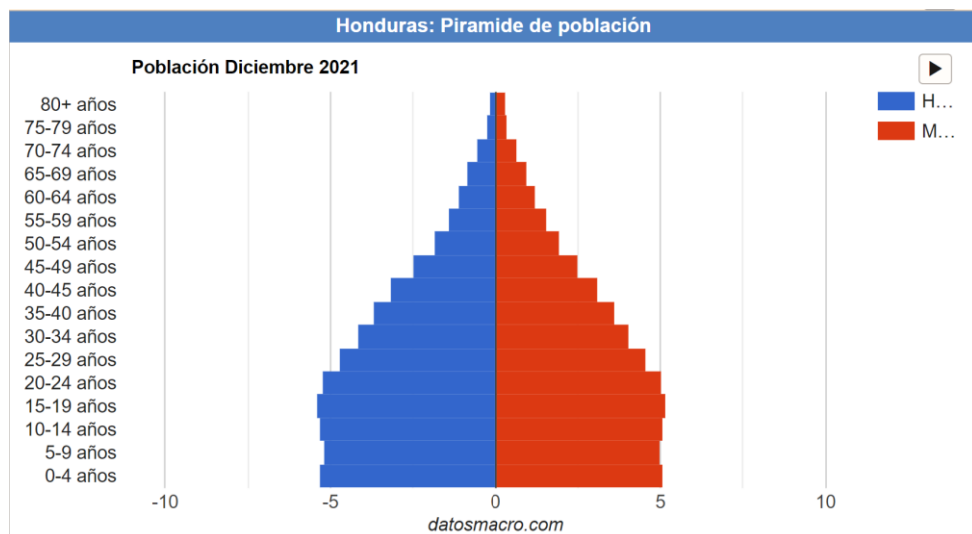
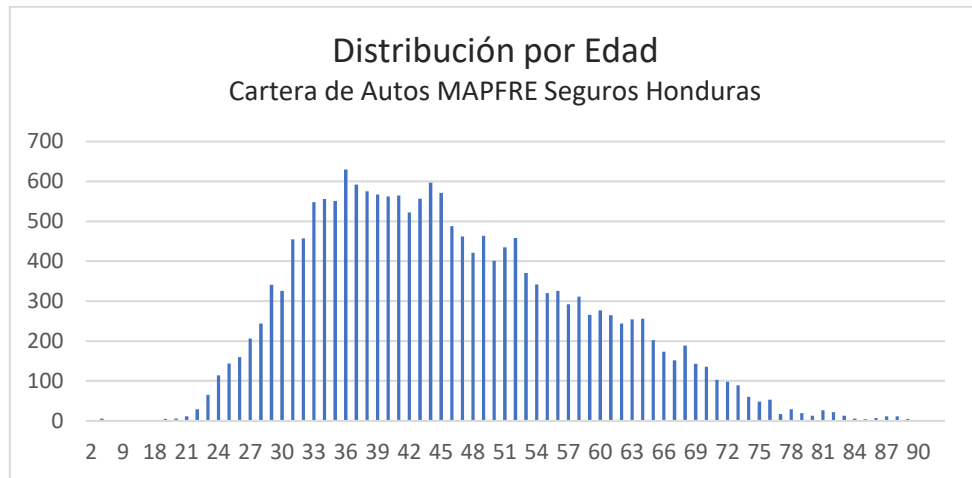


Figura 20. Comparación Distribución por Edad.

Fuente: (Expansión, 2021)

En el análisis de edad se observa que la mayoría de la población de la cartera de Autos de MAPFRE Seguros Honduras, oscila entre los 24 y 70 años de edad, siendo el asegurado de menor edad de 18 años, tiempo en que las leyes del país habilitan a conducir un automotor, por otro lado, resalta en el gráfico de distribución, asegurados con edad muy avanzada (90 años), que disponen de su póliza aún vigente. En comparación a la pirámide de edad de la población hondureña se observa una estructura similar.

3. HALLAZGO: Distribución por Siniestro

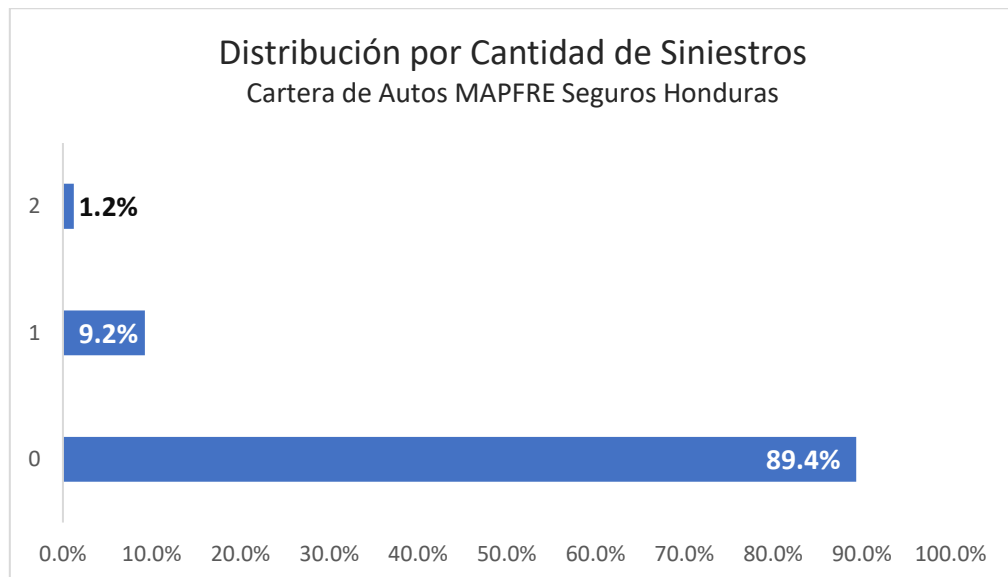
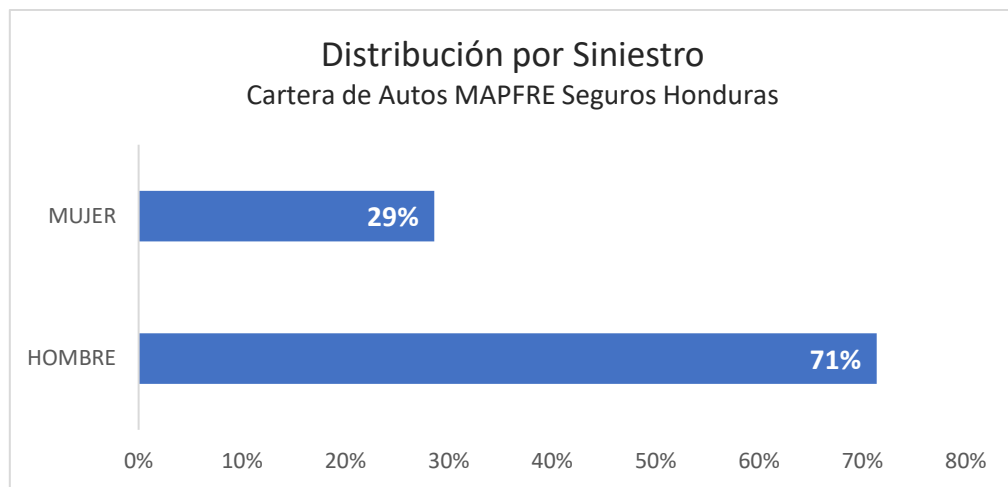


Figura 21. Comparación Distribución por Siniestro.

Fuente: (Elaboración propia, 2023)

Analizando a detalle la siniestralidad de los asegurados de pólizas de auto, resalta el sexo masculino, presenta una tendencia significativamente superior a tener accidentes automovilísticos en comparación a las mujeres, lo cual está asociado a una mayor densidad del sexo en la cartera. En otra arista es reseñable que menos del 11% de la cartera ha presentado un reclamo por siniestro a la compañía aseguradora.

4. HALLAZGO: Distribución por Suma Asegurada

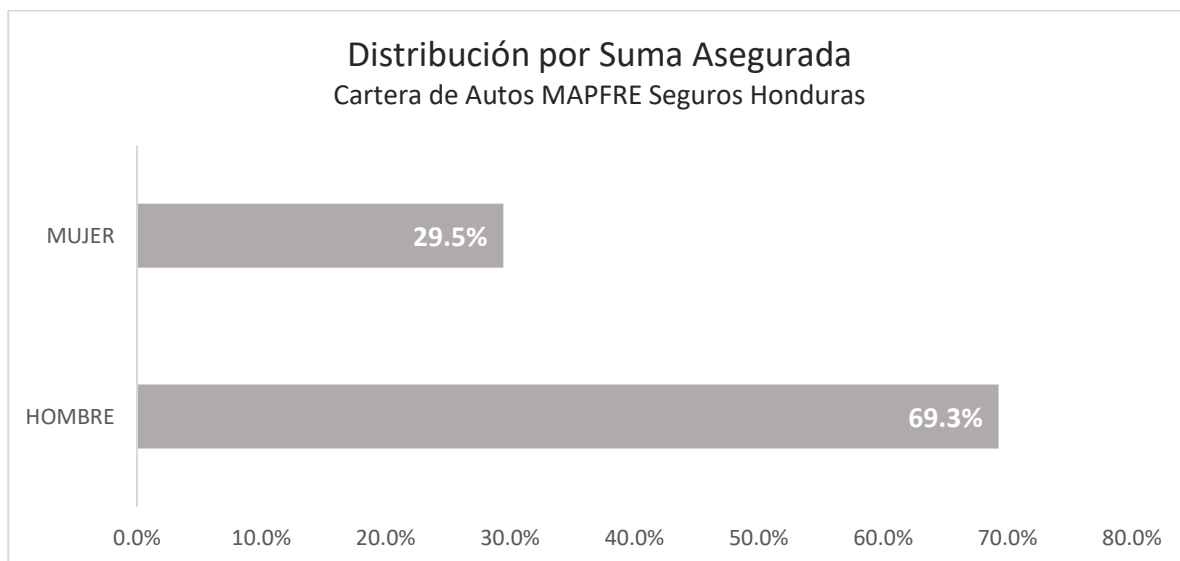
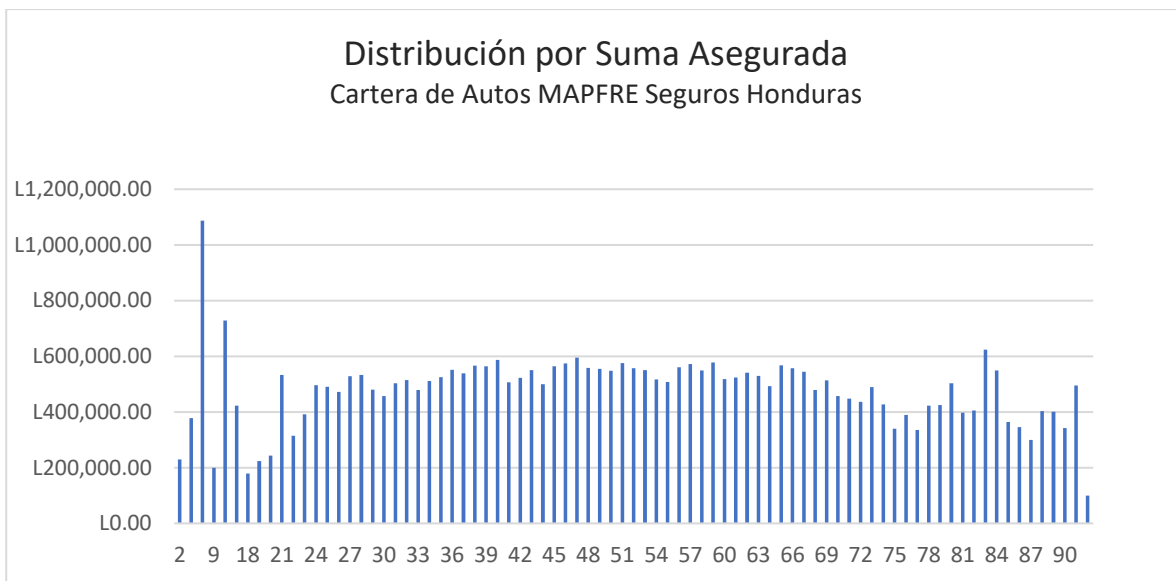


Figura 22. Comparación Distribución por Suma Asegurada.

Fuente: (Elaboración propia, 2023)

Realizando un análisis por suma asegurada, la distribución muestra que el valor máximo al cual puede ascender un vehículo es de 1.2 millones de lempiras y que el valor medio ronda los 400 mil, por suma asegurada. Si analizamos por sexo, identificamos que, los hombres representan un 69.3% del riesgo para la aseguradora y las mujeres un 29.5% respectivamente.

5. HALLAZGO: Distribución por Prima

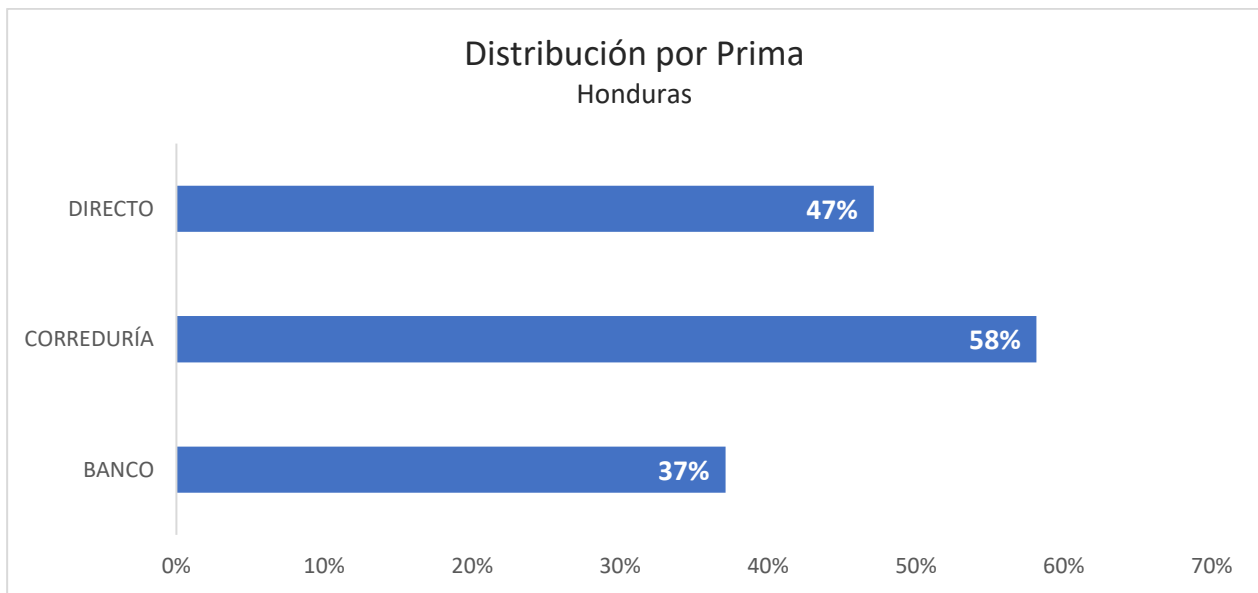
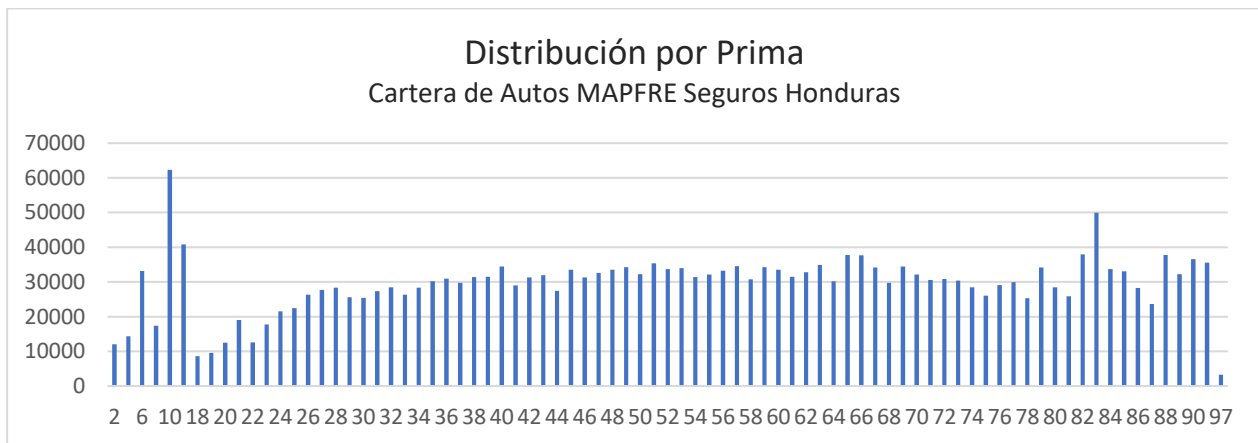


Figura 23. Comparación Distribución por Prima.

Fuente: (Elaboración propia, 2023)

El análisis de frecuencia de prima muestra que L 20,000.00, es el valor que más se repite en el proceso core de ingresos. El estudio del vector más importante de una aseguradora, refleja que el canal correduría está liderando las ventas de seguro de autos, seguido por el canal directo y cerrando con el canal bancario.

6. HALLAZGO: Distribución por Producto

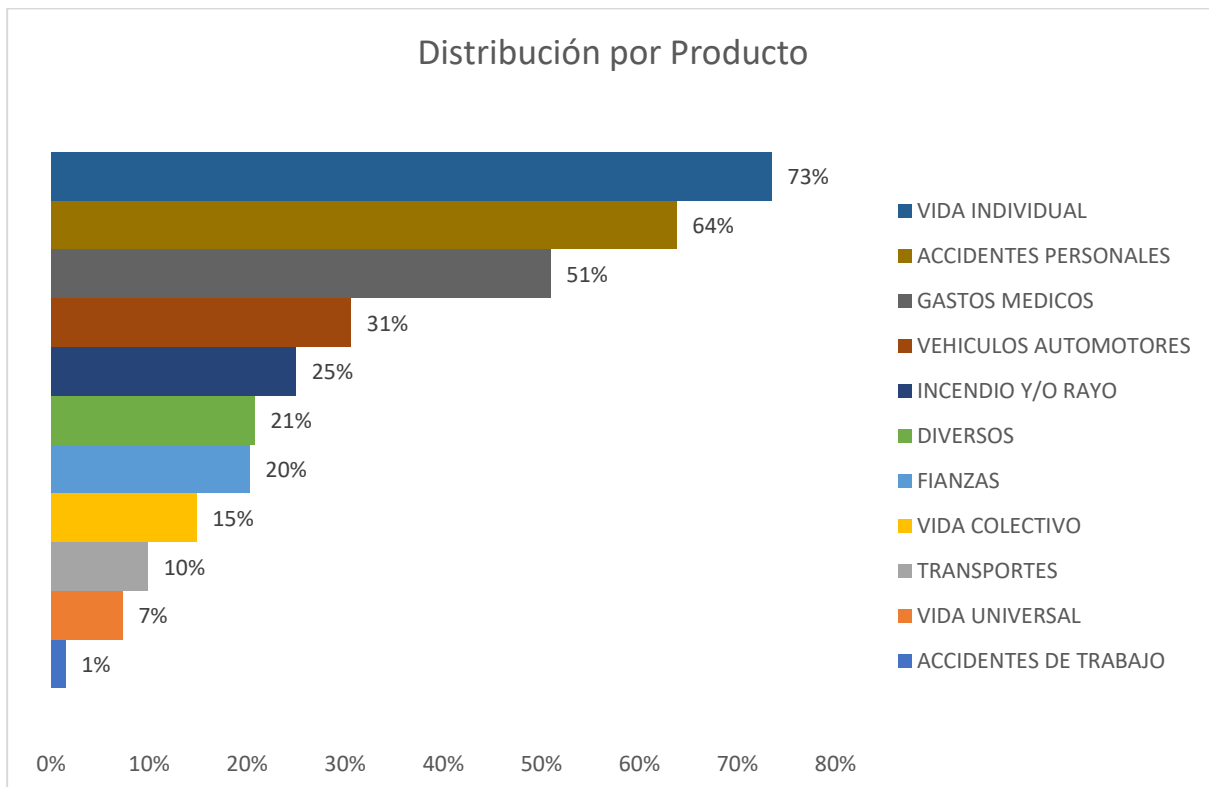


Figura 24. Comparación Distribución por Producto.

Fuente: (Elaboración propia, 2023)

En relación a la distribución de productos, tenemos a vida individual con la mayor representación en la cartera con un 73% y en cuarta posición tenemos al producto de autos con un 31%, resaltamos estos dos productos a causa de que son los objetos de estudio de la presente investigación.

4.2.2. PROCESO DE TRANSFORMACIÓN

Previo a la aplicación de un modelo matemático a un set de datos, la literatura indica que es necesario una preparación adecuada de los registros, tal cual un cocinero limpia y prepara los ingredientes antes de “cocinarlos”. Se estima el 80% de la labor de “cocinar una predicción” exitosa, está directamente relacionada a la limpieza y transformación de los datos, en los párrafos siguientes se detalla las tareas de transformación realizadas sobre cada base obtenida de MAPFRE Seguros Honduras:

4.2.2.1. BBDD_Segmentación_Demografica

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.
3. Renombrar columnas con referencias al contenido.
4. Transformar variable tipo String a Fechas.
5. Calcular edad a través de diferencia en fechas.
6. Convertir variables sexo, estado civil, tipo de persona, tipo de póliza, canal de venta, forma de pago y tipo de moneda a nomenclaturas estándar.
7. Reordenar las columnas para un tratamiento expedito.
8. Agrupar data por número de póliza y fecha de ultimo suplemento.
9. Eliminar pólizas canceladas o no renovadas.
10. Incluir variable con rangos de vigencia de póliza.
11. Eliminar pólizas de personas jurídicas (empresas).
12. Filtrar por pólizas de auto únicamente.
13. Join con BBDD_Siniestros_Autos.

4.2.2.2. BBDD_Siniestros_Autos

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.

3. Remover filas con registros duplicados.
4. Agrupar data por número de póliza.
5. Join con BBDD_Segmentación_Demografica.

4.2.2.3. BBDD_Prima

1. Agrupar data por número de póliza.
2. Join con BBDD_Siniestros_Autos.

4.2.2.4. BBDD_Suma asegurada

1. Agrupar data por número de póliza
2. Incluir variable con percentiles de suma asegurada.
3. Join con BBDD_Prima.

4.2.2.5. BBDD_Coberturas Adicionales

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.
3. Pivotear filas a columnas.
4. Join con BBDD_Suma asegurada

4.2.2.6. BBDD_Nivel de ingresos

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.

3. Convertir variables de monto de ingreso a nomenclatura estándar.
4. Eliminar pólizas con campo valor de ingreso vacío.
5. Join con BBDD_Coberturas Adicionales.

4.2.2.7. BBDD_Venta Cruzada

1. Renombrar columnas con referencias al contenido.
2. Remover guiones de variable número de identidad.
3. Remover columnas de variables superfluas al modelo.
4. Remover columnas con datos vacíos mayor al 90%.
5. Filtrar por pólizas de vida individual únicamente.
6. Agrupar data por número de documento.
7. Join con BBDD_Nivel de ingresos.

El proceso de transformación nos permite enumerar la cantidad de predictores o variables observadas para predecir la existencia de venta cruzada, según es detallado en la tabla a continuación:

TABLA 7 – DICCIONARIO DE VARIABLES

Variable	Descripción	Valores esperados
#_DOCUM	Identidad	Numérico
NUM_POLIZA	No de Póliza	Numérico
EDAD	Edad	Numérico
SEXO	Sexo	Hombre/Mujer
TIPO_PERSONA	Tipo de personaría	Natural/Jurídico
EST_CIVIL	Estado civil	Soltero/Casado/Viudo/Unión Libre
EMAIL	correo electrónico	tipo correo
NACIONALIDAD	Nacionalidad	Cadena
VAL_INGRESOS	Nivel de ingresos	De 0 a 3/De 4 a 6/De 7 a 10/De 11 a 20/De 21 a 50/De 50 en adelante
TIP_ACT_ECONOMICA	Tipo de actividad económica	1 2 3 4 5 6 7 8 9 10 11 12

TXT_NOM_TRABAJO	Nombre de lugar de trabajo	Cadena
TIP_NIVEL_INGRESO	Tipo nivel de ingreso	1/2/3/4/5/6
TXT_FONDOS	Origen de los fondos	Cadena
PROFESION	Profesión/Ocupación	Cadena
CIUDAD	Ciudad	TGU/SPS/LCB
DOMICILIO	Dirección	Cadena
TIP_SEGURO	Ramo	AUTOS/VIDA
MODALIDAD	Producto	AUTO-TOTAL/BAC
CANAL_VENTA	Canal de adquisición de seguro	Corredor/Directo/Banco
TIEMP_EMISION	Años desde la emisión inicial de póliza	0 a 12 meses/12 a 36 meses/36 a 60 meses
NUM_RENOVACIONES	Tiene renovación el cliente	Numérico
CANT_SPTO	Cantidad de suplementos que tiene	Numérico
MARCA	Marca del vehículo	Cadena
TIP_VEH	Tipo de vehículo	CAMIONETA /TURISMO/CAMION/PICK UP
USO_VEH	Uso del vehículo	COMERCIAL PARTICULAR
COBER_AD_1	Coberturas adicional 1	SI/NO
COBER_AD_2	Coberturas adicional 2	SI/NO
COBER_AD_3	Coberturas adicional 3	SI/NO
COBER_AD_4	Coberturas adicional 4	SI/NO
COBER_AD_5	Coberturas adicional 5	SI/NO
COBER_AD_6	Coberturas adicional 6	SI/NO
COBER_AD_7	Coberturas adicional 7	SI/NO
FORMA_PAGO	Forma de Pago	Contado/Fraccionado
MONEDA	Moneda de Pago	Lempira/dólar
TOTAL_PRIM	Prima total pagada	Numérico
PERC_PRIM	Percentil en que se encuentra en base a la prima pagada	1/2/3/4/5
SUMA_ASEG	Suma asegurada total	Numérico
PERC_ASEG	Percentil en que se encuentra en base a la suma asegurada	1/2/3/4/5
CANT_SINI	Cant de siniestros	Numérico
MONT_SINI	Monto total de siniestro pagado	Numérico
VENCRU	Posee producto adicional / Variable a Predecir	Si/No

Fuente: Elaboración propia, 2023

Luego de consolidar todas las bases en un set único, concluimos la preparación realizando una división de la base de datos, tendremos en un lado los registros de pólizas nuevas emitidas durante el presente año y que no disponen de venta cruzada, y por otra parte la base de pólizas con vigencias mayor a 1 año.

Obteniendo todas las variables referenciadas en el diccionario, y con el objetivo de facilitar el trabajo de lectura de datos y el cálculo a los algoritmos de aprendizaje automático, los datos fueron transformados a variables numéricas, de modo que por ejemplo la variable “sexo” toma el valor 0 en caso de mujer y 1 en caso de hombre. Se replicó esta acción para el resto de las variables, excepto aquellas que ya sostenían un valor numérico como “prima” las cuales mantienen su valor original.

4.2.3. ANÁLISIS DE CORRELACIÓN

Concluido el proceso de transformación, se procedió a realizar un análisis de correlación entre variables. La herramienta comúnmente utilizada para medir la relación e influencia entre las variables se denomina como Matriz de Correlación, aprovechada para medir de forma adicional la fuerza de la relación entre variables y puede tomar valores en su coeficiente entre menos uno y uno. El coeficiente de correlación, oscilará cerca a uno cuando la relación entre dos variables sea mayor en magnitud y comportamiento, es decir si una aumenta la otra también aumenta en la misma proporción. Por el contrario, el coeficiente oscilará cerca de -1 cuando las variables miden características opuestas o diferentes.

Ortega (2023) define como Matriz de Correlación:

“Una matriz de correlación es una herramienta estadística que muestra la intensidad y la dirección de la relación entre dos o más variables.”

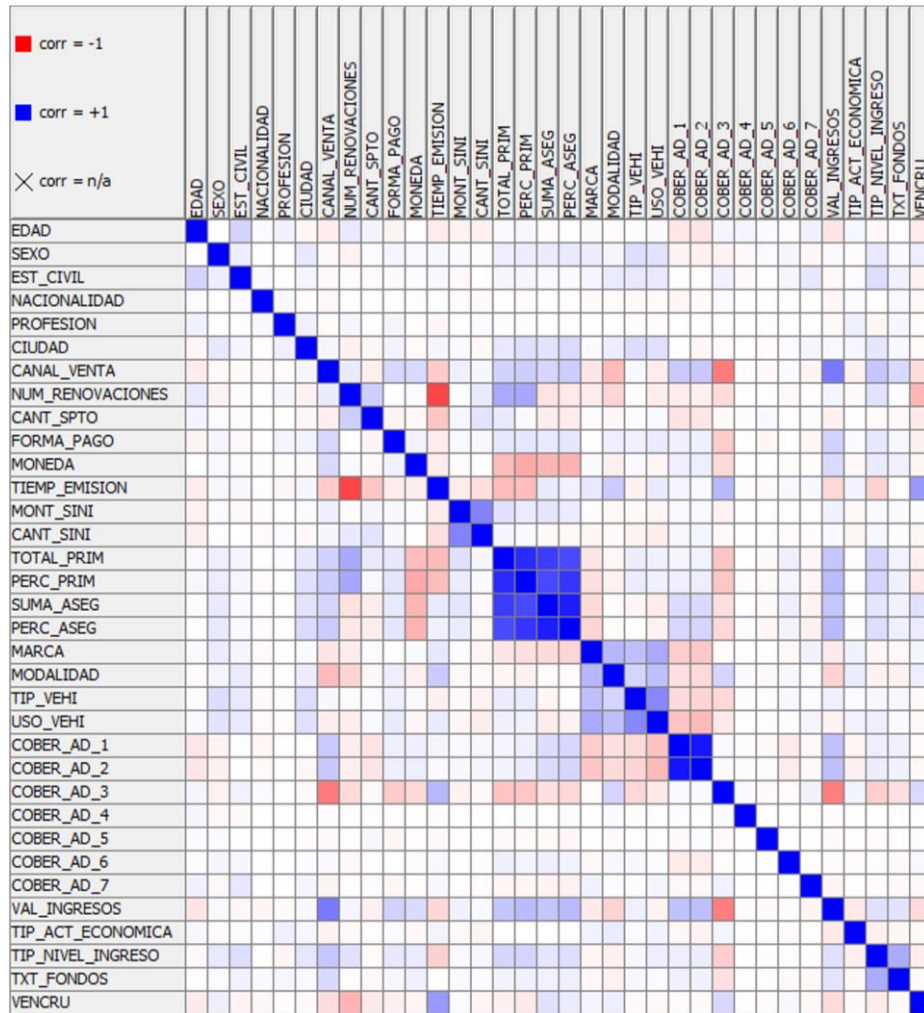


Figura 25. Matriz de Correlación.

Fuente: (Elaboración propia, 2023)

En el grafico anterior se observa que “VENCRU”, la variable a predecir, tiene en su mayoría correlaciones cercanas a cero, exceptuando algunas variables como “tiempo de emisión” que se posiciona arriba de cero, esto se traduce, a que miden características similares que están relacionadas, pero la relación que presentan es inferior a 0,5 por lo que no es una relación excesivamente fuerte.

4.2.4. APLICACIÓN DE ALGORITMO

Teniendo la base de datos completa y luego de identificar la correlación de las variables, se deberán crear dos subconjuntos de datos, uno será empleado para el entrenamiento de los algoritmos y el otro será usado para probar la habilidad predictiva de los mismos. La partición se realizó basado en el método de Machine Learning, que sugiere, realizar como la mejor estrategia una partición 80-20%, asegurando que la proporción de las clases o valores sea similar en ambos conjuntos. Se comprobó de forma efectiva que el conjunto original y los dos nuevos subconjuntos tienen la misma cantidad de registros de clientes con la variable a predecir, en este caso con clientes con venta cruzada.

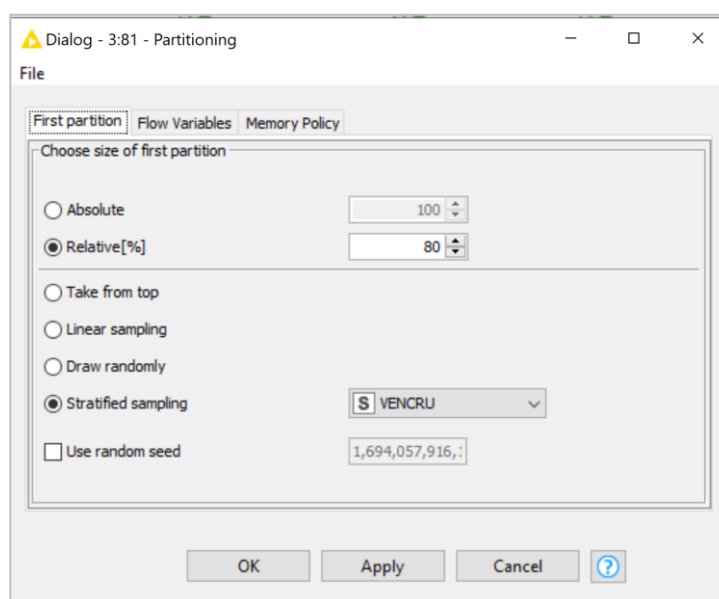


Figura 26. Partición de Datos.

Fuente: (Elaboración propia, 2023)

TABLA 8 – CARACTERÍSTICAS DE SET DE DATOS

% DE INDIVIDUOS QUE HAN REALIZADO VENTA CRUZADA		
Conjunto de datos original	Subconjunto de datos de entrenamiento	Subconjunto de datos de validación
20.34%	20.32%	20.39%

Fuente: Elaboración propia, 2023

Realizada la partición de los datos, el siguiente paso es seleccionar los parámetros óptimos en la configuración del algoritmo, denominada esta fase como el entrenamiento del algoritmo, el análisis de parámetros se realizó de forma aislada probando de forma aleatoria que parámetro es el mejor, con el resultado a continuación:

1. Algoritmo XGBoost

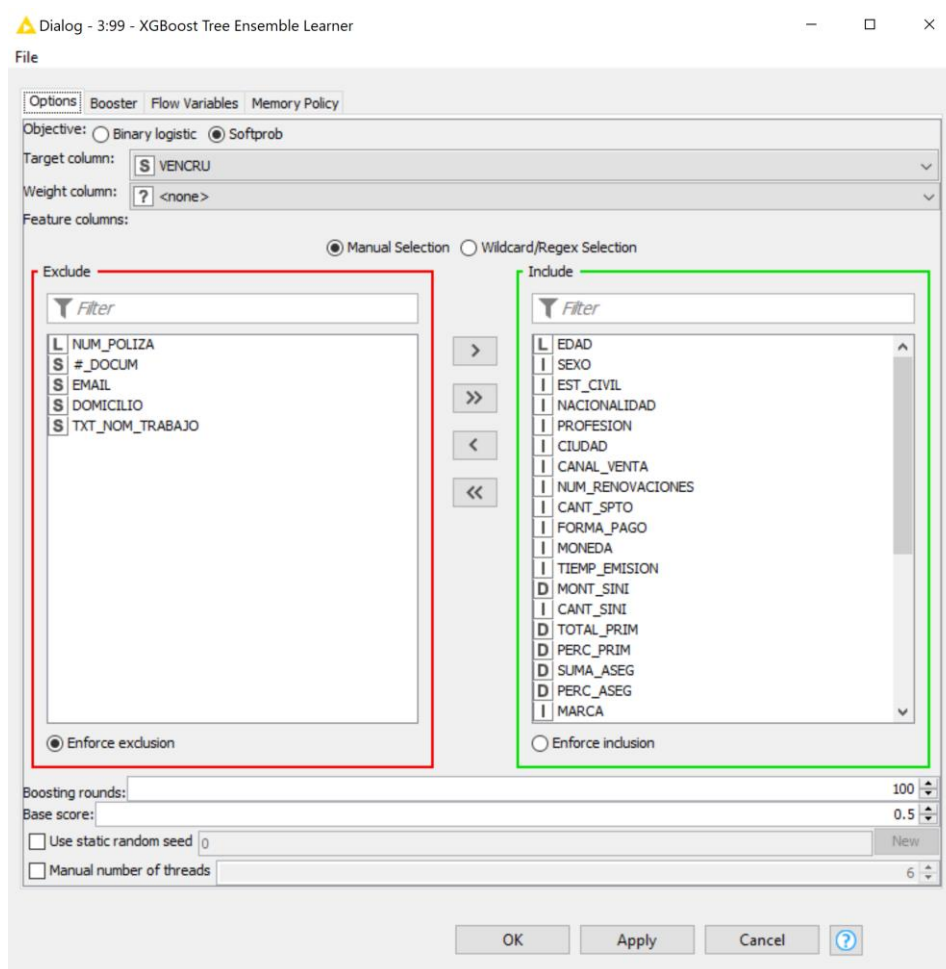


Figura 27. Entrenamiento Algoritmo XGBoost.

Fuente: (Elaboración propia, 2023)

Se ha seleccionado el target columna como VENCRU, se han excluido las variables que no tienen correlación y se ha parametrizado el boosting rounds en valor de 100.

2. Algoritmo Bosques Aleatorios

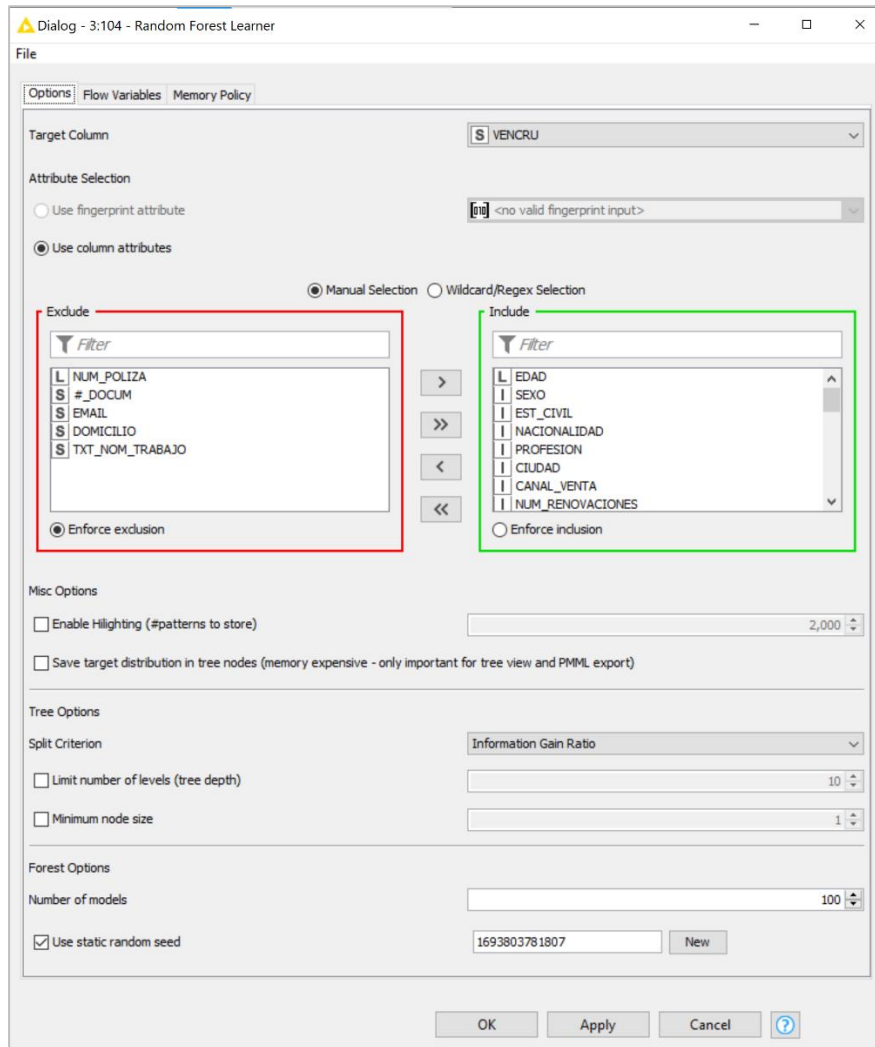


Figura 28. Entrenamiento de Algoritmo Bosques Aleatorios.

Fuente: (Elaboración propia, 2023)

Se ha seleccionado el target columna como VENCRU, se han excluido las variables que no tienen correlación y se ha parametrizado el número de modelos en valor de 100.

4.2.5. COMPARACIÓN DE RESULTADOS

Según se ha descrito en capítulo anterior, la Matriz de Confusión es la técnica en excelencia para medir los resultados de las predicciones y el rendimiento del algoritmo. El funcionamiento es fácil de entender, además posee indicadores que permiten extraer conclusiones y oportunidades de ajuste en el entrenamiento. Bajo la premisa anterior, se muestran las matrices de confusión del algoritmo XGBoost y bosques aleatorios:

1. XGBoost

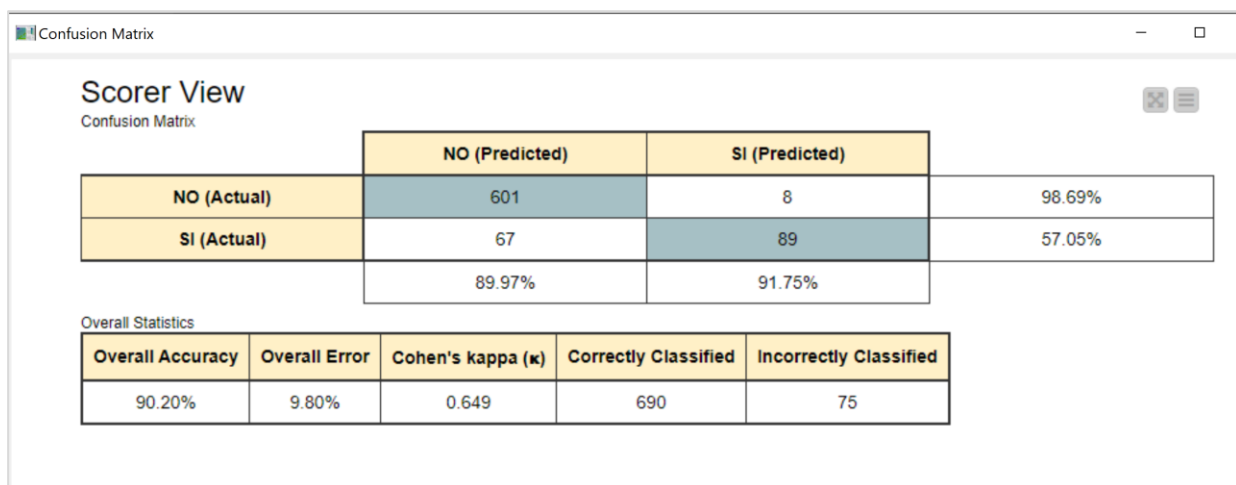
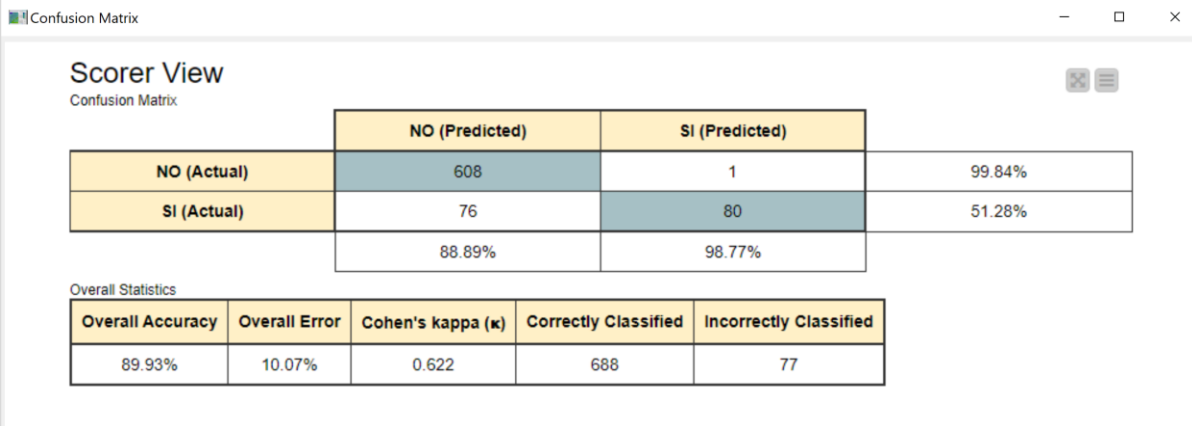


Figura 29. Matriz de Confusión XGBoost.

Fuente: (Elaboración propia, 2023)

2. Bosques Aleatorios



The screenshot shows a 'Scorer View' window with a 'Confusion Matrix' and 'Overall Statistics'.

Confusion Matrix:

	NO (Predicted)	SI (Predicted)	
NO (Actual)	608	1	99.84%
SI (Actual)	76	80	51.28%
	88.89%	98.77%	

Overall Statistics:

Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
89.93%	10.07%	0.622	688	77

Figura 30. Matriz de Confusión Bosques Aleatorios.

Fuente: (Elaboración propia, 2023)

Los indicadores a comparar para medir la eficacia de los modelos son:

Exactitud: Mide la proporción total de resultados correctos. Tanto verdaderos positivos como verdaderos negativos sobre el total de valores.

$$Exactitud = \frac{TP + TN}{Total}$$

Precisión: Mide la proporción de acierto dentro de sólo los casos positivos. A efectos más gráficos mide la dispersión de las observaciones.

$$Precisión = \frac{TP}{TP + FP}$$

Sensibilidad: Es la capacidad del modelo de discriminar los casos positivos de los negativos. Se calcula como la proporción de los verdaderos positivos predichos sobre todos los casos positivos reales.

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

F-Score: Es una mezcla de la precisión y la sensibilidad en una sola métrica y sirve para casos donde la distribución de las clases no es pareja de tal forma que el F-Score pondera los valores. Se calcula siguiendo la siguiente fórmula.

$$F1 = \frac{2 * \text{Precisión} * \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$$

TABLA 9 – COMPARACIÓN RESULTADOS DE ALGORITMOS

XGBoost		Bosques Aleatorios	
Exactitud	90.20%	Exactitud	89.93%
Precisión	91.75%	Precisión	98.77%
Sensibilidad	57.05%	Sensibilidad	51.28%
F-Score	70.36%	F-Score	67.51%

Fuente: Elaboración propia, 2023

Comparando ambos algoritmos se evidencia que el mejor es el XGBoost, ya que su porcentaje de exactitud es ligeramente mayor que el modelo de Bosques Aleatorios, sin embargo, los indicadores de Sensibilidad y F-Score muestran valores por debajo del desempeño buscado.

4.3 RESULTADOS Y ANÁLISIS DE LOS DATOS ENCONTRADOS CON OTRAS TÉCNICAS QUE LOS AUTORES ESTIMEN Y JUSTIFIQUEN QUE SON NECESARIAS.

Como hemos comentado a lo largo de la investigación, en aprendizaje automático, la medición del rendimiento es una tarea vital. En adición a la matriz de confusión, se puede utilizar otras herramientas como métricas de la predicción. La curva AUC-ROC es una de las técnicas de evaluación más importante y utilizadas para validar el desempeño de cualquier algoritmo.

La curva ROC indica que qué tan bueno es el algoritmo, a través de una representación gráfica, en lugar de un valor único como suele ser en otras métricas.

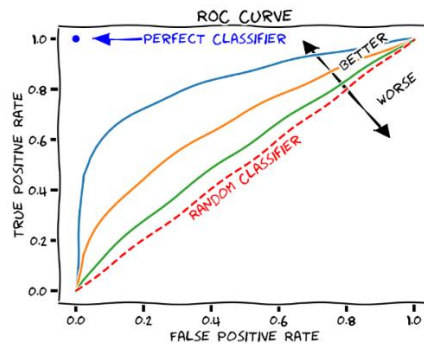


Figura 31. Diagrama ROC Curve.

Fuente: (UDRC, s. f.)

Como se evidencia en la imagen de referencia anterior, si el AUC es igual a uno, el desempeño del modelo es perfecto en el desempeño, y si fuese cero, intercambiaría los valores positivos por negativos y viceversa. Si el desempeño del modelo muestra un AUC igual 0,5 entonces no diferencia las clases de forma correcta y la predicción es aleatoria. En síntesis, entre más se acerca la curva a 1, mejor desempeño y sensibilidad tiene el modelo para clasificar correctamente. Bajo la premisa anterior se muestra a continuación la curva ROC-AUC de los modelos de XGBoost y bosque aleatorio, para el caso de venta cruzada de MAPFRE Seguros Honduras:

1. XGBoost

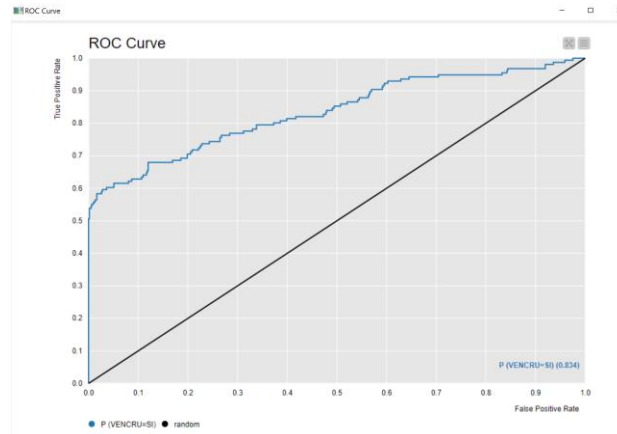


Figura 32. Resultado ROC Curve XGBoost.

Fuente: Elaboración propia, 2023

2. Bosques Aleatorios

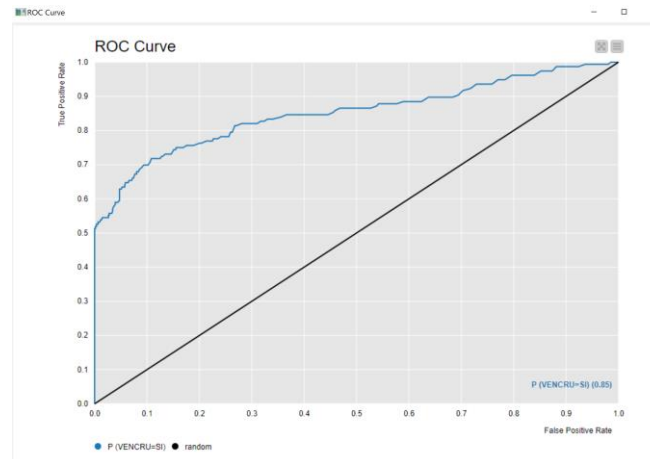


Figura 33. Resultado ROC Curve Bosques Aleatorios.

Fuente: Elaboración propia, 2023

TABLA 10 – COMPARACIÓN RESULTADOS DE CURVA ROC-AUC

XGBoost	Bosques Aleatorios
83.40%	85.00%

Fuente: Elaboración propia, 2023

Se observa, en concordancia con la matriz de confusión, que la forma de la curva ROC y el AUC que ambos modelos se comportan de forma similares, y por tanto, son capaces de predecir la venta cruzada.

4.4 ESTUDIO DE ALGORITMO CON BASE DE NUEVA EMISIÓN

4.4.1. PROYECCIÓN DE VENTAS

Luego de finalizar el entrenamiento de los algoritmos, hemos seleccionado XGBoost para estudiar los resultados a obtener con la base de datos de nueva emisión, nos referimos a “nueva emisión” a la base de pólizas con vigencia menor a 12 meses y de clientes que disponen de una única póliza de autos. Damos paso a la predicción de los candidatos a venta cruzada, específicamente candidatos a los cuales se puede vender una póliza de vida individual, obteniendo los resultados siguientes:

TABLA 11 –PREDICCIÓN DE CANDIDATOS

Descripción	Cantidad
Cantidad de Asegurados analizados	2776
Asegurados No candidatos a Venta Cruzada	2329
Asegurados candidatos a Venta Cruzada	447
Mujeres candidatas a Venta Cruzada	126
Hombres candidatos a Venta Cruzada	321

Fuente: Elaboración propia, 2023

4.4.2. PROYECCIÓN DE VENTAS

A partir de la predicción de candidatos de venta cruzada realizada en el paso previo, y tomando como base el monto promedio de prima por cliente L2642.00 en pólizas de vida individual, podemos realizar estimaciones del monto que se puede alcanzar implantando una técnica de venta cruzada:

TABLA 12 –PROYECCIÓN DE VENTAS

Descripción		Prima proyectada
Asegurados candidatos a Venta Cruzada	L	1,180,974.00*
Mujeres candidatas a Venta Cruzada	L	332,892.00
Hombres candidatos a Venta Cruzada	L	848,082.00

Fuente: Elaboración propia, 2023

*El monto de colocación representado en la tabla anterior, es el valor máximo que podría alcanzarse en el escenario de de una predicción 100% exitosa, sin embargo, los indicadores mostrados por la matriz de confusión señalan que, pueden surgir falsos positivos durante la predicción, lo que se traduce en una caída en el monto de ventas.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- 5.1.1 A través de la investigación realizada en MAPFRE Seguros Honduras, donde han sido utilizados 6,332,400 registros, se puede concluir que el análisis de datos facilita la identificación de mejoras en la información disponible en la organización, y favorece la toma de decisiones estratégicas.
- 5.1.2 En cumplimiento al primer objetivo específico, el cual es realizar el análisis de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2020-2023, ha sido alcanzado a través de técnicas de analítica de datos, las cuales han permitido concluir que las características más importantes de los registros de asegurados, son captadas por la aseguradora a través de sus procesos operativos core y permiten destacar las particularidades más significativas de la cartera, siendo estas: La mayor densidad (70%) de la cartera de pólizas de auto es masculina, no se presentan tantos siniestros como se estimaba, al contrario únicamente el 11% de la cartera presenta reclamos por siniestros, el valor medio de un vehículo asegurado asciende a 400 mil lempiras, el valor pagado por asegurado por la póliza de auto, que más se repite es por L20,000.00 lempiras.
- 5.1.3 En respuesta al objetivo específico número dos, el cual se basa en la identificación de un marco referencial de variables para la detección de los asegurados candidatos a venta cruzada, ha sido alcanzado luego de un proceso exhaustivo de transformación de bases de datos, a través del cual se ha diseñado un diccionario de 16 variables tipo demográficas y 25 variables tipo conductuales. A partir del diccionario referenciado de variables, fue posible la diagramación de la intensidad de correlación de los predictores con la variable a predecir: “venta cruzada”, dando como resultado en primer lugar el tiempo desde la primera emisión de la póliza de autos (0.4), en segundo el número de renovaciones (-0.28) que ha sufrido y en tercer lugar canal de venta (-0.14) a través del cual se colocó el producto.

- 5.1.4 El tercer objetivo específico, el cual es considerado como la cúspide de la investigación, buscaba diseñar un modelo numérico capaz de predecir candidatos a adquirir un seguro adicional a la póliza de auto, fue culminado a través de la selección y entrenamiento de los algoritmos XGBoost y Bosques aleatorios, los cuales superaron de forma satisfactoria el nivel de exactitud exigido en la hipótesis (70%), medido a través de la matriz de confusión y con porcentajes resultantes de 90.20% y 89.23% respectivamente. La capacidad de predicción y confiabilidad de los modelos fue reconfirmado a través de la curva ROC que mostro un resultado de 83.40% para el algoritmo XGBoost y de 85% para bosques aleatorios.
- 5.1.5 La investigación ha evidenciado su vigor, valor agregado y disponibilidad de uso inmediato, el entrenamiento del algoritmo XGBoost y su aplicación a la base de datos con vigencia menor a un (1) año, ha permitido pronosticar 447 candidatos a venta cruzada y proyectar un monto de colocación mayor a 1 millón de lempiras en ventas para la compañía aseguradora.
- 5.1.6 Existe oportunidad de mejora en los gestión de datos de la compañía de la aseguradora, de forma específica en la calidad de datos, y en los controles durante el registro de un cliente, así como en el proceso de limpieza de los registros en reposo en el sistema de MAPFRE Seguros Honduras, esto se concluyó durante el proceso de transformación en el cual se descartaron el 51% de los registros extraídos a causa de que no cumplían con parámetros estándar de información como ser, registros vacíos, información incompleta o valores atípicos.
- 5.1.7 Concluida la investigación es reseñable las habilidades otorgadas por la maestría de Analítica Negocios, destacando entre ellas, el pensamiento crítico para ver el problema de diferentes perspectivas, la capacidad de preparar los datos para un análisis efectivo y la capacidad de aplicar la matemática y las estadísticas de forma adecuada para resolver un problema de la vida real.

5.2 RECOMENDACIONES

- 5.2.1 Se recomienda a la unidad de proyectos de la aseguradora, implantar un proyecto de calidad de datos, que impulse controles durante el registro de un cliente y asegure el llenado integro de los datos, así como la depuración de los registros en reposo en el sistema de MAPFRE Seguros Honduras, esto basado en que durante el proceso de transformación se descarto el 51% de los registros extraídos a causa de que no cumplían con parámetros estándar de información como ser, registros vacíos, información incompleta o valores atípicos.
- 5.2.2 Se recomienda a la dirección de la aseguradora, impulsar una unidad de gobierno del dato, que persiga la definición de políticas gestión de datos, asegure controles durante toda la cadena de valor del dato y vigile el almacenamiento de los datos.
- 5.2.3 Se recomienda a la dirección de la aseguradora, implantar una unidad de inteligencia de negocios, que explote la Big data que posee la aseguradora, diseñe modelos descriptivos, predictivos y prescriptivos, generando un valor y posicionamiento a la compañía sobre la competencia.
- 5.2.4 Se recomienda a la unidad comercial de la aseguradora, impulsar una campaña de venta cruzada segmentada por canal, a partir de los candidatos predichos por los algoritmos entrenados durante la investigación (XGBoost y Bosques aleatorios). El éxito de la campaña de venta cruzada no se basa en el algoritmo por sí solo, necesita ser soportado por una potente estrategia de venta, que incluya un conocimiento amplio en las coberturas y condiciones de los productos y un script de venta finamente definido.

CAPITULO VI. APLICABILIDAD

6.1 NOMBRE DE LA PROPUESTA

Modelo predictivo de venta cruzada en productos de auto y vida en MAPFRE Seguros Honduras.

6.2 JUSTIFICACION DE LA PROPUESTA

A través de técnicas de analítica de datos, se han podido identificar algunas características importantes de los registros de los asegurados, destacando entre ellas: La mayor densidad (70%) de la cartera de pólizas de auto es masculina, la edad media de los asegurados de autos oscila entre los 24 y 70 años, no se presentan tantos siniestros como se estimaba, únicamente el 11% de la cartera presenta reclamos por siniestros, el valor medio de un vehículo asegurado asciende a 400 mil lempiras, el valor pagado por asegurado por la póliza de auto, que más se repite es por L20,000.00. Por otra parte, ha sido diseñado un marco referencial de variables para la detección de los asegurados candidatos a venta cruzada, luego de un proceso exhaustivo de transformación de bases de datos, a través del cual se ha diseñado un diccionario de 16 variables tipo demográficas y 25 variables tipo conductuales. A partir del diccionario referenciado de variables, fue posible la diagramación de la intensidad de correlación de los predictores con la variable a predecir: “venta cruzada”, dando como resultado en primer lugar el tiempo desde la primera emisión de la póliza de autos (0.4), en segundo el número de renovaciones (-0.28) que ha sufrido y en tercer lugar canal de venta (-0.14) a través del cual se colocó el producto. Como punto total en la importancia de la propuesta es que se ha diseñado un modelo numérico capaz de predecir candidatos a adquirir un seguro adicional a la póliza de auto, a través de los algoritmos XGboost y Bosques aleatorios, los cuales tienen un nivel de exactitud de 90.20% y 89.23% respectivamente. La capacidad de predicción y confiabilidad de los modelos fue reconfirmado a través de la curva ROC que mostro un resultado de 83.40% para el algoritmo XGBoost y de 85% para bosques aleatorios.

6.3 ALCANCE DE LA PROPUESTA

6.3.1. OBJETIVO GENERAL

Implementar un modelo predictivo enfocado en la detección de asegurados candidatos a adquirir un seguro de vida, como producto adicional a la póliza de autos previamente obtenida, que influya en catapultar la técnica de venta cruzada, logrando la fidelización de clientes y explotación máxima de utilidad de cartera, en MAPFRE Seguros Honduras.

6.3.2. OBJETIVOS ESPECÍFICOS

1. Realizar el análisis y transformación de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2020-2023.
2. Predecir a través de un modelo numérico entrenado, los candidatos a adquirir un seguro adicional a la póliza de autos en MAPFRE Seguros Honduras.

6.4 DESCRIPCIÓN Y DESARROLLO A DETALLE DE LA PROPUESTA

6.4.1. DESCRIPCIÓN DE LA PROPUESTA

El modelo predictivo de venta cruzada es una potente herramienta que permitirá a MAPFRE Seguros Honduras mejorar sus ratios de colocación de pólizas a clientes ya existentes, a través de un análisis de patrones de comportamiento.

A continuación, se detallan las fases que deberán ser ejecutadas para implantar el modelo de predicción:

TABLA 13 – FASES IMPLANTACION MODELO PREDICTIVO

Etapa	Acción	Descripción
I	Querys de consulta a base de datos	Se extraerá del sistema core los registros actualizados de información del cliente, que añadan valor al algoritmo matemático de predicción.

II	Análisis y transformación de datos	A través de técnicas de análisis de datos se realizará una depuración de los registros sin valor o incorrectos y se realizarán las transformaciones numéricas de los registros.
III	Entrenamiento de algoritmo	Se entrenará el algoritmo de predicción a través del set de datos depurados y transformados.
IV	Predicción de candidatos	Alcanzado un entrenamiento idóneo del algoritmo se predecirá los clientes con vigencia nueva del año 2023 propensos a adquirir un seguro de vida adicional a su póliza de auto.
V	Proyección de ventas	Se realizarán estimaciones de oportunidad de mercado basado en el monto de prima promedio de colocación de póliza de vida.

Fuente: Elaboración propia, 2023

6.4.2. DESARROLLO DE LA PROPUESTA

6.4.2.1. Querys de consulta a base de datos

A través de consultas a las bases de datos del sistema core, se han obtenido archivos en formato CSV, los cuales han sido almacenados bajo nombres asociativos a las variables almacenadas en cada colección. La cantidad de registros de cada base se amplían en el apartado siguiente:

1. BBDD_Segmentación_Demografica: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de autos, se incluyen datos demográficos y geográficos. Contiene 206,849 filas y 302 columnas.
2. BBDD_Siniestros_Autos: Incluye registros de asegurados de la compañía captados durante el proceso de reclamo por siniestro de póliza de auto, se incluyen datos conductuales. Contiene 9,207 filas y 301 columnas.

3. BBDD_Prima: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos de la prima pagada por el cliente. Contiene 2,270,532 filas y 48 columnas.
4. BBDD_Suma asegurada: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos de la cuantía de la suma asegurada. Contiene 64,514 filas y 39 columnas.
5. BBDD_Coberturas Adicionales: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos de las coberturas adicionales asumidas por el cliente. Contiene 3,485,475 filas y 102 columnas.
6. BBDD_Nivel de ingresos: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de autos, se incluyen datos del ingreso económico del cliente. Contiene 9,510 filas y 31 columnas.
7. BBDD_Venta Cruzada: Incluye registros de asegurados de la compañía captados durante el proceso de emisión de póliza de auto, se incluyen datos que describen la densidad de la cartera de productos del asegurado. Contiene 286,313 filas y 27 columnas.

6.4.2.2. Análisis y transformación de datos

Previo a la aplicación de un modelo matemático a un set de datos, la literatura indica que es necesario una preparación adecuada de los registros, nos referimos a la limpieza y transformación de los datos, en los párrafos siguientes se detalla las tareas de transformación realizadas sobre cada base obtenida de MAPFRE Seguros Honduras:

1. BBDD_Segmentación_Demografica

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.
3. Renombrar columnas con referencias al contenido.
4. Transformar variable tipo String a Fechas.
5. Calcular edad a través de diferencia en fechas.
6. Convertir variables sexo, estado civil, tipo de persona, tipo de póliza, canal de venta, forma de pago y tipo de moneda a nomenclaturas estándar.
7. Reordenar las columnas para un tratamiento expedito.
8. Agrupar data por número de póliza y fecha de ultimo suplemento.
9. Eliminar pólizas canceladas o no renovadas.
10. Incluir variable con rangos de vigencia de póliza.
11. Eliminar pólizas de personas jurídicas (empresas).
12. Filtrar por pólizas de auto únicamente.
13. Join con BBDD_Siniestros_Autos.

2. BBDD_Siniestros_Autos

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.
3. Remover filas con registros duplicados.
4. Agrupar data por número de póliza.
5. Join con BBDD_Segmentación_Demografica.

3. BBDD_Prima

1. Agrupar data por número de póliza.
2. Join con BBDD_Siniestros_Autos.

4. BBDD_Suma asegurada

1. Agrupar data por número de póliza
2. Incluir variable con percentiles de suma asegurada.
3. Join con BBDD_Prima.

5. BBDD_Coberturas Adicionales

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.
3. Pivotear filas a columnas.
4. Join con BBDD_Suma asegurada

6. BBDD_Nivel de ingresos

1. Remover columnas con datos vacíos mayor al 90%.
2. Remover columnas de variables superfluas al modelo.
3. Convertir variables de monto de ingreso a nomenclatura estándar.
4. Eliminar pólizas con campo valor de ingreso vacio.
5. Join con BBDD_Coberturas Adicionales.

7. BBDD_Venta Cruzada

1. Renombrar columnas con referencias al contenido.
2. Remover guiones de variable número de identidad.
3. Remover columnas de variables superfluas al modelo.

4. Remover columnas con datos vacíos mayor al 90%.
5. Filtrar por pólizas de vida individual únicamente.
6. Agrupar data por número de documento.
7. Join con BBDD_Nivel de ingresos.

6.4.2.3. Entrenamiento de algoritmo

Luego de consolidar todas las bases en un set único, concluimos la preparación realizando una división de la base de datos, tendremos en un lado los registros de pólizas nuevas emitidas durante el presenta año y que no disponen de venta cruzada, y por otra parte la base de pólizas con vigencias mayor a 1 año, estas últimas serán utilizadas para entrenamiento del algoritmo de predicción XGBoost.

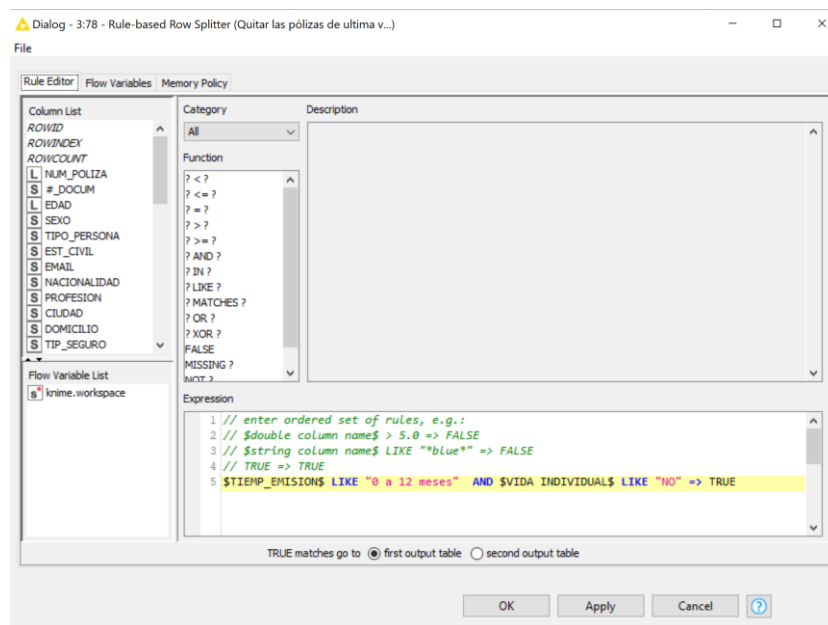


Figura 34. División de Datos.

Fuente: (Elaboración propia, 2023)

Realizada la división de los datos, el siguiente paso es seleccionar los parámetros óptimos en la configuración del algoritmo XGBoost, denominada esta fase como el entrenamiento del algoritmo:

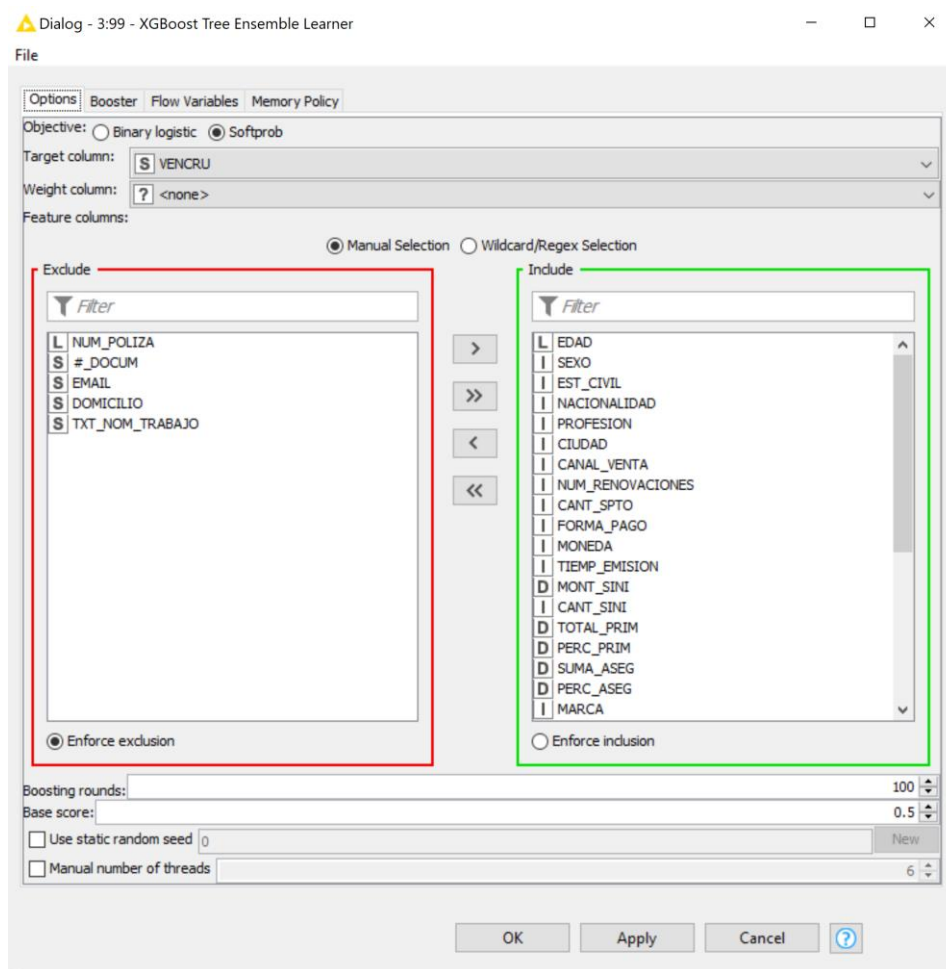


Figura 35. Entrenamiento Algoritmo XGBoost.

Fuente: (Elaboración propia, 2023)

Se ha seleccionado el target columna como VENCRU, se han excluido las variables que no tienen correlación y se ha parametrizado el boosting rounds en valor de 100.

6.4.2.4. Predicción de candidatos

Luego de finalizar el entrenamiento del algoritmo XGBoost, damos paso a la predicción de los candidatos a venta cruzada, específicamente candidatos a los cuales se puede vender una póliza de vida individual, obteniendo los resultados siguientes:

TABLA 14 –PREDICCIÓN DE CANDIDATOS

Descripción	Cantidad
Cantidad de Asegurados analizados	2776
Asegurados No candidatos a Venta Cruzada	2329
Asegurados candidatos a Venta Cruzada	447
Mujeres candidatas a Venta Cruzada	126
Hombres candidatos a Venta Cruzada	321

Fuente: Elaboración propia, 2023

6.4.2.5. Proyección de ventas

A partir de la predicción de candidatos de venta cruzada realizada en el paso previo, y tomando como base el monto promedio de prima por cliente L2642.00 en pólizas de vida individual, podemos realizar estimaciones del monto que se puede alcanzar implantando una técnica de venta cruzada:

TABLA 15 –PROYECCIÓN DE VENTAS

Descripción	Prima proyectada	
Asegurados candidatos a Venta Cruzada	L	1,180,974.00
Mujeres candidatas a Venta Cruzada	L	332,892.00
Hombres candidatos a Venta Cruzada	L	848,082.00

Fuente: Elaboración propia, 2023

6.5 CRONOGRAMA DE IMPLEMENTACIÓN

En el siguiente apartado, se detalla el cronograma de actividades y el costo de los recursos que debería acometer MAPFRE Seguros Honduras, para la implementación y uso del modelo de predicción de venta cruzada:

TABLA 16 – GANTT DE IMPLANTACIÓN DE MODELO

MODELO DE VENTA CRUZADA

LINEAS DE ACTUACIÓN	RESPONSABLE	ENERO				FEBRERO				MARZO			
		S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
Presentación de propuesta a Dirección General y Comercial	Analista de datos	■	■										
Instalación de software de minería de datos	Oficial de IT			■									
Extracción de data historica de clientes	Oficial de IT				■	■							
Análisis y limpieza de datos	Analista de datos						■	■					
Predicción de candidatos a venta cruzada	Analista de datos								■	■			
Inducción a fuerza de venta en colocación de venta cruzada	Oficial Comercial									■	■		
campaña de venta cruzada	Oficial Comercial											■	■
Seguimiento de resultados	Analista de datos												

Fuente: Elaboración propia, 2023

TABLA 17 – PERFILES DE PUESTOS Y PRESUPUESTO REQUERIDO

Estrategia	Perfil requerido	Cantidad	Costo (en base a horas por mes)	Presupuesto
Modelo de Venta Cruzada	Analista de datos	2	L 25,000.00	L 50,000.00
	Oficial de IT	1	L 5,000.00	L 5,000.00
	Oficial Comercial	3	L 10,000.00	L 30,000.00
	Total			L 85,000.00

Fuente: Elaboración propia, 2023

6.6 CONCORDANCIA DE SEGMENTO DE LA TESIS CON LA PROPUESTA

CAPITULO I		
Título de Investigación	Objetivo General	Objetivos específicos
Modelo predictivo de venta cruzada en productos de auto y vida en una aseguradora	Diseñar e implementar un modelo predictivo enfocado en la detección de asegurados candidatos a adquirir un seguro de vida, como producto adicional a la póliza de autos previamente obtenida, que influya en catapultar la técnica de venta cruzada, logrando la fidelización de clientes y explotación máxima de utilidad de cartera, en MAPFRE Seguros Honduras.	<ol style="list-style-type: none"> 1. Identificar un marco referencial de variables a considerar para el análisis de detección de asegurados candidatos a venta cruzada de póliza de autos y vida. 2. Realizar el análisis de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2022-2023. 3. Entrenar un modelo numérico de predicción de candidatos a adquirir un seguro adicional a la póliza de autos para MAPFRE Seguros Honduras.

CAPÍTULO II	CAPÍTULO III		
Teorías/Metodologías de sustento	Variables	Población	Técnicas
<ol style="list-style-type: none"> 1. Aprendizaje automático supervisado 2. Aprendizaje automático no supervisado 3. Aprendizaje automático semi supervisado 4. Aprendizaje automático reforzado 	<p>Dependientes Independientes</p>	<p>14,000 asegurados que componen el negocio de auto y vida.</p>	<ol style="list-style-type: none"> 1. Revisión sistemática de Literatura. 2. Consultas/Querys a bbdd de core de la entidad. 3. Transformación de Datos. 4. Análisis numérico y Estadístico, específicamente: XGBoost y Bosques Aleatorios. 5. Algoritmos de Aprendizaje Automático Supervisado

CAPITULO V	
Conclusiones	
5.1.1	A través de la investigación realizada en MAPFRE Seguros Honduras, donde han sido utilizados 6,332,400 registros, se puede concluir que el análisis de datos facilita la identificación de mejoras en la información disponible en la organización, y favorece la toma de decisiones estratégicas.
5.1.2	Concluida la investigación es reseñable las habilidades otorgadas por la maestría de Analítica Negocios, destacando entre ellas, el pensamiento crítico para ver el problema de diferentes perspectivas, la capacidad de preparar los datos para un análisis efectivo y la capacidad de aplicar la matemática y las estadísticas de forma adecuada para resolver un problema de la vida real.
5.1.3	En cumplimiento al primer objetivo específico, el cual es realizar el análisis de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2020-2023, ha sido alcanzado a través de técnicas de analítica de datos, las cuales han permitido concluir que las características más importantes de los registros de asegurados, son captadas por la aseguradora a través de sus procesos operativos core y permiten destacar las particularidades más significativas de la cartera, siendo estas: La mayor densidad (70%) de la cartera de pólizas de auto es masculina, no se presentan tantos siniestros como se estimaba, al contrario únicamente el 11% de la cartera presenta reclamos por siniestros, el valor medio de un vehículo asegurado asciende a 400 mil lempiras, el valor pagado por asegurado por la póliza de auto, que más se repite es por L20,000.00 lempiras.
5.1.4	En respuesta al objetivo específico número dos, el cual se basa en la identificación de un marco referencial de variables para la detección de los asegurados candidatos a venta cruzada, ha sido alcanzado luego de un proceso exhaustivo de transformación de bases de datos, a través del cual se ha diseñado un diccionario de 16 variables tipo demográficas y 25 variables tipo conductuales. A partir del diccionario referenciado de variables, fue posible la diagramación de la intensidad de correlación de los predictores con la variable a predecir: “venta cruzada”, dando como resultado en primer lugar el tiempo desde la primera emisión de la póliza de autos (0.4), en segundo el número de renovaciones (-0.28) que ha sufrido y en tercer lugar canal de venta (-0.14) a través del cual se colocó el producto.
5.1.5	El tercer objetivo específico, el cual es considerado como la cúspide de la investigación, buscaba diseñar un modelo numérico capaz de predecir candidatos a adquirir un seguro adicional a la póliza de auto, fue culminado a través de la selección y entrenamiento de los algoritmos XGBoost y Bosques aleatorios, los cuales superaron de forma satisfactoria el nivel de exactitud exigido en la hipótesis (70%), medido a través de la matriz de confusión y con porcentajes resultantes de 90.20% y 89.23% respectivamente. La capacidad de predicción y confiabilidad de los modelos fue reconfirmado a través de la curva ROC que mostro un resultado de 83.40% para el algoritmo XGBoost y de 85% para bosques aleatorios.

5.1.6 El entrenamiento del algoritmo XGBoost y su aplicación a la base de datos con vigencia menor a un (1) año, ha permitido pronosticar 447 candidatos a venta cruzada y proyectar un monto de colocación mayor a 1 millón de lempiras en ventas.

CAPITULO VI	
Nombre de la propuesta	Objetivos de la propuesta
Modelo predictivo de venta cruzada en productos de auto y vida en MAPFRE Seguros Honduras.	Objetivo general
	Implementar un modelo predictivo enfocado en la detección de asegurados candidatos a adquirir un seguro de vida, como producto adicional a la póliza de autos previamente obtenida, que influya en catapultar la técnica de venta cruzada, logrando la fidelización de clientes y explotación máxima de utilidad de cartera, en MAPFRE Seguros Honduras.
	Objetivos específicos
	1. Realizar el análisis y transformación de datos de segmentación registrado por MAPFRE Seguros referente a los asegurados de auto y vida durante el periodo 2020-2023.
	2. Predecir a través de un modelo numérico entrenado, los candidatos a adquirir un seguro adicional a la póliza de autos en MAPFRE Seguros Honduras.

BIBLIOGRAFIA

- Alberto, A. L. L. (2013). *Desarrollo de un modelo de recomendación de compra para clientes de una empresa de seguros*. <https://repositorio.uchile.cl/handle/2250/114064>
- American Marketing Association. (s. f.). Definición de Venta. En *marketing-dictionary.org*. Recuperado 28 de julio de 2023, de <https://marketing-dictionary.org/s/sales/>
- Anaya Isaza, A. (2017). Estudio comparativo de métodos espectrales para reducción de la dimensionalidad: LDA versus PCA. *INCISCOS*.
- Campo Leon, E. (2017). *Introducción a las máquinas de vector soporte (SVM) en aprendizaje supervisado* [Trabajo de Fin de Grado en Matemáticas]. Universidad de Zaragoza.
- Carlos, O. (2002). *Enfoques sobre la tecnología*. *Revista Iboamericana de Ciencia, Tecnología, Sociedad e Innovación*.
- CNBS. (2021). Perfil del Sistema Asegurador de Honduras. *Análisis del Desarrollo del Sector y Determinantes del Mercado*.
- Congreso Nacional de Honduras. (1950). *Código de Comercio* (Artículo 1127).
- Correa Rodriguez, D. (s. f.). El Contrato de Seguro. *Facultad de Derecho:1, 1*.
- DUPONT, V. (2019). La inteligencia artificial irrumpe en el mundo de las finanzas. *AFP*. <https://www.proquest.com/wire-feeds/la-inteligencia-artificial-irrumpe-en-el-mundo-de/docview/2321167435/se-2>
- Espinosa-Zúñiga, J. J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. *Ingeniería Investigación y Tecnología*, 21(3), 1-16. <https://doi.org/10.22201/fi.25940732e.2020.21.3.022>

- Fiuza Pérez, M. D. (2000). La regresión logística: una herramienta versátil. *Nefrología*, 6.
<https://www.revistanefrologia.com/es-la-regresion-logistica-una-herramienta-articulo-X0211699500035664>
- FRISS. (2022). *Insurance fraud report ES*. <https://www.friss.com/>.
<https://www.friss.com/es/insight/encuesta-de-fraude-en-seguros-2022/?submissionGuid=927f3d69-9069-4d71-818c-dbe48f424f89>
- Galindo, E. M. (31 de Octubre de 2016). METODOLOGIA DE LA INVESTIGACION, PAUTA PARA HACER TESIS. Obtenido de t.ly/RwH3
- Garita Araya, Y. (2021). Informe sectorial de seguros Centroamérica. *SCRIESGO RATING AGENCY, I*.
- Garre, M. (2007). Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software. *REICIS. Revista Española de Innovación, Calidad e Ingeniería del Software*, 3(1), 6-22.
- Gobierno de Honduras. (1994). Ley de Impuesto sobre Venta. En *Diario Oficial La Gaceta Número 27486*.
- Grasty Villanueva, T. (2014). *OPTIMIZACIÓN DE VENTAS CRUZADAS MEDIANTE INTELIGENCIA DE NEGOCIOS* [MAGÍSTER EN INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN]. Universidad de Chile.
- Guillén, S. (2022, 17 marzo). *¿Qué es una query? Definición y significado en SEO*.
<https://www.inboundcycle.com/blog-de-inbound-marketing/que-significa-query-y-porque-es-necesario-entender-este-concepto-seo>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). The elements of statistical learning. En *Springer series in statistics*. <https://doi.org/10.1007/978-0-387-21606-5>

Historia de MAPFRE. (s. f.).

<https://www.mapfre.com/statics/corporativo/grupomapfre/es/cinformativo/historia-grupomapfre.shtml>

Historia de MAPFRE Honduras - Seguros MAPFRE. (2022, 30 septiembre). MAPFRE

Honduras. <https://www.mapfre.com.hn/sobre-mapfre-honduras/historia/>

Hojas, I. M. (2021, 17 enero). Regresión logística en Python. *StatDeveloper*.

<https://www.statdeveloper.com/regresion-logistica-en-python/>

Informe de Fraude en seguros 2022 - FRISS. (s. f.). <https://www.friss.com/es/descargas/informe-de-fraude-en-seguros-2022>

Jones, E., Chonko, L., Jones, F., & Stevens, C. (2012). *Selling ASAP: Art, Science, Agility, Performance* (1.a ed.). Louisiana State University Press.

Kotler, P. (2003). *Dirección de Marketing: Conceptos Esenciales*. Pearson Educación.

Lalinde, J. D. H., Castro, J. F. E., Tarazona, M. E. P., González, J. E. F., Rangel, J. C., Sierra, C.

J. R., Torrado, M. K. A., Sierra, S. M. C., & Pirela, V. B. (2017). Sobre El Uso Adecuado Del Coeficiente De Correlación De Pearson: Definición, Propiedades Y Suposiciones.

Archivos Venezolanos de Farmacología y Terapéutica, 37(5), 587-595.

<https://bonga.unisimon.edu.co/handle/20.500.12442/246>

MARTÍNEZ, R. E. B. (2009). Árboles de decisión como herramienta en el diagnóstico médico.

Revista médica de la Universidad Veracruzana, 9, 19-24.

Misión de MAPFRE Honduras - Seguros MAPFRE. (2022, 27 septiembre). MAPFRE Honduras.

<https://www.mapfre.com.hn/sobre-mapfre-honduras/mision/>

Moral Pelaez, I. (2016). Modelos de regresión: lineal simple y regresión logística. *Revista Seden*, 14, 195-214.

- Ocaris, P. R. F. (s. f.). *LAS REDES NEURONALES y LA EVALUACIÓN DEL RIESGO DE CRÉDITO*. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-33242007000100007&lng=en&tlng=es.
- Ortega, C. (2023). Matriz de correlación: qué es, cómo funciona y ejemplos. *QuestionPro*.
<https://www.questionpro.com/blog/es/matriz-de-correlacion/>
- Pertuz, C. M. P. (2022). *Aprendizaje automático y profundo en Python*. Ra-Ma Editorial.
- ¿Qué es machine learning? | IBM. (s. f.). <https://www.ibm.com/es-es/topics/machine-learning>
- PINEDA PERTUZ, C. (2021). *Aprendizaje Automático y Profundo en Python - Una mirada hacia la inteligencia artificial* (1.a ed.). EDICIONES DE LA U LTDA.
- Ramirez-Alpizar, A., Jenkins, M., & Martinez, A. (2020). Uso de técnicas de minería de datos y aprendizaje. *Revista Ibérica De Sistemas e Tecnologías De Informação*, 97-109.
<https://www.proquest.com/scholarly-journals/uso-de-técnicas-minería-datos-y-aprendizaje/docview/2388305034/se-2>
- Republica de Honduras. (2001). Ley-de-Instituciones-de-Seguros-y-Reaseguros. En *La Gaceta*.
- Rodriguez, L. (2020). ¿Cuántas personas tienen seguros de vida y de vehículos? *elheraldo.hn*.
- Sanchez Sardaña, M. (2022). *Modelo predictivo de venta cruzada en productos de Vida y Salud* [Trabajo Fin de Master]. Universidad Carlos III de Madrid.
- Sangri Coral, A. (2014). *Introducción a la Mercadotecnia*. Grupo Editorial Patria.
- Sanz, J. B., & Vila, B. R. (2023). *Manual práctico de inteligencia artificial en entornos sanitarios*. Elsevier Health Sciences.
- Seguro de Vida- Conceptos Básicos - Office of Public Insurance Counsel*. (2022, 17 octubre). Office of Public Insurance Counsel. <https://www.opic.texas.gov/es/seguro-de-vida/conceptos-basicos/>
- Soledad, C. F. (2014). *Técnicas de venta*. Ediciones Paraninfo, S.A.

Suanya Sag, O. (2013). *El cliente social*. Profit Editorial.

THOMPSON, I. (2007). La segmentación del mercado. *Promonegocios.*, 5, 25-35.

Tocabens, B. E. (2011). Definiciones acerca del riesgo y sus implicaciones. *Revista Cubana de Higiene y Epidemiología*, 49(3), 470-481. <http://scielo.sld.cu/pdf/hie/v49n3/hie14311.pdf>

UDRC. (s. f.). https://udrc.io/news/2021/research_skills/20210407AucRocCurve.html

Unlocking the potential of deeper customer-agent-insurer relationships. (2021, 29 noviembre).

McKinsey & Company. <https://www.mckinsey.com/industries/financial-services/our-insights/unlocking-the-potential-of-deeper-customer-agent-insurer-relationships#/>

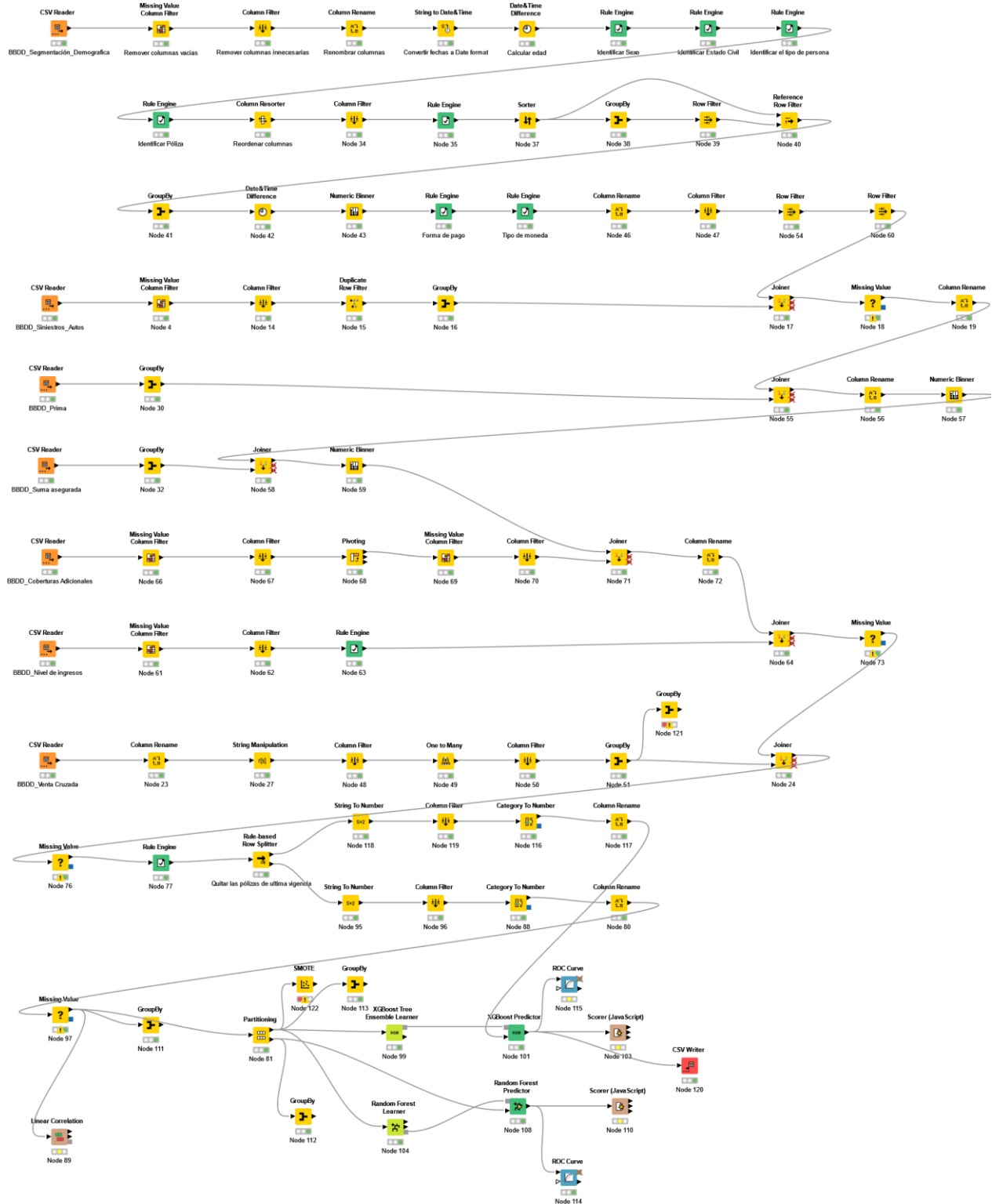
Valero, E. B., Díaz, A. F., & Gisbert, J. S. (2019). Algoritmos de machine learning para la detección del fraude en el seguro de automóviles. *Anales del Instituto de Actuarios Españoles*, 26, 23-46. <https://dialnet.unirioja.es/servlet/articulo?codigo=7749889>

venta cruzada - Fundación MAPFRE. (2020). En *Fundación MAPFRE*.

<https://www.fundacionmapfre.org/publicaciones/diccionario-mapfre-seguros/venta-cruzada/#:~:text=Diccionario%20de%20seguros%20venta%20cruzada%20%28cross%20selling%29%20Ventas,t%C3%A9cnica%20tambi%C3%A9n%20se%20persigue%20la%20fidelizaci%C3%B3n%20del%20cliente.>

ANEXOS

ANEXO 1: FLUJO DE TRABAJO PREDICCIÓN DE VENTA CRUZADA



ANEXO 2: CARTA DE AUTORIZACIÓN DE LA EMPRESA

CARTA DE AUTORIZACIÓN DE LA EMPRESA O INSTITUCIÓN

Tegucigalpa, Fco. Morazán, 29/09/23
(Ciudad), (Departamento) (Día, mes y año)

Elsy Karina Silva Valdez
(Nombre y apellidos del Director o Gerente)

Directora de Recursos Humanos
(Puesto Laboral)

MAPFRE Seguros Honduras
(Empresa o Institución)

Torre Morazán, piso 18.
(Dirección principal de la empresa o institución)

Estimado Señor(a): Elsy Karina Silva

Reciba un cordial y atento saludo. Por medio de la presente deseamos solicitar su apoyo, dado que somos alumnos de UNITEC y nos encontramos desarrollando el Trabajo de Tesis previo a obtener nuestro título de maestría en Análisis de Negocios

Hemos seleccionado como tema Modelo Predictivo de Venta Cruzada de Autos y Vida., por lo que estaríamos muy agradecidos de contar con el apoyo de la empresa que usted representa para poder desarrollar nuestra investigación. En particular, dicha solicitud se circunscribe a peticionar que se nos autorice a realizar: Inteligencia y Análisis de datos de pólizas de auto y vida, de forma generalizada y despersonalizada.

(encuestas, sondeos, etc).

A la espera de su aprobación, me suscribo de Usted.

Atentamente,

Roberto Moza
Firma, nombre y apellidos

No. de cuenta: 12213154

Firma, nombre y apellidos

No. de cuenta: _____

Por este medio, _____
(empresa / institución).

Autoriza la realización dentro de sus instalaciones el proyecto de investigación de Tesis de Postgrado antes mencionado.

Elsy Karina Silva
(Nombre y sello del Director / Gerente)

