Modelo Predictivo de Siniestros en Seguros Dentales

Cristian David González González, Javier Orlando Garay Rey, Jenny Carolina Robayo Ávila, Lissa Valentina González Castaño y Nidia Celina Niño Calixto.

Fundación Universitaria Compensar, Bogotá D.C.

Especialización en Big Data, Fundación Universitaria Compensar

Proyecto Fin de Grado Facultad de Ingeniería

Carlos Alberto Cañón R. – Docente Tutor

26 de mayo de 2023

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Modelo Predictivo de Siniestros en Seguros Dentales

Cristian David González González, Javier Orlando Garay Rey, Jenny Carolina Robayo Ávila, Lissa Valentina González Castaño y Nidia Niño Calixto

Trabajo de grado presentado como requisito parcial para optar al título de **Especialista en Big Data**

Director: Carlos Alberto Cañon R.

MBA - DBA - MGETI

Fundación Universitaria Unipanamericana – Compensar

Facultad de Ingeniería, Especialización en Big Data

Bogotá D.C., Colombia

2023

Contenido

Lista de Figuras 6
Resumen 7
Abstract 8
Introducción 9
Antecedentes y Justificación 11
Formulación del Problema 13
Pregunta del Problema 13
Objetivo General 14
Objetivos Específicos 14
Alcances y Limitaciones 14
Alcances 14
Limitaciones 15
Marco Teórico 16
Evolución y Desarrollo de los Seguros en Colombia 16
Figura 1 Línea de tiempo evolución seguros 17
¿Qué es un Seguro? 19
Seguro Odontológico o Bucal 20
Industria Aseguradora y Capitalización en Colombia – AXA 22
Incidencia del Riesgo en Salud 22
Figura 2 Consulta de Prevención 23
Resultados de la Industria Aseguradora y de Capitalización al Año 2022 23
Figura 3 Estadísticas de la Industria Aseguradora y de Capitalización 24
Industria Aseguradora y Capitalización – AXA Colpatria 25
Figura 4 Compañías de Seguros Comportamiento de las Primas Generales Emitidas 26
Cuotas Recaudadas. 26
Figura 5 Sociedades de capitalización con mayor ingreso de cuotas recaudadas 27
Siniestralidad. 27
Figura 6 Siniestros pagados en seguros generales, valor porcentual 28
Rendimientos Financieros. 28
Figura 7 Resultados financieros generales por compañías de seguros 29
Análisis Predictivo 29

Datos 30
Analítica Predictiva para Big Data: 30
Figura 8 ¿Qué diferentes tipos de análisis hay y para qué sirven? 32
¿Cómo hacer un análisis predictivo? 32
Figura 9 Proceso Análisis Predictivo 33
Beneficios del Análisis Predictivo 33
Técnicas de Análisis Predictivo 34
Predicción 34
Diseño Metodológico 36
Figura 10 Tipos de Investigación Cuantitativa 36
Instrumento de Investigación 37
Metodología de Desarrollo 37
Figura 11 Ciclo de vida de minería de datos 38
Entendimiento del Negocio (Business Understanding) 39
Comprensión de los Datos (Data Understanding) 39
Preparación de Datos (Data Preparation) 41
Modelado (Modeling) 42
Evaluación de los Datos (Data Evaluation) 43
Despliegue (Deployment) 44
Método por Análisis de Regresiones 44
Regresión Lineal 45
Regresión Lineal Simple 45
Método Machine Learning 45
Requerimientos Técnicos y Librerías 46
Librerías 46
Resultados 48
Figura 12 Diagrama Proceso Modelo Predictivo 49
Técnica Utilizada 49
Figura 13 Grafica de Regresión Lineal 50
Figura 14 Gráfica acumulada curva de regresión ajustada 51
Figura 15 Gráfico acumulado de residuales 51
Figura 16 Gráfica de Probabilidad Normal 52
Conclusiones y Recomendaciones 53

Conclusiones 53
Recomendaciones 54
Bibliografía 56

Lista de Figuras

Figura 1 Consulta de Prevención 23

Figura 2 Estadísticas de la Industria Aseguradora y de Capitalización 24

Figura 3 Compañías de Seguros Comportamiento de las Primas Generales Emitidas 26

Figura 4 Sociedades de capitalización con mayor ingreso de cuotas recaudadas 27

Figura 5 Siniestros pagados en seguros generales, valor porcentual 28

Figura 6 Resultados financieros generales por compañías de seguros 29

Figura 7 ¿Qué diferentes tipos de análisis hay y para qué sirven? 32

Figura 8 Proceso Análisis Predictivo 33

Figura 9 Tipos de Investigación Cuantitativa 36

Figura 10 Ciclo de vida de minería de datos 38

Figura 11 Diagrama Proceso Modelo Predictivo 49

Figura 12 Grafica de Regresión Lineal 50

Figura 13 Gráfica acumulada curva de regresión ajustada 51

Figura 14 Gráfico acumulado de residuales 51

Figura 15 Gráfica de Probabilidad Normal 52

Resumen

Las empresas que ofrecen seguros dentales buscan anticipar el valor total de siniestralidad al finalizar un periodo o fecha de corte, por pago de: exámenes, procedimientos, medicamentos y demás costos que incluyen un siniestro por cubrimiento de pólizas según su tipo o clase. A su vez, la relación directa en el apoyo de toma de decisiones para contrarrestar el desembolso por cubrimiento de dichos siniestros.

Este estudio se enfoca en realizar un modelo de predicción y análisis, con el fin de pronosticar los gastos a futuro mediante un modelo de regresión lineal simple y como apoyo para que compañías de seguros dentales tomen decisiones en sus presupuestos, prever hacia qué parte del cubrimiento que ofrece la póliza se deben disponer más recursos. La metodología de investigación utilizada fue cuantitativa debido a que es un modelo basado en datos históricos que permite predecir un valor antes que se finalice un periodo de tiempo determinado. Los resultados que se obtengan por procedimientos estadísticos y análisis de tendencias mediante el modelo de predicción permitirán obtener el valor de desembolso en costos por pólizas cobradas antes de que se termine periodo o fecha de corte.

Palabras clave: modelo cuantitativo, predicción, regresión, Python, seguro, siniestro, fecha de corte:

Abstract

Companies that offer dental insurance seek to anticipate the total value of claims at the end of a period or cut-off date, for payment of: examinations, procedures, medications and other costs that include a claim for coverage of policies according to their type or class. At the same time, the direct relationship in decision making to counteract the disbursement for policy coverage vs. the sales income of these in the same period. This study focuses on making a prediction and analysis model, in order to predict future expenses by means of a regression model, and as a support in dental insurance companies' decisions to forecast the budget, to foresee towards which part of the coverage offered by the policy more resources should be available. The methodology used is quantitative because it is a model based on historical data that allows predicting a value before the end of a given period of time. The results obtained through statistical procedures and trend analysis using the prediction model will make it possible to obtain the cost disbursement value for policies collected before the end of the period or cut-off date.

Introducción

El presente trabajo se enfoca en el desarrollo y puesta en marcha de un modelo de predicción y análisis de gastos a futuro mediante un modelo de regresión en Python, con el objetivo facilitar toma decisiones financieras más acertadas, especialmente para quienes manejan proyecciones de gastos e ingresos de una organización por las diferentes transacciones que se generan por seguros dentales.

> El análisis predictivo, es una forma de avanzada que examina datos o contenidos para responder a la pregunta: ¿Qué es probable que ocurra en el futuro? Utiliza técnicas estadísticas de modelización, big data y machine learning para extraer datos históricos y realizar predicciones. En el mundo empresarial es una técnica muy cotizada por los beneficios que puede reportar a la hora de, por ejemplo, identificar riesgos y oportunidades. (IBERDROLA, 2023)

> La incorporación de la salud bucal al sistema de salud produjo una transformación radical en términos de los procesos de planeación y desarrollo de políticas de salud en el país. Modificó el esquema de relaciones entre los actores del sistema y transformó la forma de ejercicio liberal independiente que prevaleció desde el origen de la profesión odontológica en el país, dando paso a la mayoritaria incorporación institucionalizada de la fuerza de trabajo en el mercado laboral, subordinado por una lógica de alta mercantilización y una posición dominante de las aseguradoras privadas (Otálvaro Castro, 2019)

La característica principal de este proyecto es consolidar la información mensual de gastos por seguros que se cubrieron por parte de los clientes que hicieron efectivo el seguro adquirido, buscando a través un modelo de predicción que se generé la información relevante de los servicios más utilizados al momento de cubrir un seguro, que hoy en día no maneja la empresa interesada en tener este tipo de predicción.

Para trabajar este proyecto es necesario conocer el manejo de la información por categorías definidas para cada seguro según su cuota de cubrimiento y valor comercial hacia el asegurado, en donde es relevante tener aplicativos o accesos más oportunos a la información, con el fin de obtener datos que se puedan clasificar y analizar de una manera más oportuna, con calidad; y a su vez trabajar estrategias para mantener a sus actuales asegurados y generar planes de acción para cautivar más clientes que aporten al ingreso de los seguros dentales.

Este proyecto se realiza con el interés de generar ideas innovadoras en las organizaciones que acceden a mejorar procesos a través de la analítica de datos, proponiendo predicciones para ser más efectivos y competitivos en el mercado demostrando que una base de datos permite, dar un valor no solo en toma de decisiones más seguras también en decisiones financieras y/o comerciales.

A su vez, con el fin de mostrar habilidades empíricas que se tienen al momento de iniciar un nuevo proceso educativo y como especialistas al adquirir nuevo conocimiento en cátedras durante el proceso educativo. En proyecto se enfoca construir una herramienta para optimizar los procesos de información que requiere una fuente vital de una organización, su información consolidada, actualizada y procesada convertida en un indicador de medición.

Antecedentes y Justificación

Las empresas aseguradoras no se encuentran ajenas al impacto del crecimiento de primas que se emiten por seguros asumidos por coberturas en siniestros, por ello buscan medios para mitigar el impacto en sus finanzas generando análisis a través de diferentes recursos y herramientas tecnológicas que les permitan gestionar este impacto en menor proporción o estar preparados a los cambios en el aseguramiento y siniestros que tienen que cubrir.

El presente proyecto tiene como objeto de estudio determinar el impacto de las coberturas en siniestros de una de las empresas aseguradoras en Colombia, AXA Falabella Colombia, de acuerdo con artículo de la Editorial La República, publicado el 30 de enero de 2023, en donde genera un top de empresas que emitieron más primas en el año 2022 basados en datos de informe publicado por Fasecolda a cierre del 2022 donde reporta resultados de empresas aseguradoras y su crecimiento en primas como el aumento en seguros que se gestionan.

Dando alcance al artículo, se busca determinar un método tecnológico que permita a la empresa aseguradora en mención, validar sus transacciones de seguros odontológicos enfocados en los siniestros ejecutados a una vigencia de 30 días y predecir el gasto a futuro en que van a ejecutar. Por ello se genera el contexto de antecedente del seguro en mención y la herramienta que permita generar la predicción, de acuerdo con los siguientes contextos:

> Los seguros dentales operan a modo de reembolso o como descuento directo. El modo de reembolso funciona al momento de pagar en el dentista, se presenta los gastos en la compañía con la cual se adquirió el seguro y esta reembolsa (devuelve) ese gasto en la cantidad por la cual se haya contratado el seguro. Por ejemplo, si se contrata un seguro dental con una cobertura 80%, y el costo dental final fue un pago de $100.000 la

> compañía de seguros te devuelve $80.000 en donde el otro 20% es costeado por el asegurado. El modo descuento directo funciona cuando se va a la clínica que presta el servicio odontológico y al momento de realizar el pago, en el mismo lugar se hará un descuento en el precio a pagar según lo establecido o la cobertura adquirida (QuePlan.Cl, 2023)

> Se realizará un modelo de análisis predictivo que es una subdisciplina del análisis de datos que usa técnicas de estadística, como aprendizaje computacional o data mining, esto con el fin de desarrollar modelos que predicen eventos futuros o conductas. Estos modelos predictivos permiten aprovechar los patrones de comportamiento encontrados en los datos actuales e históricos para identificar riesgos, oportunidades y mejoras (Espino & Martínez, 2017).

El modelo de predicción puede proyectar mediante una regresión lineal simple el valor total de siniestros que a reembolsar o pagar antes de que termine la fecha de corte, esto se va a evaluar según datos actuales y de periodos anteriores de esta forma se aprovechan los patrones de comportamiento y así la aseguradora podrá saber con cuánto dinero contar para cubrir dicha siniestralidad y tomar medidas de contingencia para equilibrar los gastos y lo recaudado.

"Este proyecto se centra en el estudio con herramientas Open Source que ofrecen las funcionalidades necesarias para llevar a cabo proyectos de análisis predictivo (Espino & Martínez, 2017)". La herramienta elegida fue Python, que es una herramienta madura que dispone de interfaces

gráficas que permiten abstraer la complejidad matemática del análisis predictivo, poniéndolo a disposición de empresas, instituciones y organizaciones.

Formulación del Problema

Las empresas de seguros dentales deben hacer el pago de: honorarios médicos odontológicos, medicamentos, procedimientos quirúrgicos, exámenes, entre otros; por los siniestros ocurridos con alguno de sus seguros por falta de un modelo que permita predecir o conocer el valor total que desembolsa la empresa AXA Colpatria Falabella por estos en seguros odontológicos y/o dentales al cierre del mes próximo.

Actualmente, se conoce el valor total de los pagos realizados una vez ejecutados los siniestros algún tiempo después de haberlo desembolsado o haber echo cierre de corte; y no se tiene ningún modelo que permita predecir los posibles gastos a futuro, por ende, no se cuenta con argumentos de toma de decisiones oportunas para disminuir los pagos de siniestros dentales.

Pregunta del Problema

¿Cómo conocer y/o predecir el valor diario y acumulado en gastos por siniestros de seguros odontológicos o dentales que se ejecutan al cierre de mes para la empresa AXA Colpatria Falabella?

Objetivo General

Analizar, diseñar y desarrollar un modelo de predicción para AXA Colpatria Falabella que permita tener información diaria y acumulada de sus futuras transacciones por siniestros odontológicos, identificando tendencias y apoyo de toma de decisiones.

Objetivos Específicos

- Realizar un diagnóstico de los datos entregados por la organización.
- Generar un diseño del modelo de predicción para regresión lineal simple.
- Proponer un modelo de predicción mediante una regresión lineal.
- Ejecutar un set de pruebas.

Alcances y Limitaciones

Alcances

- Predicción para empresa de seguro odontológico AXA Colpatria Falabella Colombia.
- Los datos para el modelo del proyecto se generan con bases abiertos.
- Se utilizará un método de regresión lineal simple.
- Originar escenarios de comparación donde se identifique que el modelo genere un 90% de eficacia entre el valor diagnosticado y el real.

Limitaciones

- Análisis predictivo enfocado solo a seguros odontológicos de la empresa en mención.
- El tiempo del modelo de predicción será máximo de 30 días, en comparaciones confidenciales de días de la semana.
- El modelo predictivo se realizará en Python.
- La información de la predicción, solo se entrega a la empresa como resultados numéricos que arroje el modelo, no incluye recomendaciones ni análisis de carácter comercial, financiero entre otros.

Marco Teórico

Evolución y Desarrollo de los Seguros en Colombia

Los primeros pasos en el desarrollo e incremento de seguros en Colombia le dieron empresas extranjeras, “las cuales fijaron sus sucursales en diferentes regiones del territorio nacional, como lo hizo “El Porvenir de las Familias” con el establecimiento de una empresa de origen chileno alrededor de 1857.” (Bravo Reyes, 2011, pág. 150).

A pesar de esta presencia en el siglo XVIII la economía se encontraba en constante lucha entre las comunicaciones y el comercio interior y externo, las exportaciones mineras, restricciones comerciales e importaciones de bienes de consumo impactaron la demanda mínima de servicios de seguros.

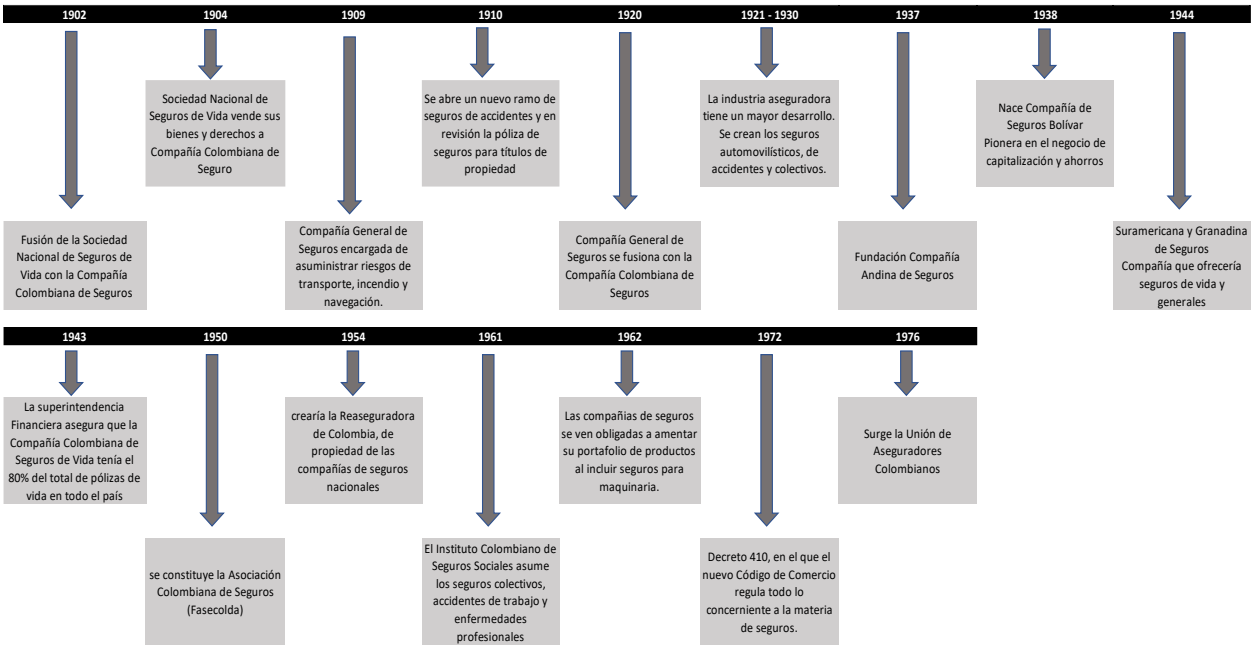
> Fue en 1.874 cuando en una reunión presidida por el jefe del estado, Santiago Pérez, se constituyó la primera sociedad anónima colombiana encargada de la explotación del seguro, la cual se denominó “Compañía Colombiana de Seguros, Colseguros”. Operando inicialmente con transporte, y cuyo propósito consistiría en asegurar las mercancías transportadas por vía fluvial a través del río Magdalena hasta Bogotá. En 1876 Colombia sufría delicados problemas de orden público y la Compañía se vería obligada a aumentar las sobreprimas. Esta situación se tornaría cada día más difícil ya que las mercancías no alcanzaban a llegar a su destino y eran devueltas, lo que obligaba a la Compañía a no aceptar nuevos amparos (Bravo Reyes, 2011).

> En 1876 Colombia sufría delicados problemas de orden público y la Compañía se vería obligada a aumentar las sobreprimas. Esta situación se tornaría cada día más difícil ya que las mercancías no alcanzaban a llegar a su destino y eran devueltas, lo que obligaba a la Compañía a no aceptar nuevos amparos.

Estas dificultades se superarían al pasar un año iniciando un proceso de negociación con tres compañías reaseguradoras de Francia ampliando el alcance de los riesgos a incendios y trasportes, en "1896 surgen las nuevas pólizas que amparan metales preciosos y reses importadas" (Bravo Reyes, 2011).

Figura 1

Línea de tiempo evolución seguros



Nota. Línea de tiempo Evolución y Desarrollo de los Seguros en Colombia de 1902 a 1930. Adaptado de [Fotografía], Bravo Reyes, 2011.

https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1169&context=gs#:~:text=En%201909%20se%20crea%20la,seguros%20para%20t%C3%ADtulos%20de%20propiedad

Con el nacimiento de la Compañía Colombiana de Seguros, Colseguros se introdujo el concepto de protección contra los riesgos como medida económica para la actividad empresarial y los seguros surgen como una herramienta de protección a la vida (Fadul Vergara, Historia de los Seguros en Colombia, 2021).

> En la legislación colombiana se encuentran estipulados una serie de regulaciones y conceptos, referente a la actividad aseguradora, las cuales se encuentran contenidas en la Constitución Política Nacional de 1991, en el marco del estado social de derecho, estipula una serie de principios y derechos los cuales regulan todo el ordenamiento jurídico, entre ellos todas las relacionadas con las actividades que desarrollan los particulares, una de ellas, la actividad aseguradora. Adicionalmente, códigos, leyes, resoluciones y decretos de protección tanto para aseguradoras como asegurados.
>
> La transformación de las empresas aseguradora tiene sus inicios en la reforma a la Ley 45/90, en la cual el Estado Nacional libera las tasas en los productos de seguros, con el propósito de buscar una mayor competencia fundamentada en el servicio, logrando operaciones internas favorables, pero donde queda la confianza e intención de sus clientes por comprar seguros como el medio comercial posible para trasladar sus riesgos.

> La situación actual de compañías en el ramo de los seguros en el país se refleja en el desarrollo y crecimiento de la actividad, gracias a que su presencia ha aumentado significativamente, contando ahora con más de 38 compañías en dicho ramo. En la actualidad podemos contratar seguros de vida por Internet o realizar los pagos de las primas con financiación; se manejan equipos modernos de inspección y la expedición de documentos se hace a velocidades increíbles. El sector asegurador solo tiene unos veintiún años de evolución y que este desarrollo se generó exclusivamente al interior de las aseguradoras colombianas (Bravo Reyes, 2011)

¿Qué es un Seguro?

Instrumentos financieros diseñados “para minimizar el impacto económico que puedan sufrir las personas como consecuencia de la ocurrencia de un suceso inesperado que pueda poner en peligro su vida o sus bienes. (Viva Seguro Fasecolda, 2018)

> Los seguros funcionan cuando las personas que adquieren seguros comparten sus riesgos en un gran fondo común para que estos sean administrados por agentes expertos, las aseguradoras, quienes compensan a los beneficiarios del seguro por los eventos que ocurran y estén amparados. Este debe ser a futuro, incierto y medible, que tenga una probabilidad de ocurrencia y que se pueda cuantificar económicamente. (Viva Seguro Fasecolda, 2018).

Las compañías de seguros se esfuerzan constantemente por garantizar la adecuada protección y seguridad de los asegurados, ofreciendo distintos tipos de seguros en función de los riesgos, buscando

opciones de cobertura y necesidades de protección específicas de cada individuo. Actualmente, ofrecen tres tipos de seguros:

- Seguros personales
- Seguros de daños o patrimoniales
- Seguros de prestación de servicios.

"En la prestación de seguros se incluyen aquellos los de actividad aseguradora en la prestación de un servicio al asegurado, regularmente ofrece una indemnización económica. Se encuentran principalmente los de asesoría o defensa jurídica, asistencia médica y odontológica, asistencia en viajes, asistencia veterinaria, servicios funerarios, reparaciones y/o arreglos al hogar (Fundación MAPFRE, 2023)".

Seguro Odontológico o Bucal

"Es un seguro voluntario con el objetivo de satisfacer necesidades en salud oral, brindando atención con los máximos estándares de calidad, ofreciendo servicios en todas las especialidades odontológicas".

"Un tipo de cobertura importante puede ayudar a las personas a gestionar el coste de los cuidados dentales. Comprender cómo funciona y qué cubre el seguro dental, ayuda en la tomar decisiones sobre su salud dental y su bienestar financiero".

Puede cubrir cuidados preventivos rutinarios, procedimientos más extensos, como empastes, extracciones y endodoncias. Además, el seguro dental puede cubrir también tratamientos de:

- Diagnóstico
- Urgencias.
- Operatoria.
- Odontopediatría.
- Ortodoncia
- Endodoncia.
- Periodoncia.
- Cirugía y rehabilitación.
- Procedimientos estéticos
- Actividades de promoción y prevención (Manatal, 2023).

La comercialización de seguros en Colombia puede ser realizada directamente por las entidades aseguradoras (compañías y cooperativas de seguros), o por medio de actores autorizados legalmente para realizar la actividad de intermediación de seguros, esto es: sociedades corredoras, agencias de seguros y agentes. Además, se permite la comercialización masiva de seguros a través de figuras como el uso de red de entidades vigiladas por la Superintendencia Financiera y el de corresponsalía

Actualmente, la venta directa de seguros es realizada a través de 24 compañías de seguros generales, 19 compañías de seguros de vida y 2 cooperativas de seguros;

> mientras en materia de intermediación de seguros operan 50 sociedades corredoras de seguros. No se conocen cifras oficiales que den cuenta del número de agencias y agentes en el mercado, pero según el documento de 2013 del Fondo Monetario Internacional que recoge las conclusiones del Financial Stability Assessment Program para Colombia, existían para ese año más de 10 mil agencias y agentes (Torresa, 2017).

De acuerdo con último estudio generado por MERCO del ranking de empresas aseguradora, "actualmente las siguientes aseguradoras en Colombia con mejor reputación han sido objeto de revisión independiente por parte de KPMG. La metodología seguida para elaborar Merco Empresas se encuentra disponible: Sura, Seguros Bolívar, Allianz Colombia, Aseguradora Solidaria" (MERCO, 2022).

Industria Aseguradora y Capitalización en Colombia – AXA

Incidencia del Riesgo en Salud

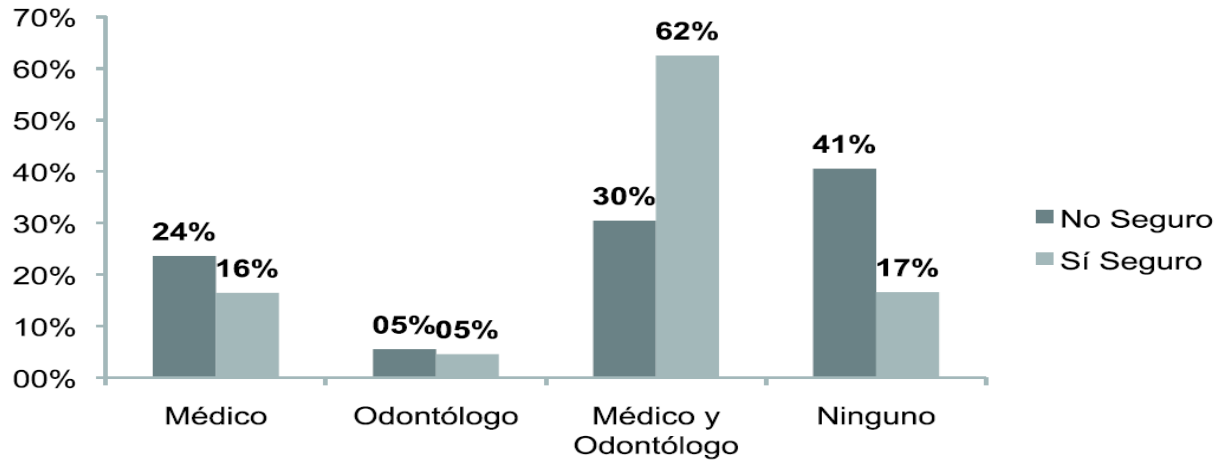
Con el fin de cubrir los costos asociados a estos eventos nocivos para la salud, las personas utilizan diferentes fuentes incluso al mismo tiempo. En particular, solo se analiza el grupo de población que se considera que tiene un seguro privado.

"Se estima que sólo el 16% de las urgencias y el 40% de las hospitalizaciones fueron cubiertas por el Sistema de Seguridad Social Público, el restante fue cubierto por el seguro privado adquirido por las personas, sin la concurrencia de recursos del sistema de seguridad social" (Mayorga, 2013).

El análisis anterior indica que del 80% de las atenciones ambulatorias y el 60% de los eventos de hospitalización de las personas con algún tipo de aseguramiento privado son cubiertas por éste, sin concurrencia de recursos del aseguramiento público. Las variables que inciden en las condiciones generales de salud pública, tales como el estado subjetivo de salud, la existencia de planes de prevención en salud, el tiempo de espera ante la solicitud de consultas de medicina general y el tiempo de incapacidad asociado a los eventos de riesgo en salud (Mayorga, 2013).

Figura 2

Consulta de Prevención



Nota. FASECOLDA ECV 2011. Porcentaje de personas que asisten a consulta de prevención, según el tipo de cobertura. Adaptado de [Fotografía], Mayorga, W, 2011 (https://revista.fasecolda.com/index.php/revfasecolda/article/view/18)

Resultados de la Industria Aseguradora y de Capitalización al Año 2022

Con base en los estados financieros y anexos transmitidos originalmente por las entidades aseguradoras a la Superintendencia Financiera, los resultados preliminares de la Industria Aseguradora y de Capitalización al cierre del año 2022, se identifica en relación con los seguros y siniestralidad que:

Primas Emitidas, para el grupo de seguros de personas en el 2022, el mercado asegurador emitió primas por un total de $ 12.085.936 millones, lo que representa un crecimiento del 18% con respecto al 2021.

En salud, el aumento en las tarifas como consecuencia del mayor costo de las atenciones médico-asistenciales, además de la mayor demanda en planes complementarios y seguros de salud, explican crecimiento del 16% en las primas.

En siniestros incurridos, que se definen como el costo final de las reclamaciones asumidas por las aseguradoras, presentaron una variación del 12% durante el 2022 con respecto al mismo período del año anterior.

Figura 3

Estadísticas de la Industria Aseguradora y de Capitalización

COMPARATIVO POR RAMOS DE SEGUROS

Comparativo diciembre-2021 vs diciembre-2022

(millones de pesos)

RAMO	PRIMAS EMITIDAS			SINIESTROS PAGADOS			PRIMAS DEVENGADAS			SINIESTROS INCURRIDOS			SINIESTRALIDAD CTA CIA		
	dic-21	dic-22	VAR (%)	dic-21	dic-22	VAR (%)	dic-21	dic-22	VAR (%)	dic-21	dic-22	VAR (%)	dic-21	dic-22	VAR (%)
Colectivo y Vida grupo	4.858.215	5.902.496	21,5%	3.157.311	2.138.495	-32,3%	4.099.509	5.027.139	22,6%	2.557.015	1.796.939	-29,7%	62%	36%	-26,6%
Salud	**2.420.219**	**2.804.697**	**15,9%**	**1.805.210**	**2.012.824**	**11,5%**	**2.332.022**	**2.637.013**	**13,1%**	**1.723.527**	**1.976.816**	**14,7%**	**74%**	**75%**	**1,1%**
Accidentes personales	1.312.512	1.564.885	19%	298.015	373.467	25,3%	1.118.456	1.298.472	16,1%	222.275	331.191	49,0%	20%	26%	5,6%
Vida Individual	1.223.982	1.349.917	10%	543.137	344.881	-36,5%	975.756	1.088.660	11,6%	444.374	316.379	-28,8%	46%	29%	-16,5%
Educativo	161.751	177.919	10%	240.283	272.184	13,3%	64.244	-12.570	-119,6%	265.840	320.597	20,6%	414%	-2550%	-2964,3%
Otros ramos de Personas	248.627	286.022	15%	274.448	264.973	-3,5%	135.391	30.611	-77,4%	267.887	254.635	-4,9%	198%	832%	634,0%
TOTAL SEGUROS DE PERSONAS	**10.225.306**	**12.085.936**	**18%**	**6.318.404**	**5.406.824**	**-14,4%**	**8.725.378**	**10.069.325**	**15,4%**	**5.480.918**	**4.996.557**	**-8,8%**	**63%**	**50%**	**-13,2%**

Salud: se incluyen los ramos de Salud y Enfermedades de alto costo.

Nota. Nájera A., A. A., & Durán R., V. A., 2023. Resultados de la industria a diciembre de 2022. Adaptado [Fotografía] Revista Fasecolda, (189), 32–41. Recuperado a partir de https://revista.fasecolda.com/index.php/revfasecolda/article/view/877

Industria Aseguradora y Capitalización – AXA Colpatria

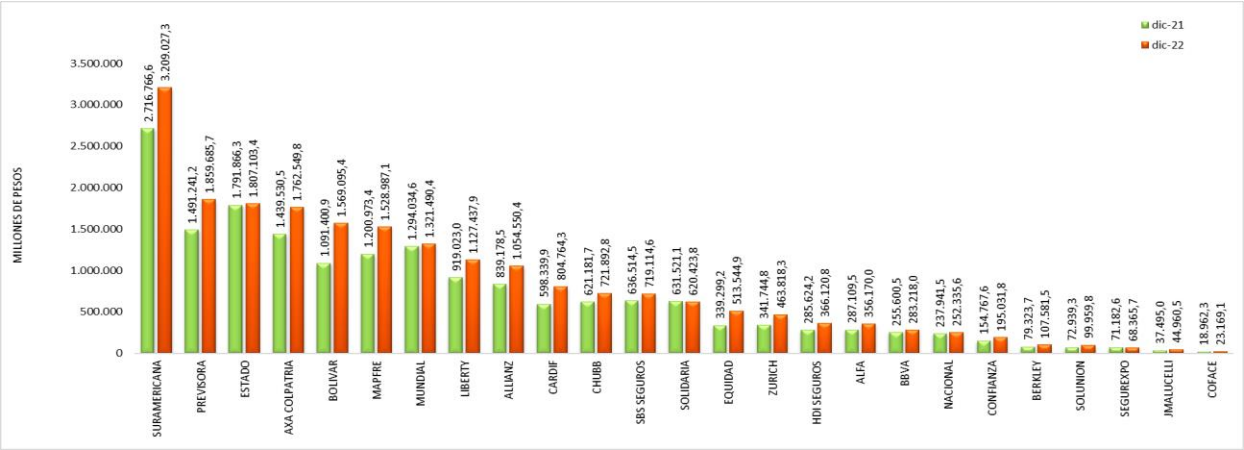
De acuerdo con el contexto del proyecto y para analizar cifras que arrojaron los estudios de 2021 de la de la Federación de Aseguradores y resultados de los movimientos financieros, de la empresa de estudio, es importante definir los términos a los que hace referencia y las cifras del año inmediatamente anterior:

> "Primas Emitidas, son valores cobrados por la compañía sobre riesgos asumidos en un determinado periodo; y la compañía aseguradora, se obliga a indemnizar o satisfacer unas prestaciones en caso de que se produzca el evento a cubrir" (ALMARIO, 2023).

De acuerdo con el boletín de Fasecolda para 2022, el sector asegurador sumo primas emitidas por más de $38,01 billones, representó un crecimiento de 19,88%: La empresa aseguradora Axa Colpatria generó un aumento del 27% en prima emitidas para el año 2022 frente al 2021, estas primas emitidas, son el costo del amparo otorgado por la aseguradora por cada póliza del aseguradora.

Figura 4

Compañías de Seguros Comportamiento de las Primas Generales Emitidas

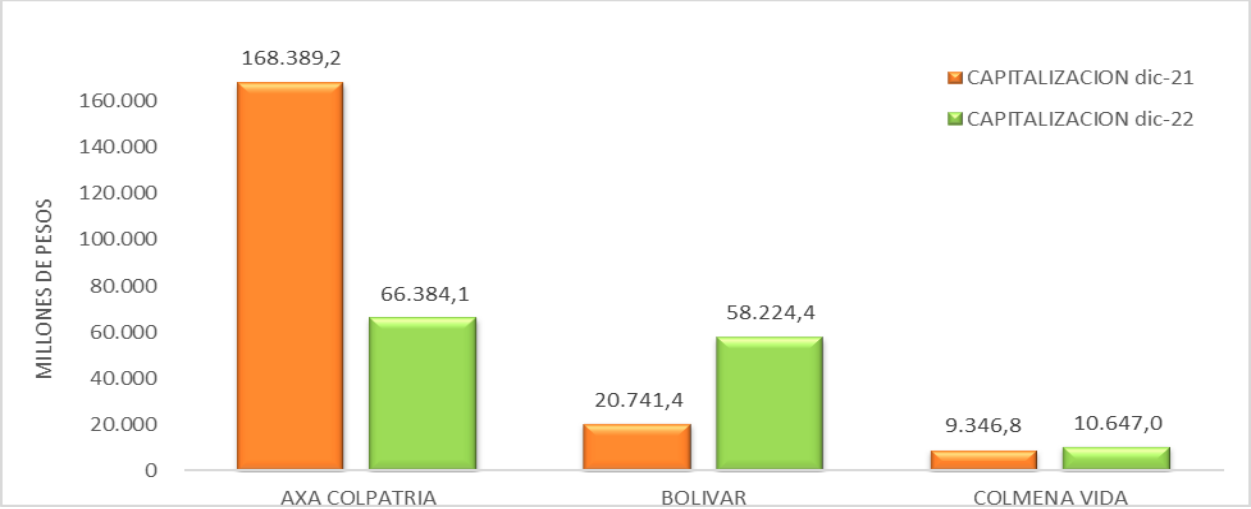


Nota. Nájera A., A. A., & Durán R., V. A., 2023. Resultados de la industria a diciembre de 2022. Adaptado [Fotografía] Revista Fasecolda, (189), 32–41. Recuperado a partir de https://fasecolda.com/fasecolda/estadisticas-del-sector/resumen-ejecutivo/

Cuotas Recaudadas. Constituyen la obligación de la aseguradora para con los suscriptores o reserva de los planes, De acuerdo con el boletín de Fasecolda para 2022 las primas emitidas, que es el costo del amparo otorgado por la aseguradora por cada póliza; la empresa aseguradora Axa Colpatria generó un aumento del 27% en prima emitidas para el año 2022 frente al 2021.

Figura 5

Sociedades de capitalización con mayor ingreso de cuotas recaudadas

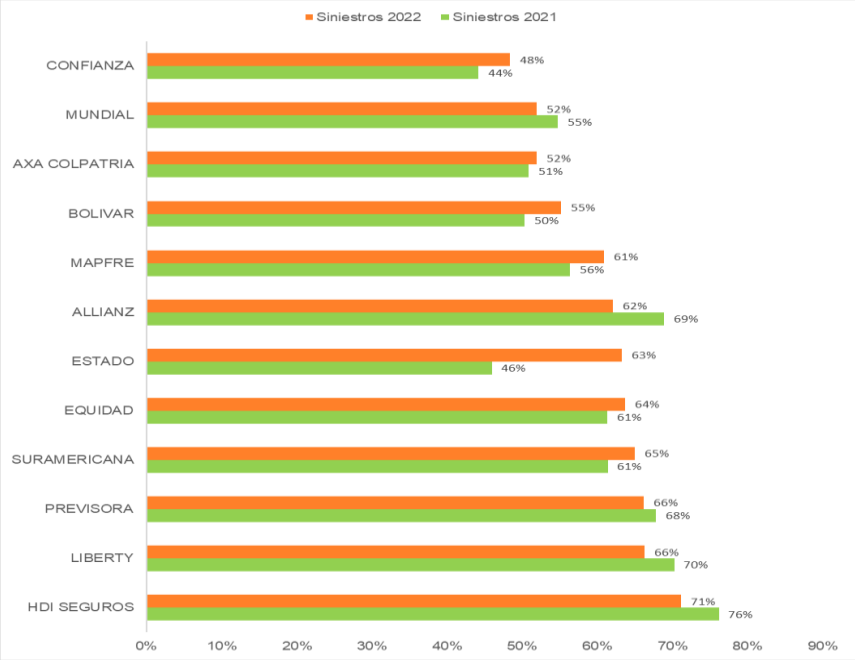


Nota. Nájera A., A. A., & Durán R., V. A., 2023. Resultados de la industria a diciembre de 2022. Adaptado [Fotografía] Revista Fasecolda, (189), 32–41. Recuperado a partir de https://fasecolda.com/fasecolda/estadisticas-del-sector/resumen-ejecutivo/

Siniestralidad. Está definida como la relación entre los siniestros incurridos, que representan el costo final de las reclamaciones por parte de las aseguradoras, y las primas devengadas, que corresponden al ingreso real por primas. Los valores reconocidos a los suscriptores de los títulos de capitalización alcanzaron la suma de $135 mil millones, es decir, un 32% menos que para el 2021.

Figura 6

Siniestros pagados en seguros generales, valor porcentual

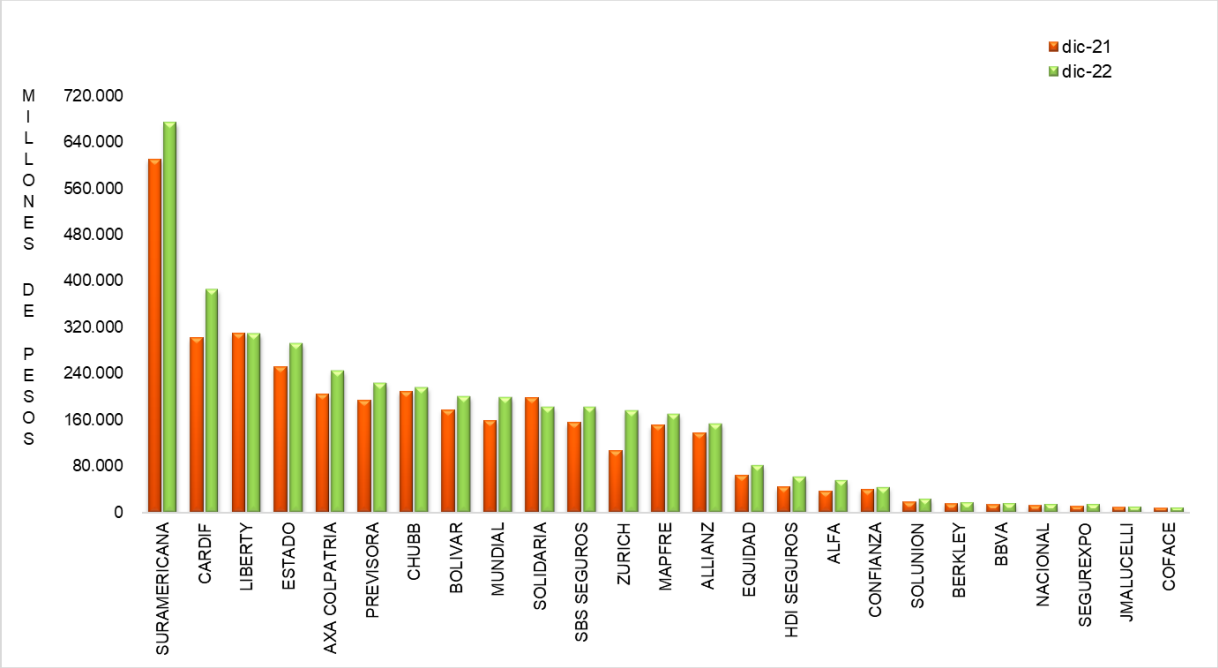


Nota. Nájera A., A. A., & Durán R., V. A., 2023. Resultados de la industria a diciembre de 2022. Adaptado [Fotografía] Revista Fasecolda, (189), 32–41. Recuperado a partir de https://fasecolda.com/fasecolda/estadisticas-del-sector/resumen-ejecutivo/

Rendimientos Financieros. Al cierre del año 2022 la industria aseguradora alcanzó un total de $8.4 billones, que le permitieron generar utilidades netas por $2.7 billones, es decir, $1.8 billón más que en el año 2021.

Figura 7

Resultados financieros generales por compañías de seguros



Nota. Nájera A., A. A., & Durán R., V. A. (2023). Resultados de la industria a diciembre de 2022. *Revista Fasecolda*, (189), 32–41. Recuperado a partir de https://revista.fasecolda.com/index.php/revfasecolda/article/view/877

/ Análisis Predictivo

El análisis predictivo, es la ciencia de predecir el comportamiento y está presente en todos los aspectos de nuestras vidas, valora el poder de los datos e influye en millones de decisiones todos los días. Los bancos, las empresas, los políticos y las organizaciones lo utilizan para convencernos que necesitamos sus productos o servicios. Predecir el comportamiento humano combate el riesgo financiero, mejora la atención médica, combate el spam, ayuda a perseguir el crimen y aumentar las ventas. La predicción se alimenta del recurso no natural más potente y actualizado; datos que trasmiten

datos residuales recopilados como resultado de las tareas rutinarias de las organizaciones (Siegel, 2013, pág. 26)

Datos

Son representaciones simbólicas (numéricas, alfabéticas, algorítmicas, etc.) de determinados atributos o variables, sean cualitativas o cuantitativas; es decir, descripciones codificadas de hechos empíricos, eventos y entidades. Los datos, son la información (valores o referencias) que las computadoras obtienen de diversas formas y que son manipuladas mediante el procesamiento de algoritmos de programación. Individualmente son irrelevantes para el usuario del sistema, pero en conjunto pueden interpretarse para obtener información completa y específica (Etecé, Editorial, 2021).

Analítica Predictiva para Big Data:

Big data, son grandes cantidades de datos, tanto estructurados como no estructurados, que son difíciles de recopilar, gestionar, procesar y analizar sin el uso de la tecnología adecuada debido a su tamaño, complejidad y velocidad de generación de nueva información. Estos datos se pueden utilizar para el aprendizaje automático en proyectos de aprendizaje automático, modelos predictivos y diversas aplicaciones analíticas (SYDLE, 2023).

> La mayoría de las compañías disponen de una gran cantidad de datos. El problema es que son datos no estructurados o semiestructurados y la toma de decisiones en ese escenario es imposible. La clave es disponer de herramientas que organicen los datos,

extraigan relaciones y hagan proyecciones sin tener conocimientos técnicos elevados (Heredia, 2015).

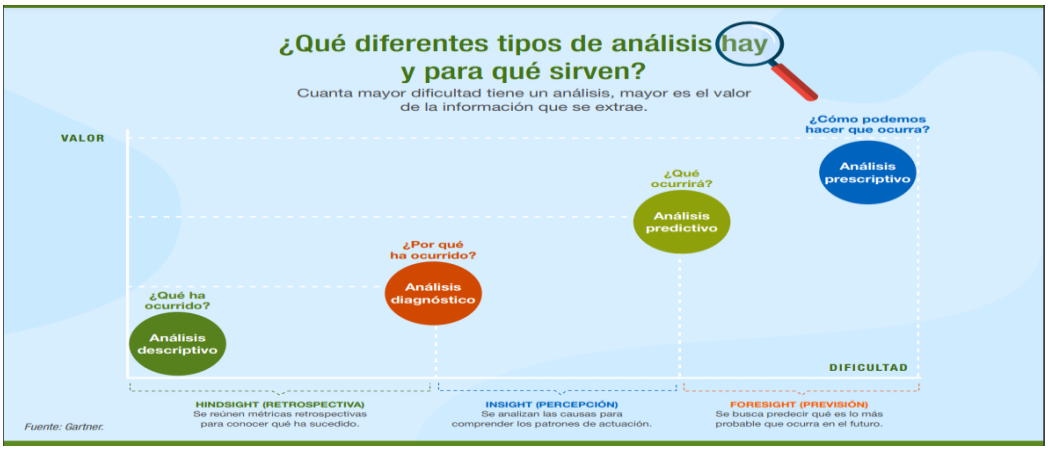
Esto se consigue con herramientas que combinen los datos no estructurados de forma sencilla, elabore resúmenes visuales y de lenguaje plano para que el analista de negocio (no de datos) entienda el valor de la información, establecer proyecciones de demanda y perfiles personales de clientes (Heredia, 2015)

Según la consultora Gartner, el análisis predictivo es una forma de análisis avanzado que examina datos o contenidos para responder a la pregunta: ¿qué es probable que ocurra en el futuro? Gracias al big data, los datos obtenidos a través de todos los sistemas conectados pueden interpretarse para obtener predicciones sobre cómo se va a comportar una persona o un grupo de población, algo también aplicable a negocios o procesos.

Hay diferencias entre el análisis descriptivo, el diagnóstico, el predictivo y el prescriptivo, pero ahora nos centraremos en estos dos últimos. **Mientras el análisis predictivo utiliza los datos recopilados para realizar predicciones sobre un supuesto**, el análisis prescriptivo toma esos datos y profundiza en ellos hasta hallar la manera de lograr que ese supuesto ocurra. Respecto a los dos primeros, el análisis descriptivo examina el hecho en sí, mientras que el análisis diagnóstico atiende a las causas.

Figura 8

¿Qué diferentes tipos de análisis hay y para qué sirven?



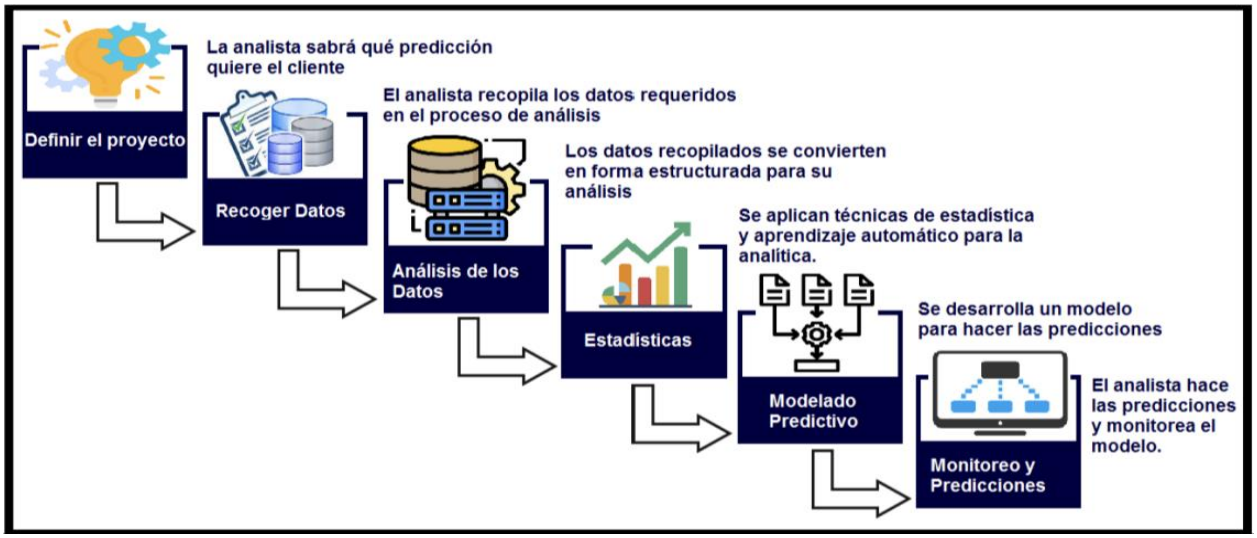
Nota. Según la consultora Gartner, el análisis predictivo es una forma de análisis avanzado para responder a la pregunta: ¿qué es probable que ocurra en el futuro? Tomada de [Fotografía], (IBERDROLA, 2023)

¿Cómo hacer un análisis predictivo?

El proceso de análisis predictivo parte de un conjunto de datos y requiere los siguientes pasos:

Figura 9

Proceso Análisis Predictivo



Nota. El proceso del análisis predictivo está conformado por seis fases. Adaptado por [FOTOGRAFIA]. (Vaibhav, 2018)

Beneficios del Análisis Predictivo

El principal beneficio del análisis predictivo es que permite a las empresas y organizaciones aprender de su experiencia pasada a partir de sus datos y tomar medidas para aplicar lo aprendido en el futuro y obtener mejores resultados. Adicionalmente, repasamos otras de sus ventajas:

- La aplicación de la inteligencia artificial, y de técnicas derivadas como el machine learning, elimina la carga del análisis manual de datos y minimiza los errores.

- Genera una ventaja competitiva en el mercado al identificar mejor y más rápido tendencias, oportunidades y desafíos, aumentando las ventas y la optimización de recursos.

- Al identificar las variables que influyen más en el comportamiento de las personas y qué les impulsa a comprar, mejora de la satisfacción de los clientes.
- Aumentan las probabilidades de éxito en el lanzamiento de productos, ya que se posee un gran conocimiento sobre las necesidades e intereses de los clientes (IBERDROLA, 2023)

Técnicas de Análisis Predictivo

Las técnicas de análisis predictivo usadas para crear modelos se basan, sobre todo, en técnicas de regresión. Mencionamos a continuación algunas de ellas: regresión lineal, elección discreta, series temporales, análisis de supervivencia o de duración, árboles de clasificación y regresión, etc. También existen una serie de métodos estadísticos avanzados para la regresión y la clasificación que usan técnicas de machine learning (IBERDROLA, 2023)

Predicción

Es el proceso que intenta determinar los valores de una o varias variables, a partir de un conjunto de datos. La predicción de valores continuos puede planificarse por las técnicas estadísticas de regresión (JAM85, DEV95, AGR96). Se pueden resolver muchos problemas por medio de la regresión lineal, y puede conseguirse todavía más aplicando las transformaciones a las variables para que un problema no lineal pueda convertirse a uno lineal (Lopez, 2006)

Python

Es un lenguaje de programación de alto nivel, interpretado, orientado a objetos y de uso generalizado con semántica dinámica, que se utiliza para la programación de propósito general. Fue creado por Guido van Rossum, nacido en 1956 en Haarlem, Países Bajos como sucesor del lenguaje ABC. Python 3 es la versión más nueva (para ser precisos, la actual) del lenguaje. Está atravesando su propio camino de evolución, creando sus propios estándares y hábitos. En el 2001 fue creada Python Software Foundation, organización sin ánimo de lucro, comunidad que tiene como objetivo desarrollar, mejorar, expandir y popularizar Python y su entorno que se mantienen por las personas reunidas en torno a PSF.

En 1999, Guido van Rossum definió sus objetivos para Python:

- Un lenguaje fácil e intuitivo tan poderoso como los de los principales competidores.
- De código abierto, para que cualquiera pueda contribuir a su desarrollo.
- El código que es tan comprensible como el inglés simple.
- Adecuado para tareas cotidianas, permitiendo tiempos de desarrollo cortos.

(STUDOCU, 2022)

Unos 20 años después, está claro que todas estas intenciones se han cumplido. Algunas fuentes dicen que Python es el lenguaje de programación más popular del mundo, mientras que otros afirman que es el tercero o el quinto. De cualquier manera, todavía

ocupa un alto rango en el top ten de PYPL PopularitY of Programming Language y el TIOBE Programming Community Index (Python, 2023)

Diseño Metodológico

El tipo de investigación utilizado es cuantitativo correlacional debido a que permite relacionar dos o más variables utilizando métodos de análisis matemáticos.

La investigación cuantitativa generalmente se implementa para concluir una relación entre dos o más variables dentro de un público objetivo. La investigación cuantitativa tiene preguntas de investigación claramente definidas y de estas, se deben obtener respuestas de una población o grupo objetivo (Marytere, 2023)

Figura 10

Tipos de Investigación Cuantitativa



Nota. Cuatro tipos diferentes de investigación cuantitativa para definir una investigación cuantitativa. Tomada de [Fotografía], por questionpro. 2023, (https://www.questionpro.com/blog/wp-content/uploads/2019/10/1118-Infografia-tipos-de-investigacion-cuantitativa.jpg)

Instrumento de Investigación

Para esta investigación se tomó como instrumento la observación cuantitativa correlacional, es una colección objetiva de datos que se centra en números y valores, los resultados de la observación cuantitativa se obtienen utilizando métodos de análisis estadísticos y numéricos. Implica la observación de cualquier entidad que pueda asociarse con un valor numérico.

Metodología de Desarrollo

Seleccionar el tipo de metodología correcto para generar una predicación, se deben tener en cuenta factores como: “el tipo de problema, es decir, clasificación o regresión según el tipo de valor pronosticado sea categórico o numérico. El propósito de integrar un lenguaje de programación como Python, consiste en integrar métodos de analítica de datos con herramientas tecnológicas, asequibles (gratis); sus aplicaciones y métodos sencillos que permitan generar cálculos predictivos a través de un lenguaje de códigos” (Márquez, 2018).

Un análisis predictivo utiliza técnicas estadísticas de modelización, big data y machine learning para extraer datos históricos y realizar predicciones. La metodología utilizada para el desarrollo del siguiente proyecto es, **CRISP-DM** (Cross-Industry Standard Process for Data Mining) (IBM, 2021)

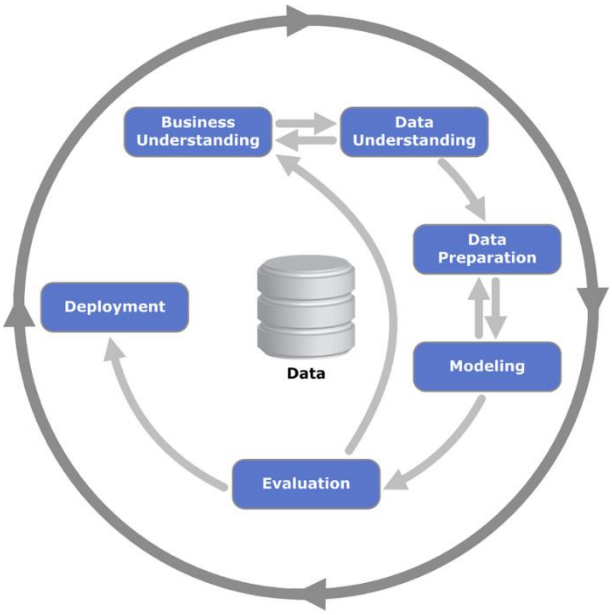
Esta modelo es común en proyectos de minería de datos, análisis y ciencia de datos permitiendo descripción de fases de un proyecto, tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. Está basado en seis fases secuenciales:

- Entendimiento del negocio: ¿Qué necesita el negocio?
- Comprensión de datos: ¿Qué datos tenemos / necesitamos? ¿Está limpio?
- Preparación de datos: ¿Cómo organizamos los datos para el modelado?
- Modelado: ¿Qué técnicas de modelado debemos aplicar?
- Evaluación: ¿Qué modelo cumple mejor con los objetivos comerciales?
- Implementación: ¿Cómo acceden las partes interesadas a los resultados?

(IBM, 2021)

Figura 11

Ciclo de vida de minería de datos



Nota. El ciclo vital del modelo contiene seis fases con flechas que indican las dependencias más importantes y frecuentes entre fases. La secuencia de las fases no es estricta. De hecho, la mayoría de los proyectos avanzan y retroceden entre fases si es necesario.

Entendimiento del Negocio (Business Understanding)

> Busca alinear los objetivos del proyecto de data mining con los objetivos del negocio, entendiendo y contextualizando la situación sobre la cual se desarrollará el análisis. Una vez establecidos y alineados los objetivos se creará un plan de proyecto que incluya las fases, tareas y actividades necesarias (Vallalta, 2023).

Esta etapa, como su nombre lo indica, busca entender y contextualizar la situación sobre la cual se desarrollará el proyecto. Para lograrlo se partió de hacer una caracterización del sistema de seguros dentales u odontológicos de Axa Falabella, que permita hacer una contextualización y entender qué lo compone, cómo se integra internamente y qué se necesita para que funcione. Al tener esto claro, se procedió a definir unos datos iniciales que me permitieran lograr analizar y modelar la predicción propuesta.

Comprensión de los Datos (Data Understanding)

> La fase de comprensión de los datos inicia con una recolección inicial de la información y procede con actividades que permiten familiarizarse con los datos (Muñoz, 2021).

> Conocer los datos, estructura y distribución, y la calidad de estos, crear algunas derivadas con base en necesidades propias del sistema y del modelo, y limpiar los registros incompletos (Vallalta, 2023).

En esta etapa se pretende comprender los datos y saber que variables vamos a utilizar para poder diseñar el modelo de predicción. Con los datos obtenidos luego de simular los siniestros en la fase de entendimiento del negocio, fue necesario hacer una selección, integración y limpieza para eliminar las variables innecesarias de las datas suministradas. Crear algunas derivadas con base en necesidades propias del sistema y del modelo, y limpiar los registros incompletos. Para completar este proceso se tienen los siguientes pasos:

- **Recopilación de los datos**: En este punto se realiza la recolección de los datos que se listaron en los recursos del proyecto. Se deben listar todos los conjuntos de datos, la ubicación, el formato, los métodos utilizados para adquirirlos, y los problemas encontrados durante la consulta. Se deben registrar los problemas encontrados y las posibles soluciones que se implementaron.

- **Descripción de los datos:** En esta parte se deben examinar las propiedades de los conjuntos de datos adquiridos. Se debe incluir el formato de la información, la cantidad de información (por ejemplo, el número de registros en la tabla) y las características. Así mismo, se debe evaluar si la información adquirida puede servir para los requerimientos del proyecto.

- **Exploración de los datos:** en esta parte se utilizan técnicas de consulta, visualización y reporte de información. Se incluye el reporte de las distribuciones de parámetros claves, la relaciones entre dos o más variables, agregaciones simples y análisis estadísticos.

- **Verificación de la calidad de los datos**: se examina la calidad de la información, se deben registrar los resultados de las verificaciones que se realizan con respecto a la validación de los datos (Muñoz, 2021).

Preparación de Datos (Data Preparation)

Cubre todas las actividades necesarias para la construcción del conjunto de datos final que se utilizará como insumo para los modelos. Es probable que la preparación de los datos se deba realizar varias veces y sin un orden prescrito. Dentro de esta fase se incluye la selección de las tablas, registros y atributos; así como la transformación y el preprocesamiento de los datos, se siguen los siguientes pasos:

- **Selección de los datos:** En este punto se deben decidir cuáles son los conjuntos de datos que se van a utilizar para la predicción del siniestro. Los criterios que se pueden utilizar son: la relevancia de la información para obtener los objetivos, la calidad, las limitaciones en el volumen y el tipo de dato.
- **Construcción de los datos:** en este punto se debe realizar la producción de atributos derivados de otros, o la transformación de nuevos atributos.
- **Formateo de los datos**: finalmente, con el propósito de evitar contratiempos en la programación, se debe asegurar la correcta asignación del tipo de formato a todos los atributos (Muñoz, 2021).

Los datos utilizados son de información abierta, se realizaron algunos cambios para ajustar los mismos; de acuerdo con, la necesidad del modelo de predicción que se proyecta ejecutar. Se registró la

información que se debe tener en cuenta al incluir y excluir para tener información limpia; se analizan cuáles son las variables dependientes e independientes. Así mismo, las fuentes de información de los siniestros registrados.

Modelado (Modeling)

Una vez se tiene un mayor entendimiento de los datos, se procede con el desarrollo y la selección de los modelos de predicción, se realiza la calibración de los parámetros de los modelos a los valores óptimos. Adicionalmente, se deben plantear las estrategias de validación (incluyendo la división de los datos en los conjuntos de entrenamiento y de prueba) y las métricas de evaluación para medir el desempeño no de los modelos. De acuerdo con, el comportamiento y características del modelo, es posible que se requiera regresar a la fase de preparación datos para un mayor refinamiento antes de proseguir con la evaluación, siguiendo los siguientes pasos:

- **Selección de la técnica de modelado**: Dependiendo del tipo de predicción, método a utilizar y el conjunto de datos definitivo, se debe seleccionar la técnica de modelado que se va a utilizar. Para esta metodología se propone contar con un modelo base de regresión lineal simple.
- **Construcción del modelo**: Se deben utilizar las herramientas definidas en los recursos del proyecto esto con el fin de poder implementar los modelos propuestos (Muñoz, 2021).

Con el análisis descriptivo se busca una caracterización más profunda que la realizada en la fase 1 (entendimiento del negocio), pues aquí se comenzaron a tener datos relevantes como métricas de desempeño de la ejecución de los procesos. Por su parte el análisis predictivo buscaba aprender el comportamiento de los datos y realizar predicciones del tiempo apoyado en técnicas de Machine learning.

Evaluación de los Datos (Data Evaluation)

> Se construyen uno o varios modelos de informacion de alta calidad. Antes de proceder con la última etapa, es importante realizar una evaluación de todos los pasos ejecutados, se realiza una interpretación de los resultados y se verifica que el modelo si alcanza los objetivos planteados. Un punto clave en esta fase es determinar si hay requerimientos que faltan por satisfacer, teniendo en cuenta:
>
> - **Evaluación de resultados:** se debe revisar el grado en el que el modelo se adhiere a los objetivos planteados y determinar si es suficiente o si se deben realizar modificaciones.
> - **Revisar el proceso:** se debe realizar un análisis de todas las actividades ejecutadas y determinar si se han reportado y aplicado todos los detalles expuestos en la presente metodología (Muñoz, 2021).

Los modelos predictivos obtenidos en la etapa anterior fueron evaluados con base en la magnitud del error de las predicciones al hacer una comparación con datos de la ejecución real en una base de pruebas. Se construyen uno o varios modelos de informacion de alta calidad, al hacer la comparación entre varios modelos, se escogió el que tuvo mejor métrica de desempeño que fue un modelo de predicción mediante una regresión lineal simple.

Despliegue (Deployment)

> El objetivo último de esta fase es realizar el despliegue de los resultados obtenidos de forma que sea propagado a los usuarios finales, así como el mantenimiento de este una vez el despliegue haya finalizado. Se debe tener en cuenta: diseñar un plan de despliegue de modelos, realizar seguimiento y mantenimiento (Vallalta, 2023).

Finalmente, la última etapa de la metodología propuesta consistió en hacer el despliegue, es decir, comenzar a utilizar el modelo predictivo construido y seleccionado. Para facilitar su utilización se propuso la creación de herramientas interactivas como los tableros de control que le permiten a los usuarios monitorear en tiempo real lo que está sucediendo en el pago de siniestros diarios en tiempo real. (repositorio.unal.edu.co)

Método por Análisis de Regresiones

> El análisis de regresión es una técnica estadística para estudiar la relación entre variables, el objetivo de la regresión es descubrir la relación funcional entre la entrada y la salida de este sistema, para poder así predecir la salida de este cuando se le presenta un dato de entrada nuevo. En un análisis de regresión simple existe una variable respuesta o dependiente ( ) y una variable explicativa o independiente ( ). El propósito es obtener una función sencilla de la variable explicativa, que sea capaz de describir lo más ajustadamente posible la variación de la variable dependiente (Cabrera, 2015).

Regresión Lineal

Este método consiste en construir a partir de los datos de entrada una función lineal (línea recta) que atraviese los puntos con la mínima variación entre cada observación y el punto predicho por la función.

Como los valores observados de la variable dependiente difieren generalmente de los que predice la función, esta posee un error. La función más eficaz es aquella que describe la variable dependiente con el menor error posible o, dicho en otras palabras, con la menor diferencia entre los valores observados y predichos (Cabrera, 2015).

Regresión Lineal Simple

El modelo de regresión lineal simple sólo está conformado por dos variables estadísticas llamadas X y Y. Considera una única variable independiente o explicativa, X y una variable dependiente o respuesta, Y asumiendo que la relación entre ambas es lineal. Para la regresión lineal simple, se asume que **X** y **Y,** se relacionan mediante la relación funcional (siendo β1 y β0 estimadores). **Fórmula $y = \beta 0 + \beta 1\, x + \varepsilon$** (Cabrera, 2015)**.**

Método Machine Learning

En este proyecto se usan métodos de Machine Learning debido a que puede ser ampliamente definida como métodos computacionales que usan la experiencia para mejorar el desempeño de las predicciones, logrando ser estas más precisas. Cuando nos referimos a experiencias hablamos específicamente de la información histórica recolectada que se utiliza para los procesos de

entrenamiento, en este caso hacemos referencia a valores de siniestros recolectados anteriormente por un periodo de 30 días atrás.

Requerimientos Técnicos y Librerías

Para el desarrollo del modelo planteado en los objetivos, fue necesario tener una data inicial, base de datos abierta para hacer el estudio de patrones y comportamiento de los datos, dicha información esta almacenada diariamente con el fin de poder hacer el análisis teniendo en cuenta un periodo anterior es decir 30 días anteriores. El modelo esta desarrollado en Python porque es eficiente y fácil de aprender; adicionalmente puede ejecutarse en diferentes plataformas. El software Python se puede obtener de descargar gratis online, se integra bien a todos los tipos de sistemas y aumenta la velocidad del desarrollo.

Librerías

Para el desarrollo y ploteo del modelo predictivo, fue necesario el uso de varias librerías que permiten el proceso de ETL (Extracción, Transformación, Carga) y de la misma forma diseñar e implementar el modelo predictivo mostrando sus respectivos resultados y graficas correspondientes:

- **Pandas**. esta librería está diseñada específicamente para la manipulación y el análisis de datos en el lenguaje Python. Es potente, flexible y fácil de usar. Gracias a Pandas, por fin se puede utilizar el lenguaje Python para cargar, alinear, manipular o incluso fusionar datos (DataScientest, 2023).

- **NumPy**. es una librería de Python especializada en el cálculo numérico y el análisis de datos, especialmente para un gran volumen de datos. Incorpora una nueva clase de objetos llamados arrays que permite representar colecciones de datos de un mismo tipo en varias dimensiones, y funciones muy eficientes para su manipulación (Aprendeconalf, 2022).

- **Datetime.** Para manejar fechas en Python se suele utilizar la librería datetime que incorpora los tipos de datos date, time y datetime para representar fechas y funciones para manejarlas (Aprendeconalf, 2020).

- **Bokeh.** Con unas pocas líneas de código Python, Bokeh crea gráficos alimentados por Javascript que proporcionan una interactividad y un dinamismo sin precedentes. El proceso de ejecución del código de Bokeh tiene dos vertientes: en primer lugar, elegimos un bloque de entre los bloques preconstruidos de Bokeh para crear nuestra visualización y, a continuación, se añade código adicional para personalizar estos bloques de forma y que se ajusten al máximo a nuestras necesidades (DataScientest, 2023).

- **Sklearn.** es uno de los open-source y bibliotecas de aprendizaje automático más populares en Python. La biblioteca scikit-learn contiene muchas herramientas eficientes para aprendizaje automático y modelado estadístico, incluyendo clasificación, regresión, agrupación, y reducción de dimensionalidad (DataScientest, 2023).

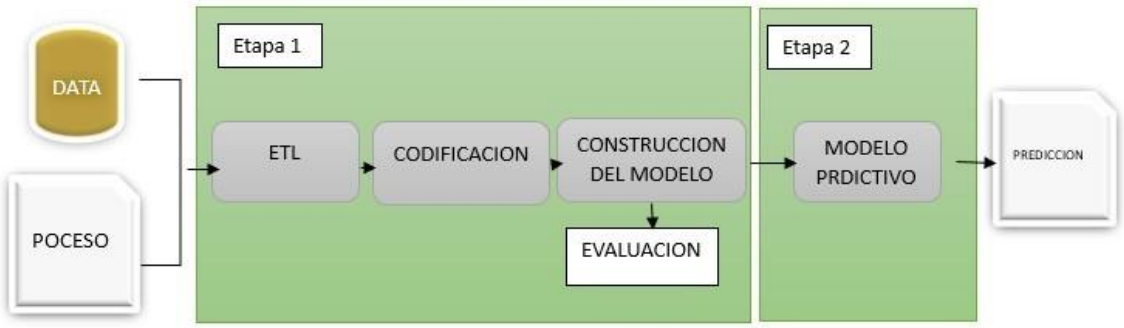
- **Webbrowser.** El módulo webbrowser proporciona una interfaz de alto nivel que permite mostrar documentos basados en web a los usuarios. En la mayoría de circunstancias, simplemente llamar a la función open () desde este módulo hará lo correcto (Python, 2023).

Resultados

Para el desarrollo del proyecto se simularon datos con una base abierta de 30 días de anterioridad de información, con el fin de poder hacer uso del método de regresión lineal simple. Con base en los registros de eventos obtenidos de la simulación, se realizó un análisis descriptivo y un análisis predictivo por medio de herramientas de la minería de procesos como lo es python y sus librerías implicadas que me permiten hacer machine learning y visualización de graficas realizadas. Aunque es importante tener en cuenta que el comportamiento de los siniestros realizados diariamente se ve afectado por la cantidad ingresada en comparación a los 30 días anteriores, el análisis descriptivo permitió principalmente caracterizar el comportamiento del proceso de siniestralidad y los otros componentes que lo puedan afectar.

Figura 12

Diagrama Proceso Modelo Predictivo



Nota: El diagrama que identifica el proceso utilizado para poder realizar el desarrollo del modelo de predicción, desde la extracción de la data hasta la entrega del modelo.

Técnica Utilizada

Se utilizo el método de regresión lineal simple donde se realiza un algoritmo para tomar el ultimo siniestro obtenido en la data y se divide por el último día del siniestro cargado obteniendo de esta forma la variable (**m)**, se multiplico dicha variable por 30 debido a que es la cantidad de días que se van a tener en cuenta para la predicción de cierre de mes y se obtiene la variable **(y)** cuya variable es la dependiente y la independiente corresponde a los días del mes que corresponden a la variable (**x)** o variable independiente.

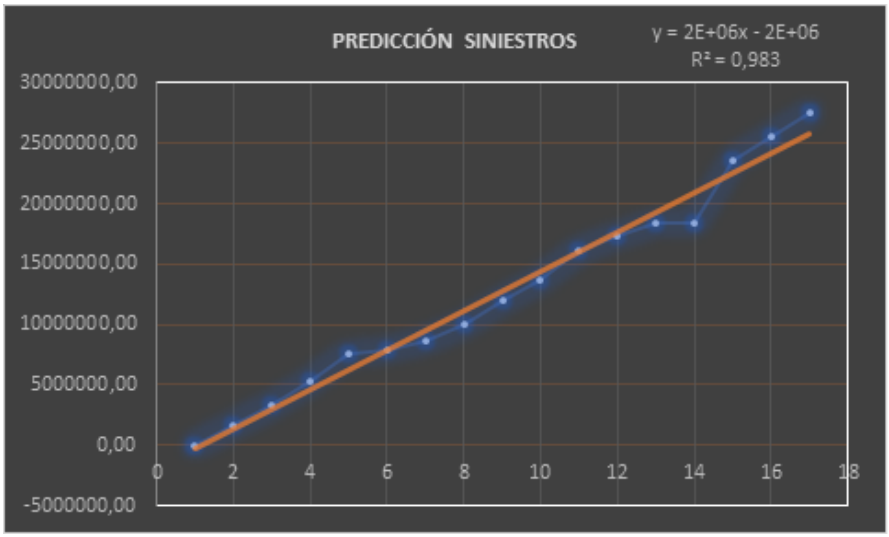
El modelo fue realizado en Python donde se realizó el proceso y se cumplieron las fases que nos brinda el método (CRISP-DM), haciendo la extracción de los datos y la transformación de los mismos para generar nuevas variables que fueron de utilidad para poder realizar el algoritmo y entrenar el modelo, una vez realizado el modelo se realizaron las respectivas pruebas mediante el coeficiente de determinación o más conocido cono R2 para validar que porcentaje de confiabilidad tenía el modelo

desarrollado, esta prueba se realizó utilizando la librería (sklearn.metrics r2_score) en donde se obtuvo un porcentaje del 0.97.

Se realizaron otras pruebas adicionales recreando el mismo modelo de predicción, pero esta vez fue realizado en Excel para validar que el diagrama de dispersión con la línea de tendencia fuera igual o similar a la generada por el modelo realizado en Python, y se evidencia que las gráficas son similares por tal motivo el modelo si cumple y puede ser aplicado a la solución del problema.

Figura 13

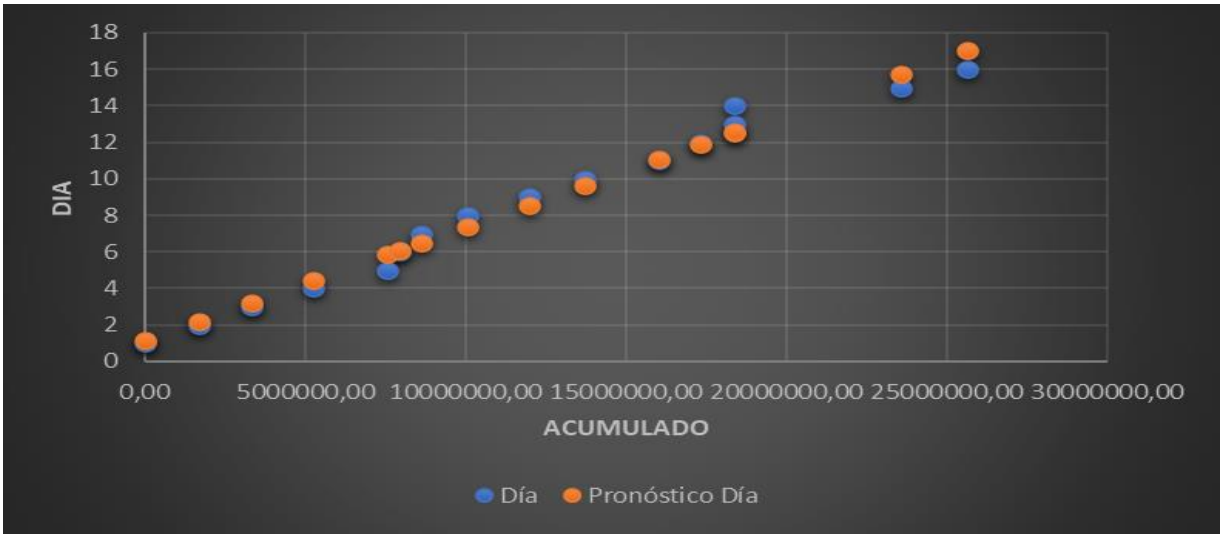
Grafica de Regresión Lineal



Nota: Grafica de dispersión con regresión lineal donde se logra evidenciar la tendencia y el modelo realizado. Acumulado de curva regresión ajustada [FOTOGRAFIA].

Figura 14

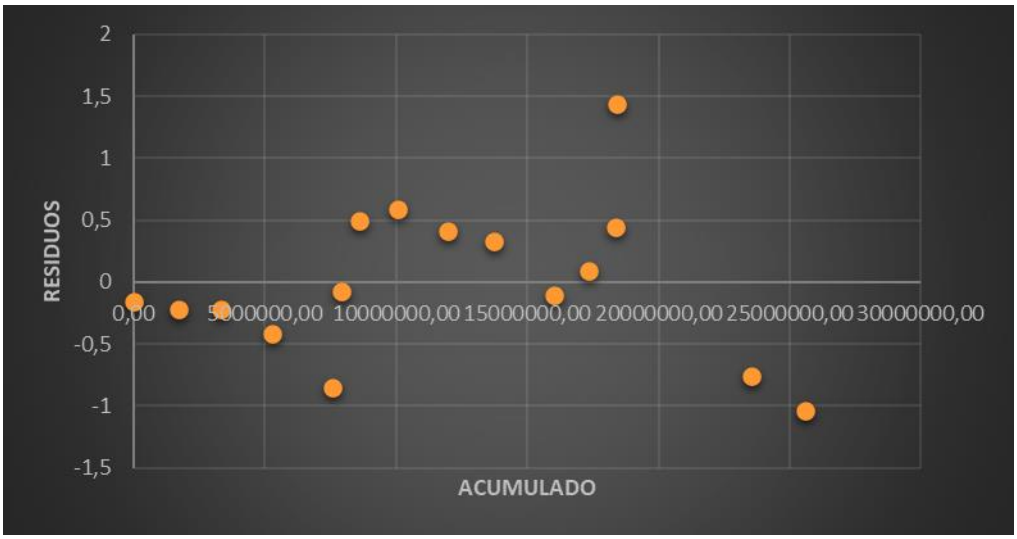
Gráfica acumulada curva de regresión ajustada



Nota. se evidencia la relación entre una variable predictora (X) y una variable de respuesta (Y), así como se puede evaluar si la regresión lineal se ajusta a los datos.

Figura 15

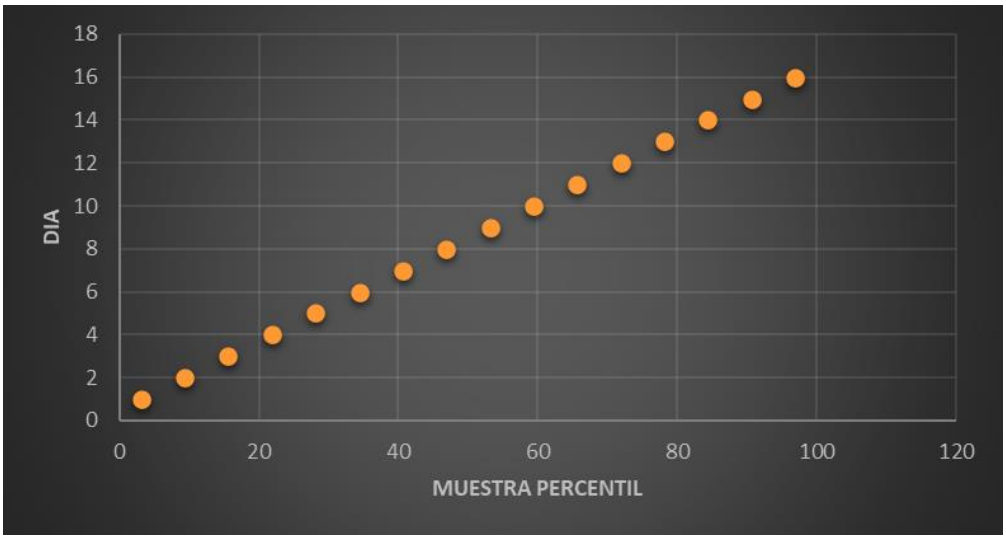
Gráfico acumulado de residuales



Nota: gráfica que se utiliza para examinar la bondad de ajuste en regresión y ANOVA. Con el fin de examinar las gráficas de residuos ayuda a determinar si se cumplen los supuestos de los mínimos cuadrados ordinarios.

Figura 16

Gráfica de Probabilidad Normal



Nota: se usa para ayudar a juzgar si una muestra de datos numéricos proviene o no de una distribución normal.

Conclusiones y Recomendaciones

Conclusiones

La evaluación de los datos generados durante el proceso de eficacia de siniestros dentales se realiza a través del diagnóstico de datos, donde se identifica la alta y baja significancia, proporcionando un punto de partida para el desarrollo de un modelo predictivo.

La capacidad de procesar y administrar datos masivos ha aumentado en los últimos años, y el análisis predictivo se ha convertido en un método viable para que las organizaciones descubran patrones que pueden usarse para formular reglas, rastrear el comportamiento de una o más variables. El modelo de regresión lineal que se utilizó en este proyecto es una base de datos de días anteriores con el fin de predecir el costo del siniestro odontológico.

El alcance del proyecto se limitó al desarrollo de un modelo predictivo para el cliente AXA Colpatria Falabella, a través de la implementación de la metodología desarrollada CRISP-DM; se puede entrenar para predecir otros valores finales durante un periodo determinad; considerando que la suma de los valores debe ingresarse en una fecha determinada. De esta forma los responsables de implementar el proceso pueden contar con tablas y gráficos como apoyo para toma de decisiones sobre los gastos de reclamos registrados hasta el momento para satisfacer las necesidades y requerimientos que existen en un momento determinado.

El modelo de regresión lineal desarrollado para la compañía de seguros dentales

Axa Colpatria Falabella, mostró una alta confiabilidad al superar el 95% de precisión de las predicciones en comparación con el valor real. Confirmando que es un modelo de alta efectividad para relaciones lineales entre las variables involucradas. Sin embargo, en algún momento pueden presentarse situaciones, factores o variables externas que pueden influir en los resultados del modelo generando, datos para futuras investigaciones y nuevos desarrollos en el modelo.

Recomendaciones

A continuación, se presentan como recomendaciones algunos aspectos que ayudan aclarar puntos claves para poder realizar proyectos similares a futuro:

- Entender claramente los requerimientos y objetivos del proyecto desde la visión del negocio, definiendo el plan del proyecto y estimación inicial de herramientas y técnicas.

- Analizar datos importantes en el proceso de selección, depuración y transformación con el fin de poder definir una data con información que se pueda utilizar en la construcción de otros modelos que sean realmente necesarios para la construcción de modelos.

- Utilizar técnicas adecuadas para el diseño y desarrollo del modelo con el fin de hacer un entrenamiento mediante machine learning de una forma eficiente y poder hallar el modelo que más se acerque a la solución del problema.

- Evaluar los modelos realizados y tomar una decisión adecuada si realmente son útiles en el proceso o toca replantear el diseño de este para poder cumplir con el objetivo.

- Con este modelo predictivo se observa que la data funciona para otros análisis diferentes a los planteados inicialmente, la predicción se puede aplicar a proyecciones con otros objetivos a partir de los datos que conforman la data, por ejemplo: análisis predictivo por día de la semana durante el mes de la predicción general.

Bibliografía

ALMARIO, M. A. (30 de Enero de 2023). *Sura Vida, Alfa Vida y Bolívar lideran top de aseguradoras que más emitieron primas.* Obtenido de https://www.larepublica.co/finanzas/sura-vida-alfa-vida-y-bolivar-lideran-top-de-aseguradoras-que-mas-emitieron-primas-3532325

Aprendeconalf. (15 de Septiembre de 2020). *La Librería Datetime.* Obtenido de ALSO ON APRENDE CON ALF: https://aprendeconalf.es/docencia/python/manual/datetime/

Aprendeconalf. (12 de Mayo de 2022). *La Librería Numpy*. Obtenido de Also On Aprende Con Alf: https://aprendeconalf.es/docencia/python/manual/numpy/

Bravo Reyes, J. H. (01 de 12 de 2011). Una mirada Histórica sobre los Seguros y sus inicios en Colombia. *Gestión & Sociedad Ciencia Unisalle, 4*(2), página 150-152. Obtenido de https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1169&context=gs#:~:text=En%201909%20se%20crea%20la,seguros%20para%20t%C3%ADtulos%20de%20propiedad.

Cabrera, F. P. (2015). *Desarrollo de un modelo basado en Machine Learning para la predicción de la demanda de habitaciones y ocupación en el sector hotelero.* Obtenido de Repositorio UTB: https://repositorio.utb.edu.co/handle/20.500.12585/1020

DataScientest. (1 de Septiembre de 2023). *Scikit-Learn : Descubre la biblioteca de Python dedicada al Machine Learning*. Obtenido de DataScientest: https://datascientest.com/es/scikit-learn-decubre-la-biblioteca-python

Espino, C., & Martínez, X. (17 de Enero de 2017). *Análisis Predictivo.* Recuperado el Abril de 2023, de Universitat Oberta de Catalunya: https://openaccess.uoc.edu/bitstream/10609/59565/6/caresptimTFG0117mem%C3%B2ria.pdf

Etecé, Editorial. (05 de 08 de 2021). *Enciclopedia Concepto*. Obtenido de Editorial Etecé: https://concepto.de/dato-en-informatica/

Fadul Vergara, W. (2021). Historia de los Seguros en Colombia. (182), 68. Recuperado el 8 de Abril de 2023, de Revista Fasecolda: https://revista.fasecolda.com/index.php/revfasecolda/article/view/725/692

Fadul Vergara, W. (2021). Historia de los Seguros en Colombia. *Revista Fasecolda*(182), página 68. Obtenido de FASECOLDA: https://revista.fasecolda.com/index.php/revfasecolda/article/view/725/692

Fundación MAPFRE. (2023). *¿Qué es el seguro?* Obtenido de Fundación MAPFRE: https://segurosypensionesparatodos.fundacionmapfre.org/seguros/definicion-seguro-asegurar/

Heredia, R. (12 de Junio de 2015). *El ranking de las mejores soluciones de análisis predictivo para empresas*. Recuperado el Abril de 2023, de E Banking News: https://www.ebankingnews.com/noticias/el-ranking-de-las-mejores-soluciones-de-analisis-predictivo-para-empresas-0032505

IBERDROLA. (2023). *ANALISIS PREDICTIVO*. Recuperado el 2023, de IBERDROLA S.A.: https://www.iberdrola.com/innovacion/analisis-predictivo

IBERDROLA. (2023). *Análisis predictivo*. Recuperado el Abril de 2023, de IBERDROLA S.A.: https://www.iberdrola.com/innovacion/analisis-predictivo

IBERDROLA. (2023). *Análisis Predictivo*. Recuperado el Abril de 2023, de IBERDROLA S.A.: https://www.iberdrola.com/innovacion/analisis-predictivo

IBM. (17 de 8 de 2021). *Conceptos básicos de ayuda de CRISP-DM*. Obtenido de https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

IBM. (17 de 08 de 2021). *Conceptos básicos de ayuda de CRISP-DM*. Recuperado el Abril de 2023, de CRISP-DM: https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview

Lopez, J. M. (2006). *TÉCNICAS DE ANÁLISIS DE DATOS*. Recuperado el Abril de 2023, de Universidad Carlos III de Madrid: https://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf

Manatal. (2023). *Seguro Gental*. Obtenido de Manatal: https://www.manatal.com/es/glossary/dental-insurance

Márquez, A. C. (11 de 2018). Monitoreo predictivo de procesos comerciales: una encuesta. *11*(6), 962-977. doi:10.1109/TSC.2017.2772256

Marquez-Chamorro. (1 de Diciembre de 2018). *Predictive Monitoring of Business Processes: A Survey*. Obtenido de https://ieeexplore.ieee.org/document/8103817

Marytere, N. (2023). *Tipos de Investigación Cuantitativa*. Recuperado el Abril de 2023, de Questionpro: https://www.questionpro.com/blog/es/tipos-de-investigacion-cuantitativa/

Mayorga, W. (2013). Los seguros de salud según la Encuesta de Calidad de Vida. *Revista FASECOLDA*(151), 65–70. Obtenido de https://revista.fasecolda.com/index.php/revfasecolda/article/view/18

MERCO. (2022). *Ranking de Empresas*. Obtenido de MERCO: https://www.merco.info/co/ranking-merco-empresas

Muguira, A. (s.f.). *QuestionPro*. Obtenido de https://www.questionpro.com/blog/es/observacion-cuantitativa/#:~:text=La%20observaci%C3%B3n%20cuantitativa%20es%20una%20colecci%C3%B3n%20objetiva%20de,obtienen%20utilizando%20m%C3%A9todos%20de%20an%C3%A1lisis%20estad%C3%ADsticos%20y%20num%C3%A9ricos.

Muñoz, J. V. (2021). *Evaluación de Modelos de MachineLearning para la Predicción deCrímenes en la Ciudad de Medellín.* Recuperado el Abril de 2023, de https://docplayer.es/227413314-Evaluacion-de-modelos-de-machine-learning-para-la-prediccion-de-crimenes-en-la-ciudad-de-medellin.html

Narvaez, M. (2023). *Tipos de investigación cuantitativa*. Recuperado el 2023, de Question Pro: https://www.questionpro.com/blog/es/tipos-de-investigacion-cuantitativa/

Otálvaro Castro, G. J. (6 de Abril de 2019). Políticas de salud bucal en Colombia. Tendencias y puntos críticos para la garantía del derecho a la salud. *Universitas Odontologica, 38*(80). doi:https://doi.org/10.11144/Javeriana.uo38-80.psbc

Python. (2023). Obtenido de Python Software Foundation: https://www.python.org/doc/essays/blurb/

Python. (2023). *Controlador de navegador web conveniente.* Obtenido de Python Software Foundation: https://docs.python.org/es/3/library/webbrowser.html

QuePlan.Cl. (2023). *Seguros de Salud*. Recuperado el Abril de 2023, de QuePlan.Cl: https://www.segurosparaestarseguro.com/seguro-medico-de-reembolso/

repositorio.unal.edu.co. (s.f.). Obtenido de https://repositorio.unal.edu.co/bitstream/handle/unal/80976/1037649240.2021.pdf?sequence=1&isAllowed=y

repositorio.utb.edu.co. (s.f.). Obtenido de https://repositorio.utb.edu.co/bitstream/handle/20.500.12585/1020/0068209.pdf?sequence=1

Siegel, E. (2013). Analítica Predictiva. En E. Siegel, *ANALÍTICA PREDICTIVA: PREDECIR EL FUTURO UTILIZANDO BIG DATA* (págs. p26-53). Anaya Multimedia.

STUDOCU. (2022). *PE1 Módulo 1 Introducción a Python*. Recuperado el 2024, de STUDOCU: https://www.studocu.com/co/document/universidad-de-pamplona/programacion/pe1-modulo-1-introduccion-a-python-y-a-la-programacion-de-computadora-external/40273580

SYDLE. (01 de 22 de 2023). *Big Data: definición, importancia y tipos*. Obtenido de SYDLE: https://www.sydle.com/es/blog/big-data-definicion-importancia-y-tipos-614b791388e600016afa7fc3/

Torresa, N. L. (2017). *Documento Técnico Comercialización de seguros*. Obtenido de URF: https://www.urf.gov.co/webcenter/ShowProperty?nodeId=%2FConexionContent%2FWCC_CLUSTER-106153%2F%2FidcPrimaryFile&revision=latestreleased

Vaibhav, K. G. (2018). *Analisis Predictivo: una revisión de tendencias y técnicas.* Obtenido de http://181.39.139.68:8080/bitstream/handle/123456789/1832/MODELO~4.PDF?sequence=1&isAllowed=y

Vallalta, R. J. (2023). *CRISP-DM: Una Metodología para Minería de Datos en Salud*. Recuperado el Abril de 2023, de IA Health Data Miner: https://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/

Viva Seguro Fasecolda. (2018). *¿Qué son los seguros?* Obtenido de Viva Seguro Fasecolda: https://vivasegurofasecolda.com/seguros/abc-del-seguro/