



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

TEMA:

**Gestión de riesgos en negocios internacionales: una perspectiva avanzada
utilizando business intelligence y machine learning**

AUTORES:

León Negrete, Victoria Elizabeth

Solano Jaramillo, Scarlet Michelle

**Trabajo de integración curricular previo a la obtención del título de
LICENCIADA EN NEGOCIOS INTERNACIONALES**

TUTOR:

Ing. Carrera Buri, Félix Miguel Mgs.

Guayaquil, Ecuador

23 de agosto del 2024



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación fue realizado en su totalidad por **León Negrete, Victoria Elizabeth y Solano Jaramillo, Scarlet Michelle**, como requerimiento para la obtención del título de **LICENCIADA EN NEGOCIOS INTERNACIONALES**.

Guayaquil, a los 23 días del mes de agosto del año 2024

f. _____

Ing. Carrera Buri, Félix Miguel Mgs.

TUTOR

f. _____

Ing. Hurtado Cevallos, Gabriela Elizabeth

DECANO O DIRECTOR DE CARRERA



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, León Negrete, Victoria Elizabeth

DECLARO QUE:

El Trabajo de Integración Curricular, **Gestión de Riesgos en Negocios Internacionales: Una Perspectiva Avanzada Utilizando Business Intelligence Y Machine Learning**, previo a la obtención del título de Licenciadas en Negocios Internacionales, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 23 días del mes de agosto del año 2024

EL AUTOR:

f. _____

León Negrete, Victoria Elizabeth



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, Solano Jaramillo, Scarlet Michelle

DECLARO QUE:

El Trabajo de Integración Curricular, **Gestión de Riesgos en Negocios Internacionales: Una Perspectiva Avanzada Utilizando Business Intelligence Y Machine Learning**, previo a la obtención del título de Licenciadas en Negocios Internacionales, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 23 días del mes de agosto del año 2024

EL AUTOR:

f. _____

Solano Jaramillo, Scarlet Michelle



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

AUTORIZACIÓN

Yo, León Negrete, Victoria Elizabeth

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Gestión de Riesgos en Negocios Internacionales: Una Perspectiva Avanzada Utilizando Business Intelligence Y Machine Learning**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a 23 días los del mes de agosto del año 2024

EL AUTOR:

f. _____

León Negrete, Victoria Elizabeth



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

AUTORIZACIÓN

Yo, Solano Jaramillo, Scarlet Michelle

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Gestión de Riesgos en Negocios Internacionales: Una Perspectiva Avanzada Utilizando Business Intelligence Y Machine Learning**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a 23 días los del mes de agosto del año 2024

EL AUTOR:

f. _____

Solano Jaramillo, Scarlet Michelle

REPORTE COMPILATIO



Ing. Carrera Buri, Félix Miguel Mgs.

AGRADECIMIENTOS

Agradezco principalmente a Dios por permitirme cumplir con éxito esta etapa en la cual he experimentado muchos altibajos, pero con su ayuda las he logrado superar con éxito. A mis padres,

Narcisa y Arturo, por ser mi guía durante toda mi vida, acompañar mis caídas y celebrar mis triunfos; sin ustedes, esto no habría sido posible. A mi hermano Alex, que ha sido mi soporte en los momentos donde sentía que no daba más y siempre estuvo ahí con sus palabras de aliento, recordándome que siempre puedo llegar a más.

También a aquella familia que me brindo la universidad: Scarlet, Rose, Juan, Estefano, Sebastián, Adriana y Mata. Con ustedes, todo este largo trayecto fue mucho más llevadero entre risas, enojos y lágrimas. Gracias por cada momento compartido, por cada momento donde me brindaron su apoyo y amor para poder culminar esta etapa juntos.

Quiero agradecer de manera especial a Scarlet y Rose que se convirtieron en mis hermanas a lo largo de estos 4 años los cuales espero que se multipliquen en muchos más. Han sido personas que han iluminado mi vida siendo ese pedazo de calma dentro de la tormenta, con quienes he compartido mis mejores risas y mis peores llantos. Gracias por nunca dejarme caer.

Adicional, agradezco a Romina por ser esa persona que nunca dejo que desconfiara de mí, que siempre me brindo su tiempo y apoyo, por no dejarme derrumbar en tiempos de angustias, su amistad significa mucho para mí.

Así también quiero agradecer a Ivan Ponce, una persona que fue pieza fundamental dentro de todo este proyecto, que nos apoyó de inicio a fin y nos daba ánimos cuando sentíamos que todo se nublaba, que siempre llegaba con un “ustedes pueden con todo” y compartió sus conocimientos de la manera más grata.

- VICTORIA ELIZABETH LEON NEGRETE

DEDICATORIA

Dedico este proyecto a Dios que me ha dado la sabiduría necesaria para poder llevarlo a cabo. Así también a mi familia la cual ha sido pieza fundamental dentro de toda esta trayectoria. Por último, pero no menos importante, la dedico a mí misma la cual con propósito, resiliencia he logrado culminar todo.

- VICTORIA ELIZABETH LEON NEGRETE

AGRADECIMIENTOS

En primer lugar, quiero agradecer a Dios por su guía y fortaleza. A mis padres, que han sido mi mayor sustento y apoyo en todo momento: a mi papá, Miguel Solano, por brindarme la oportunidad de estudiar esta carrera y por su orientación en los momentos de incertidumbre, siempre recordando que la vida se vive día a día y que, con paciencia, todo se irá acomodando. Su habilidad para transmitir calma cuando más lo necesitaba ha sido invaluable. A mi mamá, Angela Jaramillo, por consolarme en los momentos de lágrimas y estrés, por saber exactamente qué decir y cuándo ofrecerme ese abrazo reconfortante que me ayudaba a recomponerme. Has sido mi pilar emocional durante todo este trayecto. Y a mi abuelita Rosa, por sus palabras llenas de sabiduría y amor, recordando siempre que tengo la fortaleza para enfrentar todo lo que la vida me presente. A mis 3 hermanos Sebas, Ashley y Kelly por hacer el proceso más ligero con sus bromas y risas, contribuyendo a que este camino fuera más ameno y lleno de alegría.

Agradezco a los “main characters” (Juan, Adri, Estefano, Sebas, Mata) de estos cuatro años, quienes hicieron el camino más llevadero y agradable. Gracias por explicarme las cosas cuando no entendía, por estar presentes tanto en los días buenos como en los no tan buenos, y por demostrarme que se pueden construir amistades genuinas en la carrera, basadas en la generosidad y el apoyo mutuo.

Quiero hacer un agradecimiento especial a mis dos mejores amigas, Victoria y Rosa Emilia, quienes se volvieron familia para mí. Quiero agradecer profundamente a Victoria, no solo por permitirme estar a su lado en uno de los proyectos más importantes de nuestras vidas, sino también por la maravillosa amistad que me ha brindado. Agradezco su apoyo incondicional, especialmente en los momentos en que más lo necesitaba, con palabras precisas que supieron levantarme. Estoy segura de que serás una de las mejores profesionales.

Y a Rosa Emilia, no puedo dejar de agradecerle por ser ese ser de luz que irradia un amor genuino desde el inicio de nuestra amistad. Ha estado presente en cada etapa de este camino. Su apoyo inquebrantable me ha demostrado que juntas podemos superar cualquier obstáculo y alcanzar todo lo que nos proponemos. Ambas llegaron a mi vida en esta nueva etapa, y espero que nuestra amistad perdure a lo largo de todas las etapas futuras. Finalmente, quiero agradecer a Iván Ponce por su paciencia infinita al explicar conceptos complejos, siempre con la habilidad de hacerlo de manera clara y sin que me sintiera incómoda. Su apoyo y comprensión han sido esenciales para mi proceso de aprendizaje. Además, deseo expresar mi profundo agradecimiento a Félix Carrera, quien ha sido un mentor ejemplar y un profesor inspirador. Su pasión por la enseñanza y su dedicación han sido una fuente constante de motivación para seguir aprendiendo y superando mis propios límites.

- SCARLET MICHELLE SOLANO JARAMILLO

DEDICATORIA

Dedico esta tesis a mí misma, por mi capacidad, resiliencia y por la confianza que he mantenido en mí misma a lo largo de este proceso. A mis padres, quienes han estado a mi lado en todo momento, brindándome su apoyo incondicional. Nunca me quejaré de ser una "niña de mami y papi", pues esto solo subraya lo afortunada que soy por tenerlos a mi lado. Su amor y respaldo han sido mi mayor fortaleza. Y, por último, pero no menos importante, a mi hermano Sebas, quien es el niño de mis ojos desde antes que llegara a este mundo y quien se ha convertido en uno de los apoyos más genuinos que tengo en casa.

Siempre cuidándome con un abrazo o animándome con mi comida favorita.

- SCARLET MICHELLE SOLANO JARAMILLO



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

TRIBUNAL DE SUSTENTACIÓN

f. _____

Ing. Carrera Buri, Félix Miguel Mgs.

TUTOR

f. _____

Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.

DECANO O DIRECTOR DE CARRERA

f. _____

Mgs. Verónica Correa

COORDINADOR DEL ÁREA O DOCENTE DE LA CARRERA



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

CALIFICACIÓN

León Negrete, Victoria Elizabeth

Solano Jaramillo, Scarlet Michelle

Contenido:

Introducción	2
Problemática	8
Justificación	11
Alcance	16
Objetivo General	18
Objetivo Específico	18
Capitulo 1	19
Marco Teórico	19
Valor en Riesgo	19
Pros y contras del VaR	20
Valor en Riesgo Condicional (CVaR)	22
Introducción a las Máquinas de Vectores de Soporte	24
Riesgo Empírico vs. Riesgo Estructural	24
Bases de SVM	26
Operaciones Básicas en SVM	26
Optimización para la Separación Óptima	28
Resolviendo el Problema de Optimización	28
Vectores de Soporte y Clasificación	29
Hiperplanos Óptimos en la Clasificación con SVM	30
Vectores de soporte	31
Hiperplano de Margen Suave	32

Modelos no linealmente separables	34
Transformación a Espacios de Mayor Dimensionalidad	35
Condición de Mercer	37
Problemas No Linealmente Separables	39
Restricciones y Variables de Holgura.....	39
Objetivos de Optimización.....	40
Problema Primal y Dual de las Máquinas de Vectores de Soporte	41
Métodos para Clasificación Multiclase con SVM.....	43
Problema de Aprendizaje Supervisado.....	45
Problema de Aprendizaje No Supervisado	46
Marco conceptual	47
Riesgo	47
Gestión de Riesgos	48
Negocios Internacionales	50
Toma de decisiones basada en riesgos.....	51
Machine learning	52
Machine Learning aplicado en los negocios	54
Aprendizaje Supervisado	54
Análisis No Supervisado	55
Aprendizaje por Refuerzo.....	57
SVM (Máquinas de vectores de soporte)	57
Máquina de vectores de soporte para clasificación binaria	58
Parámetro de penalización ©.....	60
Marco Legal	61

CAPITULO 2.....	62
Metodología	62
Discusión	97
Conclusión.....	100
Referencias.....	102
Anexos	105

Índice de Figuras

Figura 1: Porcentaje más frecuente en seguros.....	82
Figura 2: Histograma de Edad	83
Figura 3: Histograma de Valor de Póliza.....	84
Figura 4: Valor de Póliza Anual según Tipo de Cobertura.....	86
Figura 5: Edad por Tipo de Cobertura	87
Figura 6: Valor de Póliza Anual por Edad.....	89
Figura 7: Tipo de Cobertura según Número de Beneficiarios	90
Figura 8: Valor de Póliza Anual por Edad.....	91
Figura 9: Cobertura por Edad	92
Figura 10: Cobertura por Valor de Póliza Anual	93
Figura 11: Número de Beneficiarios por Edad	94
Figura 12: Número de Beneficiarios por Cobertura	94

Resumen

Este estudio se centra en la gestión de riesgos en los negocios internacionales, un tema de creciente relevancia en un mundo globalizado. A medida que las empresas expanden sus operaciones más allá de las fronteras nacionales, enfrentan desafíos como la volatilidad de los mercados, las diferencias culturales y las variaciones en los entornos políticos y económicos. El objetivo principal de esta investigación es desarrollar un sistema de clasificación automática que optimice la identificación, evaluación y mitigación de riesgos en este contexto.

Para alcanzar este objetivo, se realizó una revisión exhaustiva de la literatura, estableciendo un marco teórico sólido sobre la intersección entre la tecnología y la gestión de riesgos. Además, se implementó una metodología basada en técnicas de aprendizaje automático, específicamente en clasificación supervisada, demostrando su efectividad en la mejora de la toma de decisiones y la gestión de datos.

Los resultados indican que la integración de herramientas tecnológicas en la gestión de riesgos no solo mejora la eficiencia operativa de las empresas, sino que también les permite adaptarse mejor a un entorno global complejo.

Abstract

This study focuses on risk management in international business, a topic of increasing relevance in a globalized world. As companies expand their operations beyond national borders, they face challenges such as market volatility, cultural differences, and variations in political and economic environments. The primary objective of this research is to develop an automated classification system that optimizes the identification, assessment, and mitigation of risks in this context.

To achieve this goal, a comprehensive literature review was conducted, establishing a solid theoretical framework on the intersection of technology and risk management. Additionally, a methodology based on machine learning techniques, specifically supervised classification, was implemented, demonstrating its effectiveness in improving decision-making and data management.

The results indicate that integrating technological tools into risk management not only enhances the operational efficiency of companies but also enables them to better adapt to a complex global environment.

Introducción

El dinamismo constante del mundo moderno encuentra una de sus manifestaciones más intensas en el ámbito de los negocios internacionales. En la era actual de la globalización, la competencia global ha alcanzado niveles sin precedentes, obligando a las empresas a adaptarse rápidamente a un entorno en perpetuo cambio. Este fenómeno no sólo redefine las estrategias empresariales, sino que también exige una comprensión profunda y multidisciplinaria de varios factores que influyen en el comercio global (Bohórquez & Enrique, 2011).

A nivel macroentorno, la economía global enfrenta desafíos como: las fluctuaciones en los mercados financieros internacionales, las tensiones comerciales entre grandes potencias económicas y las crisis sanitarias globales. La volatilidad de los mercados financieros ha desplazado a factores como la modificación de los precios de las materias primas y el comportamiento cíclico de la economía real como principales preocupaciones. Esta volatilidad es en parte inevitable, debido a la naturaleza de las transacciones de capital y las imperfecciones del mercado. El nivel de riesgo financiero en un mundo con creciente libertad para los movimientos de capital está altamente condicionado por las políticas económicas aplicadas por los Estados. El proceso de propiciar la coordinación macroeconómica, especialmente entre los países con mayor peso en los mercados globales, se enfrenta a las diversas necesidades que deben afrontar las economías en distintos momentos.

En el ámbito continental, específicamente en América Latina, las micro, pequeñas y medianas empresas (mipymes) enfrentan varios desafíos clave. Estas empresas suelen tener acceso limitado a fuentes de financiamiento y a menudo

desconocen otras alternativas, lo que resulta en restricciones debido a las asimetrías de información en los mercados financieros. Estudios en la región han mostrado que un alto porcentaje de organizaciones tienen problemas de endeudamiento, reflejando una mala administración financiera. Además, las condiciones financieras inestables afectan negativamente la toma de decisiones y el control de información en las pymes (Luis, 2008). Muchas pymes también carecen de una base adecuada, lo que limita su capacidad de adaptación y crecimiento en un entorno competitivo.

En el caso específico de Ecuador, el entorno político y económico del país juega un papel crucial. Las variaciones políticas, como cambios de gobierno o modificaciones en las políticas económicas y fiscales, pueden generar amenazas significativas para las entidades que operan en el país. La estabilidad política y económica de Ecuador es fundamental para atraer inversiones extranjeras y fomentar el comercio internacional. Factores como la dolarización de la economía ecuatoriana, las políticas comerciales del gobierno y la infraestructura local impactan directamente en la competitividad de las compañías en el mercado global (Negash & Gray, 2008).

El estudio de los negocios internacionales es, por tanto, una disciplina compleja que abarca una vasta gama de conocimientos. Es imperativo considerar aspectos económicos, que dictan las reglas del mercado y el flujo de capitales; factores sociales y culturales, que afectan la conducta del consumidor y las prácticas empresariales; así como variables políticas y legales, que imponen regulaciones y políticas comerciales. Además, los aspectos éticos y administrativos juegan un papel crucial en la toma de decisiones corporativas, mientras que los contextos históricos y culturales, incluyendo la religión y las costumbres locales, moldean las interacciones y la diplomacia (*Riesgos Financieros y Operaciones Internacionales*, s. f.).

En relación con esto, el entorno político se convierte en nuestro eje central, en el cual las variaciones de los países donde se planea iniciar operaciones pueden generar vulnerabilidad significativa para la empresa. Esta vulnerabilidad incluye cambios de gobierno que traen consigo ideologías políticas que no favorecen el libre comercio, así como marcos jurídicos específicos de esos países que pueden impedir ciertos procedimientos aceptados en el país de origen de la empresa. Analizar esta inseguridad política es fundamental al momento de tomar decisiones sobre posibles operaciones en cualquier país (Ruiz, 2023).

En esta situación, la gestión de riesgos ha evolucionado de un enfoque empírico a un marco integral de evaluación y mitigación, con el control interno como componente esencial para el cumplimiento de objetivos en un entorno global. Las instituciones que deseen prosperar en este entorno globalizado deben ser ágiles, adaptables, innovadoras y capaces de navegar por la complejidad y la incertidumbre de operar a nivel mundial. Al comprender y medir los peligros potenciales, las mencionadas instituciones pueden tomar decisiones informadas que minimicen el impacto negativo y maximicen las oportunidades. (Maricela & María, 2019.).

La gestión del riesgo pretende centrar todos sus esfuerzos en aquellos elementos generadores de condiciones de vulnerabilidad que exponen a las comunidades a circunstancias arriesgadas y que de no intervenir efectivamente sobre ellas, podrían, consiguientemente, conducir a desastres; esta se articula entonces, a evaluaciones preventivo. Bajo este aspecto, la importancia de afrontar desafíos que se enfrentan para cumplir las metas se difunde cada día más a nivel mundial, como propuesta disciplinada y estructurada que alinea las estrategias, los procesos, las personas, las tecnologías y los

conocimientos, y se sitúa en el entorno de los sistemas de control interno de las organizaciones (Maricela & María, 2019.).

Sin embargo, cabe señalar que la gestión de riesgos y el control interno no siempre han sido percibidos de esta manera, pues han cambiado a medida que se han transformado el entorno y las estructuras organizacionales. El control interno, conocido desde tiempos remotos, ha evolucionado significativamente. Inicialmente basado en acciones empíricas, se formalizó tras la Revolución Industrial, cuando el aumento de la complejidad empresarial y la división de tareas llevaron a la implementación de mecanismos internos para proteger activos y detectar desfalcos.

En Ecuador, el control interno ha seguido estas evoluciones, adaptándose a cambios legales y normativos desde la Ley Orgánica de Administración Financiera y Control de 1977, y adoptando las Normas Técnicas de Control Interno en línea con COSO desde 2002. Por lo que la gestión de riesgos se considera ahora como un componente esencial del control interno, crucial para asegurar el cumplimiento de los objetivos organizacionales en un entorno global y tecnológico.

Es así que en la era digital actual, la búsqueda de “nuevos mercados” en un contexto de globalización aumenta la necesidad de evaluar el riesgo país y la amenaza de insolvencia en operaciones de moneda extranjera; dado esto la interconexión entre mercados así como rápida circulación de comunicación han incrementado la exposición a riesgos financieros, operativos y geopolíticos, en consecuencia, las empresas deben estar preparadas para enfrentar estos desafíos y adaptarse en el menor tiempo posible.

Ante esta realidad, el Business Intelligence (BI) emerge como concepto clave para la gestión de riesgos. Al hacer uso de estas, las organizaciones pueden tomar decisiones informadas y estratégicas, mitigando riesgos y capitalizando oportunidades en un mercado global altamente competitivo y en constante cambio. Aunque existen diversas definiciones de BI, todas comparten el enfoque en el análisis de datos e información (Lönnqvist & Pirttimäki, 2006).

Por su parte, el BI permite a las organizaciones recopilar, analizar y visualizar grandes volúmenes de datos de manera eficiente, ofreciendo información valiosa sobre las operaciones comerciales y los mercados globales. Su capacidad se basa en la recopilación de datos precisos y relevantes, que luego se analizan para ofrecer una visión integral de la empresa y su entorno. Este proceso continuo de análisis y evaluación abarca todos los aspectos que influyen en una organización en su mercado, considerando factores tanto internos como externos. Donde el mencionado conocimiento soporta decisiones críticas y de alto perfil en una empresa.

Una forma de visualizar el BI es imaginando una pirámide; donde cada lado de la pirámide representa un aspecto de la “inteligencia”: Información, comunicación y recurso humano. Cada uno de estos elementos se desarrolla sobre capacidades únicas de cada empresa; entre mayor evolución exista, mayor será el grado de eficiencia y eficacia. Por lo tanto, la inteligencia de mercados es la integración de estos tres puntos en su máximo grado de desarrollo.

La inteligencia de negocios o Business Intelligence (BI) es un concepto general que incluye las aplicaciones, la infraestructura, las herramientas y las mejores prácticas que permiten obtener acceso y análisis de la información para mejorar y optimizar tanto

el rendimiento como la toma de decisiones en la organización (Ahmed, Aziz, Tezel, & Riaz, 2018). BI está estrechamente relacionado con otras áreas que le proveen la información necesaria para cumplir a cabalidad su propósito.

La relevancia de tener una gran cantidad de datos ha sido reconocida por empresas, gobiernos e inclusive por el mundo académico. Entre más fuentes de datos tenga una empresa, más claros serán sus modelos de decisión y visualización de su entorno. La inteligencia empresarial depende críticamente de la información, expresada en forma de métricas analíticas y opiniones. Podemos desarrollar conocimiento a partir de prácticamente cualquier punto en que la empresa interactúe con el mundo exterior y su propio universo (Salvador & De Barcelona Departament D'Econometria Estadística I Economia Aplicada, 2020).

El Business Intelligence (BI) proporciona a las organizaciones una visión integral y en tiempo real de sus operaciones globales. Esto incluye la capacidad de monitorear indicadores clave de rendimiento (KPI) a nivel internacional, identificar áreas de vulnerabilidad y oportunidades, y ajustar estrategias comerciales de manera ágil y eficiente. La combinación de BI y Machine Learning (ML) permite a las empresas no sólo reaccionar ante los cambios del mercado, sino también anticiparse a ellos, mejorando su capacidad de respuesta y resiliencia en un entorno cada vez más complejo y dinámico (Negash & Gray, 2008).

Problemática

La evolución constante de los mercados globales, junto con la inconstancia de los factores económicos y políticos, demandan que las empresas adopten un enfoque más sofisticado y proactivo para enfrentar el mundo laboral. La gestión de riesgos se encuentra ante desafíos significativos derivados de la falta de datos confiables y la predominancia de juicios valorativos sobre información contrastada, además de otras dificultades inherentes a esta gestión.

Un aspecto crítico es la incapacidad de las empresas para reconocer, evaluar y reducir adecuadamente estos desafíos que surgen en el contexto de los negocios internacionales. Esta incapacidad se manifiesta en diversos aspectos. Primero, es considerablemente complicado reconocer y analizar los desafíos en mercados que son intrincados y cambiantes. Lo que genera una falta de visibilidad y comprensión integral de estos problemas y complica la realización de decisiones oportunas y estratégicas. Por otro lado, las ineficiencias en la gestión de la cadena de suministro global representan otro desafío. La falta de herramientas para monitorear y optimizar esta cadena puede resultar en retrasos en entregas, problemas de calidad e incumplimientos normativos, afectando directamente la operación y reputación de la empresa (Del Carmen Becerra et al., 2021).

La incapacidad para detectar fraudes es otro aspecto crítico. La complejidad de las transacciones internacionales aumenta el riesgo de fraudes el cual puede dañar la confianza de los stakeholders. Las dificultades para cumplir con regulaciones internacionales también son notables. Las normativas complejas y cambiantes en

diferentes países representan un reto constante para las empresas, por lo que incrementa el riesgo de sanciones y problemas legales (Antonio & Pilar, 2011).

Dentro de los aspectos críticos mencionados en la gestión de negocios, una de las fallas más perjudiciales es la aplicación limitada de la gestión tradicional de riesgos en los negocios internacionales, debido a la falta de herramientas y metodologías adecuadas para identificar, analizar y reducir los riesgos de manera efectiva. Los métodos tradicionales, que se basan en el análisis manual de datos y la experiencia subjetiva, pueden ser lentos, inexactos y no escalables.

Frente a la ausencia o falta de datos relevantes, los agentes económicos suelen dar más peso a las opiniones, actuando más en función de juicios valorativos y expectativas que de información verificada. Este comportamiento gregario, donde los agentes económicos siguen las tendencias del mercado, genera una variabilidad excesiva. La falta de información fiable puede llevar a decisiones de inversión poco racionales, lo que incrementa las posibilidades de sufrir daños y tener un impacto adverso en el rendimiento financiero de las empresas (Valdivia, 2020).

En respuesta a estos desafíos, el Business Intelligence (BI) es uno de los medios más recientes para darle solución a los obstáculos que se dan día a día en el ámbito de los negocios. Sin embargo, es importante reconocer que estas herramientas, a pesar de su sofisticación, tienen una capacidad predictiva limitada, lo que significa que no pueden eliminar por completo la incertidumbre inherente a los negocios internacionales (Smith, 2018, p. 45)

Un aspecto crucial a tener en cuenta es la calidad de los datos empleados para entrenar modelos de BI, ya que esto incide directamente en su capacidad predictiva. No obstante, la obtención de datos completos y precisos es de hecho de los mayores desafíos, debido a la diversidad de fuentes, la falta de estandarización y la existencia de sesgos culturales o económicos (Johnson, 2019, p. 78).

Es importante reconocer que el análisis cuantitativo proporcionado por estas herramientas debe ir de la mano con el análisis cualitativo realizado por expertos humanos. Aunque los datos pueden mostrar la mayor parte de lo necesario, la experiencia y el conocimiento contextual de los expertos son esenciales para interpretarlos correctamente y comprender los matices del entorno global. De hecho, según Davenport y Harris (2007), 'la verdadera ventaja competitiva se logra cuando las empresas combinan análisis cuantitativos avanzados con el juicio cualitativo de los expertos' (p. 65)."

En este aspecto, resulta fundamental recalcar que los expertos deben encargarse de actualizar periódicamente los modelos predictivos con nuevos datos y eventos relevantes para mantener su precisión y relevancia. Por lo tanto, establecer un proceso continuo de monitoreo y evaluación es trascendental para garantizar que los modelos sigan siendo útiles para la gestión de riesgos.

En conclusión, la gestión de riesgos en negocios internacionales requiere un enfoque integral que combine el poder del BI con el análisis experto y la comprensión profunda del entorno global y las necesidades de los clientes internacionales. La implementación de estrategias efectivas para abordar las limitaciones de la capacidad

predictiva, el análisis de datos dispersos y no estructurados, y la construcción de relaciones sólidas con los clientes permitirá a las empresas tomar decisiones informadas, mitigar riesgos y aumentar sus posibilidades de éxito en el mercado global.

Justificación

En una economía global competitiva, el rendimiento de la innovación es ampliamente reconocido como un importante factor de éxito. La perspectiva de la innovación esboza una visión clara de las posibilidades futuras, y es por ello por lo que la adaptación es clave. Las empresas deben introducir continuamente nuevos productos y servicios que creen valor para clientes y organizaciones, al mismo tiempo que navegan por la volatilidad de los mercados, las tensiones comerciales, las crisis globales y la inestabilidad política (León, 2023).

Según el estudio de Scuotto y Santoro (2021), el rendimiento de la innovación en las empresas se basa en tres dimensiones principales: el desarrollo de la innovación incremental (la capacidad de la empresa para mejorar productos y servicios existentes), el desarrollo de la innovación radical (la capacidad de la empresa para crear productos y servicios completamente nuevos) y la apertura de nuevos mercados (la oportunidad de introducirse en mercados no explotados).

Además, varios estudios han destacado los beneficios y oportunidades que el Business Intelligence (BI) puede ofrecer a las empresas en términos de innovación, eficiencia, productividad, calidad y satisfacción del cliente, creando una ventaja

competitiva significativa. Espiñeira, Sheldon y Asociados (2008) afirman que BI permite a la empresa disponer de información correcta en el momento justo para la toma de decisiones, pues al contar con información centralizada se evita solicitar información a diferentes departamentos; se pueden evaluar diferentes escenarios al mismo tiempo, adelantando las posibles decisiones estratégicas y convirtiendo tendencias negativas en acciones positivas; se definen indicadores que permiten medir el desempeño del negocio; y se agrupa información de diferentes áreas en un único lugar.

Para que un proyecto de BI resulte exitoso, es crucial estandarizar sistemas y procesos. Generalmente, las empresas tienen diferentes sistemas en sus áreas de negocio, y cada una de ellas se siente cómoda y confiada con sus propias herramientas. Para abordar este desafío, es necesario involucrar a cada área, considerando sus opiniones y necesidades, al mismo tiempo que se estandarizan los procesos y herramientas de trabajo que utilizan. Además, es fundamental enfocarse en la usabilidad al evaluar las herramientas de los softwares de BI. Por lo tanto, las herramientas deben contar con módulos de autoservicio y presentar información y reportes de manera comprensible para todos los usuarios (Santoro, 2021).

Las herramientas que provee el BI junto a un esencial entrenamiento de parte de los colaboradores en las empresas, es lo que permitirá aprovechar los beneficios de su innovación. La formación continua en el uso de herramientas de BI es esencial para maximizar su impacto en la innovación empresarial. Capacitar a los empleados en las últimas tecnologías y metodologías de BI no solo mejora la eficiencia operativa, sino que también fomenta una cultura de innovación constante (Esteban Guerrero-Valero, 2020). Al mantenerse actualizados con las tendencias emergentes y las mejores

prácticas en BI, las empresas pueden explorar nuevas oportunidades de mercado y responder de manera más ágil a los cambios en el entorno competitivo.

José Luis García (2022) afirma, que en este sentido, la innovación abierta, que se basa en la utilización de conocimiento externo durante el proceso de innovación y la expansión de mercados, encuentra en el BI una fuente esencial de innovación. Esto implica el uso de entradas y salidas de conocimiento que facilitan y aceleran la innovación dentro de la organización. Así, se puede afirmar que hay una conexión inherente entre BI y la innovación abierta, dado que ambos aprovechan fuentes externas de conocimiento para alcanzar la excelencia empresarial y mejorar el rendimiento en la innovación.

Los avances en los algoritmos de ML, inspirados en la capacidad del sistema cognitivo para observar, analizar, aprender y tomar decisiones respecto a problemas complejos, han acelerado la integración de la inteligencia artificial (IA) en la toma de decisiones en las organizaciones. Esta integración permite a las empresas no sólo reaccionar más rápidamente a los cambios del entorno, sino también anticipar y mitigar riesgos de manera más eficaz (Pedro Sánchez, 2023)

Davenport (2010) argumenta que la gestión del conocimiento y el capital intelectual están estrechamente vinculados con la generación de datos, información y conocimiento en sí. No obstante, esta relación va más allá de la simple recolección de datos externos. Lo que ha impulsado nuevas tendencias en BI, big data (datos a gran escala) y los continuos avances en las tecnologías de la información. Según Rodríguez y

Guerra (2008), sin información es prácticamente imposible desarrollar herramientas para gestionar el conocimiento contenido en las fuentes de manera evidente, explícita u oculta.

Laudon y Laudon (2011) sostienen que la esencia de la inteligencia y el análisis de negocios radica en la capacidad de integrar todos los flujos de información de una empresa en un conjunto de datos coherente a nivel organizacional. Utilizando técnicas de modelado, herramientas de análisis estadístico y otras metodologías, se busca comprender estos datos de tal manera que los gerentes puedan tomar decisiones más acertadas y realizar planes más efectivos. Este enfoque refuerza cómo BI puede transformar datos dispersos en conocimientos valiosos.

Con todo esto, se puede aseverar que las bondades del BI se logran mediante el análisis de datos, lo que tiene una relación positiva con el rendimiento de la analítica de bases de datos y la ventaja competitiva. Esto es aplicable a cualquier tipo de empresa, ya que permite a las empresas explorar los patrones imprevistos de sus clientes, negocios y mercados, al tiempo que mejoran su comprensión del comportamiento de los clientes (Davenport, 2010).

En la economía actual, solo sobreviven aquellas organizaciones que generan nuevos conocimientos, los difunden y los integran rápidamente en nuevos procesos, productos y servicios. Estas organizaciones reconocen la información como un recurso indispensable para ampliar su competitividad y mejorar la calidad de sus servicios a los clientes, operando eficazmente en el mercado global (Casate, 2007).

Además, en un entorno tan competitivo, la supervivencia de las empresas depende del procesamiento de la información correcta en el momento adecuado, ya que esto analiza información valiosa sobre la empresa, impulsando una mayor eficiencia y fortaleciendo la innovación empresarial, un término relacionado con productividad, rentabilidad, calidad, eficacia y eficiencia.

Poseer una gran cantidad de datos sin la herramienta adecuada para analizarlos y extraer información valiosa de ellos incrementa significativamente el tiempo que los empleados invierten en datos vacíos (Davenport, 2013). Los datos son simplemente observaciones. La información se compone de datos situados en un contexto, y el conocimiento resulta de la información procesada mediante la experiencia, la reflexión y otras prácticas que permiten un entendimiento más profundo (Erickson y Rothberg, 2014, párr. 6).

De acuerdo con Jones-Farmer (2016), las tecnologías de la información seguirán evolucionando, por lo que es de suma importancia desarrollar competencias, desde la educación superior, relacionadas con el manejo de datos a gran escala y el procesamiento de estos, apoyándose en herramientas relacionadas con estas actividades. Todo esto con el objetivo de proporcionar información suficiente y en tiempo real a quienes toman las decisiones organizacionales, potenciando así el conocimiento que es difícilmente imitable por la competencia.

Por último, se les ha denominado a los datos, “el nuevo petróleo” y las organizaciones de todo el mundo se esfuerzan por explotar los repositorios de datos que tienen en su poder, en un intento por extraer información procesable que les pueda dar

una ventaja sobre la competencia. Trasladando este razonamiento a un país, este se vuelve más competitivo y atractivo para la inversión extranjera siempre y cuando las condiciones macroeconómicas reflejan un ambiente propicio para tales efectos. Sin embargo, desde el punto de vista económico, la gestión se identifica con el manejo de activos intelectuales, tales como la protección y comercialización de las innovaciones y la propiedad intelectual, con el fin de usar la tecnología para conseguir ventajas competitivas en el mercado (Perdomo, 2014). El McKinsey Global Institute muestra en sus investigaciones que los datos pueden crear un valor significativo para la economía mundial, generando alta productividad y competitividad en las empresas (Manyika et al., 2011).

Alcance

Esta investigación se centra en proporcionar una comprensión profunda y detallada de cómo las tecnologías avanzadas, específicamente Business Intelligence, pueden revolucionar la gestión de riesgos en el ámbito de los negocios internacionales. Al explorar este campo, se pretende ofrecer un enfoque innovador y eficiente para la identificación, evaluación y mitigación de riesgos en un entorno globalizado.

Por esto, la investigación está dirigida a: directivos y ejecutivos de empresas internacionales responsables de la gestión de riesgos, planificación estratégica y operaciones internacionales; analistas de riesgo y profesionales de BI que trabajan en la evaluación de riesgos y en la implementación de soluciones de BI; académicos e investigadores interesados en la gestión de riesgos y tecnologías emergentes; y

estudiantes de negocios y tecnología, especialmente aquellos en programas de MBA, gestión de riesgos, data science y BI.

Ante la ineficacia de los métodos tradicionales de gestión de riesgos para enfrentar los desafíos actuales, surge la necesidad de implementar herramientas, conceptos y conocimientos más avanzados para reducir los desafíos. Además, la creciente disponibilidad de datos exige el uso de herramientas avanzadas del BI para analizar y extraer información valiosa. Una gestión efectiva de riesgos puede tener un impacto significativo en la sostenibilidad y rentabilidad de las empresas, haciendo que la mejora en las herramientas y conocimientos para la gestión de riesgos sea crucial. Proporcionar a los profesionales y estudiantes habilidades avanzadas y conocimientos sobre BI en este contexto fomenta el desarrollo de competencias relevantes y competitivas en el mercado laboral actual.

Hay valiosas herramientas y metodologías para diversos profesionales y sectores dentro de esta investigación. Para los profesionales de la gestión de riesgos, los ejecutivos y directivos de empresas con operaciones internacionales, será imprescindible ya que les ayudará a gestionar de manera más efectiva los riesgos inherentes a la globalización, donde su aplicación permite mejorar la precisión en la predicción de riesgos, la toma de decisiones estratégicas y la resiliencia organizacional. Los analistas de BI pueden diseñar sistemas de BI más robustos para la gestión de riesgos, ofreciendo insights más precisos y accionables a las organizaciones.

Académicos, investigadores y estudiantes de postgrado encuentran en esta investigación un recurso valioso para la enseñanza, la investigación y la formación académica en la intersección de la tecnología y la gestión de riesgos. Las herramientas tecnológicas de gestión empresarial pueden crear soluciones más efectivas y adaptadas a las necesidades de la gestión de riesgos en negocios internacionales. Policymakers y reguladores pueden diseñar políticas más efectivas para mitigar riesgos en comercio y finanzas internacionales. Emprendedores y startups pueden encontrar estrategias para gestionar riesgos desde el inicio y promover un crecimiento sostenible. El público en general puede comprender cómo las herramientas tecnológicas transforman la gestión de riesgos en empresas globales.

Objetivo General

Desarrollar un sistema de clasificación automática para la gestión de riesgos en los negocios internacionales.

Objetivo Específico

- 1) Ofrecer una base teórica, conceptual y legal sólida para el modelo de clasificación empleado en el análisis de datos.
- 2) Explicar la metodología de una clasificación supervisada para la implementación de conceptos de aprendizaje automático.
- 3) Evaluar el modelo y su eficiencia a través de los resultados de la clasificación de riesgos, determinando si el riesgo es alto, medio o bajo.

Capítulo 1

Marco Teórico

Dentro de la amplia ciencia de datos, una rama de la inteligencia artificial conocida como aprendizaje automático (Machine Learning) se destaca por su capacidad de aprendizaje autónomo y generación de resultados sin programación explícita. Según Rouhiainen (2018), "gracias al aprendizaje automático, muchos de los dispositivos que verás en el futuro obtendrán experiencia y conocimientos a partir de la forma en que son utilizados para poder ofrecer una experiencia al usuario personalizada". Los avances en esta área han permitido la creación de sistemas inteligentes con habilidades cognitivas similares a las humanas, que influyen en nuestra vida comercial y personal y moldean las interacciones en los mercados electrónicos de diversas maneras (Janiesch et al., 2020).

Valor en Riesgo

El Valor en Riesgo (VaR) de una variable aleatoria X con un nivel de confianza $\alpha \in]0,1[$ se define como el menor α -percentil de dicha variable. En otras palabras, el $\text{VaR}_\alpha(X)$ es un umbral tal que la probabilidad de que X sea menor que ese umbral es igual a α . Este concepto es ampliamente utilizado en regulaciones financieras, como Basilea I y Basilea II, para medir la magnitud de los cambios diarios en el valor de los activos (López Domínguez, 2022)

Para variables aleatorias que siguen una distribución normal, el VaR es proporcional a la desviación estándar de la variable. Si F_x es la función de distribución acumulativa de X , entonces el VaR se puede definir matemáticamente como:

$$VaR_\alpha(X) = \min\{z | F_x(z) \geq \alpha\}$$

Esta definición, basada en el trabajo de Rockafellar y Uryasev, muestra cómo se puede calcular el VaR utilizando la función de distribución acumulativa.

A pesar de su facilidad de cálculo e intuición, el VaR tiene ciertas limitaciones matemáticas. Dependiendo del nivel de confianza α , para distribuciones discretas, el VaR puede ser una función discontinua y no convexa. Esto implica que pequeñas variaciones en los datos pueden llevar a grandes cambios en el valor calculado del VaR, lo que puede complicar su interpretación y uso en ciertas aplicaciones prácticas (López Domínguez, 2022).

Pros y contras del VaR

➤ Ventajas

El Valor en Riesgo (VaR) es una herramienta de gestión de riesgos relativamente sencilla de entender. La intuición detrás del percentil α de una distribución es fácilmente comprensible, y el VaR ofrece una interpretación clara: cuánto se puede perder con un cierto nivel de confianza. El VaR proporciona un único número que mide el riesgo, definido por un nivel de confianza específico, como $\alpha = 0.95$

Se pueden comparar dos distribuciones al examinar sus VaR para el mismo nivel de confianza. Especificar el VaR para todos los niveles de confianza define completamente

la distribución. En este sentido, el VaR es superior a la desviación estándar. A diferencia de la desviación estándar, el VaR se enfoca en una parte específica de la distribución determinada por el nivel de confianza, lo cual es a menudo lo que se necesita. Esto ha hecho que el VaR sea popular en la gestión de riesgos en sectores como las finanzas, la energía nuclear, el espacio aéreo, la ciencia de materiales y diversas aplicaciones militares.

Una de las propiedades importantes del VaR es la estabilidad de los procedimientos de estimación. Como el VaR no considera la cola de la distribución, no se ve afectado por pérdidas extremadamente altas en la cola, que suelen ser difíciles de medir. El VaR se estima mediante modelos paramétricos; por ejemplo, el VaR de covarianza basado en la suposición de una distribución normal es muy conocido en finanzas, utilizando modelos de simulación como el histórico o el de Monte Carlo, o mediante aproximaciones basadas en la expansión de Taylor de segundo orden (Serraino, 2016).

➤ **Contras**

El VaR no tiene en cuenta las propiedades de la distribución más allá del nivel de confianza. Esto implica que el $\text{VaR}_\alpha(X)$ puede aumentar dramáticamente con un pequeño aumento en α . Para estimar adecuadamente el riesgo en la cola, puede ser necesario calcular varios VaR con diferentes niveles de confianza. El hecho de que el VaR no considere la cola de la distribución puede llevar a la asunción involuntaria de riesgos elevados.

Por ejemplo, en un entorno financiero, vender opciones muy fuera del dinero puede resultar en recibir una prima sin ninguna pérdida al vencimiento la mayoría de las veces,

pero un gran movimiento adverso del mercado puede provocar una pérdida extremadamente alta, que el VaR no puede capturar.

El control de riesgos mediante el VaR puede dar lugar a resultados indeseables en caso de distribuciones sesgadas. Por ejemplo, la optimización de una cartera utilizando VaR puede resultar en una cola significativamente más larga en comparación con la optimización utilizando el CVaR (Conditional Value at Risk), lo cual se mide por la pérdida máxima de esas carteras. Además, el VaR es una función no convexa y discontinua para distribuciones discretas, lo que hace que la optimización del VaR sea un problema computacionalmente desafiante. Aunque existen códigos, como el PSG, que pueden trabajar con funciones VaR de manera eficiente, optimizar carteras con una función de rendimiento VaR y dar forma a las distribuciones de la cartera con múltiples restricciones de VaR puede ser complejo (Uryasev, 2016).

Valor en Riesgo Condicional (CVaR)

El Valor en Riesgo Condicional (CVaR), introducido por Rockafellar y Uryasev, es una medida percentil alternativa de riesgo que proporciona una visión más completa del riesgo extremo en comparación con el Valor en Riesgo (VaR). Para variables aleatorias con funciones de distribución continua, el CVaR es igual a la expectativa condicional de X dado que X excede el VaR con un nivel de confianza $\alpha \in]0,1[$. En el caso general, para variables aleatorias con funciones de distribución posiblemente discontinuas, el CVaR se define como la media de la α -distribución generalizada de cola.

Matemáticamente, el CVaR de X con nivel de confianza α , denotado como $\text{CVaR}_\alpha(X)$, no es simplemente el promedio de los resultados que superan el $\text{VaR}_\alpha(X)$. Para distribuciones generales, puede ser necesario dividir un átomo de probabilidad. Por ejemplo, cuando la distribución se modela mediante escenarios, el CVaR se puede obtener promediando una fracción de estos escenarios (Sarykalin, 2017).

Contrariamente a la creencia popular, el $\text{CVaR}_\alpha(X)$, no es igual a un promedio simple de los resultados que exceden el $\text{VaR}_\alpha(X)$. Para explicar esta idea con más detalle, introducimos definiciones alternativas de CVaR. Denotamos $\text{CVaR}_{\alpha^+}(X)$, llamado "CVaR superior", como la expectativa condicional de X dado que $X > \text{VaR}_\alpha(X)$:

$$\text{CVaR}_{\alpha^+}(X) = E[X | X > \text{VaR}_\alpha(X)]$$

El $\text{CVaR}_\alpha(X)$, puede definirse alternativamente como el promedio ponderado del $\text{VaR}_\alpha(X)$ y $\text{CVaR}_{\alpha^+}(X)$ como sigue. Si $F_X(\text{VaR}_\alpha(X)) < 1$, lo que significa que existe la posibilidad de una pérdida mayor que $\text{VaR}_\alpha(X)$, entonces:

$$\text{CVaR}_\alpha(X) = 1 - \alpha \int_0^\alpha \text{VaR}_u(X) du$$

mientras que si $F_X(\text{VaR}_\alpha(X)) = 1$, lo que significa que $\text{VaR}_\alpha(X)$ es la mayor pérdida que puede ocurrir, entonces:

$$\text{CVaR}_\alpha(X) = \text{VaR}_\alpha(X)$$

La definición de CVaR como en la ecuación anterior muestra que el CVaR no se define únicamente como una expectativa condicional. La función $CVaR_{\alpha}(X)$, denominada "CVaR inferior", coincide con el $CVaR_{\alpha}(X)$ para distribuciones continuas; sin embargo, para distribuciones generales, puede ser discontinua y no convexa. La construcción del $CVaR_{\alpha}$ como promedio ponderado del $VaR_{\alpha}(X)$ y $CVaR_{\alpha+}(X)$ es una innovación importante en la medición del riesgo financiero (Sarykalin, 2017).

Introducción a las Máquinas de Vectores de Soporte

Las máquinas de vectores de soporte representan un avance en la teoría de los sistemas de aprendizaje. La teoría del aprendizaje estadístico proporciona un nuevo marco para modelar algoritmos de aprendizaje, fusiona el campo del aprendizaje automático y la estadística, e inspira nuevos algoritmos que superan las dificultades enfrentadas por el aprendizaje convencional. Estos métodos tienen un diseño modular inherente que simplifica su implementación y análisis y permite la inserción de conocimientos específicos del dominio. La máquina de vectores de soporte (Support Vector Machine, SVM) es un algoritmo novedoso de aprendizaje con muestras pequeñas. A diferencia del principio tradicional de minimización del riesgo empírico, las SVM se basan en el principio de minimización del riesgo estructural (Soman et al., 2009a).

Riesgo Empírico vs. Riesgo Estructural

El análisis del riesgo empírico y el riesgo estructural es crucial para entender cómo los modelos de aprendizaje automático, como las Máquinas de Vectores de Soporte (SVM), gestionan la capacidad de ajuste y generalización.

➤ **Riesgo Empírico:**

El riesgo empírico se refiere a la pérdida promedio sobre el conjunto de datos de entrenamiento. Formalmente, para un conjunto de datos $S = \{(x_i, y_i)\}_{i=1}^n$ y una función de predicción f , el riesgo empírico

$LS(f)$ se define como:

$$LS(f) = \frac{1}{n} \sum_{i=1}^n \ell(f(x_i), y_i)$$

Donde ℓ es la función de pérdida que mide el error entre la predicción $f(x_i)$ y la etiqueta real y_i . El riesgo empírico se centra en minimizar el error en los datos disponibles, pero puede llevar a un sobreajuste si el modelo se adapta demasiado a las particularidades del conjunto de entrenamiento (Steinwart & Christmann, 2008).

➤ **Riesgo Estructural**

En contraste, el riesgo estructural incorpora un término de regularización que penaliza la complejidad del modelo, equilibrando la capacidad del modelo para ajustarse a los datos de entrenamiento con su capacidad para generalizar a nuevos datos no vistos. El riesgo estructural se formula como:

$$R(f) = LS(f) + H(f)$$

Donde:

$$\frac{\neq}{S} \quad I$$

$LS(f)$ es el riesgo empírico, que representa la pérdida promedio en el conjunto de entrenamiento.

$H(f)$ es la función de regularización, que penaliza la complejidad del modelo para evitar el sobreajuste.

λ es un hiper parámetro que controla la importancia relativa del término de regularización, determinando la simplicidad o complejidad del modelo final (Liu, 2019; Steinwart & Christmann, 2008).

Bases de SVM

Las Máquinas de Soporte Vectorial (SVM) se utilizan para resolver problemas de clasificación binaria. En este contexto, disponemos de un conjunto de ejemplos de entrenamiento, y el objetivo es generalizar este conocimiento para clasificar nuevos datos no vistos. Esto significa que, cuando se presenta un nuevo ejemplo, debemos determinar si pertenece a una de las dos posibles clases, etiquetadas como +1 o -1. En esencia, seleccionamos la clase que más se asemeja a los ejemplos de entrenamiento (Morales, 2021).

Operaciones Básicas en SVM

Algunas de las operaciones fundamentales en SVM incluyen:

Producto Punto: Dado un vector x y un vector w , el producto punto se calcula como:

$$\omega \cdot x = \sum_{i=1}^n w_i x_i$$

donde w_i representa el i -ésimo elemento del vector w y x_i representa el i -ésimo elemento del vector x . La interpretación geométrica del producto punto es que calcula el coseno del ángulo entre los vectores w y x si estos están normalizados a una longitud de

Cálculo de la Norma: El producto punto también se utiliza para calcular la longitud o norma de un vector x , denotada como $\|x\|$

$$\|x\| = \sqrt{x \cdot x}$$

Distancia entre Vectores: La distancia entre dos vectores x y y : se puede calcular como la longitud del vector diferencia $x - y$:

$$\|x - y\| = \sqrt{(x - y) \cdot (x - y)}$$

Estas operaciones son fundamentales en el funcionamiento de las SVM, ya que se utilizan para definir y optimizar los hiperplanos de separación en el espacio de características (Escalante, 2021).

Support Vector Machines (SVM) en la clasificación de problemas linealmente separables

En el contexto de la gestión de riesgos, las Support Vector Machines (SVM) son algoritmos poderosos para clasificar problemas que pueden ser separados linealmente. Según Wang (2017), el entrenamiento de una SVM en un conjunto de datos de entrenamiento implica encontrar un hiperplano óptimo que maximice el margen de separación entre dos clases, donde x_i es el vector de características y y_i es la etiqueta de clase.

Optimización para la Separación Óptima

La superficie de separación óptima puede ser definida mediante un problema de optimización con restricciones. Este problema se expresa buscando minimizar la siguiente función:

$$\frac{1}{2} \|w\|^2$$

sujeto a las restricciones:

$$y_i(w \cdot x_i + b) \geq 1, \forall i$$

Aquí, w es el vector de pesos y b es el sesgo. Para resolver este problema, se define la función de Lagrange:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2$$

donde $\alpha_i > 0$ son los coeficientes de Lagrange.

Resolviendo el Problema de Optimización

El objetivo es minimizar w y b en la función de Lagrange. Esto se logra al buscar las derivadas parciales respecto a w y b y establecerlas a cero. Al hacerlo, el problema original se transforma en el problema dual siguiente:

$$\max_{\alpha} \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j (x_i \cdot x_j)$$

sujeto a las restricciones:

$$\sum_{i=1}^n a_i y_i = 0$$

$$a_i \geq 0$$

Vectores de Soporte y Clasificación

Si α_i es la solución óptima, el vector de pesos óptimos w se expresa como una combinación lineal de los vectores de entrenamiento:

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

De acuerdo con las condiciones de Kuhn-Tucker, la solución óptima también debe cumplir:

$$w = \sum_{i=1}^n \alpha_i y_i x_i$$

Para la mayoría de las muestras, α_i será cero. Solo aquellos valores de α_i diferentes de cero corresponden a los vectores de soporte, que son una pequeña fracción de todas las muestras. La función de clasificación óptima se define como:

$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i (x_i \cdot x) + b\right)$$

Hiperplanos Óptimos en la Clasificación con SVM

El método de hiperplanos óptimos, introducido por Vapnik en 1982, es crucial para la clasificación de datos sin errores en el contexto de las máquinas de soporte vectorial (SVM) (Cortes & Vapnik, 1995). Este método busca separar un conjunto de datos de entrenamiento etiquetados mediante un hiperplano que maximice la separación entre las diferentes clases.

Formalmente, un conjunto de datos de entrenamiento etiquetados puede representarse como:

$$\{(x_i, y_i)\}_{i=1}^n \quad (17)$$

donde x_i son los vectores de características y y_i son las etiquetas correspondientes. Este conjunto se considera linealmente separable si existe un vector w y un escalar b que satisfacen las siguientes desigualdades:

$$y_i(w \cdot x_i + b) \geq 1 \quad \forall i$$

Estas desigualdades pueden expresarse de manera más compacta como:

$$y_i(w \cdot x_i + b) \geq 1$$

El hiperplano óptimo, denotado como (w_0, b_0) , es aquel que maximiza el margen entre las dos clases, definido como la distancia más corta desde el hiperplano hasta los puntos más cercanos de cualquiera de las dos clases. Este margen se maximiza resolviendo el siguiente problema de optimización:

$$\text{Minimizar } \frac{1}{2} w \cdot w$$

sujeto a

$$y_i(w \cdot x_i + b) \geq 1$$

Este problema se convierte en un problema de programación cuadrática donde el objetivo es minimizar $w \cdot w$ sujeto a las restricciones de clasificación. La solución a este problema proporciona el hiperplano que separa las clases con el mayor margen posible, ayudando a mejorar la capacidad de generalización del modelo y reduciendo el riesgo de sobreajuste.

Vectores de soporte

Estos vectores juegan un papel crucial en la clasificación y son esenciales para construir un modelo robusto y eficiente en la gestión de riesgos.

En el contexto de SVM, los vectores x_i que cumplen la condición $y_i(w \cdot x_i + b) = 1$ se conocen como vectores de soporte. El vector w_0 que define el hiperplano óptimo puede representarse como una combinación lineal de estos vectores de entrenamiento:

$$w_0 = \sum_{i=1}^n \alpha_i y_i x_i$$

donde $\alpha_i \geq 0$. Para encontrar los parámetros α , se resuelve el siguiente problema de programación cuadrática:

$$W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

sujeto a las restricciones:

$$\sum_{i=1}^n \alpha_i y_i = 0$$

$$\alpha_i \geq 0$$

Este problema se puede resolver de manera eficiente dividiendo los datos de entrenamiento en porciones pequeñas y resolviendo el problema de programación cuadrática para cada porción. Luego, se ajustan iterativamente los vectores de soporte para construir el hiperplano óptimo para el conjunto completo de datos de entrenamiento (Cortes & Vapnik, 1995).

La utilidad de los vectores de soporte radica en su capacidad para definir de manera precisa el límite de decisión entre clases, lo que es crucial para la precisión y generalización del modelo en la gestión de riesgos y otras aplicaciones del aprendizaje supervisado. (Cortes & Vapnik, 1995).

Hiperplano de Margen Suave

El hiperplano de margen suave es una técnica crucial en las Máquinas de Soporte Vectorial (SVM) cuando los datos de entrenamiento no se pueden separar sin errores. En tales casos, se busca encontrar un hiperplano que separe los datos de entrenamiento con un número mínimo de errores.

Para manejar esta situación, se introducen variables no negativas $\xi_i \geq 0$ para $i=1, \dots, \ell$

El objetivo es minimizar el siguiente funcional:

$$\min_{\omega, \xi} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{\ell} \xi_i$$

sujeto a las restricciones:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

Para valores suficientemente pequeños de ξ_i , este funcional describe el número de errores de entrenamiento. Minimizar este funcional permite identificar un subconjunto mínimo de errores de entrenamiento (y_i', x_i')

Si estos datos se excluyen del conjunto de entrenamiento, el resto puede ser separado sin errores mediante la construcción de un hiperplano óptimo de separación (Cortes & Vapnik, 1995).

El problema puede reformularse minimizando el siguiente funcional:

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \varepsilon_i$$

donde

$F(u)$ es una función convexa monótona. Para $F(u) = u^2$, el problema sigue siendo de programación cuadrática. El vector w que determina el hiperplano se puede expresar como una combinación lineal de los vectores de soporte x_i :

$$\omega = \sum_{i=1}^l a_i y_i x_i$$

Para encontrar $A = (\alpha_1, \dots, \alpha_\ell)$, se resuelve el problema dual maximizando: sujeto a las restricciones:

$$\sum_{i=1}^{\ell} a_i y_i = 0$$

$$0 \leq a_i \leq C$$

En esta formulación, ξ es un escalar adicional que representa el valor máximo permitido de α , $\alpha_{\max} = \max(\alpha_1, \dots, \alpha_{\ell})$. Esto hace que el funcional $W(A)$ no sea cuadrático debido al término con α_{\max} . La maximización de:

$W(A)$ bajo estas restricciones pertenece a los problemas de programación convexa. Para construir un clasificador de margen suave, se puede resolver el problema de programación convexa en el espacio ℓ -dimensional de los parámetros.

A o el problema de programación cuadrática en el espacio dual $\ell+1$ de los parámetros A y ξ . En los experimentos, los hiperplanos de margen suave se construyen resolviendo el problema dual de programación cuadrática (Cortes & Vapnik, 1995).

Modelos no linealmente separables

Es fundamental contar con modelos que puedan manejar datos complejos y no linealmente separables. Las Máquinas de Soporte Vectorial (SVM) ofrecen una solución eficiente mediante la utilización de funciones de transformación conocidas como funciones kernel, las cuales permiten proyectar los datos en un espacio de mayor dimensionalidad donde se puedan separar linealmente (Gonzalez, 2023).

Transformación a Espacios de Mayor Dimensionalidad

Cuando los datos no pueden ser separados linealmente en su espacio original, se aplica una función de transformación o función kernel. Esta transformación incrementa la dimensionalidad de los vectores de entrada, permitiendo así que los datos sean separados linealmente mediante un hiperplano en el nuevo espacio.

Para resolver un problema de clasificación con SVMs se necesita la matriz de kernel K (Gramm matrix) que es la siguiente:

$$K = \begin{matrix} k(x_1, x_1) & k(x_1, x_2) & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & k(x_2, x_n) \\ k(x_n, x_1) & k(x_n, x_2) & k(x_n, x_n) \end{matrix}$$

Donde:

- $k(x_i, x_j)$ es una medida de similitud entre patrones.
- K es $n \times n$ es la dificultad computacional y de almacenamiento.

Los principales tipos de funciones kernel son:

- **Kernel Polinómico**

El kernel polinomial es una función utilizada en métodos de aprendizaje automático, particularmente en las máquinas de vectores de soporte (SVM). Su principal utilidad es permitir que los algoritmos de aprendizaje trabajen en un espacio de características transformado, lo que puede hacer más fácil la separación de datos no linealmente separables en el espacio original (*Kernel Functions*, 2020)

Este no solo considera las características dadas de las muestras de entrada para determinar su similitud, sino también combinaciones de estas. El espacio de características (implícito) de un kernel polinómico es equivalente al de la regresión polinómica, pero sin el crecimiento combinatorio en el número de parámetros a aprender (*Wayback Machine*, 2013). El mismo se define como:

$$k(x, y) = (\gamma \langle x, y \rangle + \lambda)^d = (\gamma x^T y + \lambda)^d$$

Donde los parámetros γ y λ servirán para ajustar los coeficientes del kernel polinómico. Cuando $\gamma=1$ y $\lambda=0$ el kernel se denomina *homogéneo* (*Wayback Machine*, 2013).

En este caso, el espacio sobre el que se calcula el producto escalar, o sea, $\phi(x)$ es de dimensiones:

$$\binom{k+d}{d} = \frac{(k+d)!}{d! k!}$$

Donde k es la dimensión original del vector x . Esto quiere decir:

$$\phi: \mathbb{R}^k \rightarrow \mathbb{R}^{\binom{k+d}{d}}$$

- **Kernel Radial o Gaussiano (RBF)**

El kernel radial, también conocido como kernel de función de base radial (RBF, por sus siglas en inglés). La principal ventaja del kernel RBF es su capacidad para manejar problemas donde los datos no son linealmente separables, permitiendo que los modelos aprendan patrones complejos y no lineales en los datos (*Kernel Functions*, 2020). Este se define como:

$$k(x, y) = e^{-\frac{\|x - y\|^2}{2\sigma^2}}$$

Donde γ es igual a $\frac{1}{2\sigma^2}$ γ es un parámetro que medirá la varianza del kernel. Por otra parte, $\|x - y\|^2$ representa la distancia euclídea al cuadrado entre x e y .

En el caso del kernel RBF, el espacio sobre el que se calculará el producto escalar entre los vectores es un espacio de dimensiones infinitas el cual se define de la siguiente manera (Santamaria, 2018):

$$k(x, y) = \Phi(x)^T \Phi(y) = e^{-\frac{\|x - y\|^2}{2\sigma^2}}$$

La distancia entre $\Phi(x)$ y $\Phi(y)$ es:

$$d(\Phi(x), \Phi(y)) = \sqrt{\|\Phi(x) - \Phi(y)\|^2} = \sqrt{2 \left(1 - e^{-\frac{\|x - y\|^2}{2\sigma^2}}\right)}$$

$$= \sqrt{2(1 - k(x, y))}$$

Cada uno de estos kernels tiene propiedades particulares que los hacen adecuados para diferentes tipos de problemas de clasificación (Escalante, 2022).

Condición de Mercer

Un aspecto crucial para que una función sea considerada un kernel válido es que debe satisfacer la condición de Mercer. Esta condición establece que para toda función $g(x)$, la siguiente integral debe ser no negativa:

$$\iint K(x,y)g(x)g(y)dx dy \geq 0$$

La condición de Mercer asegura que la función kernel sea definida positiva, lo cual es esencial para que el kernel funcione correctamente en las SVM y para que el modelo sea estable y eficiente (Morales, 2021).

Problemas No Linealmente Separables

Para abordar problemas no linealmente separables con Máquinas de Vectores de Soporte (SVM), la estrategia principal consiste en mapear los vectores de entrada a un espacio de características de alta dimensión y construir una superficie de separación óptima en ese espacio. En casos donde el conjunto de entrenamiento no es linealmente separable, algunas muestras de entrenamiento no pueden cumplir con las condiciones establecidas. Para manejar estos casos, se introducen variables de holgura ($\xi_i \geq 0$). (Cortes & Vapnik, 1995; Wang, 2017).

Restricciones y Variables de Holgura

Las restricciones en un problema de clasificación no linealmente separable se definen de la siguiente manera:

1. **Variable de Holgura ($\xi_i \geq 0$):** Esta variable se introduce para manejar errores de clasificación. Cuando ocurre una clasificación incorrecta, ξ_i indica el límite superior del número de errores de clasificación en el conjunto de entrenamiento, describiendo así el grado de correctitud en la clasificación.

Objetivos de Optimización

En este contexto, se persiguen dos objetivos:

1. **Maximizar el margen:** Se busca que el margen de separación entre las clases sea lo más grande posible.
2. **Minimizar los errores de clasificación:** Se pretende que el número de errores de clasificación sea lo más pequeño posible.

Estos objetivos se combinan en una única función objetivo mediante la introducción del parámetro de penalización C , que pondera adecuadamente ambos objetivos. Bajo las restricciones establecidas, se busca minimizar la siguiente función:

$$\min_{i=1}^n \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i \right\}$$

Donde C es el parámetro de penalización que controla el grado de penalización por las muestras clasificadas incorrectamente, logrando un equilibrio entre la proporción de muestras mal clasificadas y la complejidad del modelo.

Solución del Problema de Optimización

La solución óptima del problema de optimización se encuentra en el punto de silla de la función de Lagrange, definida como:

$$L(w, b, \alpha, \varepsilon) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i - \sum_{i=1}^n \alpha_i [y_i(w * x_i + b) - 1 + \varepsilon_i] - \sum_{i=1}^n \beta_i \varepsilon_i$$

Donde α y β son los vectores de multiplicadores de Lagrange. La condición $\alpha_i \geq 0$ se modifica en comparación con el caso linealmente separable, permitiendo una formulación más flexible para manejar errores de clasificación (Cortes & Vapnik, 1995; Wang, 2017).

Problema Primal y Dual de las Máquinas de Vectores de Soporte

➤ Problema Primal

El problema primal de las Máquinas de Vectores de Soporte (SVM) se refiere al problema original de optimización que busca encontrar un hiperplano óptimo que separe las clases maximizando el margen y minimizando el número de errores de clasificación (Wang, 2017) La formulación del problema primal es la siguiente:

$$\min_{w, b, s} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

sujeto a las restricciones:

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i=1, \dots, n$$

Donde:

w es el vector de pesos,

b es el sesgo,

ξ_i son las variables de holgura que permiten errores de clasificación,

C es el parámetro de penalización que controla el equilibrio entre maximizar el margen y permitir errores de clasificación,

y_i son las etiquetas de clase,

$\phi(x_i)$ es una función de mapeo a un espacio de características de mayor dimensión.

➤ Problema Dual

El problema dual es una reformulación del problema primal utilizando la teoría de multiplicadores de Lagrange. La solución del problema dual es equivalente a la del problema primal. La formulación del problema dual es la siguiente:

$$\max_{\alpha} \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i \alpha_j y_i y_j K(x_i, x_j)$$

sujeto a las restricciones:

$$0 \leq a_i \leq C, \sum_{i=1}^n a_i y_i = 0$$

Donde:

α es el vector de multiplicadores de Lagrange,

$K(x_i, x_j)$ es la función kernel que define el producto punto en el espacio de características,

C es el parámetro de penalización,

y_i son las etiquetas de clase.

Métodos para Clasificación Multiclase con SVM

Algoritmos para Clasificación Multiclase

Support Vector Machines (SVM) fueron inicialmente diseñadas para resolver problemas de clasificación binaria. Sin embargo, para aplicar SVM a problemas de clasificación multiclase, es necesario transformar el problema en múltiples problemas de clasificación binaria. Existen dos enfoques principales para resolver la clasificación multiclase basada en SVM (Cervantes, 2020b).

Algoritmo Uno-contra-Resto (One-vs-Rest)

El método Uno-contra-Resto es uno de los enfoques más comunes para la clasificación multiclase con SVM. En este método, se construye una serie de clasificadores binarios, donde cada clasificador distingue una clase de las demás. Para construir el primer clasificador, las muestras de la clase objetivo se etiquetan como positivas, mientras que las muestras de todas las demás clases se etiquetan como negativas (Zhang, 2012).

El problema de optimización que se resuelve para el primer clasificador SVM

es:

$$\min_{w_1, b_1, s_1} \frac{1}{2} \|w_1\|^2 + C \sum_{i=1}^n \varepsilon_i$$

sujeto a las restricciones:

$$y_k^{(ij)}(w_k) + b_j \geq 1 - \varepsilon, \quad \varepsilon \geq 0, \quad \forall k = 1, \dots, n$$

donde

w es el vector de pesos, b es el sesgo, ξ_i son las variables de holgura, C es el parámetro de penalización, $y_i(1)$ son las etiquetas binarias para la primera clase, y $\phi(x_i)$ es el mapeo a un espacio de características de mayor dimensión.

Al resolver este problema, se obtiene un clasificador SVM para la primera clase. De manera similar, se construyen clasificadores para las otras clases.

Algoritmo Uno-contra-Uno (One-vs-One)

El método Uno-contra-Uno fue propuesto como una alternativa al Uno-contra-Resto. En este enfoque, se construyen clasificadores binarios para cada par único de clases. Para cada par de clases i y j , las muestras de la clase i se etiquetan como positivas y las de la clase j como negativas (Zhang, 2012).

El problema de optimización que se resuelve para el clasificador entre las clases i y j es:

$$\min_{w_{ij}, b_{ij}, s_{ij}} \frac{1}{2} \|w_{ij}\|^2 + C \sum_{k=1}^n \xi_{ijk}$$

Sujeto a restricciones:

$$y_k^{(ij)}(w_k) + b_j \geq 1 - \varepsilon, \quad \varepsilon \geq 0, \quad \forall k = 1, \dots, n$$

donde

w_{ij} es el vector de pesos, b_{ij} es el sesgo, ξ_{ijk} son las variables de holgura, C es el parámetro de penalización, $y_k(ij)$ son las etiquetas binarias para el par de clases i y j y $\phi(x_k)$ es el mapeo a un espacio de características de mayor dimensión. Después de resolver estos problemas de optimización, se obtienen $k(k-1)/2$ clasificadores SVM.

Durante la clasificación, cada clasificador vota por una clase, y la clase con el mayor número de votos se asigna como la clase predicha.

Comparación de Métodos

El método Uno-contra-Resto ofrece simplicidad en la implementación con un número reducido de clasificadores (k), pero cada clasificador tiene que considerar todas las muestras de entrenamiento, lo que puede incrementar el tiempo de entrenamiento. Por otro lado, el método Uno-contra-Uno maneja mejor el desequilibrio en los datos y puede ofrecer mayor precisión, aunque requiere entrenar un número mayor de clasificadores $k-1/2$ (Zhang, 2012).

Problema de Aprendizaje Supervisado

En la gestión de riesgos, el aprendizaje supervisado se utiliza para predecir eventos adversos, evaluar la probabilidad de ocurrencia de riesgos específicos y optimizar estrategias de mitigación. Esto se logra mediante la creación de modelos predictivos que identifican patrones en los datos históricos y los utilizan para realizar predicciones precisas sobre eventos futuros. (Soman et al., 2009, p. 18).

El problema general del aprendizaje supervisado se puede formular de la siguiente manera: Para cualquier valor de m un conjunto de aprendizaje $I^\#$ (una realización de $L \times S$) sin conocimiento previo de las funciones $P(\cdot)$, $Y(\cdot)$ o $A(\cdot)$ el objetivo es encontrar una función $h(\cdot)$ definida en A que minimice el error de predicción esperado, definido por:

$$\text{Err}(a) = \text{Exr}\{L(rA(A))\}$$

donde es una función de pérdida que mide la discrepancia entre sus dos argumentos. En otras palabras, el objetivo del aprendizaje es encontrar una función de los atributos de entrada que, en el mejor de los casos, prediga el resultado del atributo de salida. La bondad del ajuste se evalúa según una función de pérdida predefinida. En la práctica, esta función de pérdida depende del tipo de valores del atributo de salida, como se indica en los siguientes dos tipos de problemas: riesgo empírico y riesgo estructural (Soman et al., 2009, p. 18)

Problema de Aprendizaje No Supervisado

Este enfoque desempeña un papel crucial en diversos contextos donde los datos se presentan de manera abundante al investigador. Según Celebi y Aydin (2016), los datos a gran escala y sin etiquetas suelen ser agrupados en clústeres, permitiendo su categorización mediante métodos o algoritmos. En este marco, los algoritmos deben aprender de los datos proporcionados para clasificarlos en clústeres con propiedades similares.

"La agrupación de datos (o clustering) es un enfoque no lineal basada en datos que se puede utilizar para dividir conjuntos de datos espaciales y temporales" (Borzooei et al., 2020). A través del clustering, es posible categorizar los datos según sus características, empleando algoritmos. Los más comunes son el k-means y el método de mezcla gaussiana.

Aquí podemos observar la fórmula del método gaussiano para la clusterización de datos:

$$p(x|\lambda) = \sum_{i=1}^M \omega_i g(x|\mu_i, \Sigma_i)$$

Marco conceptual

Riesgo

El estudio del riesgo ha evolucionado junto con las sociedades humanas y las transformaciones tecnológicas y científicas. Anteriormente, el riesgo era una noción subyacente en las decisiones diarias, pero con el tiempo se ha convertido en un concepto central en la investigación científica y en las políticas públicas. Como señalan Yassi et al. (1999), "El riesgo se define como la probabilidad de que ocurra un evento no deseado y la evaluación de las consecuencias de ese evento" (p. 15).

Esta definición abarca tanto los riesgos naturales, como los desastres naturales, como los riesgos inducidos por el ser humano, como los accidentes industriales o los

riesgos de salud pública. La percepción del riesgo varía según las culturas, las condiciones socioeconómicas y las experiencias personales, lo que influye en cómo se gestionan y se comunican los riesgos en diferentes contextos.

En términos económicos, el riesgo se asocia con la incertidumbre y la variabilidad en los resultados financieros o económicos. Las decisiones de inversión y financiamiento están intrínsecamente vinculadas a la evaluación de riesgos y a las estrategias de mitigación. Como menciona el Glosario de Términos en Salud Ambiental (2005), "Los modelos de evaluación de riesgos financieros y económicos ayudan a las empresas y a los gobiernos a tomar decisiones informadas sobre inversiones, políticas y regulaciones" (p. 45).

Gestión de Riesgos

La gestión de riesgos se concibe como un proceso integral y dinámico que abarca la identificación, evaluación y gestión de las diversas formas de riesgo que enfrentan las localidades y la sociedad en su conjunto. Este enfoque va más allá de la simple reducción de la vulnerabilidad o la mitigación de las amenazas, ya que implica tomar decisiones colectivas sobre los niveles y formas de riesgo que se pueden considerar aceptables en períodos específicos, así como los cambios necesarios para evitar las consecuencias negativas de los daños que puedan ocurrir (Hampton, 2009).

Históricamente, se observa que los riesgos han sido asumidos sin la información adecuada para medir la magnitud de los posibles daños, resultando en costos mucho más altos de lo anticipado. Esto ha llevado a la quiebra de instrumentos de seguridad como empresas de seguros, o a pérdidas irreparables en comunidades y regiones. Un modelo de gestión de riesgos busca construir información mínima que permita calcular y prever los riesgos, así como desarrollar reservas financieras, sociales, psicológicas y emocionales que aseguren la supervivencia ante posibles impactos adversos (Krysiak, 2011).

Este proceso implica la colaboración de diversos sectores de la sociedad para construir información y determinar las tareas necesarias para desarrollar reservas de recursos y opciones de respuesta. Además, requiere el diseño y la implementación rentable de instrumentos de gestión que no solo aseguren la supervivencia de las organizaciones aseguradoras, sino también de la sociedad y sus individuos.

La gestión de riesgos, en definitiva, no se limita a reducir la vulnerabilidad, sino que implica un proceso social de interacción y diálogo continuo para evaluar los cambios progresivos en los riesgos y los instrumentos de mitigación social frente a daños probables. Este enfoque no solo involucra costos financieros, sino que busca acuerdos sociales para manejar los impactos productivos sin sacrificar beneficios inmediatos (Hampton, 2009).

Negocios Internacionales

Los negocios internacionales abarcan un conjunto diverso de transacciones comerciales y actividades económicas entre individuos, empresas y gobiernos de diferentes países, con el propósito de satisfacer necesidades y alcanzar objetivos específicos a nivel global. Este campo de estudio y práctica administrativa involucra el intercambio de bienes, servicios, capital, personal, tecnología, información y datos a través de las fronteras nacionales (Czinkota, Rivoli y Ronkainen, 1989:18).

Desde una perspectiva amplia, las transacciones de negocios internacionales incluyen la exportación e importación de productos, la prestación de servicios como consultorías y banca, la inversión directa en industrias extranjeras, y la transferencia de tecnología a través de licencias y acuerdos de propiedad intelectual (Robock y Simmonds, 1983:3). Estas actividades no solo facilitan la expansión de mercados para las empresas, sino que también promueven la integración económica global, la cooperación internacional y el intercambio cultural.

Las empresas que participan en negocios internacionales enfrentan desafíos significativos que no se presentan en el ámbito doméstico. Estos desafíos incluyen la adaptación a diferentes sistemas legales y regulaciones, la gestión de riesgos cambiarios y políticos, y la comprensión de las variaciones en los entornos culturales y sociales de los países en los que operan (Daniels y Radebaugh, 1989:6).

Además, el movimiento de personas en negocios internacionales implica la movilización de personal capacitado para operar en diferentes contextos culturales y económicos. La transferencia de tecnología y conocimiento entre países juega un papel crucial en el desarrollo y la innovación industrial, permitiendo a las empresas acceder a nuevas tecnologías y procesos productivos.

A diferencia del comercio internacional, que se centra en la compra y venta de bienes y servicios entre países, los negocios internacionales abarcan un espectro más amplio de actividades que incluyen la inversión directa en el extranjero, la cooperación en investigación y desarrollo, y la gestión de cadenas de suministro globales (Mercado, 1986:17). Esto implica que las empresas deben no solo entender las dinámicas del mercado global, sino también adaptarse continuamente para competir y crecer en un entorno económico globalizado y dinámico.

Toma de decisiones basada en riesgos

El uso de herramientas informáticas para la toma de decisiones basada en riesgos ha sido un tema de estudio significativo en el campo de los sistemas de información desde los años 70, especialmente en lo que respecta a los sistemas de apoyo a la decisión. Un ejemplo notable es el trabajo de Warenski , que utilizó inteligencia artificial para el análisis de riesgos de préstamos en la industria del papel y la pulpa mediante modelado financiero. Otim también ha realizado análisis recientes sobre la evaluación del valor y el riesgo en las inversiones en tecnología de la información,

destacando la complejidad y la multiplicidad de partes interesadas que conllevan estos proyectos, lo que requiere la consideración de la política organizacional.

Kozhikode y Li exploraron cómo el pluralismo político influye en la expansión de los bancos comerciales en India, integrando la gestión de riesgos en su análisis. En el ámbito industrial, la toma de decisiones no solo implica múltiples partes interesadas, sino también diversos criterios. Silvestri propuso una técnica de evaluación de riesgos basada en múltiples criterios para analizar la seguridad en la manufactura. Lakemond ofrece un método para la evaluación temprana de riesgos en el desarrollo de productos, permitiendo identificar desafíos potenciales desde las primeras etapas del proceso.

Machine learning

Arthur Samuel, pionero en el Machine Learning y la inteligencia artificial, definió el Machine Learning como "el campo de estudio que permite a las computadoras aprender sin estar explícitamente programadas para ello" (McClendon & Meghanathan, 2015). En los años 50, Samuel desarrolló un programa que aprendía a jugar a las damas, jugando miles de partidas contra sí mismo para determinar las mejores y peores posiciones en el tablero (Ng, 2012).

Tom Mitchell, otro líder en este campo, propuso en 1997 una definición formal: "un programa aprende de la experiencia E respecto a una tarea T y un rendimiento P, si el rendimiento medido por P en la tarea T mejora con la experiencia E". Por ejemplo, en

el caso del programa de damas de Samuel, la tarea T es jugar a las damas, el rendimiento P se mide por el porcentaje de partidas ganadas y la experiencia E es el entrenamiento jugando contra sí mismo (Mitchell, 1997).

El Machine Learning ha sido definido de varias maneras desde perspectivas matemáticas, algorítmicas y estadísticas. Un enfoque integrador lo describe como "un conjunto de técnicas dentro de la inteligencia artificial que utiliza métodos estadísticos para buscar patrones y crear máquinas inteligentes capaces de aprender y tomar decisiones basadas en datos empíricos de diversas fuentes" (Álogos, 2018). En este sentido, el Machine Learning es una subdisciplina de la inteligencia artificial que busca que las máquinas realicen tareas que requerirían inteligencia humana (Negnevitsky, 2002).

Dentro de la inteligencia artificial, existen varios subcampos que se especializan en diferentes aspectos del sistema cognitivo humano: el Machine Learning se centra en el reconocimiento de patrones y el aprendizaje; los sistemas expertos se enfocan en la toma de decisiones; los sistemas de búsqueda y planificación determinan la mejor secuencia de pasos para resolver problemas; el reconocimiento del habla y el procesamiento del lenguaje natural analizan y tratan la información recibida mediante el lenguaje humano; la percepción se refiere a la comprensión del entorno, y la robótica convierte las decisiones en acciones físicas (Russell & Norvig, 2010).

Machine Learning aplicado en los negocios

En el contexto empresarial actual, hay una creciente demanda de aplicaciones inteligentes que emplean técnicas de aprendizaje automático (machine learning) para optimizar la toma de decisiones. IBM, una multinacional líder en tecnología y consultoría, ha desarrollado numerosas de estas aplicaciones tanto para su uso interno como para sus clientes. Estas aplicaciones están diseñadas para realizar análisis de negocios y mejorar decisiones operativas y estratégicas (Apte, 2010).

La evolución del uso de datos en las empresas ha avanzado del análisis descriptivo al predictivo y, más recientemente, al análisis prescriptivo. El análisis descriptivo ofrece una visión retrospectiva del negocio, respondiendo preguntas sobre lo que ocurrió y dónde. El análisis predictivo, por otro lado, proporciona una visión futura, abordando preguntas sobre lo que podría suceder y qué pasaría si las tendencias actuales continúan. Estas avanzadas capacidades analíticas mejoran significativamente la competitividad empresarial (Apte, 2010).

Aprendizaje Supervisado

El aprendizaje supervisado es una técnica de machine learning donde la máquina aprende a partir de ejemplos proporcionados por un operador. Según la Asociación para el Progreso de la Dirección (2019), este método implica que el algoritmo de aprendizaje automático recibe un conjunto de datos conocidos que contienen tanto las entradas

como las salidas deseadas. El objetivo del algoritmo es encontrar un método para mapear estas entradas a las salidas correspondientes.

Durante este proceso, el operador ya conoce las respuestas correctas al problema. El algoritmo, entonces, identifica patrones en los datos, aprende de estas observaciones y comienza a hacer predicciones. A medida que el algoritmo realiza predicciones, el operador lo corrige, permitiendo que el algoritmo mejore su precisión y rendimiento progresivamente. Este ciclo de predicción y corrección continua hasta que el algoritmo alcanza un alto nivel de precisión y rendimiento (Rojas, 2020).

Es importante destacar que el aprendizaje supervisado se divide principalmente en dos categorías: clasificación y regresión. En la clasificación, el objetivo es asignar datos a categorías discretas, mientras que en la regresión, el objetivo es predecir valores continuos. Este enfoque es ampliamente utilizado en diversas aplicaciones prácticas, como la detección de fraudes, el diagnóstico médico y el análisis de mercado.

Análisis No Supervisado

El Machine Learning comprende varios tipos de aprendizaje, entre los cuales se encuentra el aprendizaje no supervisado. Este método permite que los algoritmos aprendan a partir de los datos proporcionados sin necesidad de una guía específica. En lugar de seguir instrucciones predefinidas, los algoritmos identifican nuevos patrones y extraen información debido a su diseño estructurado previamente (Rojas, 2020).

En el aprendizaje no supervisado, los algoritmos intentan reconocer patrones en los datos desde el primer momento en que interactúan con ellos. Una vez entrenados, estos algoritmos pueden analizar nuevos datos no etiquetados y, mediante los patrones aprendidos, determinar si los nuevos datos encajan en las clasificaciones previas. Este proceso se conoce como clasificación. Además, el algoritmo puede predecir nuevos datos basándose en los patrones previamente identificados (Gago, 2017).

Para aplicar el aprendizaje no supervisado, es necesario dividir los datos en conjuntos de entrenamiento y prueba. Los datos de entrenamiento establecen los parámetros que el algoritmo seguirá posteriormente, aprendiendo la lógica detrás de las relaciones entre los datos. Esto permite la clasificación de datos con patrones similares. Los datos de prueba, por su parte, evalúan el rendimiento del algoritmo. Sin embargo, en el aprendizaje no supervisado, el algoritmo debe comprender la lógica detrás de los datos sin etiquetas, permitiendo que el modelo obtenga los patrones necesarios para encajar con los datos de prueba y producir resultados predictivos. Un riesgo es el sobreajuste (overfitting), que ocurre cuando el modelo se ajusta demasiado a los datos de entrenamiento y falla al enfrentar datos nuevos que no tienen relación con los de entrenamiento, resultando en un mal rendimiento del modelo (Data Science Team, 2022).

Aprendizaje por Refuerzo

El aprendizaje por refuerzo es un enfoque dentro del machine learning que se centra en procesos de aprendizaje regulados. Según la Asociación para el Progreso de la Dirección (2019), en este tipo de aprendizaje, los algoritmos reciben un conjunto de acciones, parámetros y valores finales. El objetivo es que el algoritmo explore diferentes opciones y posibilidades, monitorizando y evaluando cada resultado para determinar cuál es el óptimo (Jordan, M. I., & Mitchell, T. M. , 2015).

Este método enseña a la máquina mediante el proceso de ensayo y error. A medida que el algoritmo interactúa con su entorno, aprende de experiencias pasadas y comienza a adaptar su enfoque en respuesta a la situación para lograr el mejor resultado. Este aprendizaje es continuo y dinámico, permitiendo que el algoritmo mejore su desempeño con el tiempo

SVM (Máquinas de vectores de soporte)

El algoritmo de máquinas de vectores de soporte (Support Vector Machine) es un clasificador discriminante cuyo objetivo es establecer un hiperplano en un espacio N-dimensional (donde N representa el número de variables independientes) que maximice la separación entre los datos de ambas clases (Gandhi, 2018). Este hiperplano actúa como una frontera de decisión que clasifica los datos, asignando una clase a cada lado de este.

Los vectores de soporte son los datos que están más cerca del hiperplano, afectando su posición y orientación. Encontrar un hiperplano que divida los datos no siempre tiene una única solución. La solución óptima es aquella que maximiza la distancia entre dos líneas paralelas situadas simétricamente a cada lado de la frontera de decisión, conteniendo en ellas a los vectores de soporte. Esta distancia se denomina margen. Los modelos con un margen grande reducen el error de generalización, mientras que los modelos con un margen pequeño son menos propensos al sobreajuste (overfitting) (Gandhi, 2018). En el algoritmo de vectores de soporte, si la salida lineal es mayor que 1 se asigna a una clase y si es menor que -1 se asigna a la otra (Gandhi, 2018). Por lo tanto, el rango que actúa como margen es $[-1,1]$.

Para problemas de clasificación no lineales, se emplean técnicas de mapeo y kernelización. Estas técnicas reorganizan los datos mediante funciones matemáticas, ya sea en el mismo plano o en planos de dimensiones superiores, para hacerlos separables (Román, 2019). Esta es una de las principales ventajas de los vectores de soporte frente a otros modelos, junto con su alta precisión y baja necesidad de potencia computacional (Gandhi, 2018). Además, estos principios pueden aplicarse para resolver problemas de regresión (Support Vector Regression).

Máquina de vectores de soporte para clasificación binaria

Una SVM de clasificación binaria busca encontrar el hiperplano que mejor divide el conjunto de datos de entrada en dos clases, asignando una etiqueta de clase a cada punto de datos en función de su posición con respecto a este hiperplano. Los

puntos de datos que están más cerca del hiperplano y que influyen en su posición se conocen como "vectores de soporte", ya que son críticos para la determinación del hiperplano de separación óptimo. Además, estas pueden utilizar funciones kernel para mapear los datos de entrada a un espacio de características de mayor dimensión, lo que permite encontrar hiperplanos no lineales en el espacio original (Goh, K.-S., Chang, E., & Cheng, K.-T., 2003).

Hiperplano:

Los hiperplanos se utilizan en machine learning, especialmente en técnicas como las máquinas de soporte vectorial (SVM), para separar datos en diferentes categorías. En este contexto, los hiperplanos también pueden ser utilizados para simular cambios de conceptos debido a su capacidad de ajustar suavemente su orientación y posición al modificar la magnitud de sus pesos.

El uso de hiperplanos para simular cambios de conceptos es particularmente útil en situaciones donde los datos pueden evolucionar con el tiempo. Al ajustar la orientación y posición de los hiperplanos, los modelos de machine learning pueden adaptarse a nuevas tendencias y patrones sin necesidad de un rediseño completo (Blanco, I. F, 2016).

Vectores de soporte

Los "vectores de soporte" son los puntos de datos de entrenamiento que se encuentran más cerca del hiperplano de separación óptimo y que influyen en la determinación de dicho hiperplano. Estos puntos de datos son críticos para la construcción del clasificador SVM, ya que definen la posición y orientación del hiperplano de separación entre las clases.

Esta definición destaca la importancia de los vectores de soporte en el proceso de entrenamiento de una SVM, ya que son los puntos clave que determinan la frontera de decisión entre las clases en el espacio de características de alta dimensión (Tasé, R. O.R.2016)

Parámetro de penalización ©

El parámetro de penalización en una SVM es un factor que controla la influencia de la regularización en la optimización del hiperplano de separación. Este parámetro, comúnmente denotado como C , equilibra la minimización del error de clasificación y la maximización del margen entre las clases. Un valor más alto de C permite a la SVM clasificar correctamente más puntos de datos de entrenamiento, pero puede resultar en un modelo más propenso al sobreajuste. Por otro lado, un valor más bajo de C favorece un margen más amplio pero puede llevar a un mayor error de clasificación en los datos de entrenamiento (Cristianini, N., & Shawe-Taylor, J., 2000).

Marco Legal

En el ámbito legal, en el Ecuador existen normativas importantes a su vez se han generado varios planes alineados a una normativa legal entre los cuales se destacan: La Ley Orgánica de Telecomunicaciones con El Plan Nacional de Telecomunicaciones y Tecnologías de Información del Ecuador 2016-2021. El Convenio Marco de Cooperación Interinstitucional firmado con los diferentes entes como el Ministerio de Telecomunicaciones y de La Sociedad de la Información, la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación. Los Planes Nacionales para el Desarrollo (2013-2017). La Ley Orgánica de Telecomunicaciones. El Código Orgánico de La Economía Social de los Conocimientos, Creatividad e Innovación.

Así como los artículos que contemplan el desarrollo de la investigación científica, y de la innovación tecnológica, de las TIC y diversas áreas del conocimiento que están en la Constitución de la República del Ecuador de los cuales se puede destacar los (Art. 281, Art. 385, Art. 387, Art. 423, Art. 388) que permiten alinearse y ser partícipes de la adopción de estas tecnologías emergentes para el cambio y adopción de una cultura de transformación digital para un mejor desarrollo a nivel empresarial en el ámbito local con proyección internacional, lo que se convierte en un impacto positivo.

Adicional a esto la Constitución de la República del Ecuador (2008) en su Art. 313 también señala que “El Estado se reserva el derecho de administrar, regular, controlar y gestionar los sectores estratégicos, de c41 sostenibilidad ambiental, precaución, prevención y eficiencia” (pág. 133). Y por último la Constitución de la República del Ecuador (2008) agrega que en su Art. 321 de la Carta Magna, “el Estado reconoce y garantiza el derecho a la propiedad en sus formas pública, privada, comunitaria, estatal, asociativa, cooperativa, mixta, y que deberá cumplir su función social y ambiental” (pág. 149)

CAPITULO 2

Metodología

Support Vector Machine (SVM)

Las Máquinas de Soporte Vectorial (SVM) fueron presentadas por primera vez en 1992 por Boser, Guyon y Vapnik en la conferencia COLT-92. Las SVM son métodos de aprendizaje supervisado que se utilizan para la clasificación y la regresión, y pertenecen a la familia de clasificadores lineales generalizados. En esencia, una SVM es una herramienta de predicción que emplea la teoría del aprendizaje automático para maximizar la precisión predictiva, al tiempo que evita el sobreajuste a los datos de entrenamiento.

Las SVM funcionan utilizando un espacio de hipótesis de funciones lineales en un espacio de características de alta dimensión. El modelo se entrena mediante un algoritmo de aprendizaje basado en la teoría de optimización, que implementa un sesgo de aprendizaje derivado de la teoría del aprendizaje estadístico. Esta metodología permite a las SVM identificar el hiperplano discriminante que maximiza el margen entre las clases, mejorando así la capacidad de generalización del modelo.

La popularidad inicial de las SVM se observó en la comunidad de NIPS y, con el tiempo, se han convertido en una parte integral de la investigación en aprendizaje automático a nivel mundial. Las SVM ganaron notoriedad en tareas de reconocimiento de escritura a mano, donde utilizaron mapas de píxeles como entrada y lograron una precisión comparable a redes neuronales sofisticadas con características elaboradas.

Los fundamentos teóricos de las SVM fueron desarrollados por Vapnik y ganaron popularidad debido a sus características prometedoras, como un mejor rendimiento empírico. La formulación de las SVM se basa en el principio de Minimización del Riesgo Estructural

(SRM), que ha demostrado ser superior al principio tradicional de Minimización del Riesgo Empírico (ERM) utilizado por las redes neuronales convencionales. El SRM minimiza un límite superior del riesgo esperado, mientras que el ERM se enfoca en minimizar el error en los datos de entrenamiento. Esta diferencia dota a las SVM de una mayor capacidad de generalización, que es esencial en el aprendizaje estadístico.

La capacidad de generalización de un modelo SVM entrenado de manera óptima se basa en la relación entre los vectores de soporte y el conjunto de datos de entrenamiento. El principio mencionado indica que un modelo SVM con menos vectores de soporte tiende a ser más compacto y simple, lo que generalmente conduce a un mejor rendimiento. Este enfoque es consistente con el principio de la Navaja de Ockham, que favorece las explicaciones más simples y menos complejas.

Para aplicar esta teoría en la metodología de SVM, se debe prestar atención al número de vectores de soporte en el modelo entrenado. Un menor número de vectores de soporte sugiere una mejor capacidad de generalización, ya que el modelo será menos propenso a sobreajustarse a los datos de entrenamiento. En situaciones donde los datos no son perfectamente separables, es crucial ajustar el modelo para encontrar un equilibrio que minimice el error de clasificación fuera de la muestra, sin depender exclusivamente de la cantidad de vectores de soporte.

En esta tesis, se emplea la metodología basada en SVM para abordar problemas específicos de clasificación. La elección de SVM se justifica por su robustez, precisión y capacidad de generalización, lo que las convierte en una herramienta ideal para el análisis y la interpretación de los datos en una aplicación práctica.

El método de Máquinas de Vectores de Soporte (SVM, por sus siglas en inglés) es un clasificador basado en discriminantes cuyo objetivo es encontrar el hiperplano óptimo que separe las diferentes clases de muestras. Este proceso se lleva a cabo resolviendo un problema de optimización cuadrática que se formula de la siguiente manera:

$$\min_{\omega, b} J = \frac{1}{2} \|\omega\|^2 + C \sum \xi_i$$

sujeto a las restricciones:

$$\min_{\omega, b} J = \frac{1}{2} \|\omega\|^2 + C \sum \xi_i$$

$$\xi_i \geq 0$$

Aquí, $X = \{x_i\}$ representa el conjunto de vectores de características y $Y = \{y_i\}$ el conjunto de etiquetas de clase del conjunto de datos de entrenamiento. El vector ω es el vector de pesos del hiperplano de decisión, mientras que b es el sesgo del modelo. Las variables ξ_i son las variables de holgura que indican la distancia de una instancia particular desde su lado correcto de la frontera de decisión y son distintas de cero para los ejemplos que violan la restricción $y_i(\omega * x_i + b) \geq 1$. El parámetro C controla la compensación entre la

maximización del margen y la minimización del error, siendo un valor mayor de C más enfocado en reducir el error.

Para casos donde los datos no son linealmente separables, SVM utiliza funciones kernel para realizar el cálculo del producto punto en un espacio de alta dimensión mapeado. La tarea de aprendizaje implica resolver el problema dual del anterior:

$$\max_{\alpha} L(\alpha) = \sum a_i - \frac{1}{2} \sum a_i a_j y_i y_j K(x_i, x_j)$$

sujeto a:

$$\sum a_i y_i = 0$$

Aquí, $K(\cdot, \cdot)$ es la función kernel $a_i y$ son las constantes multiplicativas de Lagrange asociadas con cada punto de datos del entrenamiento. En el punto óptimo de esta función, a_i toma valores de 0, entre 0 y C , o igual a C . Los vectores de entrada donde a_i es mayor que 0 se denominan vectores de soporte y son cruciales para la clasificación, ya que definen la frontera de decisión, mientras que el resto de los datos pueden ser ignorados.

El hiperplano de decisión óptimo se expresa como:

$$\omega = \sum a_i y_i \phi(x_i)$$

donde ϕ es una función de mapeo que satisface la relación $K(x_i; x_j) = \phi(x_i) * \phi(x_j)$

Para clasificar una nueva instancia de prueba x , SVM utiliza la siguiente función de decisión:

$$f(x) = \sum_{x_i \in SV} a_i y_i K(x, x_i) + b$$

donde SV es el conjunto de vectores de soporte. El valor de $f(x)$ indica la distancia de la instancia de prueba respecto al hiperplano de decisión, y el signo del valor determina la etiqueta de clase, ya sea positiva o negativa.

Máquinas de Vectores de Soporte Lineales y Conocimiento Previo

El problema de clasificar m puntos en un espacio n -dimensional puede representarse mediante la matriz A de tamaño $m * n$). Cada punto A_i se clasifica en una de dos clases, A^+ o A^- , según lo indicado por la matriz diagonal D con unos positivos o negativos en su diagonal. La máquina de vectores de soporte de programación lineal (LP-SVM) con kernel lineal se utiliza para resolver este problema, formulado como el siguiente programa lineal con el parámetro $\nu > 0$:

$$\min_{(\omega, \gamma, y) \in R^{n+1}} \{ \nu e^T y + \frac{1}{\|\omega\|} \text{ sujeto a } D(A\omega - \gamma e) + y \geq e, y \geq 0 \}$$

Aquí, $\|\cdot\|_1$ representa la norma $L1$, y es un vector de variables de holgura que miden el error empírico, y (ω, γ) determinan un plano separador. Alternativamente, este problema puede reformularse como:

$$\min_{(\omega, \gamma, y, t) \in R^{n+1+2n}} \{ \nu e^T t \text{ sujeto a } D(A\omega - \gamma e) + y \geq e, t \geq \omega \geq -t, y \geq 0 \}$$

donde e es un vector de unos. El vector ω es la normal a los planos limitantes:

$$x^T \omega = \gamma + 1, \quad x^T \omega = \gamma - 1$$

Estos planos separan los puntos de las clases A^+ y A^- , respectivamente. La constante γ establece su posición relativa al origen. Si las clases son linealmente separables, $\gamma = 0$, el plano $x^T \omega = \gamma + 1$ delimita todos los puntos de A^+ y el plano $x^T \omega = \gamma - 1$ delimita todos los puntos de A^- :

$$A_i \omega \geq \gamma + 1, \quad \text{para } D_{ii} = 1, \\ A_i \omega \leq \gamma - 1, \quad \text{para } D_{ii} = -1$$

El plano $x^T \omega = \gamma$, a medio camino entre los planos limitantes, actúa como un plano separador que separa completamente las clases si $\gamma = 0$, de lo contrario, lo hace aproximadamente.

Maximizar la distancia entre los planos, o "margen", mediante la norma

$L1 \|\omega\|_1$, mejora la capacidad clasificatoria de generalización del modelo. Si las clases no son separables linealmente, los planos limitantes proporcionan un "margen suave" determinado por la variable de error y :

$$A_i \omega + y_i \geq \gamma + 1, \quad \text{para } D_{ii} = 1, \\ A_i \omega - y_i \leq \gamma - 1, \quad \text{para } D_{ii} = -1.$$

Minimizar la norma $L1$ del error y con el peso v proporciona un plano separador aproximado:

$$x \in A^+ \text{ si } \text{signo}(x^T \omega - \gamma) = 1, \\ x \in A^- \text{ si } \text{signo}(x^T \omega - \gamma) = -1.$$

Consideremos ahora la información previa, donde todos los puntos x que cumplen las desigualdades lineales $Bx \leq b$ pertenecen a la clase A^+ . Esto implica que el conjunto de

conocimiento previo $\{x|Bx \leq b\}$ debe estar en el lado A^+ del plano limitante $x^T\omega = \gamma + 1$. Si esta implicación no se puede satisfacer exactamente, se pueden introducir variables de error de holgura. La implicación se puede reescribir como:

$$Bx \leq b \implies x^T\omega \geq \gamma + 1$$

Si esta implicación se mantiene, se deduce que $Bx \leq b$ y $x^T\omega < \gamma + 1$ no tienen solución simultáneamente, lo cual se puede comprobar mediante la condición:

$$B^T u + \omega = 0, \quad b^T u + \gamma + 1 \leq 0, \quad u \geq 0$$

Esta equivalencia se puede demostrar utilizando el teorema de Farkas, y es fundamental para el enfoque basado en conocimiento previo en las máquinas de vectores de soporte.

Proceso de R

Para iniciar el proceso, se realizó la carga de la base de datos utilizando el archivo CSV proporcionado. Este archivo se encuentra en la ruta `../DATA/Seguros CSV.csv`. La función `read.csv2` se seleccionó debido a que los campos en el archivo están delimitados por punto y coma (;) además que, esta función carga un archivo con formato tabular y genera un dataframe, asignando las líneas a casos y los campos a variables dentro del archivo., El código utilizado fue:

```
Seguros = read.csv2("../DATA/Seguros CSV.csv")
```

Una vez cargados los datos en el data frame denominado Seguros, se realizó una depuración inicial, eliminando las columnas que no eran relevantes para el análisis posterior. En este caso, el operador dentro de `[, -c]` (indica que se excluyeron las columnas) 1, 2, 4, 5, 7, 9 y 10, como se muestra en el siguiente código:

```
Seguros <- Seguros [, -c (1,2,4,5,7,9,10)]
```

Luego de ello se convirtió en la columna “VALOR_DE_POLIZA._ANUAL” del data frame Seguros a un tipo de dato numérico. Debido a que esta columna inicialmente estaba en un formato de factor, esta conversión asegura que los valores se tratan como números para facilitar cálculos posteriores.

```
Seguros$VALOR_DE_POLIZA._ANUAL<-  
as.numeric(Seguros$VALOR_DE_POLIZA._ANUAL)
```

La última modificación que le realizamos a la base es:

```
segurosCL <- na.omit(Seguros)
```

Lo cual creó un nuevo data frame llamado segurosCL que es una versión de Seguros sin valores faltantes (NA). La función na.omit elimina todas las filas que contienen cualquier valor faltante, garantizando que el data frame resultante segurosCL no tenga valores NA.

Una vez que la base de datos está lista, es crucial descargar e instalar una serie de paquetes necesarios. Para poder hacer uso de ellos, los instalamos mediante la función “install.packages”. Después, es necesario cargar las bibliotecas correspondientes mediante la función “**library**”. Este proceso garantiza que todas las herramientas y funcionalidades requeridas estén disponibles para el análisis y modelado de datos.

Paquetes por instalar:

```
install.packages("e1071")
```

```
install.packages("ggplot2")
```

```
install.packages("lattice")
```

```
install.packages("caret")
```

```
library(e1071)
```

```
library(ggplot2)
```

```
library(lattice)
```

```
library(caret)
```

A continuación, se presenta una descripción detallada de cada biblioteca utilizada y su propósito específico en el contexto del análisis de este proyecto.

- **e1071:** Este paquete está diseñado para ofrecer una experiencia intuitiva en el ajuste de modelos y la predicción de nuevos datos. Es esencial para implementar Support Vector Machines (SVM). La interfaz del paquete e1071 permite ajustar modelos y realizar predicciones utilizando dos enfoques principales: la interfaz de vectores/matrices y la interfaz de fórmulas. De manera automática, el paquete selecciona el modo adecuado en función del tipo de variable dependiente (y). Si “y” es un factor, el paquete activa el modo de clasificación; si “y” es numérico, se utiliza el modo de regresión. En casos donde “y” se omite, el paquete asume que la tarea es la detección de novedades.

- **Ggplot2:** Es un paquete de visualización de datos que permite crear gráficos complejos de una manera sencilla y elegante. Este paquete se basa en la gramática de gráficos y es ampliamente utilizada para generar visualizaciones que ayudan a entender mejor los patrones y tendencias en los datos.
- **Lattice:** El paquete de extensión Lattice es una implementación de los gráficos Trellis para R. Es un sistema de visualización de datos de alto nivel, potente y elegante, con énfasis en datos multivariantes. Está diseñado para satisfacer la mayoría de las necesidades gráficas típicas con un ajuste mínimo, pero también se puede extender fácilmente para manejar requisitos no estándar. La interfaz de usuario de Lattice consta principalmente de varias funciones genéricas de alto nivel, cada una diseñada para crear un tipo particular de visualización por defecto. Aunque estas funciones producen salidas diferentes, comparten muchas características comunes, reflejadas en varios argumentos que afectan las visualizaciones de manera similar. Es relevante llamar a este paquete del modelado de SVM, donde la comprensión y visualización de datos multidimensionales es crucial.
- **Caret:** El paquete “caret” (acrónimo de Classification And Regression Training) es fundamental para la optimización y simplificación del proceso de entrenamiento de modelos complejos de regresión y clasificación en R. Este paquete facilita la construcción y evaluación de modelos al proporcionar herramientas integradas para ajustar parámetros, seleccionar características y evaluar el rendimiento del modelo.

Al utilizar “caret”, se pueden realizar evaluaciones exhaustivas mediante remuestreo, seleccionar el modelo óptimo entre varios parámetros y estimar el rendimiento del modelo a partir de un conjunto de entrenamiento. Su capacidad para manejar una amplia gama de técnicas de remuestreo y métodos de evaluación permite una personalización detallada del proceso de modelado, asegurando que se obtenga el mejor rendimiento posible. La función “train”, una de las principales del paquete, es especialmente valiosa, ya que permite ajustar modelos de manera eficiente y evaluar su desempeño de forma sistemática.

Implementación del Modelo de Support Vector Machine (SVM)

Para la implementación del modelo de Support Vector Machine (SVM) en el análisis de datos, se siguió el siguiente procedimiento:

1) Transformación de variable clasificatoria a factor

```
segurosCL$TIPO_DE_COBERTURA <- factor(segurosCL$TIPO_DE_COBERTURA)
```

La variable **TIPO_DE_COBERTURA**, que representa la variable respuesta del modelo, se convirtió en un factor para asegurar que el modelo de clasificación pueda interpretarla correctamente. Este paso es esencial para tratar la variable como categórica en lugar de continua.

```
segurosCL$TIPO_DE_COBERTURA <- factor(segurosCL$TIPO_DE_COBERTURA)
```

2) *Se planta una semilla*

```
set.seed(123)
```

Se estableció una semilla para garantizar la reproducibilidad de los resultados. La función “set.seed()” en R es fundamental para la generación de números aleatorios, ya que asegura que los resultados sean consistentes en análisis estadísticos y simulaciones. En el contexto de nuestro análisis, el uso de “set.seed(123)” garantiza que cualquier partición aleatoria de los datos sea replicable en futuras ejecuciones del código. Aunque se pueden seleccionar diversas semillas, en este contexto de clasificación, la semilla “123” se ha elegido como la más apropiada para asegurar la consistencia de los resultados.

3) *Entrenamos los datos*

Para entrenar el modelo, utilizamos la función *createDataPartition()* del paquete *caret*, que permite crear particiones de datos para entrenamiento y prueba, garantizando que las distribuciones de las clases en ambos conjuntos sean representativas del conjunto de datos original. Específicamente, se empleó el siguiente comando:

```
train <- createDataPartition(segurosCL$TIPO_DE_COBERTURA, p=0.7, list = FALSE)
```

En este contexto, el parámetro *p* se establece en 0.7, lo que significa que el 70% de los datos se destinan al conjunto de entrenamiento, mientras que el 30% restante se reserva para la evaluación del modelo. Este enfoque es crucial para mantener una distribución equilibrada de las clases, especialmente cuando la variable dependiente (*TIPO_DE_COBERTURA*) es categórica.

El parámetro *list = FALSE* controla el formato de la salida de la función. Al establecer *list = FALSE*, *createDataPartition()* devuelve una matriz o un vector con los índices de las observaciones que componen el conjunto de entrenamiento, lo que facilita la indexación de los datos para la creación del conjunto de entrenamiento. Si *list* fuera *TRUE*, la función devolvería una lista con índices para cada partición.

Posteriormente, se entrenó un modelo de Support Vector Machine (SVM) utilizando un kernel lineal. La fórmula *segurosCL\$TIPO_DE_COBERTURA* indica que *TIPO_DE_COBERTURA* es la variable dependiente y que todas las demás variables en el conjunto de datos *segurosCL* se utilizan como variables predictoras.

4) Armar la máquina de soporte vectorial

En este paso, se procede a ajustar un modelo de Máquina de Soporte Vectorial (SVM) utilizando el conjunto de datos de entrenamiento. Para ello, se emplea la función *svm()* del paquete **e1071** en R. El modelo se ajusta mediante la siguiente instrucción:

```
MSV <- svm(TIPO_DE_COBERTURA~., data = segurosCL[train,], kernel = "linear")
```

En esta expresión, *TIPO_DE_COBERTURA ~.* indica que *TIPO_DE_COBERTURA* es la variable dependiente que se pretende predecir, mientras que el símbolo “.” especifica que se utilizarán todas las demás variables del conjunto de datos como predictores. La opción *data = segurosCL[train,]* establece que el modelo debe ajustarse usando el conjunto de datos de entrenamiento, que se selecciona a partir de los índices generados en el paso anterior. La opción *kernel = "linear"* define el tipo de núcleo del modelo como lineal, lo que implica que el modelo intentará encontrar un hiperplano lineal para separar las clases en el espacio de características.

Una vez ajustado el modelo, se utiliza la función `summary()` para obtener un resumen detallado del ajuste:

summary(MSV)

Este resumen proporciona información crucial sobre el modelo, incluyendo el número de vectores de soporte, los coeficientes del modelo y la función del margen, lo cual es fundamental para evaluar la calidad del ajuste y la complejidad del modelo.

Para evaluar el rendimiento del modelo en el conjunto de datos de entrenamiento, se genera una matriz de confusión que compara las predicciones del modelo con las clases verdaderas. La instrucción utilizada es:

```
table(segurosCL[train,]$TIPO_DE_COBERTURA, fitted(MSV), dnn = c("Actual",  
"Predicho"))
```

En este código, `segurosCL[train,]$TIPO_DE_COBERTURA` representa las clases verdaderas de las observaciones en el conjunto de datos de entrenamiento. Este vector contiene las etiquetas reales que se están tratando de predecir. Por otro lado, `fitted(MSV)` se encarga de devolver las predicciones realizadas por el modelo SVM ajustado (MSV) para el mismo conjunto de datos. Estos valores corresponden a las clases que el modelo ha predicho para cada observación en el conjunto de entrenamiento.

La función `table()` crea una tabla de contingencia, también conocida como matriz de confusión, que muestra la frecuencia de las combinaciones de los valores reales y predichos.

En esta matriz, el argumento `dnn` (**que significa `dimnames names`**) = `c("Actual", "Predicho")` se utiliza para etiquetar las dimensiones de la tabla. En la matriz de confusión resultante:

"Actual" se refiere a las clases verdaderas, es decir, las etiquetas reales de las observaciones.

"Predicho" se refiere a las clases predichas por el modelo.

Estas etiquetas facilitan la interpretación de la matriz, permitiendo una clara comparación entre las clases verdaderas y las predicciones del modelo. La matriz de confusión proporciona una visión detallada de cuántas predicciones fueron correctas y cuántas fueron incorrectas, organizadas por clase. Esta evaluación es crucial para medir la precisión del modelo y para identificar posibles áreas de mejora.

5) *Realizar la predicción*

Una vez que el modelo de Máquina de Soporte Vectorial (SVM) ha sido ajustado utilizando el conjunto de datos de entrenamiento, procedemos a utilizar dicho modelo para realizar predicciones sobre el conjunto completo de datos. Esto se logra mediante la función **`predict()`** en R, como se muestra a continuación:

```
PRED <- predict(MSV,segurosCL)
```

En este código, la función **`predict()`** se emplea para generar predicciones basadas en el modelo SVM previamente ajustado (MSV). El argumento `MSV` representa el modelo que contiene la información y los parámetros aprendidos durante el proceso de entrenamiento. El conjunto de datos completo **`segurosCL`** se utiliza para obtener las predicciones, abarcando tanto las observaciones de entrenamiento como las de prueba.

El resultado de esta operación es un vector de predicciones, **PRED**, donde cada elemento corresponde a la clase predicha por el modelo para cada observación en el conjunto de datos. Estas predicciones son fundamentales para evaluar el desempeño del modelo y su capacidad de generalización a datos no vistos durante el entrenamiento.

La función **predict()** permite aplicar el modelo ajustado a nuevos datos, facilitando la obtención de predicciones que pueden ser comparadas con las clases verdaderas para medir la precisión y eficacia del modelo en el conjunto de datos completo.

6) *Armamos la matriz de confusión*

Para evaluar el rendimiento del modelo de Máquina de Soporte Vectorial (SVM) en el conjunto de datos completo, se construyó una matriz de confusión. Esta matriz permite comparar las predicciones del modelo con las clases verdaderas de las observaciones, proporcionando una herramienta esencial para la evaluación de modelos de clasificación. El proceso se realizó mediante el siguiente código:

```
MSVMC <- table(segurosCL$TIPO_DE_COBERTURA, PRED, dnn = c("Actual",  
"Predicho"))
```

MSVMC

En este código, **segurosCL\$TIPO_DE_COBERTURA** representa las clases verdaderas para todas las observaciones en el conjunto de datos completo, mientras que **PRED** contiene las clases predichas por el modelo SVM para las mismas observaciones. La función **table()** crea una tabla de contingencia, también conocida como matriz de confusión, que muestra la frecuencia de las combinaciones de los valores reales y predichos.

El argumento `dnn = c("Actual", "Predicho")` se utiliza para etiquetar las dimensiones de la matriz, donde **"Actual"** se refiere a las clases verdaderas y **"Predicho"** se refiere a las clases predichas por el modelo. Esta matriz permite analizar detalladamente el rendimiento del modelo, identificando verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.

La matriz de confusión es fundamental para medir la precisión del modelo, así como otras métricas de rendimiento relevantes, y proporciona una visión clara de las áreas en las que el modelo puede necesitar mejoras.

7) *Visualización de Resultados*

Para analizar y comprender mejor el rendimiento del modelo de Máquina de Soporte Vectorial (SVM), se realizaron varias visualizaciones de los resultados obtenidos. Utilizando la función `plot()` en R, se crearon gráficos que muestran la relación entre las variables del conjunto de datos y las predicciones del modelo.

El código utilizado para estas visualizaciones es el siguiente:

- EDAD

```
plot(MSV, data = segurosCL[train,], EDAD ~ COBERTURA)
```

```
plot(MSV, data = segurosCL[train,], EDAD ~ NUMERO_DE_BENEFICIARIOS) plot(MSV,
```

```
data = segurosCL[train,], EDAD ~ VALOR_DE_POLIZA._ANUAL)
```


- COBERTURA

```
plot(MSV, data = segurosCL[train,], COBERTURA ~ EDAD)
```

```
plot(MSV, data = segurosCL[train,], COBERTURA ~ NUMERO_DE_BENEFICIARIOS)
```

```
plot(MSV, data = segurosCL[train,], COBERTURA ~ VALOR_DE_POLIZA._ANUAL)
```

- NUMERO_DE_BENEFICIARIOS

```
plot(MSV, data = segurosCL[train,], NUMERO_DE_BENEFICIARIOS ~ EDAD)
```

```
plot(MSV, data = segurosCL[train,], NUMERO_DE_BENEFICIARIOS ~ COBERTURA)
```

```
plot(MSV, data = segurosCL[train,], NUMERO_DE_BENEFICIARIOS ~  
VALOR_DE_POLIZA._ANUAL)
```

En este código, la función **plot()** se emplea para crear gráficos que visualizan las relaciones entre diferentes variables del conjunto de datos.

Primero, se visualiza la relación entre **VALOR_DE_POLIZA._ANUAL** y otras variables como **COBERTURA**, **NUMERO_DE_BENEFICIARIOS**, y **EDAD**. La variable dependiente (**VALOR_DE_POLIZA._ANUAL**) se coloca en el eje y, mientras que las variables independientes (**COBERTURA**, **NUMERO_DE_BENEFICIARIOS**, y **EDAD**) se colocan en el eje x en gráficos separados.

Luego, se visualiza la relación de **EDAD** con las variables **COBERTURA**, **NUMERO_DE_BENEFICIARIOS**, y **VALOR_DE_POLIZA._ANUAL**. Aquí, **EDAD** es la variable dependiente en el eje y y las otras variables se colocan en el eje x en gráficos separados.

Posteriormente, se visualiza la relación de **COBERTURA** con las variables **EDAD**, **NUMERO_DE_BENEFICIARIOS**, y **VALOR_DE_POLIZA._ANUAL**. En este caso, **COBERTURA** es la variable dependiente en el eje y y las otras variables se colocan en el eje x.

Finalmente, se visualiza la relación de **NUMERO_DE_BENEFICIARIOS** con las variables **EDAD**, **COBERTURA**, y **VALOR_DE_POLIZA._ANUAL**. Aquí, **NUMERO_DE_BENEFICIARIOS** es la variable dependiente en el eje “y” y las otras variables se colocan en el eje x.

El propósito de estas visualizaciones es proporcionar una representación gráfica de cómo las predicciones del modelo SVM se relacionan con las variables del conjunto de datos, especialmente en relación con la variable de respuesta, que en este caso es el tipo de cobertura (nacional o internacional). Estas visualizaciones permiten evaluar la efectividad del modelo en la clasificación de los datos y proporcionan una comprensión más profunda de los factores que influyen en el riesgo asociado a cada tipo de cobertura.

Durante el proceso de visualización de los resultados del modelo de Máquina de Soporte Vectorial (SVM), se emplearon varias funciones y operadores clave para manipular y graficar los datos.

Primero, en la línea de código `segurosCL[train,]`, se está realizando una subasignación del dataframe `segurosCL`. La notación `segurosCL[train,]` se utiliza para seleccionar un subconjunto del dataframe, donde `train` es un vector de índices de filas generado previamente mediante la función `createDataPartition()`. La coma, dentro de los corchetes indica que se seleccionan todas las columnas de las filas especificadas por `train`. Esto es útil para trabajar únicamente con los datos de entrenamiento al entrenar y evaluar el modelo SVM.

Por otro lado, el operador ~ se utiliza para especificar fórmulas en R, tanto en el modelado como en la graficación de datos. En el contexto del modelado, como en `MSV <- svm(TIPO_DE_COBERTURA ~ ., data = segurosCL[train,], kernel = "linear")`, la fórmula `TIPO_DE_COBERTURA ~ .` indica que `TIPO_DE_COBERTURA` es la variable dependiente y el punto `.` representa todas las demás variables del dataframe `segurosCL[train,]` como variables independientes. Esto permite incluir todas las variables relevantes en el modelo de manera eficiente.

En el contexto de la visualización, como en `plot(MSV, data = segurosCL[train,], VALOR_DE_POLIZA._ANUAL ~ COBERTURA)`, la fórmula `VALOR_DE_POLIZA._ANUAL ~ COBERTURA` especifica que `VALOR_DE_POLIZA._ANUAL` debe ser graficado en el eje y y `COBERTURA` en el eje x. Este enfoque facilita la creación de gráficos que ilustran la relación entre la variable dependiente y una o más variables independientes, lo cual es esencial para interpretar y comunicar los resultados del modelo de manera clara y efectiva.

8) *Paso la clasificación a la base*

En el paso final del análisis, se integra las predicciones generadas por el modelo de Máquina de Soporte Vectorial (SVM) en el conjunto de datos “Seguroscl”, se asignan estas predicciones a una nueva columna en el dataframe. El proceso se lleva a cabo mediante el siguiente código:

```
segurosCL$PRED <- predict(MSV, segurosCL)
```

Utilizando la función predict(), se aplican las predicciones del modelo a todo el conjunto de datos segurosCL. El resultado de esta predicción se almacena en una nueva columna llamada PRED dentro del dataframe. Este proceso permite añadir una dimensión adicional a los datos, proporcionando las clasificaciones generadas por el modelo para cada observación en el conjunto de datos. De este modo, es posible analizar y comparar las predicciones del modelo con las etiquetas reales para una evaluación más completa del rendimiento del modelo en la clasificación de riesgos asociados a los diferentes tipos de cobertura.

Análisis de Resultados:

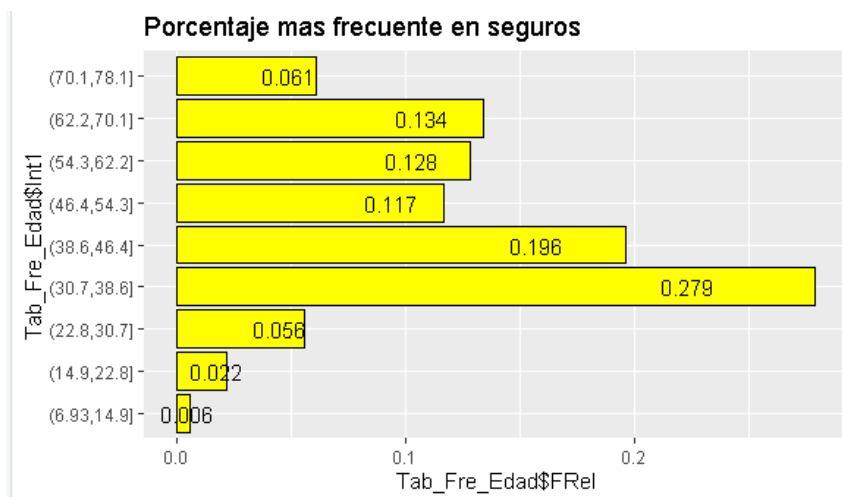


Figura 1: Porcentaje más frecuente en seguros

El gráfico presentado es un histograma de frecuencias que ilustra la distribución de edades de las personas que han contratado un seguro. En este gráfico, cada barra horizontal representa un rango de edad y su longitud refleja el porcentaje de contrataciones dentro de ese intervalo. El eje X, que es horizontal, muestra la frecuencia relativa de seguros por rango de edad, con valores que van de 0 a 0.2, indicando que el porcentaje máximo representado es del 20%. El eje Y, vertical, representa los rangos de edad, agrupados en intervalos como (14.9, 22.8], que abarca de 14.9 a 22.8 años

Las barras del gráfico revelan que el rango de edad más frecuente para la contratación de seguros es (30.7, 38.6], con una frecuencia relativa de 0.279, lo que equivale al 27.9% de los seguros, seguido por el rango (38.6, 46.4], con una frecuencia relativa de 0.196, es decir, el 19.6%. Otros rangos de edad incluyen: (62.7, 70.8] con un 13.4%, (54.3, 62.2] con un 12.8%, (46.4, 54.3] con un 11.7%, (70.1, 78.1] con un 6.1%, (22.8, 30.7] con un 5.6%, (14.9, 22.8] con un 2%, y (6.93, 14.9] con un 0.6%.

El análisis indica que el grupo de edad con mayor porcentaje de contratación de seguros está entre 30.7 y 38.8 años, lo que sugiere una mayor propensión a asegurar entre adultos jóvenes y de mediana edad. Esto puede deberse a una mayor conciencia de la necesidad de protección financiera, responsabilidades adquiridas, o una mayor capacidad adquisitiva. Además, se observa una disminución en la contratación de seguros tanto en los rangos de edad más jóvenes (14.9-22.8) como en los más altos (70.1-78), con una concentración notable de seguros en el grupo de edad entre 30 y 60 años. En conclusión, los adultos jóvenes y de mediana edad son los que más frecuentemente contratan seguros, mientras que las frecuencias son más bajas en los extremos etarios.

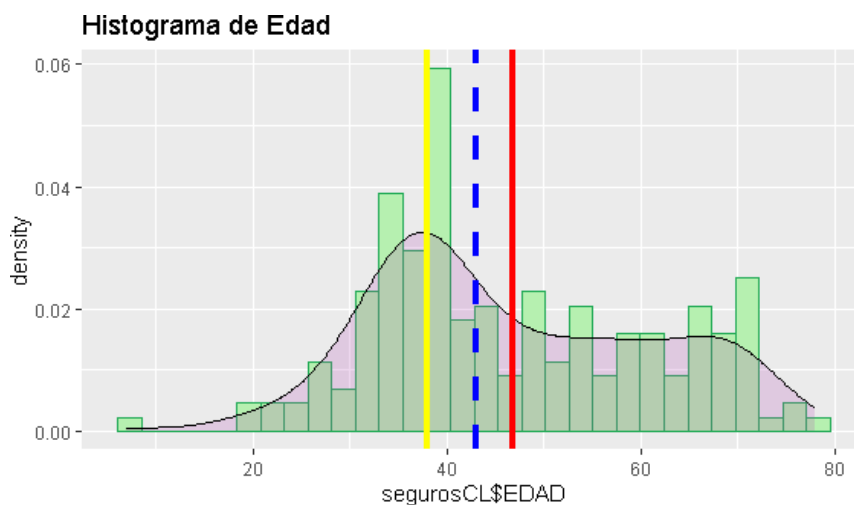


Figura 2: Histograma de Edad

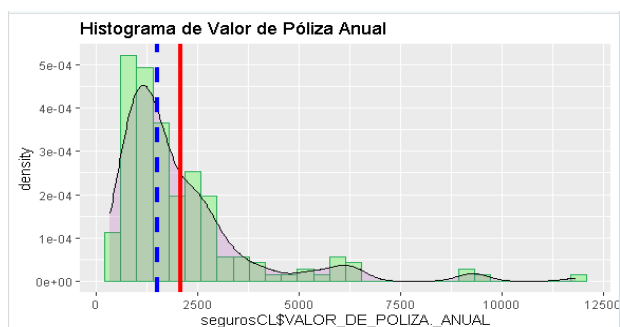
En el gráfico titulado "Histograma de Edad," se presenta la distribución de edades de los asegurados. El eje X muestra las edades, divididas en intervalos específicos según la variable segurosCL\$EDAD, mientras que el eje Y representa la densidad de la distribución de estas edades.

La curva negra superpuesta al histograma es una estimación de la densidad de probabilidad, proporcionando una visión más clara de la distribución subyacente de los datos. Al analizar los elementos del gráfico, se puede observar que la media de las edades se encuentra aproximadamente entre 46 y 47 años, y la mediana está entre los 43 y 47 años, lo que sugiere una distribución cercana a la normalidad. La moda se sitúa en los 38 años.

La distribución presenta una ligera asimetría positiva, lo que indica que la mayoría de los asegurados tienen edades cercanas al promedio, con una menor proporción en los extremos de edad, especialmente en los más jóvenes. La desviación estándar varía entre 14 y 44, con un límite superior de 60 años y un límite inferior de 42 años. El rango de edad confiable para la adquisición del seguro se sitúa entre los 32 y 60 años.

Finalmente, la línea roja vertical que indica la media de la distribución se encuentra alrededor de los 40 años, lo que sugiere que el grupo etario predominante en este conjunto de datos se aproxima a esta edad.

Figura 3: Histograma de Valor de Póliza



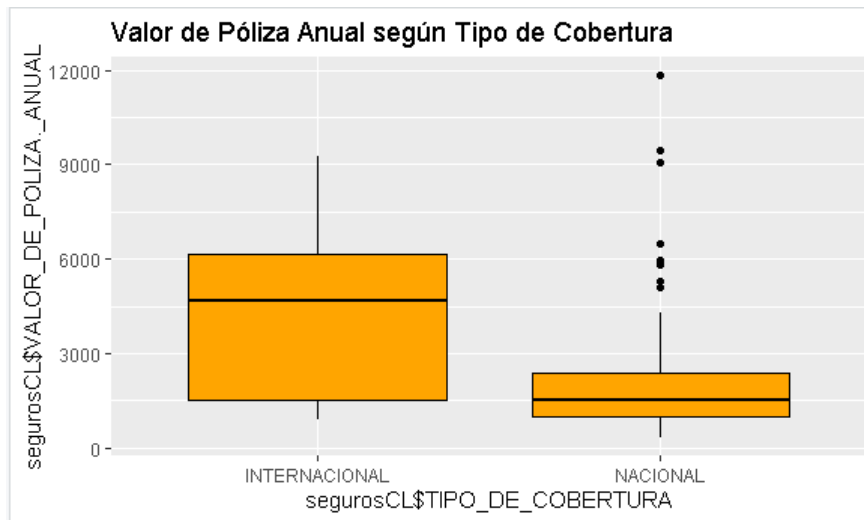
En el gráfico, se visualiza la distribución de los valores anuales de las pólizas de seguro. El eje X representa los valores anuales de las pólizas según la variable `segurosCL$VALOR_DE_POLIZA_ANUAL`, mientras que el eje Y refleja la densidad de la distribución de estos valores.

La distribución observada muestra una tendencia hacia la normalidad, aunque se aprecia una disminución hacia la derecha. Esta disminución podría estar relacionada con la reducción del poder adquisitivo a medida que el costo de la póliza aumenta. La distribución es asimétrica positiva, lo que indica que la mayoría de las pólizas tienen un valor relativamente bajo, pero también existen algunas pólizas con valores significativamente más altos.

Las barras verdes del histograma reflejan la frecuencia de los valores de las pólizas en intervalos específicos, mientras que la línea negra de densidad suaviza la distribución, proporcionando una visión más clara de la densidad de probabilidad de los valores de las pólizas. La línea roja vertical en el gráfico representa el valor promedio de las pólizas, ofreciendo una referencia clara del valor típico de una póliza en este conjunto de datos.

La mayoría de los valores de las pólizas anuales se concentran en el rango inferior de la escala, con una notable cantidad de pólizas valoradas en menos de \$2500. La distribución muestra una cola larga hacia la derecha, lo que sugiere la existencia de algunas pólizas con valores significativamente más altos, aunque son menos comunes. La línea roja indica que la media del valor de las pólizas se sitúa alrededor de \$2500.

Figura 4: Valor de Póliza Anual según Tipo de Cobertura



El gráfico titulado "Valor de Póliza Anual según Tipo de Cobertura" muestra la distribución de los valores anuales de las pólizas de seguro, diferenciadas por tipo de cobertura: Internacional y Nacional.

En el eje X (horizontal) se representan los tipos de cobertura, "INTERNACIONAL" y "NACIONAL", mientras que en el eje Y (vertical) se muestran los valores anuales de las pólizas de seguro, según la variable `segurosCL$VALOR_DE_POLIZA_ANUAL`.

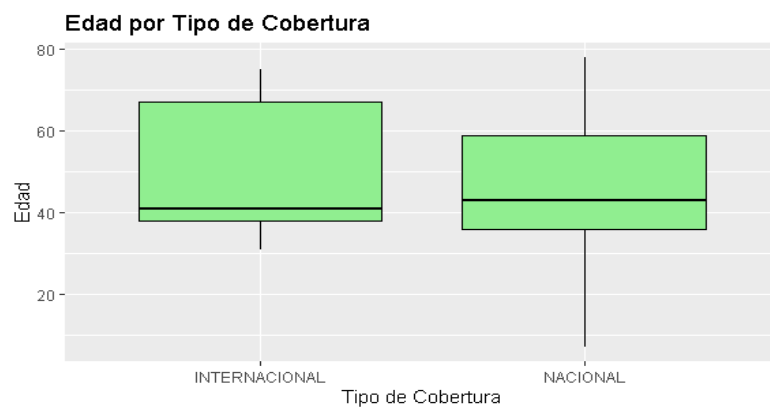
Los boxplots del gráfico proporcionan una visión clara de la distribución de los valores de las pólizas para cada tipo de cobertura. En el caso de las pólizas con cobertura Internacional, la mediana del valor anual está alrededor de los \$6000. El rango intercuartílico (IQR), que va desde el primer cuartil (Q1) hasta el tercer cuartil (Q3), se encuentra aproximadamente entre \$4000 y \$8000. Es notable que no se identifican valores atípicos (outliers) significativos, lo que sugiere una distribución más normal y concentrada de los valores.

Por otro lado, las pólizas con cobertura Nacional tienen una mediana alrededor de los \$2000. El rango intercuartílico (IQR) está aproximadamente entre \$1000 y \$3000. Sin embargo, se observan varios valores atípicos (outliers) que se extienden por encima de los \$5000, indicando la existencia de algunas pólizas con valores significativamente más altos. Esto refleja que hay un grupo de asegurados con mayor capacidad de pago, aunque no forman parte del conjunto típico de datos.

El gráfico de boxplots confirma visualmente estas diferencias, mostrando una distribución más normal y concentrada para las pólizas Internacionales, mientras que las pólizas Nacionales presentan una mayor dispersión y una cola superior más pronunciada.

En conclusión, las pólizas de seguro con cobertura Internacional tienden a tener valores anuales más altos en comparación con las pólizas con cobertura Nacional. La mediana y el rango intercuartílico son mayores para las pólizas Internacionales, lo que indica una distribución más uniforme y concentrada. En contraste, las pólizas Nacionales presentan una mayor dispersión, con varios valores atípicos elevados. Esto sugiere que, aunque la mayoría de las pólizas Nacionales son de menor valor, existe un grupo de asegurados que paga significativamente más, posiblemente reflejando la obligatoriedad de ciertos seguros de menor valor para un amplio sector de la población.

Figura 5: Edad por Tipo de Cobertura

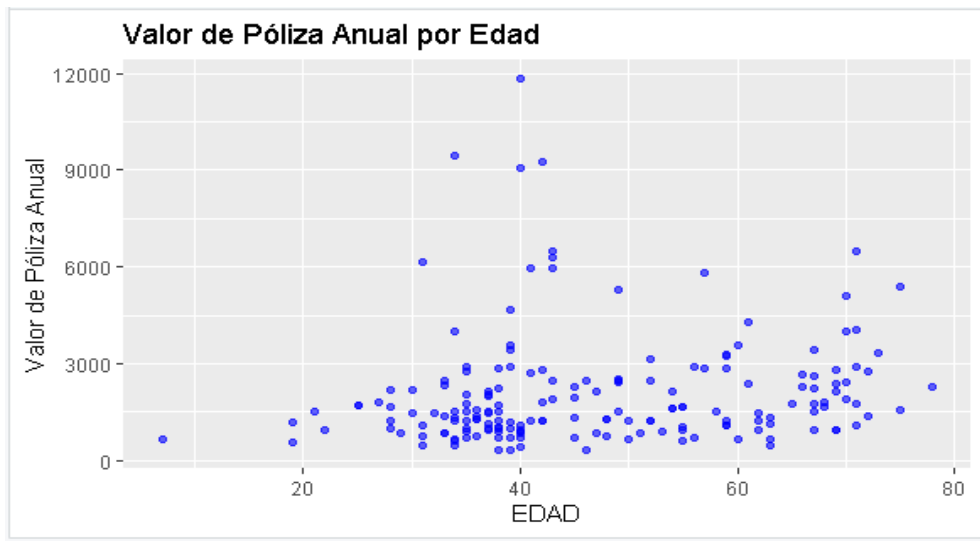


Dentro del gráfico de “Edad por Tipo de Cobertura” que se presenta en este análisis se permite examinar la distribución de la edad de los asegurados, diferenciada por tipo de cobertura: Internacional y Nacional. La mediana de la edad en la cobertura internacional se ubica en torno a los 40 años. Por otro lado, la mediana de la edad en la cobertura nacional es similar a la internacional ya que también se sitúa alrededor de los 40 años, indicando que este tipo de cobertura es más accesible o preferida por personas de adultez media.

El rango intercuartílico (IQR) de la cobertura internacional, que va cerca de 40 a 70 años, refleja una dispersión de 30 años, mientras que en la cobertura nacional, el IQR se sitúa entre un rango aproximado a 40 y 60 años. Sin embargo, el rango de edad total en la cobertura internacional es más amplio, abarcando de 30 a 80 años, mientras que en la cobertura nacional se limita a un rango de 10 a 80 años. Esta diferencia sugiere que la cobertura internacional puede estar dirigida a un grupo demográfico más diverso en términos de edad.

Es importante destacar que no se identifican valores atípicos (outliers) en ninguno de los grupos, lo que indica una distribución uniforme de los datos sin desviaciones extremas. En resumen, el análisis de este gráfico sugiere que la cobertura internacional es más común entre los asegurados de mayor edad, mientras que la cobertura nacional se concentra en un rango de edad más joven. Estas diferencias en la distribución etaria podrían estar relacionadas con las distintas necesidades y preferencias de los asegurados según su edad, y tienen implicaciones importantes para la segmentación del mercado y la estrategia de producto de las aseguradora.

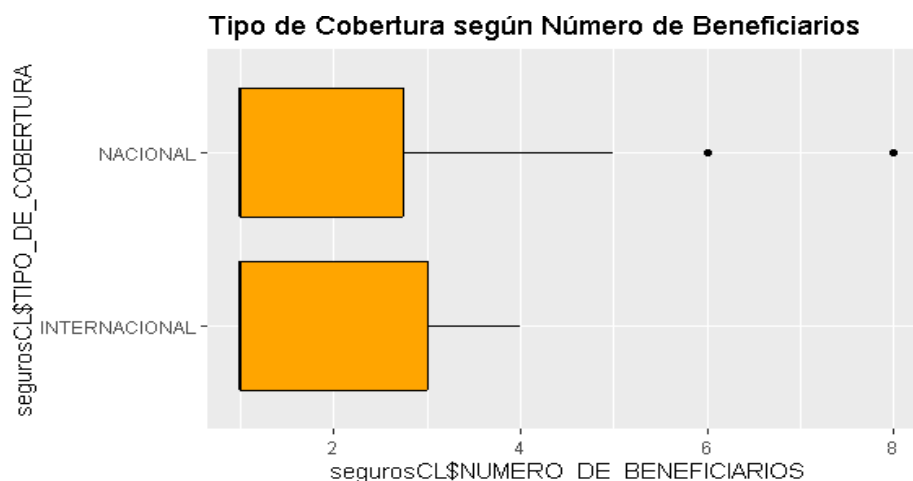
Figura 6: Valor de Póliza Anual por Edad



El gráfico de dispersión analiza la relación entre la edad de los asegurados y el valor de la póliza anual. La mayor concentración de puntos se observa en el rango de pólizas con valores entre 0 y 4000, lo que indica que la mayoría de los asegurados paga un valor moderado por sus pólizas. Sin embargo, también se observa la presencia de varios valores atípicos, con pólizas que superan los 6000 y alcanzan hasta 12000. Estos valores extremos, aunque menos frecuentes, son significativos y no están ligados a una edad específica, sino que se distribuyen a lo largo de diferentes edades.

Un punto crucial que destaca en este gráfico es la relación entre la edad y el costo de las pólizas: a medida que la edad de los asegurados aumenta, especialmente a partir de los 40 años, se observa una mayor dispersión en los valores de las pólizas, sugiriendo que el costo tiende a aumentar con la edad. Aunque no hay una tendencia uniforme de incremento con la edad, los asegurados más jóvenes (por debajo de 40 años) tienden a pagar valores de póliza más bajos, mientras que, en edades superiores, algunos asegurados pagan cantidades significativamente más altas, reflejando una mayor variabilidad y la influencia de factores adicionales.

Figura 7: Tipo de Cobertura según Número de Beneficiarios



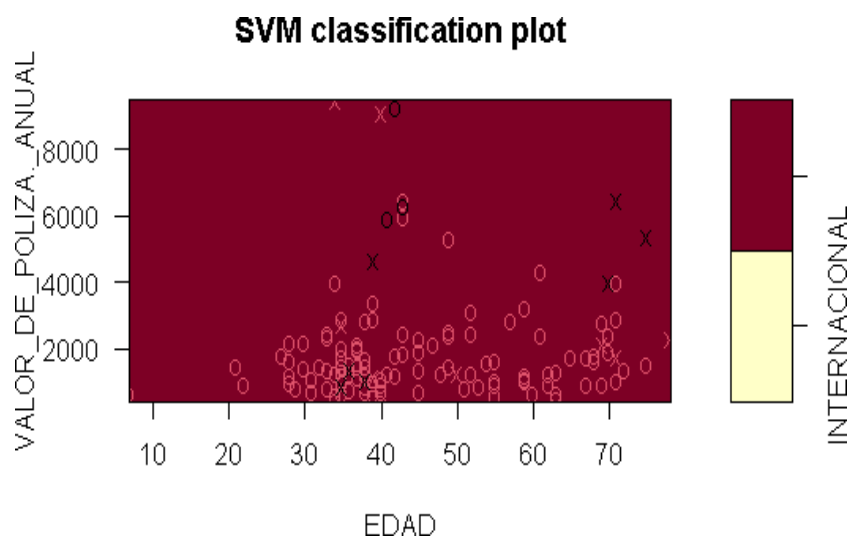
El gráfico de “Tipo de Cobertura según Número de Beneficiarios” muestra la distribución del número de beneficiarios para dos tipos de cobertura: Nacional e Internacional.

En cuanto a la mediana, ambos tipos de cobertura presentan un valor igual de

1. El rango intercuartílico (IQR) para la cobertura Nacional se extiende de 1 a 5, y se observan valores atípicos en 6 y 8 beneficiarios. Por otro lado, en la cobertura Internacional, el IQR se ubica entre 1 y 4, sin valores atípicos identificados. Esto sugiere que la cobertura Nacional presenta algunos casos con un número significativamente mayor de beneficiarios en comparación con la cobertura Internacional.

Esta variabilidad podría reflejar necesidades específicas de ciertos clientes, lo que indica que la compañía de seguros debería investigar estos casos extremos en la cobertura Nacional para entender mejor las razones detrás de estos valores atípicos. Aunque no se observan diferencias significativas en la mediana entre los tipos de cobertura, la presencia de valores atípicos en la cobertura Nacional podría señalar una mayor variabilidad en comparación con la cobertura internacional.

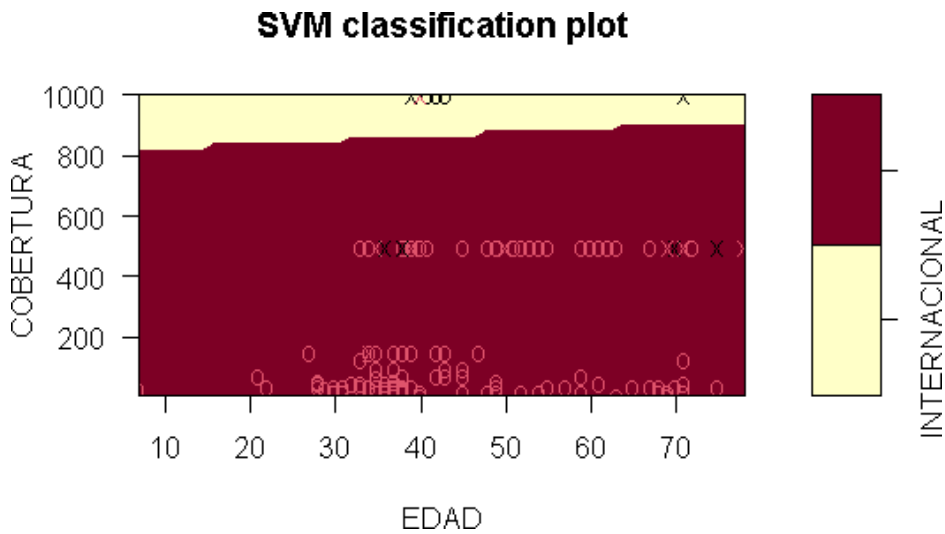
Figura 8: Valor de Póliza Anual por Edad



En este gráfico, bajo la misma estructura donde se sigue categorizando pólizas de seguros como Nacionales o Internacionales en función de dos variables: "Edad" y "Valor de Póliza Anual". La interpretación del gráfico revela que la clasificación del modelo SVM está significativamente influenciada por la edad en el caso de los seguros nacionales. En este contexto, el valor de la póliza tiende a aumentar a medida que las personas envejecen, lo que refleja una relación directa entre la edad y la preferencia por adquirir seguros nacionales. A medida que las personas envejecen, se perciben como más vulnerables y, por lo tanto, más propensas a asegurar su bienestar a través de pólizas más costosas.

Sin embargo, este solo es significativo para seguros nacionales, las personas toman los seguros mediante que uno envejece, al contrario del seguro internacional que nada tiene que ver con pólizas basadas en la edad.

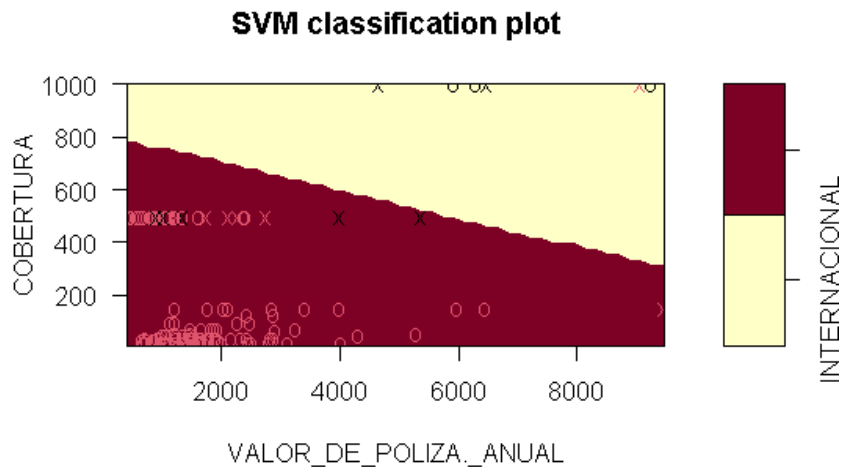
Figura 9: Cobertura por Edad



En el gráfico, se analizan las variables "cobertura" y "edad", revelando una clara distinción entre seguros nacionales e internacionales. Se observa que la mayoría de las pólizas con un nivel de cobertura superior a aproximadamente 800 están clasificadas como internacionales. Esto sugiere que los seguros internacionales suelen estar asociados a coberturas más altas, posiblemente debido a los mayores riesgos y costos asociados con la cobertura en el extranjero.

Por otro lado, aquellas pólizas con una cobertura inferior se clasifican como nacionales, lo que podría indicar que los seguros nacionales tienden a ser más económicos o que los asegurados optan por niveles de cobertura más bajos. Este análisis sugiere que el nivel de cobertura es un factor determinante en la diferenciación entre estos dos tipos de seguros. El modelo SVM ha identificado una relación muy fuerte entre el nivel de cobertura y el tipo de seguro.

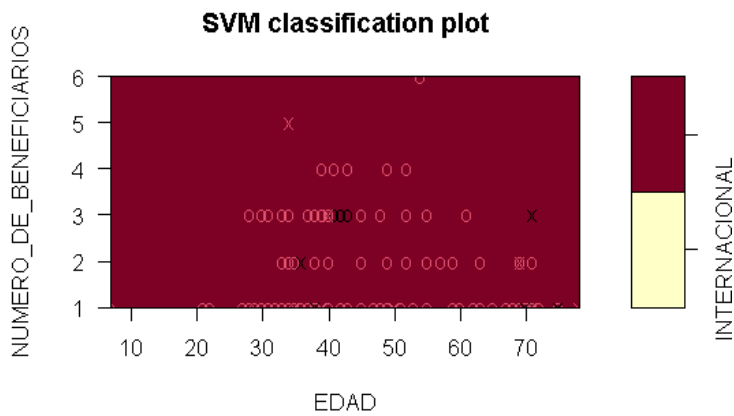
Figura 10: Cobertura por Valor de Póliza Anual



El gráfico presenta la clasificación de pólizas de seguro en dos categorías: "Internacional" y "Nacional", en función de dos variables clave: el valor anual de la póliza y la cobertura ofrecida.

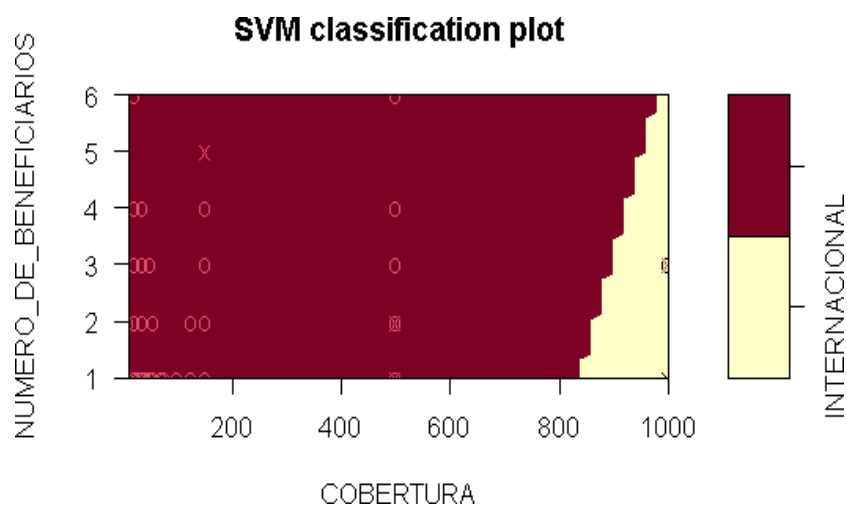
El análisis del gráfico revela una tendencia clara de que a medida que la cobertura de la póliza aumenta, también tiende a incrementarse el valor anual de la misma, lo que conduce a que estas pólizas sean clasificadas predominantemente como "Internacional". En contraste, las pólizas con menor cobertura y menor valor anual son clasificadas principalmente como "Nacional". Esto sugiere que las pólizas que ofrecen una mayor cobertura requieren un mayor costo anual, y para alcanzar niveles más altos de cobertura, es necesario recurrir a pólizas de tipo internacional.

Figura 11: Número de Beneficiarios por Edad



Dentro de este gráfico las variables consideradas son la edad de los asegurados y el número de beneficiarios asociados a la póliza. Específicamente, casi todos los datos caen en la región rojo oscuro, lo que indica que el modelo SVM asigna una gran mayoría de pólizas a la categoría "nacional" sin una diferenciación significativa en base a la edad o al número de beneficiarios. Sin embargo, se puede observar que de 30 a 70 existe una constante entre el número de beneficiarios siendo este de 2 a 3 y teniendo ciertos valores atípicos de 5 beneficiarios.

Figura 12: Número de Beneficiarios por Cobertura



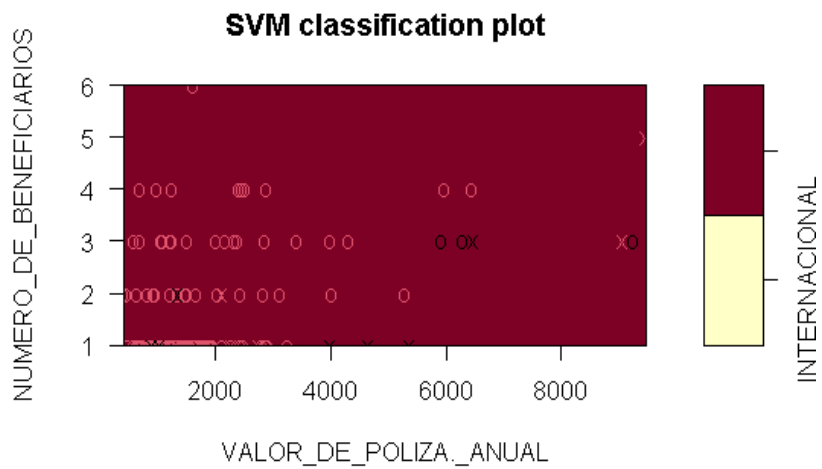
El gráfico presenta la relación entre la cobertura y el número de beneficiarios en las pólizas de seguro, y cómo estas variables influyen en la clasificación de las pólizas como "Internacional" o "Nacional".

El gráfico revela un patrón de que a medida que la cobertura aumenta, disminuye el número de beneficiarios asociados a la póliza. Esto se evidencia en la concentración de pólizas con menor cobertura (situadas a la izquierda del gráfico) que cuentan con un mayor número de beneficiarios, mientras que las pólizas con mayor cobertura (situadas hacia la derecha) tienden a tener un número reducido de beneficiarios.

El modelo SVM clasifica la mayoría de las pólizas con alta cobertura y un menor número de beneficiarios como "Internacional", mientras que aquellas con cobertura baja, pero más beneficiarios, tienden a ser clasificadas como "Nacional".

Con esto se puede observar una relación inversa entre la cobertura y el número de beneficiarios, donde las pólizas con coberturas más amplias tienden a tener menos beneficiarios y son clasificadas como internacionales.

Figura 14: Número de Beneficiarios por Valor de Póliza Anual



El gráfico nos presenta la relación entre el valor anual de las pólizas de seguros y el número de beneficiarios cubiertos. Se puede observar que la mayoría de las pólizas se concentran en un rango de valor anual entre 0 y 6000, y cubren principalmente entre 1 y 4 beneficiarios. Como nos muestra el gráfico, la nacional tienden a ofrecer una mayor cobertura de beneficiarios a un costo anual más bajo, en contraste con las pólizas de cobertura Internacional, que podrían asociarse con un menor número de beneficiarios y un costo anual más elevado. Esto sugiere que, si la prioridad es asegurar a toda la familia, las pólizas nacionales podrían ser una opción más asequible y adecuada.

Discusión

En el contexto de nuestro proyecto, es crucial garantizar la exactitud de los resultados y la rigurosidad de las metodologías utilizadas. Por lo tanto, presentaremos de manera concisa un artículo científico titulado "Improving the Performance of the Support Vector Machine in Insurance Risk Classification: A Comparative Study", el cual ofrece una visión valiosa sobre la aplicación y optimización de las Máquinas de Vectores de Soporte (SVM) en la clasificación de riesgos de seguros. Este estudio es relevante para nuestra investigación, ya que proporciona una comparación detallada de las metodologías para mejorar el rendimiento de los modelos SVM en el contexto de seguros.

El artículo examina el uso de SVM para la clasificación de riesgos en seguros, destacando la importancia de abordar los desafíos asociados con los datos incompletos y la necesidad de optimizar los parámetros del modelo para lograr una mayor precisión. La investigación presenta dos enfoques principales: la imputación de valores faltantes utilizando redes autoasociativas y la optimización de parámetros a través de algoritmos genéticos, junto con la reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA). Estos métodos buscan mejorar la calidad de los datos y ajustar el modelo SVM para superar las limitaciones de los datos incompletos, demostrando que una combinación de técnicas puede llevar a una mejora significativa en el rendimiento del modelo.

En nuestro proyecto, la aplicación de SVM mantiene el mismo enfoque de seguros, donde se centra en la clasificación de pólizas en categorías de internacional y nacional, utilizando variables como el valor anual de la póliza, la edad del asegurado, el número de beneficiarios y el nivel de cobertura.

Al igual que en el estudio comparativo, nuestra metodología se enfoca en asegurar la integridad de los datos, incluyendo la imputación de valores faltantes para minimizar el riesgo de sesgos y garantizar un análisis más fiable en una base de datos completa.

Los resultados obtenidos, analizados a través de gráficos de dispersión y boxplots, ofrecen una comprensión detallada de cómo las variables predictoras mencionadas anteriormente permiten evaluar el impacto de cada variable sobre la variable de respuesta, que en este caso determina si una póliza es más adecuada para el mercado nacional o internacional. Al observar cómo estas variables afectan la clasificación, se puede deducir si es más viable trabajar con seguros nacionales o internacionales, basándose en patrones emergentes y diferencias significativas entre las categorías analizadas. Este enfoque confirma la validez de las técnicas empleadas y subraya la importancia de una gestión cuidadosa de los datos para obtener resultados precisos y confiables.

Los principios y generalizaciones derivadas de ambos estudios indican que la calidad de los datos y la optimización del modelo son fundamentales para el rendimiento de los SVM en la clasificación de riesgos. Tanto el estudio comparativo como nuestra investigación destacan la necesidad de abordar problemas de datos incompletos y utilizar técnicas avanzadas para mejorar la precisión de los modelos. Estos hallazgos subrayan que, independientemente de las variaciones en las técnicas específicas, el objetivo común es mejorar la capacidad predictiva y la confiabilidad de los modelos SVM en contextos complejos.

Al presentar el artículo “Improving the Performance of the Support Vector Machine in Insurance Risk Classification: A Comparative Study”, buscamos respaldar la implementación de SVM en nuestra tesis y demostrar que, como en el estudio comparativo, nuestro enfoque también se basa en principios sólidos de manejo de datos y optimización del modelo.

Ambos estudios reflejan la versatilidad de SVM y proporcionan valiosas perspectivas para mejorar la clasificación en el área de seguros, destacando la importancia de técnicas rigurosas para la gestión de riesgos.

Conclusión

Esta investigación ha establecido una base teórica que subraya la importancia de la gestión de riesgos en los negocios internacionales, destacando que la integración de la tecnología es clave para la sostenibilidad y el crecimiento empresarial en un entorno globalizado. Se han expuesto las principales herramientas para mitigar riesgos y se ha resaltado la necesidad de un marco conceptual claro que guíe su implementación. Por ello, un enfoque disciplinado y estructurado en la gestión de riesgos, alineando estrategias, procesos y tecnologías, es esencial para alcanzar los objetivos en el ámbito internacional.

Además, el desarrollo de un sistema de clasificación automática ha demostrado la eficacia del aprendizaje automático en la optimización de la toma de decisiones, enriqueciendo el conocimiento existente al proponer una metodología innovadora y rigurosa para la gestión de riesgos.

Por último, el análisis realizado muestra que la compañía de seguros enfrenta serias limitaciones en su capacidad para competir en los mercados internacionales, lo que incrementa los riesgos al intentar expandirse. Esta situación ha llevado a la empresa a centrarse en el mercado nacional, que es más estable y manejable. Sin embargo, si bien esta estrategia es segura a corto plazo, podría no ser suficiente a largo plazo si no se desarrollan las competencias necesarias para la competencia internacional.

Este resultado evidencia que la falta de internacionalización no solo limita el potencial de crecimiento de las compañías de seguros, sino que también las vuelve vulnerables frente a competidores que ya han logrado éxito en el mercado global. Para mitigar este riesgo y asegurar la competitividad futura, es indispensable que las aseguradoras consideren invertir en una estrategia

de internacionalización más agresiva, creando alianzas estratégicas y mejorando su propuesta de valor global, incluyendo la adopción de nuevas tecnologías.

Aunque los desafíos son significativos, diversificar las inversiones en mercados extranjeros no solo reduce los riesgos inherentes al mercado interno, sino que también abre nuevas oportunidades de crecimiento. Esto requiere que la aseguradora reconsidere sus estrategias actuales, enfocados en superar los obstáculos que frenan la expansión internacional y en aprovechar las oportunidades que surjan en este proceso.

Referencias

- Algoritmos de aprendizaje automático para análisis y predicción de datos Resumen Abstract Palabras clave Keyword Introducció. (16 de 04 de 2018). Recuperado el 24 de August de 2024, de REDICCES: http://redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
- Betancourt, G. (2005). LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). Scientia et Technica. Obtenido de <https://moodle2.utp.edu.co/index.php/revistaciencia/article/view/6895/4139>
- Business intelligence y la toma de decisiones financieras: una aproximación teórica. (s.f.). Recuperado el 24 de August de 2024, de Redalyc: <https://www.redalyc.org/pdf/5177/517751547010.pdf>
- Cortes, C. V. V. (1995). Support-Vector Networks. Kluwer Academic Publishers, Boston. Manufactured in The Netherlands., 277-279.
- Corvalan, J. (20 de 09 de 2017). Inteligencia artificial: retos, desafíos y oportunidades - Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia*. Recuperado el 24 de August de 2024, de Redalyc: <https://www.redalyc.org/journal/5340/534057837015/html/>
- Fuzzy sigmoid kernel for support vector classifiers. (2004). Science Direct, 62. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0925231204003352>
- Geometric Interpretation of v-SVM Classifiers (2024), de A Geometric Interpretation of v-SVM Classifiers: https://proceedings.neurips.cc/paper_files/paper/1999/file/7fea637fd6d02b8f0adf6f7dc36aed93-Paper.pdf
- Holos. (2016). Nuevo algoritmo multclasificador para flujos de datos con cambios de concepto. 2. Obtenido de <https://www.redalyc.org/pdf/4815/481554865003.pdf>
- Inteligencia artificial: retos, desafíos y oportunidades - Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia*. (s.f.). Recuperado el 24 de August de 2024, de Redalyc: <https://www.redalyc.org/journal/5340/534057837015/html/>
- Krogh, G. (02 de 2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. Obtenido de Science Direct: https://www.sciencedirect.com/science/article/pii/S0148296320306512?ref=pdf_download&fr=RR-2&rr=8b85c70648938de5

- León, O. (2023). Impacto de las capacidades de análisis de big data en la innovación empresarial. Obtenido de <https://www.redalyc.org/journal/2913/291376231005/html/>
- León, O. (2023, mayo 06). Impacto de las capacidades de análisis de big data en la innovación empresarial. Redalyc. Retrieved May 29, 2024, from <https://www.redalyc.org/journal/2913/291376231005/html/>
- León, O. (21 de 11 de 2022). Impacto de las capacidades de análisis de big data en la innovación empresarial. Recuperado el 24 de August de 2024, de Redalyc: <https://www.redalyc.org/journal/2913/291376231005/html/>
- Miller, K. D. (1992). A Framework for Integrated Risk Management in International Business. *Journal of International Business Studies*, 23(2), 311–331. doi: 10.1057/palgrave.jibs.8490270 Recuperado de <https://sci-hub.se/https://doi.org/10.1057/palgrave.jibs.8490270>
- Morales, E., González, J., & Escalante, H. J. (s.f.). Máquinas de soporte vectorial. Recuperado de <https://ccc.inaoep.mx/~hugojair/Courses/MachineLearning15/svm2.pdf>
- Murillo, M., & Caceres, G. (12 de 2013). Business intelligence y la toma de decisiones financieras: una aproximación teórica. Recuperado el 24 de August de 2024, de Redalyc: <https://www.redalyc.org/pdf/5177/517751547010.pdf>
- Ordoñez, D. (06 de 2022). Modelo predictivo basado en Machine Learning dirigido a PYMES de venta, caso de estudio Bluefields. Obtenido de <https://revistasnicaragua.cnu.edu.ni/index.php/Interculturalidad/article/view/7530/9389>
- Rengifo, R. (01 de 04 de 2024). Capacidades de análisis de big data en las empresas: una revisión sistemática de la literatura. Obtenido de ResearchGate: https://www.researchgate.net/publication/380144656_Capacidades_de_analisis_de_big_data_en_las_empresas_una_revision_sistemica_de_la_literatura
- Repositorio Digital de Ciencia y Cultura de El Salvador REDICCES: Algoritmos de aprendizaje automático para análisis y predicción de datos. (15 de Octubre de 2018). Recuperado el 24 de August de 2024, de REDICCES: <http://redicces.org.sv/jspui/handle/10972/3626>
- Rodrigo, J. A. (2 de January de 2021). Máquinas de Vector Soporte (SVM) con Python. Recuperado el 24 de August de 2024, de Cienciadedatos.net: <https://cienciadedatos.net/documentos/py24-svm-python>
- Rojas, E. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. ProQuest. Obtenido de <https://www.proquest.com/docview/2388304894?sourcetype=Scholarly%20Journals>

- Sánchez Anzola, N. (2015). Máquinas de soporte vectorial y redes neuronales artificiales en la predicción del movimiento. Recuperado de <https://core.ac.uk/download/pdf/230113657.pdf>
- Shetewy, N. (10 de 2022). Impact of financial development and internet use on export growth: New evidence from machine learning models. *ScienceDirect*, 61. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0275531922000319?via%3Dihub#preview-section-cited-by>
- Soman, K., Loganathan, R., & Ajay, V. (2011). *Machine Learning with SVM and Other Kernel Methods*. PHI Learning Pvt. Ltd., 2009.
- Support Vector Machines. (s.f.). Recuperado el 24 de August de 2024, de Stanford Engineering Everywhere: <https://see.stanford.edu/materials/aimlcs229/cs229-notes3.pdf>
- The Normal Distribution - Mathematics A-Level Revision. (2024). Recuperado el 24 de August de 2024, de Revision Maths: https://revisionmaths.com/advanced-level-maths-revision/statistics/normal-distribution#google_vignette
- Tovar, C. (02 de 05 de 2017). Investigación sobre la aplicación de business intelligence en la gestión de las pymes de Argentina. Obtenido de https://www.palermo.edu/economicas/cbrs/pdf/pbr15/PBR_15_05_Tovar.pdf
- Weiße, A., Wellein, G., Alvermann, A., & Fehske, H. (2006). The kernel polynomial method. *Reviews of Modern Physics*, 78(1), 275–306. doi:10.1103/revmodphys.78.275
Recuperado de <https://sci-hub.se/https://doi.org/10.1103/RevModPhys.78.275>

Anexos

Base de datos

NOMBRE DE CLIENTE	PLAN	VALOR_DE_P OLIZA ANUAL	FECHA DE NACIMIENTO	EDAD	Numero de Poliza	COBERTUR/ VIGENCIA	COMPANIA	# DE BENEFICIA RIOS	TIPO DE COBERTU RA	
JONATHAN OMAR CAMINO GARCIA	BMI SUPPORT	\$ 136.656,00	26/1/1988	36	A123G10598	500	44577	BMI	2	INTERNACIONAL
JONATHAN OMAR CAMINO GARCIA	ASISPRO 20	\$ 1.296,00	26/1/1988	36	318689	20	44577	ASISKEN	2	NACIONAL
CAROLINA LILIBETH MOREIRA PONCE	BMI SUPPORT	\$ 9.018,00	31/1/1989	35	A122G11693	500	44577	BMI	1	INTERNACIONAL
GALO IZQUIERDO AUTORADIADOR	BUPA	\$ 1.183.104,00	22/10/1983	40	22263262	10	44593	BUPA	8	NACIONAL
LEONARDO DAVID CONSTANTINE LEON	BMI GMM	\$ 100.992,00	6/3/1987	37	GM22G25912	500	44593	BMI	3	NACIONAL
LEONARDO DAVID CONSTANTINE LEON	MEDIKEN 20K	\$ 203.208,00	6/3/1987	37	GM22G25912	20	44593	MEDIKEN	3	NACIONAL
BORIS GEOVANNY GARCIA RODRIGUEZ	BMI HOSPICARE	\$ 169.104,00	19/8/1985	38	HC22G06958	75	44608	BMI	1	NACIONAL
MARITZA GALINKA IZQUIERDO VALLEJO	PLAN 80	\$ 13.086,00	11/6/1961	63	146326	10	44593		1	NACIONAL
JOHANNA MELISSA GALECIO WILLEGAS	ASISKEN 20K	\$ 174.108,00	9/6/1989	35	27735	20	44612	ASISKEN	1	NACIONAL
RODRIGO GONZALO CAAMAÑO PERROTA	ASISKEN 20K	\$ 165.276,00	19/5/1956	68	27655	20	44612	ASISKEN	1	NACIONAL
FARUTH NAVIL CORTEZ VERA	BMI IDEAL	\$ 594.612,00	12/4/1983	41	A122G12034	1000	44621	BMI	3	INTERNACIONAL
AGUIRRE MORENO ANDY FABRICIO	BMI SIGMA	\$ 40.116,00	2/8/1989	34	S122G36751	150	44621	BMI	3	NACIONAL
NANCY MIRELLA FREIRE YAGUAL	BMI GMM	\$ 95.988,00	22/1/1955	69	GM21G17234	500	44621	BMI	1	NACIONAL
JONATHAN RICHARD DURAN CRUZ	BMI GMM	\$ 69.588,00	27/8/1985	38	GM22G26712	500	44636	BMI	3	NACIONAL
JUAN PABLO BURGOS MARENGO	PLAN 80	\$ 417,00	14/4/1984	40	GM20G16189	500	44621	BMI	2	NACIONAL
TOMAS DANIEL BAZAN VALVERDE	BMI GMM	\$ 5.976,00	15/8/1989	34	GM22G26744	500	44636	BMI	3	NACIONAL
JORGE EDUARDO BURGOS VENEGAS	PLAN 80	\$ 960,00	14/4/1984	40	146406	10	44635		1	NACIONAL
MARIANITA GARDENIA LLERENA MONTES	BMI GMM	\$ 145.487,00	17/9/1961	62	GM22Q26977	500	44652	BMI	1	NACIONAL
JENNIFER ROXANA AUREA SORIANO	BMI SIGMA	\$ 213.876,00	31/10/1986	37	S122G37024	150	44645	BMI	1	NACIONAL
LEON ZAMBRANO JENNY COLOMBIA	BMI GMM	\$ 60.744,00	13/8/1988	55	GM22G27051	500	44652	BMI	1	NACIONAL
ALCIVAR AREVALO FREDDY ENRIQUE	BMI GMM	\$ 2.480,00	23/9/1974	49	GM24G45980	20	44652	BMI	4	NACIONAL
ALCIVAR AREVALO FREDDY ENRIQUE	MEDIKEN 20K	\$ 24.258,00	23/9/1974	49	GM24G45980	500	0	MEDIKEN	4	NACIONAL
ALCIVAR FERNANDEZ FREDDY ENRIQUE	BMI GMM	\$ 84.396,00	2/8/1976	47	GM22G27063	500	44652	BMI	1	NACIONAL
HUGO OSWALDO VANONI BRAVO	BMI IDEAL	\$ 648.192,00	22/4/1953	71	GM21G18702	1000	44667	BMI	3	INTERNACIONAL
MARIA DEL ROCIO IÑIGA URGILLES	BMI GMM	\$ 125.844,00	29/9/1971	52	GM22G27285	500	44667	BMI	3	NACIONAL
MARIA ANGELICA YEPEZ PIZA	BMI HOSPICARE	\$ 325.632,00	11/5/1965	59	HC22G07142	75	44667	BMI	1	NACIONAL
CHECA BARROS ALEXANDRA VICENTA	BMI GMM	\$ 108.852,00	5/4/1965	59	GM22G27301	500	44667	BMI	1	NACIONAL
CHECA BARROS ALEXANDRA VICENTA	MEDIKEN 20K	\$ 1.071,00	5/4/1965	59	GM22G27301	20	44666	MEDIKEN	1	NACIONAL
ZULLY VANESSA GONZALEZ CEVALLOS	MEDIKEN 45K	\$ 122.688,00	12/7/1986	38	GM21G18702	20	44659	MEDIKEN	3	NACIONAL
HUGO OSWALDO VANONI BRAVO	ASISKEN 20K	\$ 408.152,00	22/4/1953	71	28345	20	44671	ASISKEN	2	NACIONAL
HUGO OSWALDO VANONI BRAVO	ASISKEN 50K	\$ 10.734,00	22/4/1953	71	28345	50	44666	ASISKEN	1	NACIONAL
PAOLA MARCELA LOPEZ ECHAVARRIA	BMI GMM	\$ 75.456,00	13/4/1976	48	GM22G27576	500	44682	BMI	1	NACIONAL
MORENO ENCALADA MARIA MARGARITA	BMI GMM	\$ 2.910,00	28/2/1968	56	GM22G27611	500	44682	BMI	1	NACIONAL
RICHARD CARLOS SANTOS MORENO	BMI SUPPORT	\$ 104.472,00	21/2/1986	38	A122G12762	500	44682	BMI	1	INTERNACIONAL
PAOLA MARCELA LOPEZ ECHAVARRIA	MEDIKEN 20K	\$ 126.708,00	13/4/1976	48		20	44682	MEDIKEN	1	NACIONAL
RODRIGUEZ UZCATEGUI SANDRA CAROLINA	BMI INFINITY	\$ 276.756,00	1/1/1989	35	IK22G02710	500	44682	BMI	1	NACIONAL
CARLOS SAMANIEGO SANTOS RICAURTE	PLAN 80	\$ 95.988,00	4/1/1955	69	146602	10	44682		2	NACIONAL
MACIAS QUINTON UBALDO ELADIO	BMI SUPPORT	\$ 537.876,00	12/9/1948	75	A122G12991	500	44697	BMI	1	INTERNACIONAL
GUILLEN SALGUERO ZAY DORA	BMI SIGMA	\$ 284.684,00	26/5/1965	59	S122G37922	150	44713	BMI	1	NACIONAL
LUIS GILBERTO GARCIA RODRIGUEZ	BMI HOSPICARE	\$ 193.272,00	26/10/1978	45	HC39S86524	75	44713	BMI	1	NACIONAL
CHARLES MILTON GARCIA RODRIGUEZ	BMI HOSPICARE	\$ 18.258,00	13/5/1982	42	HC23685874	75	44713	BMI	1	NACIONAL
ISABEL MONSERRATE PALMA BERMEO	BMI GMM	\$ 1.248,00	19/3/1972	52	GM22G29858	500	44713	BMI	4	NACIONAL
ISABEL MONSERRATE PALMA BERMEO	MEDIKEN 20K	\$ 246.132,00	19/3/1972	52		20	44713	MEDIKEN	4	NACIONAL
ROGER DANIEL ECHANQUE TOSCANO	BMI IDEAL	\$ 613.205,00	12/5/1993	31	A1165895425	1000	44713	BMI	4	INTERNACIONAL
JACINTO RODOLFO HOJAS CAMPOVERDE	BMI SIGMA	\$ 647.448,00	20/7/1981	43	S122G38216	150	44728	BMI	4	NACIONAL
JACINTO ORLANDO HOJAS CAMPOVERDE	BMI SIGMA	\$ 598.344,00	20/7/1981	43	S122G38216	150	44728	BMI	4	NACIONAL
KIANG COELLO JO YEN MICHAEL	BMI SIGMA	\$ 947.352,00	16/9/1989	34	S122G38212	150	44728	BMI	5	NACIONAL
SABRINA IVANOVA GUEVARA MORA	PLAN 3D	\$ 75.336,00	28/2/1988	36	4151593	30	44728	SALUD	1	NACIONAL
MELL IVANOVA QUINONEZ GUEVARA	PLAN 3D	\$ 66.684,00	21/5/2017	7	4151593	30	44728	SALUD	1	NACIONAL
LILIA ELECTRA VERA ZAMBRANO	PLAN 4D	\$ 178.484,00	22/2/1959	65	4151664	40	44728	SALUD	1	NACIONAL
ISABEL DE LA CRUZ ACUÑA VALDIVIEZO	BMI GMM	\$ 228.768,00	14/2/1946	78	GM22G28971	500	44743	BMI	1	NACIONAL
RIVERA BENAVIDES BERNARDO	BMI SIGMA	\$ 16.956,00	22/6/1999	25	S122G38379	150	44743	BMI	1	NACIONAL
RIVERA BENAVIDES LORENZO	BMI SIGMA	\$ 179.652,00	6/11/1996	27	S122G38452	150	44743	BMI	1	NACIONAL
RAUL SANTIAGO COLMONT VELEZ	BMI IDEAL	\$ 27.012,00	3/4/1983	41	A122G13775	1000	44758	BMI	4	NACIONAL
ADOLFO LAUTANO CHONG JIE	BMI SIGMA	\$ 285.902,00	19/4/1986	38	S122G38501	150	44758	BMI	3	NACIONAL
JENNIFFER MICHELLE PADILLA PARRA	BMI IDEAL	\$ 359.286,00	20/4/1985	39	A122G13861	1000	44757	BMI	1	INTERNACIONAL
EVELYN GUADALUPE GOMEZ CRESPO	BMI GMM	\$ 34.056,00	16/12/1984	39	GM22G29208	500	44758	BMI	3	NACIONAL
ADRIANA NATHALY ZAMBRANO PARRALES	BUPA35K	\$ 151.008,00	9/2/1987	37	MT23G06801	35	44757	BUPA	3	NACIONAL
MARTHA ADRIANA FREIRE LABORDE	BMI GMM	\$ 70.968,00	26/8/1978	45	GM22G29214	500	44758	BMI	3	NACIONAL
EVELYN GUADALUPE GOMEZ CRESPO	BUPA35K	\$ 98.592,00	26/12/1984	39	GM22G29208	35	44757	BUPA	4	NACIONAL
EDUARDO MANUEL YCAZA CEPEDA	BUPA35K	\$ 192.192,00	1/8/1953	70	235698	35	44774	BUPA	1	NACIONAL
BLANCA FIEDAD JAUREGUI RENDON	BUPA35K	\$ 177.216,00	1/12/1956	67	22268170	35	44774	BUPA	1	NACIONAL

BLANCA PIEDAD JAUREGUI RENDON	BMI GMM	\$ 96.036,00	1/12/1956	67	GM12G29343	500	44774	BMI	1	NACIONAL
JAJME ALBERTO AUREA VILLACRES	BMI SUPPORT	\$ 150.204,00	21/7/1957	67	A122G13983	1000	44774	BMI	1	INTERNACIONAL
JAJME ALBERTO AUREA VILLACRES	BMI SIGMA	\$ 262.363,00	21/7/1957	67	A122G13983	150	44774	BMI	1	NACIONAL
MOGROVEJO MORALES GABRIEL ANDRES	BMI SIGMA	\$ 16.956,00	26/1/1999	25	S122G38733	150	44774	BMI	1	NACIONAL
ZENAIDA ELENA ALCIVAR BOWEN	BMI SIGMA	\$ 345.084,00	29/11/1956	67	S122G38730	150	44774	BMI	1	NACIONAL
DARIANNY NARCISA VALVERDE PEREZ	ASISKEN20K	\$ 65.292,00	2/5/1990	34	292165	20	44783	ASISKEN	1	NACIONAL
GONZALO JOSE CAAMAÑO PEREZ	BMI SIGMA	\$ 123.012,00	7/6/1982	42	S122G38857	150	44789	BMI	1	NACIONAL
MARIA AUXILIADORA NUQUES CEDEÑO	BMI SIGMA	\$ 279.144,00	7/6/1982	42	S122G38857	150	44805	BMI	1	NACIONAL
JOSE RICARDO HERNANDEZ AGUIRRE	BMI SIGMA	\$ 342.156,00	5/4/1985	39	S122Q39205	150	44805	BMI	3	NACIONAL
DANIEL DAVID GONZALEZ CASTILLO	BMI GMM	\$ 117.444,00	14/3/1985	39	GM12G30265	500	44805	BMI	1	NACIONAL
SABANDO JUSTILLOS ZOILA FATIMA	MEDIKEN 100K	\$ 22.878,00	11/6/1979	45		100	44805	MEDIKEN	1	NACIONAL
FALCONI CHOEZ ALBERTO MARCELINO	BMI IDEAL	\$ 927.588,00	12/12/1981	42	A122G14579	1000	44820	BMI	3	INTERNACIONAL
VANESSA LILIANA NARANJO MEJIA	BMI SIGMA	\$ 24.588,00	9/9/1977	46	S122Q39156	150	44805	BMI	1	NACIONAL
VANESSA LILIANA NARANJO MEJIA	BMI GMM	\$ 33.144,00	9/9/1977	46	S122Q39156	500	44805	BMI	1	NACIONAL
CUEVA SANCHEZ LUZ GRACIELA	BMI IDEAL	\$ 629.424,00	5/6/1981	43	A122G14566	1000	44820	BMI	3	INTERNACIONAL
CARRASCO RIVERA EDDIE	BMI HOSPICARE	\$ 189.612,00	10/8/1980	43	GM23G3748	75	44820	BMI	1	NACIONAL
LARA GONZALEZ RODRIGO OCTAVIO	BMI GMM	\$ 333.552,00	23/1/1951	73	GM12G30383	500	44820	BMI	3	NACIONAL
IVAN RIVERA GARCIA	BMI GMM	\$ 10.608,00	17/5/1969	55	GM12G30565	500	44820	BMI	3	NACIONAL
SANTIAGO VINICIO VINTIMILLA LEON	BMI GMM	\$ 226.032,00	16/9/1956	67	GM12G30778	500	44835	BMI	1	NACIONAL
PATRICIA GUADALUPE CHOEZ REYES	BMI GMM	\$ 125.184,00	13/3/1982	62	GM12G30865	500	44835	BMI	1	NACIONAL
CIELO MARLU MONTALVAN MONTES	BMI GMM	\$ 175.416,00	17/11/1952	71	GM12G31037	500	44850	BMI	1	NACIONAL
NEYI DELICIA SORNOZA ALVARADO	BMI GMM	\$ 65.904,00	10/2/1964	60	GM12G31156	500	44850	BMI	1	NACIONAL
MARIA BELEN TORRES SARMIENTO	BMI SIGMA	\$ 205.716,00	22/10/1988	35	S122G39567	150	44850	BMI	2	NACIONAL
NUBIA LEONOR SARMIENTO MALDONADO	BMI SIGMA	\$ 35.646,00	25/4/1964	60	S122G39566	150	44850	BMI	1	NACIONAL
MARTHA MARITZA PAREDES SABANDO	BMI SIGMA	\$ 213.732,00	18/2/1977	47	S122G39574	150	44850	BMI	1	NACIONAL
ANDREA CRISTINA JIMENEZ DELGADO	BUPA 35K	\$ 1.500,00	5/9/1988	35	PRCP-22-274524	35	44864	BUPA	2	NACIONAL
MIGUEL AGUSTIN NUÑEZ ALMEIDA	BMI GMM	\$ 6.618,00	9/4/1961	63	GM12G31432	500	44866	BMI	2	NACIONAL
DAVID MAURICIO BERMEZ ZAPATA	BUPA 35K	\$ 218.808,00	1/8/1995	28	PRCP-22-275823	35	44864	BUPA	3	NACIONAL
JESUS DANIEL BERMEZ ZAPATA	BUPA 35K	\$ 218.808,00	12/4/1994	30	PRCP-22-275783	35	44864	BUPA	3	NACIONAL
VICTOR ANDRES BERMEZ ZAPATA	BUPA 35K	\$ 234.204,00	5/3/1991	33	PRCP-22-275781	35	44864	BUPA	3	NACIONAL
JENNIFER DAYANA LOMAS LEON	BMI GMM	\$ 34.356,00	16/6/1966	38	GM12G31729	500	44881	BMI	3	NACIONAL
DAYANARA DEL ROCIO LOPEZ VELEZ	BMI HOSPICARE	\$ 150.612,00	29/12/2002	21	HC22G07636	75	44881	BMI	1	NACIONAL
JENNIFER DAYANA LOMAS LEON	BUPA 35K	\$ 151.008,00	16/6/1986	38	PRCP-22-276039	35	44881	BUPA	2	NACIONAL
JORGE RAFAEL VALAREZO ULLOA	BUPA 35K	\$ 289.536,00	20/6/1985	39	PRCP-22-277040	35	44880	BUPA	4	NACIONAL
JORGE RAFAEL VALAREZO ULLOA	BMI GMM	\$ 89.348,00	20/6/1985	39	GM12G31992	500	44881	BMI	4	NACIONAL
HAMILTON STALYN SUAREZ MARIN	BMI GMM	\$ 124.572,00	17/4/1974	50	GM12G32170	500	44896	BMI	1	NACIONAL
PAMELA CAVAGNARO ROMERO	BUPA 35K	\$ 134.784,00	24/8/1989	34	GM12G32372	35	44896	BUPA	1	NACIONAL
MARIA ALEJANDRA LEON UBILLA	BUPA 35K	\$ 95.196,00	1/6/2002	22	PRCP-22-278562	35	44896	BUPA	1	NACIONAL
NATHALIA CABRERA RODRIGUEZ	BUPA 35K	\$ 1.248,00	16/3/1990	34	2782365	35	44896	BUPA	1	NACIONAL
ELVIRA YEPEZ PACHECO	BUPA 35K	\$ 151.008,00	2/1/1975	49	PRCP-22-277660	35	44896	BUPA	1	NACIONAL
PAOLA JACKELINE OCHOA PACHECO	BMI GMM	\$ 94.836,00	23/8/1968	55	PRCP-22-277456	500	44911	BMI	2	NACIONAL
PAMELA CAVAGNARO ROMERO	BMI GMM	\$ 46.284,00	24/8/1989	34	GM12G32372	500	44911	BMI	2	NACIONAL
PAOLA JACKELINE OCHOA PACHECO	BUPA 35K	\$ 167.232,00	23/8/1968	55	PRCP-22-277456	35	44896	BUPA	2	NACIONAL
EVELYN CABALLERO DORADO	BUPA 35K	\$ 83.196,00	9/1/1991	33	OPTCP-22-277406	35	44896	BUPA	2	NACIONAL
SUSANA TAMAYO INSUASTI	BUPA 35K	\$ 1.560,00	25/7/1949	75	GM11G25422	35	44910	BUPA	1	NACIONAL
ROSA VIOLETA GRANDA MATAMOROS	BUPA 35K	\$ 1.812,00	24/4/1956	68	PRCP-22-279160	35	44918	BUPA	1	NACIONAL
JUAN FRANCISCO SILVA GALLARDO	BMI GMM	\$ 122.352,00	26/3/1982	42	GM12G32914	500	44927	BMI	3	NACIONAL
MARIA ANTONIETA GARCIA ESPINOZA	BUPA 35K	\$ 2.436,00	2/10/1953	70	569836	35	44942	BUPA	1	NACIONAL
ARREAGA ESPINOZA FRANCISCO XAVIER	BMI GMM	\$ 2.283,00	16/9/1957	66	GM12G33129	500	44942	BMI	3	NACIONAL
JONATHAN DAVID GOMEZ ROSALES	BUPA 35K	\$ 11.532,00	19/12/1960	63	PRCP-23-279918	35	44949	BUPA	1	NACIONAL
JUAN ROBERTO CASTELLO VILLANUEVA	BUPA 35K	\$ 285.276,00	13/2/1967	57	GM12G33898	35	44958	BUPA	2	NACIONAL
JUAN ROBERTO CASTELLO VILLANUEVA	BMI GMM	\$ 104.513,00	13/2/1967	57	PRCP-23-280712	500	44958	BMI	2	NACIONAL
DANIELA VILLAGOMEZ VIZCAINO	BUPA 35K	\$ 148.932,00	22/1/1987	37	OPTCP-23-280713	55	44958	BMI	1	NACIONAL
LUIS CARLOS HARO MERA	BMI GMM	\$ 105.198,00	8/1/1986	38	GM12G33817	500	44973	BMI	1	NACIONAL
MONICA MARYDRI OREA ARGUINZONES	BMI SIGMA	\$ 300.282,00	18/1/1984	40	S123G40856	150	44973	BMI	1	NACIONAL
VALENTINA BEJARANO LUGO	BUPA 35K	\$ 123.132,00	11/4/1996	28	OPTCP-23-281535	55	44980	BUPA	1	NACIONAL
GIULIANO ROBERTO NOGALES ENEA	BUPA 35K	\$ 165.348,00	20/7/1996	28	PRCP-23-281518	35	44980	BUPA	1	NACIONAL
TATIANA MARIUXI GALECIO VILLEGAS	BUPA 35K	\$ 73.848,00	19/2/1993	31	PRBS-23-281579	35	44985	BUPA	1	NACIONAL
JUAN FRANCISCO SILVA GALLARDO	BMI GMM	\$ 129.996,00	26/3/1982	42	GM12G32914	500	44986	BMI	1	NACIONAL
DANIEL FERNANDO CORONADO ARRIETA	BMI GMM	\$ 99.408,00	28/11/1988	35	GM12Q34193	500	44986	BMI	4	NACIONAL
JAZMIN MOREIRA	BMI SIGMA	\$ 357.587,00	29/3/1985	39	S121G31622	150	45017	BMI	3	NACIONAL
KAREN INGRID AUHING FERRIN	BMI GMM	\$ 47.113,00	9/9/1991	32	GM12G34252	500	45017	BMI	1	NACIONAL
ERICKA BARATAU	SALUD 5D	\$ 1.149,00	10/3/1987	37	FUN-4165323	100	45022	SALUD	1	NACIONAL
SOLANGE ESTEPHANIA SALLINAS VASQUEZ	BMI IDEAL	\$ 628.597,00	7/1/1981	43	A123G16559	1000	45032	BMI	2	INTERNACIONAL
KAREN INGRID AUHING FERRIN	BUPA 35K	\$ 47.113,00	9/9/1991	32	GM12G34252	35	45036	BUPA	1	NACIONAL
ADOLFO CHONG DIAZ	BMI GMM	\$ 324.656,00	19/4/1986	38	S122G38501	500	45032	BMI	2	NACIONAL

VALENTINA BEJARANO LUGO	BMI GMM	\$	48.050,00	11/4/1996	28	GM23G34430	500	45032	BMI	3	NACIONAL
WASHINGTON NICOLAS SANCHEZ ROJAS	BMI GMM	\$	75.629,00	20/9/1960	63	GM23G34409	500	45032	BMI	2	NACIONAL
WASHINGTON NICOLAS SANCHEZ ROJAS	ASISKEN 30K	\$	2.385,00	10/2/1955	69	568978	30	45036	ASISKEN	1	NACIONAL
CRISTINA LISSETTE SORIANO SORNOZA	BUPA 125K	\$	245.803,00	27/8/1990	33	27623	125	45036	BUPA	2	NACIONAL
CYNTHIA RAQUEL JOONG PARRAGA	BMI GMM	\$	225.603,00	3/3/1986	38	456989	500	45047	BUPA	1	NACIONAL
PATRICIO ELIAS HUAYAMABE MITE	ASISKEN 20K	\$	156.983,00	7/7/1988	36	568963	20	45061	ASISKEN	1	NACIONAL
EDUARDO PATRICIO ARGUDO VINCENT	BMI GMM	\$	240.522,00	26/2/1963	61	GM24G42924	500	45323	BMI	3	NACIONAL
MELANIE MARIANA TORRES REYES	ASISKEN 50K	\$	136.988,00	24/2/1991	33	568924	50	45066	ASISKEN	1	NACIONAL
EMILY VIVIANA CAMPOS GAONA	ASISKEN 20K	\$	5.856,00	15/11/2004	19	31952	20	45311	ASISKEN	1	NACIONAL
KARINA LISSETTE GORDILLO UBIDIA	ASISKEN 30K	\$	88.608,00	26/2/1986	38	31940	30	45311	ASISKEN	1	NACIONAL
NAGIB MIGUEL CHAGERBEN HINOSTROSA	ASISKEN SUMMA	\$	249.024,00	10/6/1981	43	323526	100	45321	ASISKEN	1	NACIONAL
MARIA LUISA POVEDA MARTILLO	ASISKEN 20K	\$	281.292,00	3/5/1955	69		20	45321	ASISKEN	1	NACIONAL
LUIS CARLOS HARO MERA	ASISKEN 30K	\$	98.052,00	8/1/1986	38		30	45332	ASISKEN	1	NACIONAL
JORGE GELLIBER ORTEGA LOPEZ	BMI GMM	\$	73.056,00	8/4/1968	56	270173-3	500	45328	BMI	1	NACIONAL
CAROLINA LILIBETH MOREIRA PONCE	ASISKEN 20K	\$	98.304,00	31/1/1989	35	A122G11693	20	45342	ASISKEN	1	NACIONAL
JOSE MANUEL GUTIERREZ BOZZA	BMI GMM	\$	89.964,00	23/12/1970	53	GM24G43366	500	45338	BMI	1	NACIONAL
KARLA YESSSENIA TORRES REYES	BMI SIGMA	\$	158.035,00	28/5/1993	31	S124Q43882	150	45331	BMI	1	NACIONAL
CARLOS HUMBERTO AYALA BAIKAL	BUPA 55K	\$	2.760,00	25/9/1951	72	OPTBS-24-298329	55	45352	BUPA	1	NACIONAL
CARLOS HUMBERTO AYALA BAIKAL	BMI GMM	\$	135.588,00	25/9/1951	72	OPTBS-24-298329	500	45352	BMI	1	NACIONAL
KATHY LIZ VELASQUEZ CAMPOZANO	BMI GMM	\$	95.976,00	19/1/1962	62	14225	500	45352	BMI	1	NACIONAL
CHRISTIAN JAVIER TOALONGO MORENO	ASISKEN 20K	\$	109.524,00	26/5/1993	31	PRCP-23-284594	20	45332	ASISKEN	3	NACIONAL
ANDREA DEL ROCIO ALMEIDA QUEZADA	ASISKEN SUMMA SOI	\$	151.356,00	18/12/1989	34	32176	100	45332	ASISKEN	1	NACIONAL
MARITZA ARACELY TOALA RAMOS	SALUD SKY	\$		2/9/1986	37	FUN-4185567	70	45331	SALUD	1	NACIONAL
ARIANNA MONTERO PANCHANA	SALUD SKY	\$	49.104,00	2/7/1993	31	FUN-4185628	70	45338	SALUD	1	NACIONAL
BERTHA INELDA PANCHANA DEL PEZO	SALUD RESPALDO	\$	49.104,00	6/10/1960	63	PRCP-23-281771	500	45338	SALUD	1	NACIONAL
MICHELLE ESTEFANIA AVILES TOALA	SALUD SKY	\$	291.708,00	25/8/1988	35	FUN-4185567	70	45342	SALUD	1	NACIONAL
JOHN GLUBIS JIJON PINTO	BMI GMM	\$	582.492,00	2/2/1967	57	GM24G43674	500	45352	BMI	2	NACIONAL
CARLA ALEJANDRA CORDOVA CARRION	BMI GMM	\$	152.943,00	6/7/1966	58	PRCP-22-269628-05	500	45036	BMI	1	NACIONAL
JULIO ALBERTO MERCHAN MERCHAN	BMI GMM	\$	511.104,00	25/6/1954	70	HC19G02995	500	45352	BMI	1	NACIONAL
DANIEL ALEJANDRO ANDRADE ESPINEL	ASISKEN 20K	\$	6.888,00	14/3/1989	35	32249	20	45371	ASISKEN	1	NACIONAL
MELINA NINFA GOMEZ ALLIERI	ASISKEN 30K	\$	88.608,00	11/4/1984	40	PRCP-23-287709	30	45342	ASISKEN	2	NACIONAL
MARCIA DEL CONSUELO ROMERO SARAGURO	ASISKEN 20K	\$	312.996,00	15/5/1972	52	32325	20	45346	ASISKEN	2	NACIONAL
PEDRO PABLO TRIVIÑO GOMEZ	ASISKEN 20K	\$	120.888,00	4/6/2005	19	323524028223	20	44226	ASISKEN	1	NACIONAL
ELLIOTT ALBERTO MONCAYO CORDOVA	ASISKEN 20K	\$	125.231,00	15/1/1990	34	32331	20	45351	ASISKEN	2	NACIONAL
MANUEL BENJAMIN LITUMA CORAISACA	BMI GMM	\$	12.672,00	10/11/1975	48	45367	500	45367	BMI	3	NACIONAL
FERNANDO MIGUEL HERNANDEZ MEDINA	BMI IDEAL	\$	465.559,00	13/5/1985	39		1000	45367	BMI	1	INTERNACIONAL
ANTONIO PAUL LEON VARGAS	BMI GMM	\$	85.272,00	6/2/1973	51	GM24G43981	500	45367	BMI	1	NACIONAL
DANIEL ALFREDO PULLEY ITURRALDE	ASISKEN 50K	\$	1.260,00	20/10/1987	36	32434	50	45367	ASISKEN	1	NACIONAL
EVELYN ANDREA CABALLERO DORADO	BMI GMM	\$	83.736,00	9/1/1991	33	731697	500	45483	BMI	2	NACIONAL
ENRIQUE JOSE GIRALDO BARBERY	EXPIRIENCE SALUD	\$	908.976,00	17/12/1983	40	FUN-4187709	1000	45372	SALUD	3	NACIONAL
MAURICIO RAMIRO SALAZAR GUILLEN	BMI GMM	\$	167.184,00	7/11/1968	55	GM24G44298	500	45367	BMI	2	NACIONAL
MILTON ALBERTO TOBAR VEGA	BUPA 125K	\$	290.604,00	3/5/1953	71	S120G29171	125	45366	BUPA	1	NACIONAL
CATHERINE PAULLETTE GUILINDRO	ASISKEN 50K	\$	146.664,00	9/9/1991	32	S121G31992	50	45371	ASISKEN	1	NACIONAL
MARTHA ADRIANA FREIRE LABORDE	ASISKEN 20K	\$	135.444,00	26/8/1978	45	GM22G29214	20	45371	ASISKEN	2	NACIONAL
CRISTHIAN OMAR VELASCO ACUÑA	ASISKEN 20K	\$	251.748,00	7/4/1975	49	14074001001	20	45381	ASISKEN	4	NACIONAL
ANA RAMONA RODRIGUEZ RODRIGUEZ	ASISKEN SUMMA CUR	\$	610.863,00	26/7/1956	68		100	45371	ASISKEN	1	NACIONAL
ANA RAMONA RODRIGUEZ RODRIGUEZ	ASISKARE	\$			124		500	45381	ASISKEN	1	NACIONAL
YESSSENIA DAYANARA VELASQUEZ VARGAS	ASISKEN 30K	\$	94.332,00	2/6/1993	31	32475	30	45361	ASISKEN		NACIONAL
KEVIN RAFAEL CORRAL PARRAGA	ASISKEN 50K	\$	103.512,00	17/8/1995	28		50	45381	ASISKEN		NACIONAL
MARIA FERNANDA GOMEZ FIGUEROA	BMI INNOVA	\$		10/7/1991	33		80		BMI		NACIONAL
MARIANA NARCISIA CERCAO DE MEZA	ASISKEN 50K	\$	236.592,00	26/6/1962	62		50	45392	ASISKEN		NACIONAL
LUIS EDUARDO MOREIRA PONCE	BMI IDEAL	\$	3.909,00	6/10/1985	38	A122G12677	1000	45398	BMI		INTERNACIONAL
WENDY PATRICIA INSUASTI CORONEL	BMI IDEAL	\$	189.276,00	25/10/1983	40	GM23G34645	1000	16/04/2024	BMI		INTERNACIONAL
HUGO OSWALDO VANONI BRAVO	BMI GMM	\$	209.762,00	22/4/1953	71	A122G12563	500	45398	BMI		NACIONAL
MIRNA VIOLETA PONCE TOALA	BMI SUPPORT	\$	275.016,00	4/8/1958	65	A124G19855	500	45398	BMI		INTERNACIONAL
DOMENICA STEFFI TALANO ALVARADO	BMI GMM	\$	126.588,00	23/4/1994	30	GM24G45017	500	45402	BMI		NACIONAL
JONATHAN OMAR CAMINO GARCIA	BMI SUPPORT	\$	1.629,00	26/1/1988	36		500	45413	BMI		INTERNACIONAL
JOSE EDUARDO SANCHEZ APOLO	BMI GMM	\$	1.260,00	11/7/1962	62	GM24G45208	500	45413	BMI		NACIONAL
ERIKA ALEXANDRA QUINTAPALLO AYALA	BMI GMM	\$	168.089,00	11/4/1983	41	IV24Q00532	500	45413	BMI		NACIONAL
GENESIS STEFANIA PEÑA ALCIVAR	ASISKEN 50K	\$	116.616,00	24/12/1990	33	32872	50	45392	ASISKEN		NACIONAL
SILVIA CAROLINA FREIRE LABORDE	ASISKEN 20K	\$	14.484,00	18/3/1980	44	4164563	20	20/04/2024	ASISKEN		NACIONAL
SILVIA CAROLINA FREIRE LABORDE	ASISKARE	\$	53.952,00	18/3/1980	44	4164563	500	20/04/2024	ASISKEN		NACIONAL
LUIS ANGEL TORRES REYES	ASISKEN 50K	\$	76.932,00	26/1/2002	22		50	45402	ASISKEN		NACIONAL
ANGELA MARIA LOPEZ VERA	SALUD SKY	\$	407.856,00	15/5/1967	57	FUN-4189621	70	45404	SALUD		NACIONAL
MERY DE JESUS SALDAÑA ZAMBRANO	BMI GMM	\$	105.936,00	19/2/1969	55		500		BMI		NACIONAL
MARIA MAGDALENA MAZA MERINO	BMI GMM	\$	139.667,00	25/5/1961	63	S116G04572	500	44667	BMI		NACIONAL

ROSIANEY FERRIN MOREIRA	SALUD SKY 70	\$	280.308,00	3/8/1960	63	FUN-4190360	70	45411	SALUD		NACIONAL
ADRIANA FERRIN FERRIN	SALUD PRO150	\$	280.308,00	3/3/1991	33	FUN-4190360	150	45411	SALUD		NACIONAL
ANDRES ANTONIO LANATA ALAVA	ASISPRO30K	\$	34.044,00	20/11/1982	41	31591	30	45412	ASISKEN		NACIONAL
CHRISTIAN WILMER GUARNIZO CAICEDO	ASISKEN 20K	\$	3.432,00	23/4/1976	48	23090	20	45412	ASISKEN		NACIONAL
BLANCA ROSARIO BERMEJO ALVAREZ	BUPA55K	\$	268.212,00	8/12/1957	66	OPTCP-24-299613	55	45413	BUPA	1	NACIONAL
KLEBER JOSE IZQUIERDO CASTRO	BMI IDEAL	\$	190.602,00	21/6/1991	33	AID24G20086	1000	16/05/2024	BMI		INTERNACIONAL
MARTHA GRACIELA MORA VALVERDE	BMI GMM	\$	212.232,00	15/11/1954	69	GM24G45627	500	45428	BMI	2	NACIONAL
JHONNY ROBERTO AREVALO TOALA	BMI GMM	\$	2.160,00	18/5/1970	54	GM24G45627	500	45428	BMI	2	NACIONAL
KAREN GARCIA HUAYAMABE	SALUD PRO150	\$	2.640,00	25/5/1987	37	53670	150	45428	SALUD		NACIONAL
GALO ANTONIO IZQUIERDO MIRANDA	BMI GMM	\$	7.032,00	22/10/1983	40	GM24G45662	500	45428	BMI	3	NACIONAL
URSULA LISSETH SILVA INTRIAGO	BMI GMM	\$	48.708,00	17/3/1991	33	GM24G45758	500	45428	BMI		NACIONAL
DENNIS ROLANDO CASTRO YULAN	BMI GMM	\$	131.916,00	30/10/1985	38	GM24G45693	500	45428	BMI		NACIONAL
WILMER ALBERTO ROSADO RIOS	BMI IDEAL	\$	380.794,00	14/5/1985	39	AID24G20112	1000	45428	BMI		INTERNACIONAL
MARCELO FELIPE NORIEGA MUÑOZ	BMI IDEAL	\$	348.613,00	7/5/1984	40	AID24G20151	1000	16/05/2024	BMI		INTERNACIONAL
JARAMILLO PARRALES ANGELA MARIA	ASISKEN 50K	\$	235.668,00	5/2/1974	50	SI22Q37575	50	45422	ASISKEN		NACIONAL
STEFFANI ALEJANDRA YANEZ SANCHEZ	ASISKEN SUMMA C	\$	1.143,00	10/2/1990	34	33474	100	45432	ASISKEN		NACIONAL
ANGEL AMABLE YANEZ VINUEZA	ASISKEN 50K	\$	480.732,00	31/3/1962	62	10055	50	45432	ASISKEN		NACIONAL
ANGEL AMABLE YANEZ VINUEZA	ASISKEN 500K	\$	273.228,00	31/3/1962	62	10056	500	45432	ASISKEN		NACIONAL
ANDREA VIVIANA DUTAN FLORES	ASISKEN 20K				124	33411	20		ASISKEN		NACIONAL
LILIBETH CECILIA BRAVO VELASQUEZ	ASISKEN 20K	\$	145.392,00	14/11/1993	30		20	45442	ASISKEN	1	NACIONAL
REINALDO RHOA RIVAS	SALUD 70 PLUS				124		70		SALUD	2	NACIONAL
ANDREA SOFIA SOTO ANDRADE	SALUD 70 PLUS				124		70		SALUD	1	NACIONAL
MARY DEL ROCIO CHALAN AGUILAR	BUPA 55K	\$	530.784,00	16/6/1975	49	OPTCP-24-300139	55	45444	BUPA	2	NACIONAL
MARCELA YAMILETTE LOMAS LEON	ASISPRO 50K	\$	101.568,00	18/6/1996	28	33497	50	45444	ASISKEN	1	NACIONAL
JULIO RAFAEL MAYORGA RODRIGUEZ	BMI SUPPORT	\$	401.584,00	18/2/1954	70	SI19C18809	500	45459	BMI	1	INTERNACIONAL
GEOVANNY EDILBERTO FIGUAVE MENDEZ	ASISKEN SUMMA SOI	\$	83.736,00	4/6/1984	40	MT00715009	100	45442	ASISKEN	1	NACIONAL
JORGE LUIS GUERRERO CAICEDO	ASISKEN 20K	\$	65.568,00	17/4/1974	50	33646	20	45442	ASISKEN	1	NACIONAL
EMILIO JOSE LEON SILVA	SALUD 50 PLUS	\$	431.364,00	20/9/1962	61	32289-1	50	45436	SALUD	3	NACIONAL
LUIS FERNANDO ZAMBRANO PARRAGA	BMI GMM			19/1/1984	40	GM24G46435	500		BMI	1	NACIONAL
DIANA ALEXANDRA ROBLES CAMPOVERDE	SALUD 5D	\$	93.708,00	21/10/1986	37	FUN-4193143	100	45457	SALUD	1	NACIONAL
ALEKSEI STRUGANOV	ASISKEN SUMMA SOI	\$	123.888,00	29/11/1988	35	33815	100	45463	ASISKEN	1	NACIONAL
KAREN NATHALY OCHOA OCHOA	ASISKEN 30K	\$	85.284,00	24/4/1995	29	PRCP-22-277040	30	45463	ASISKEN	1	NACIONAL
ANTHONY LEONARDO PROAÑO MOREIRA	MEDIKEN	\$	109.008,00	8/5/1984	40	14519	20	45458	MEDIKEN	1	NACIONAL
FELIPE GUSTAVO BARRIGA GOMEZ	BMI GMM	\$	122.196,00	9/4/1965	59	772681	500	45474	BMI	2	NACIONAL
FELIPE GUSTAVO BARRIGA GOMEZ	BUPA 55K	\$	330.192,00	9/4/1965	59	OPTCP-24-300738	55	45474	BUPA	2	NACIONAL
LOMAS LEON ANDREA ESTEFANIE	ASISKEN 30K	\$	1.980,00	9/2/1987	37		30	45432	ASISKEN	1	NACIONAL
DANIEL EDUARDO POLO GAVILANES	ASISKARE 500K	\$	124.632,00	4/8/1982	41		500	45473	ASISKEN	4	NACIONAL
JOHNNY LORENZO CABRERA VASQUEZ	ASISKEN 20K	\$	161.556,00	3/7/1970	54	33924	20	45463	ASISKEN	6	NACIONAL
JOHNNY LORENZO CABRERA VASQUEZ	ASISKARE 500K	\$	161.556,00	3/7/1970	54	33924	500	20/06/2024	ASISKEN	6	NACIONAL
GENESIS SILVANA CASTRO ALMEIDA	BMI GMM	\$	34.356,00	26/10/1999	24		500	45474	BMI		NACIONAL
CINTHYA LISSETTE MARTINEZ RUIZ	ASISKEN 20K	\$	15.012,00	7/4/1988	36	33890	20	45473	ASISKEN		NACIONAL

Script

```
# Gráfico de barras para el tipo de cobertura
GR1 <- ggplot(Tab_Fre_Tipo_Cobertura, aes(x=Tab_Fre_Tipo_Cobertura$var1, y=Tab_Fre_Tipo_Cobertura$Fre1)) +
  geom_bar(stat = "identity", fill="pink") +
  geom_text(aes(label=paste0(Tab_Fre_Tipo_Cobertura$Fre1)), position = position_stack(vjust = 0.8)) +
  labs(title = "Seguros por Tipo de Cobertura")

GR1

### VARIABLE NUMÉRICA edad ###
x1 <- segurosCL$EDAD
# Aplicar la regla de Sturges
k1 <- nclass.sturges(x1)
# calcular la amplitud
Int1 <- cut(x1, breaks = k1)
Int1

## Construcción de Tabla de Frecuencias ####
Fre_edad <- as.data.frame(table(Int1))
Fre_edad
Tab_Fre_Edad <- transform(Fre_edad,
  Fre1 = round(prop.table(Fre_edad$Freq), 3),
  Facu = cumsum(Fre_edad$Freq),
  FacuR = round(cumsum(prop.table(Fre_edad$Freq)), 3))
Tab_Fre_Edad
```

```

# Gráfico de barras para la edad##
G2 <- ggplot(Tab_Fre_Edad, aes(x=Tab_Fre_Edad$Int1, y=Tab_Fre_Edad$FRE1)) +
  geom_bar(stat = "identity", fill="yellow", colour="black", size=0.5) +
  geom_text(aes(label=paste0(Tab_Fre_Edad$FRE1)), position = position_stack(vjust = 0.8)) + coord_flip() +
  labs(title = "Porcentaje mas frecuente en seguros")

G2

#### Medidas de tendencia central ####
mean(segurosCL$EDAD) # media
median(segurosCL$EDAD) # mediana
library(modeest)
mfv(segurosCL$EDAD) # moda
library(moments)
skewness(segurosCL$EDAD) # coeficiente de asimetría
var(segurosCL$EDAD) # varianza
sd(segurosCL$EDAD) # desviación estándar
sd(segurosCL$EDAD)/mean(segurosCL$EDAD) # coeficiente de variación
kurtosis(segurosCL$EDAD) # curtosis

# Gráfico de normalidad para la edad
GR3 <- ggplot(segurosCL, aes(x=segurosCL$EDAD)) +
  geom_histogram(aes(y=..density..), fill="#B6F0B0", colour="#23AB57") +
  geom_density(alpha=.2, fill="#AF41B2") +
  geom_vline(aes(xintercept=mean(segurosCL$EDAD)), color="red", size=1.5) +
  geom_vline(aes(xintercept=median(segurosCL$EDAD)), color="blue", size=1.5, linetype="dashed") +
  geom_vline(aes(xintercept=mfv(segurosCL$EDAD)), color="yellow", size=1.5) +
  labs(title = "Histograma de Edad")

GR3

# medidas de tendencia central poliza
mean(segurosCL$VALOR_DE_POLIZA._ANUAL) # media
median(segurosCL$VALOR_DE_POLIZA._ANUAL) # mediana
library(modeest)
mfv(segurosCL$VALOR_DE_POLIZA._ANUAL) # moda
library(moments)
skewness(segurosCL$VALOR_DE_POLIZA._ANUAL) # coeficiente de asimetría
var(segurosCL$VALOR_DE_POLIZA._ANUAL) # varianza
sd(segurosCL$VALOR_DE_POLIZA._ANUAL) # desviación estándar
sd(segurosCL$VALOR_DE_POLIZA._ANUAL)/mean(segurosCL$VALOR_DE_POLIZA._ANUAL) # coeficiente de variación
kurtosis(segurosCL$VALOR_DE_POLIZA._ANUAL) # curtosis

#grafico normalidad valor polizas#
GR4 <- ggplot(segurosCL, aes(x=segurosCL$VALOR_DE_POLIZA._ANUAL)) +
  geom_histogram(aes(y=..density..), fill="#B6F0B0", colour="#23AB57") +
  geom_density(alpha=.2, fill="#AF41B2") +
  geom_vline(aes(xintercept=mean(segurosCL$VALOR_DE_POLIZA._ANUAL)), color="red", size=1.5) +
  geom_vline(aes(xintercept=median(segurosCL$VALOR_DE_POLIZA._ANUAL)), color="blue", size=1.5, linetype="dash") +
  labs(title = "Histograma de Valor de Póliza Anual")

GR4

```

```

## Gráficas Especiales ##
# Categorica con Categorica#boxplot##

GRES1 <- ggplot(segurosCL, aes(x= segurosCL$NUMERO_DE_BENEFICIARIOS, y=segurosCL$TIPO_DE_COBERTURA)) +
  geom_boxplot(fill="orange", colour="black") +
  labs(title = "Tipo de Cobertura según Número de Beneficiarios")
GRES1

# Categorica con Numerica
GRES2 <- ggplot(segurosCL, aes(x= segurosCL$TIPO_DE_COBERTURA, y=segurosCL$VALOR_DE_POLIZA._ANUAL)) +
  geom_boxplot(fill="orange", colour="black") +
  labs(title = "Valor de Póliza Anual según Tipo de Cobertura")

GRES2

# Gráfico de edad por tipo de cobertura
GRES3<- ggplot(segurosCL, aes(x=TIPO_DE_COBERTURA, y=EDAD)) +
  geom_boxplot(fill="lightgreen", colour="black") +
  labs(title = "Edad por Tipo de Cobertura",
        x = "Tipo de Cobertura",
        y = "Edad")

GRES3

#NUM NUM#
GRES4 <- ggplot(segurosCL, aes(x=EDAD, y=VALOR_DE_POLIZA._ANUAL)) + geom_point(color="blue", alpha=0.6) +
  labs(title = "Valor de Póliza Anual por Edad",
        x = "EDAD",
        y = "Valor de Póliza Anual")
GRES4

```



```

Seguros <- read.csv2("Seguros CSV.csv")
Seguros <- Seguros[,-c(1,2,4,5,7,9,10)]
Seguros$VALOR_DE_POLIZA._ANUAL <- as.numeric(Seguros$VALOR_DE_POLIZA._ANUAL)

|
segurosCL <- na.omit(Seguros)

library(e1071)
library(ggplot2)
library(lattice)
library(caret)

# Paso 1.- Transformo la variable clasificatoria a factor
segurosCL$TIPO_DE_COBERTURA <- factor(segurosCL$TIPO_DE_COBERTURA)

#Paso 2.- Para un arbol o cualquier derivado de un arbol de decision, se planta una semilla
set.seed(123)

#Paso 3.- Entrenamos los datos
train <- createDataPartition(segurosCL$TIPO_DE_COBERTURA, p=0.7, list = FALSE)

#Paso 4.- Armo la máquina de soporte vectorial
MSV <- svm(TIPO_DE_COBERTURA~., data = segurosCL[train,], kernel = "linear")
summary(MSV)

table(segurosCL[train,]$TIPO_DE_COBERTURA, fitted(MSV), dnn = c("Actual", "Predicho"))

#Paso 5.- Hacemos la prediccion
PRED <- predict(MSV,segurosCL)

#Paso 6.- Armamos la matriz de confusion
MSVMC <- table(segurosCL$TIPO_DE_COBERTURA, PRED, dnn = c("Actual", "Predicho"))
MSVMC

#Paso 7.- Graficamos

#VALOR_DE_POLIZA._ANUAL
plot(MSV, data = segurosCL[train,], VALOR_DE_POLIZA._ANUAL ~ COBERTURA)
plot(MSV, data = segurosCL[train,], VALOR_DE_POLIZA._ANUAL ~ NUMERO_DE_BENEFICIARIOS)
plot(MSV, data = segurosCL[train,], VALOR_DE_POLIZA._ANUAL ~ EDAD)

#EDAD
plot(MSV, data = segurosCL[train,], EDAD ~ COBERTURA)
plot(MSV, data = segurosCL[train,], EDAD ~ NUMERO_DE_BENEFICIARIOS)
plot(MSV, data = segurosCL[train,], EDAD ~ VALOR_DE_POLIZA._ANUAL)

#COBERTURA
plot(MSV, data = segurosCL[train,], COBERTURA ~ EDAD)
plot(MSV, data = segurosCL[train,], COBERTURA ~ NUMERO_DE_BENEFICIARIOS)
plot(MSV, data = segurosCL[train,], COBERTURA ~ VALOR_DE_POLIZA._ANUAL)

#NUMERO_DE_BENEFICIARIOS
plot(MSV, data = segurosCL[train,], NUMERO_DE_BENEFICIARIOS ~ EDAD)
plot(MSV, data = segurosCL[train,], NUMERO_DE_BENEFICIARIOS ~ COBERTURA)
plot(MSV, data = segurosCL[train,], NUMERO_DE_BENEFICIARIOS ~ VALOR_DE_POLIZA._ANUAL)

#Paso 1a clasificación a la base
segurosCL$PRED <- predict(MSV,segurosCL)

```

```

# Cargar datos
Seguros = read.csv2 ("../DATA/scarlet CSV.csv")

Seguros <- Seguros[,-c(1,2,4,5,7,9,10)]

Seguros$VALOR_DE_POLIZA._ANUAL <- as.numeric(Seguros$VALOR_DE_POLIZA._ANUAL)

segurosCL <- na.omit(Seguros)

# Tabla de frecuencia para el valor de póliza anual
table(Seguros$valor_poliza_anual)

##### TABLA DE FRECUENCIA VARIABLE CATEGÓRICA #####
## TIPO DE COBERTURA - tipo_cobertura ##
table(segurosCL$TIPO_DE_COBERTURA)

Fre_tipo_cobertura <- as.data.frame(table(segurosCL$TIPO_DE_COBERTURA))
Fre_tipo_cobertura
Tab_Fre_Tipo_Cobertura <- transform(Fre_tipo_cobertura,
                                   FREl = round(prop.table(Fre_tipo_cobertura$Freq), 3),
                                   FACu = cumsum(Fre_tipo_cobertura$Freq),
                                   FACuR = cumsum(round(prop.table(Fre_tipo_cobertura$Freq), 3)))

Tab_Fre_Tipo_Cobertura
install.packages("ggplot2")
library(ggplot2)

```



**Presidencia
de la República
del Ecuador**



**Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes**



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **León Negrete, Victoria Elizabeth** con C.C: # **0930299631** y **Solano Jaramillo, Scarlet Michelle**, C.C: **0931369458**, autoras del trabajo de titulación: **Gestión de riesgos en negocios internacionales: una perspectiva avanzada utilizando business intelligence y machine learning**, previo a la obtención del título de **Licenciada en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaramos tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizamos a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, a los 23 días del mes de agosto del año 2024

LOS AUTORES

f. _____

León Negrete, Victoria Elizabeth

C.C: 0930299631

f. _____

Solano Jaramillo, Scarlet Michelle

C.C: 0931369458



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Gestión de riesgos en negocios internacionales: una perspectiva avanzada utilizando business intelligence y machine learning.		
AUTOR(ES)	León Negrete, Victoria Elizabeth Solano Jaramillo, Scarlet Michelle		
REVISOR(ES)/TUTOR(ES)	Ing. Carrera Buri, Félix Miguel Mgs		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Facultad de Economía y Empresas		
CARRERA:	Negocios Internacionales		
TÍTULO OBTENIDO:	Licenciado en Negocios Internacionales		
FECHA DE PUBLICACIÓN:	23 de agosto del 2024	No. DE PÁGINAS:	111
ÁREAS TEMÁTICAS:	Gestión de riesgos, Negocios internacionales, Globalización		
PALABRAS CLAVES/KEYWORDS:	Riesgos en negocios, Mundo globalizado, Operacionalización		
RESUMEN/ABSTRACT: Este estudio se centra en la gestión de riesgos en los negocios internacionales, un tema de creciente relevancia en un mundo globalizado. A medida que las empresas expanden sus operaciones más allá de las fronteras nacionales, enfrentan desafíos como la volatilidad de los mercados, las diferencias culturales y las variaciones en los entornos políticos y económicos. El objetivo principal de esta investigación es desarrollar un sistema de clasificación automática que optimice la identificación, evaluación y mitigación de riesgos en este contexto. Para alcanzar este objetivo, se realizó una revisión exhaustiva de la literatura, estableciendo un marco teórico sólido sobre la intersección entre la tecnología y la gestión de riesgos. Además, se implementó una metodología basada en técnicas de aprendizaje automático, específicamente en clasificación supervisada, demostrando su efectividad en la mejora de la toma de decisiones y la gestión de datos. Los resultados indican que la integración de herramientas tecnológicas en la gestión de riesgos no solo mejora la eficiencia operativa de las empresas, sino que también les permite adaptarse mejor a un entorno global complejo.			
ADJUNTO PDF:	<input type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Teléfono: +593997532718 +593992345671	E-mail: victoria.leon@cu.ucsg.edu.ec scarlet.solano@cu.ucsg.edu.ec	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE)::	Nombre: Freire Quintero Cesar Enrique		
	Teléfono: +593-990090702		
	E-mail: cesar.freire@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			