



---

# MODELO DE RESERVAS COMERCIALES, POR MEDIO DE CHAIN LADER.

---

Por: Abraham Guerrero.



DECEMBER 3, 2023

## Introducción:

En el ámbito de las compañías de seguros, hay un sector que se especializa en lidiar con la incertidumbre inherente a eventos aleatorios imprevistos, comúnmente llamados siniestros. Estas empresas ofrecen productos de seguro para respaldar financieramente a individuos, familias o entidades en caso de que ocurra algún evento inesperado, como terremotos u otras situaciones impredecibles. Los seguros involucran pagos regulares del cliente a la aseguradora a cambio de cobertura financiera en caso de que se materialice un siniestro.

Un aspecto crucial en esta industria es la evaluación de probabilidades de riesgo al vender productos de seguro. Las aseguradoras, al considerar las características de sus clientes, pueden ofrecer servicios más específicos y adaptados a las necesidades individuales. Posteriormente, el cálculo de reservas, esencial para cubrir posibles siniestros, se realiza anualmente mediante el método Chain-Ladder, basado en patrones históricos y la suposición de que las tendencias pasadas continuarán en el futuro. Este trabajo se centra en proponer un nuevo método de cálculo de reservas, integrando históricos de la empresa y nuevos modelos estadísticos o de aprendizaje automático. En particular, se aplica a una aseguradora que se especializa en seguros de responsabilidad comercial de autos, destinados a empresas que utilizan vehículos en sus operaciones y cubren daños a terceros durante estas actividades comerciales.

## Comprensión del Negocio:

### *objetivos empresariales:*

Con la creciente adopción de métodos estadísticos y de aprendizaje automático para la estimación y predicción, las compañías de seguros, en particular las especializadas en seguros comerciales de automóviles, buscan aprovechar estas técnicas para mejorar la estimación de reservas. Este enfoque apunta a minimizar errores, garantizar eficiencia de recursos, aumentar ganancias, construir confianza del cliente, mantener estabilidad económica, cumplir con regulaciones y atraer inversionistas.

### Evaluación de la situación y riesgos del proyecto:

La aseguradora cuenta con un historial completo de datos sobre siniestros y seguros de autos comerciales. Cuenta con profesionales calificados para llevar a cabo el proyecto de minería de datos, que busca mejorar la estimación de la reserva. Los mayores riesgos incluyen la implementación de un método menos eficaz que el tradicional y la posibilidad de que la estimación de la reserva no mejore, lo que podría afectar la solvencia económica. El plan de contingencia implica descartar el modelo de minería de datos si los objetivos no se cumplen.

## Definición de Objetivos en Minería de Datos:

1. Establecimiento de Metas para la Minería de Datos Aprovechando datos históricos de reservas de la aseguradora, se desarrollará un modelo estadístico o de machine learning para llevar a cabo el cálculo y la estimación de triángulos destinados al cálculo de reservas anuales. Las predicciones generadas se someterán a una validación comparándolas con las reservas reales de cada año.
2. Parámetros de Éxito en Minería de Datos El objetivo es obtener un modelo que supere en rendimiento al convencional método Chain Ladder, exhibiendo excelentes resultados en diversas métricas de evaluación. Dichas métricas incluyen el Error Cuadrático Medio, el Error Absoluto Medio, el Coeficiente de Determinación y el Error Porcentual Absoluto Medio. Estas medidas son pertinentes ya que se están pronosticando reservas, es decir, valores continuos.

- **Planificación del Proyecto:**

Tarea	Tiempo estimado	Recursos necesarios
Entendimiento del problema	1 semana	Acceso a libros, papers y datos de la compañía
Obtención de los datos granulares por cada uno de los siniestros de la compañía	1 semana	Determinación de los distintos productos de la compañía
Centralización de los datos y agrupación de los mismos por categoría.	1 semana	Se requiere un entendimiento para la construcción de matrices de desarrollo
Construcción de las matrices de desarrollo con pagos incrementales y acumulados.	1/2 semana	Automatización con python
Aplicación de Chain - Ladder y GML para predecir los impactos de las reclamaciones.	1/2 semana	Mecanismos de validación

## Descripción de los datos:

En relación con la minería de datos, se dispone de una extensa base de datos con registros provenientes de diversas aseguradoras, enfocados en el sector de seguros de autos comerciales. En total, se cuentan con 15,800 registros y 13 variables, aunque solo 6 columnas o atributos resultan verdaderamente relevantes para el propósito del modelo.

Las variables abarcan distintos aspectos, incluyendo información temporal que refleja años específicos de eventos relevantes. Entre las variables clave, se encuentra el código de cada aseguradora, mientras que el resto se compone principalmente de valores numéricos que representan pérdidas, reservas y primas, a excepción de una variable binaria.

En términos de codificación, el único atributo que podría presentar solapamientos es el código de cada aseguradora, aunque, en la práctica, esto rara vez sucede en la empresa. El conjunto de datos incorpora diversas variables, tales como GRCODE y GRNAME, que representan el código y nombre de la compañía NAIC, respectivamente. Además, variables como AccidentYear y DevelopmentYear indican el año de ocurrencia y desarrollo de reclamaciones, con un rango que va desde 1988 hasta 1997.

Otras variables importantes incluyen IncurLoss\_, CumPaidLoss\_, BulkLoss\_, PostedReserve97\_, EarnedPremDIR\_, EarnedPremCeded\_, EarnedPremNet\_, y Single, cada una ofreciendo detalles cruciales sobre pérdidas, reservas, primas y características de la aseguradora. Es relevante destacar que el conjunto de datos se compone de información tanto de aseguradoras individuales como de grupos, y un indicador binario llamado Single se utiliza para distinguir entre entidades únicas y aseguradoras pertenecientes a un grupo.

## Modelado:

La selección de técnicas de modelado es un paso crucial en el proceso. Para empezar, el campo de mayor interés, *IncurLoss\_C*, debe ser transformado a un formato numérico para cumplir con los requisitos del análisis. El objetivo puntual del modelo es estimar la reserva o la parte inferior de un triángulo para el cálculo de reservas, lo cual establece claramente la dirección y propósito del modelado.

En términos de requisitos de tamaño, aunque los modelos no exigen un tamaño específico, se reconoce que contar con más datos puede mejorar la efectividad de los modelos de regresión lineal al permitir la identificación más precisa de patrones. Aunque no es imprescindible que los modelos ofrezcan resultados fácilmente presentables, los enfoques planteados aseguran que esta característica esté presente.

La elección de las técnicas de modelado adecuadas es otra consideración esencial. En este caso, la implementación de los modelos requerirá tres conjuntos de datos: entrenamiento, validación y prueba. Es importante destacar que el método de Chain-Ladder, al ser un modelo determinístico, no necesita este tipo de división. Aunque el conjunto actual de datos es suficiente para obtener resultados confiables, la posibilidad de obtener más datos en el futuro se percibe como una mejora potencial.

En cuanto a la calidad de los datos, se confirma que los modelos requieren un cierto nivel de calidad, y este requisito se ha logrado en la etapa de preparación de los datos. Además, los datos disponibles son considerados del tipo adecuado para el modelo, eliminando la necesidad de ajustes adicionales. Este análisis proactivo asegura que el proceso de modelado se base en fundamentos sólidos y datos fiables.

## Supuestos del Modelo:

Los supuestos del modelo desempeñan un papel crucial en la interpretación y confiabilidad de los resultados. En el caso del método de Chain-Ladder, se destaca su carencia de supuestos matemáticos rigurosos, basándose simplemente en la disponibilidad de información sobre el año del siniestro, años de reclamación y el monto de pérdidas (Incor\_loss\_C).

En contraste, los modelos de regresión lineal normal imponen varios supuestos esenciales para su validez:

- **Linealidad:** Se presume que la relación entre variables independientes y la variable dependiente es lineal, reflejando cambios proporcionales.
- **Independencia de errores:** Los errores deben ser independientes, sin relación entre las predicciones de diferentes observaciones.
- **No multicolinealidad:** Se asume la ausencia de multicolinealidad perfecta entre las variables independientes.

Los modelos Ridge y Lasso, además de cumplir con los supuestos de regresión lineal normal, introducen particularidades:

- **Regularización:** Ridge y Lasso incorporan términos de regularización (L2 y L1, respectivamente) para prevenir el sobreajuste y reducir la varianza del modelo.
- **Selección de características (Lasso):** Lasso puede realizar la selección automática de características al forzar algunos coeficientes a cero, eliminando variables menos relevantes.

## Criterios de Bondad para los Modelos:

La evaluación de la bondad de ajuste se realiza mediante el Mean Absolute Percentage Error (MAPE), aplicado al conjunto de prueba. Este indicador, calculado como la suma de las diferencias absolutas entre las predicciones y los valores reales dividida por el número de observaciones, determinará el rendimiento de cada modelo. La elección del mejor modelo recae en aquel que obtenga el MAPE más pequeño.

Aunque existen otros métodos de bondad de ajuste basados en la determinación de la distribución de probabilidad de los errores del modelo, este trabajo se enfocará principalmente en el MAPE debido a la naturaleza determinística del método Chain-Ladder, que excluye la aplicación de estos métodos de verificación.

## **Diseño de Prueba y Evaluación de Modelos:**

En el diseño de prueba, se ha optado por la implementación del método Cross-Validation para evaluar y comparar modelos de regresión. Este proceso se lleva a cabo de la siguiente manera:

### **1. Selección de Conjuntos:**

- Se eligen tres conjuntos fundamentales: entrenamiento, validación y prueba. Una aseguradora se designa como el conjunto de prueba inicial. Una segunda aseguradora se asigna al conjunto de validación, y las restantes conforman el conjunto de entrenamiento inicial.
- Se entrena un modelo con estos conjuntos iniciales.

### **2. Evaluación y Selección de Modelos:**

- Los modelos se evalúan utilizando el conjunto de validación, aplicando la métrica MAPE para determinar el rendimiento.
- El mejor modelo se selecciona basándose en el menor MAPE en el conjunto de validación.
- El modelo elegido se aplica al conjunto de prueba, y se calcula el MAPE para evaluar su rendimiento en datos no vistos.

### **3. Iteración de Conjuntos:**

- Se itera el proceso asignando una nueva aseguradora al conjunto de prueba, otra al conjunto de validación y las restantes al conjunto de entrenamiento.
- Se repiten los pasos 1 y 2 hasta que todas las aseguradoras hayan sido parte de los tres conjuntos en algún momento.

### **4. Comparación con Chain-Ladder:**

- Se selecciona el modelo que ha tenido el menor MAPE en los diferentes conjuntos de prueba a lo largo de todas las iteraciones.
- Se compara su rendimiento con el método tradicional Chain-Ladder
- utilizando la métrica MAPE.



## Construcción del modelo:

**Configuración de Parámetros:** En el proceso de construcción de modelos, se ha optado por utilizar un parámetro de regularización de 0.001 para los modelos Ridge y Lasso. Esta elección se basa en análisis previos que indican que este valor es el óptimo para lograr un rendimiento eficiente.

**Descripción del Modelo:** En cuanto a las conclusiones significativas del modelo final de Lasso, se observa que este ha demostrado un comportamiento superior al método tradicional, aunque es crucial considerar que esta mejora puede estar relacionada con la cantidad limitada de aseguradoras en la base de datos. Experiencias anteriores han revelado que al aumentar la cantidad de aseguradoras en el entrenamiento, los modelos pierden la capacidad de caracterizar de manera individual el comportamiento de cada aseguradora.

Respecto a los problemas de ejecución, la clase que realiza los cálculos del Chain-Ladder es la parte más demorada. Dado que los algoritmos de los modelos de regresión dependen de esta clase y el Cross-Validation a su vez depende de los algoritmos de los modelos de regresión, se evidencia la razón por la cual el Cross-Validation presenta demoras significativas en su ejecución.

**Evaluación del Modelo:** Considerando la implementabilidad de los resultados, se destaca que los resultados del modelo son fácilmente implementables en el contexto empresarial.

Al analizar el impacto de los resultados en los criterios de éxito establecidos durante la fase de entendimiento empresarial, se concluye que el modelo cumple con los objetivos empresariales al proporcionar una forma mejorada de estimar la reserva de las compañías de seguros de automóviles comerciales.

Se destaca que la opinión de otros analistas de la compañía respecto a los modelos construidos es favorable, sugiriendo que la aplicación de estos modelos ha sido bien recibida y apreciada dentro del equipo.

## Conclusión:

En el transcurso de este proyecto, siguiendo la metodología CRISP-DM, hemos logrado identificar y abordar eficazmente los desafíos asociados a la estimación de reservas para pólizas de autos comerciales. Tras un exhaustivo proceso de evaluación comercial y de minería de datos, hemos determinado que el modelo de **regresión Ridge** ha demostrado ser la opción más eficiente. Con una métrica MAPE del 1.399% y coeficientes óptimos, este modelo superó tanto a otras regresiones como al método tradicional de Chain-Ladder.

Mirando hacia el futuro, nuestros próximos pasos incluirán la implementación del modelo final en un entorno operativo real. Este proceso nos permitirá observar su comportamiento en situaciones del mundo real y, al mismo tiempo, realizar evaluaciones continuas para asegurar su eficacia a lo largo del tiempo. La meta última es que este modelo se establezca como una herramienta exitosa para la estimación precisa de reservas en el ámbito de las pólizas de autos comerciales. A medida que avanzamos, estaremos atentos a oportunidades de refinamiento y mejora, asegurando así la continua eficiencia y relevancia de nuestro enfoque en el campo de la estimación de reservas para la industria de seguros.