

Oscar Pulido

## Aplicaciones del aprendizaje de máquinas en Actuaría y Finanzas

### Primera versión Fase de Comprensión del negocio para el problema de estimación de provisiones

El objetivo de este documento es presentar la etapa de entendimiento del negocio contemplada dentro de la metodología CRISP-DM para resolver el problema de estimación de provisiones o reservas de una compañía aseguradora. Particularmente, se desea ahondar en el uso de la metodología Chain-Ladder para la constitución de la reserva de siniestros IBNR (Incurred But Not Reported), que compone la Reserva de Siniestros Pendientes, de acuerdo al capítulo IV del decreto 2973 de 2013 de la legislación colombiana con respecto a la actualización del régimen de reservas para las compañías aseguradoras. (Estatuto Orgánico del sistema financiero, Decreto 2973 de 2013, diciembre 20 de 2013).

### Determinación de objetivos de negocio

#### Antecedentes empresariales

#### Tarea 1: Determinación de la estructura organizacional

En primer lugar, se mostrará la estructura organizacional de la compañía aseguradora:

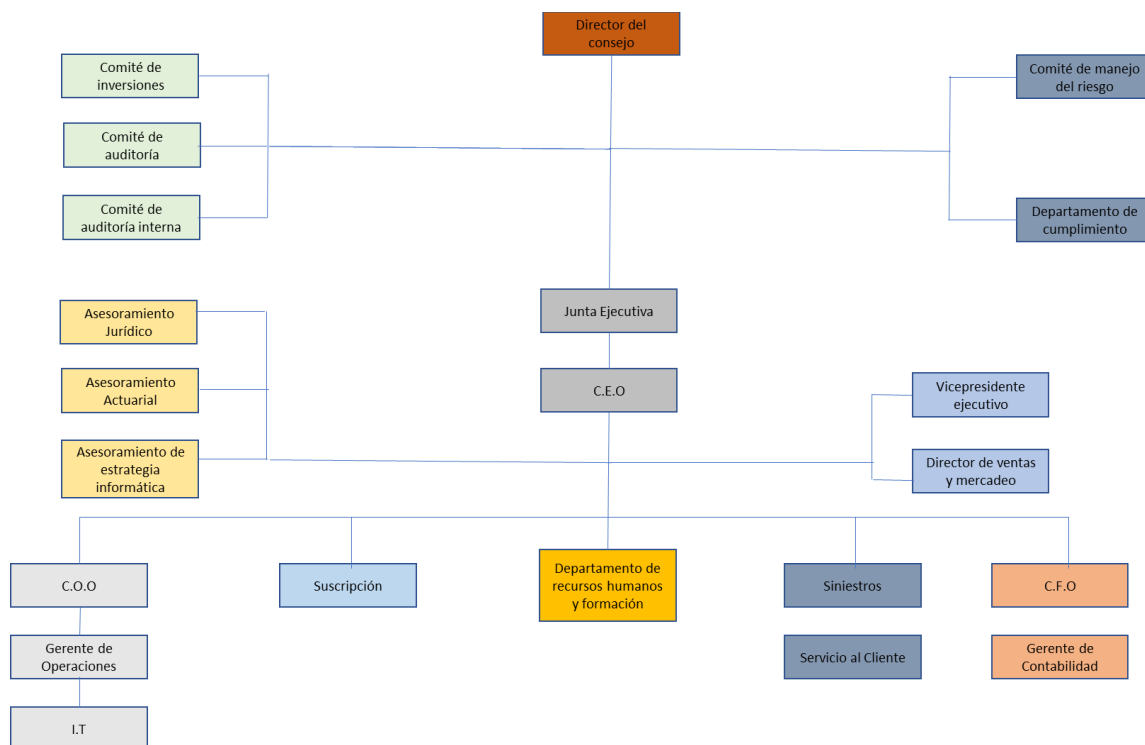


Diagrama realizado con base en <https://www.pacificcrosshealth.com/en/company-information/organization-structure/>

Para los propósitos de este proyecto, se determina que las áreas mas importantes son: Asesoramiento actuarial, Asesoramiento de estrategia informática, I.T, Siniestros, Contabilidad y el Comité de Manejo del Riesgo, además de los directivos. De manera que los miembros claves del proyecto se encuentran en estas áreas. Los stakeholders del proyecto no se muestran en el gráfico, pero el grupo está conformado por aquellos accionistas que financiarán el proyecto. Las áreas que estarán afectadas en mayor medida por este proyecto son la de siniestros, actuaría, contabilidad y I.T.

## Tarea 2: Determinación del área de problema

Una vez identificada la estructura organizacional de la compañía se pasa determinar el área del problema. El área de problema en este proyecto es la de siniestros, que en esta compañía aseguradora es la encargada de constituir la Reserva de Siniestros Pendientes, y, por lo tanto, la reserva IBNR. Si bien existen más áreas involucradas, el área que verá mayores transformaciones es esta.

El problema que enfrenta actualmente el área de Siniestros de la compañía aseguradora (y las compañías en general) es que los métodos actualmente utilizados para estimar el IBNR parecen ser insuficientes para determinar la reserva que realmente se debe constituir y así poder cumplir con los requisitos del regulador, de los accionistas y de las áreas que requieren de esta información para cumplir con sus objetivos de negocio. Actualmente, la compañía no contempla metodologías diferentes a las tradicionales, pues se ha venido usando lo mismo desde que se tiene conocimiento y el regulador lo aprueba así, sin embargo, valdría la pena pensar en metodologías que involucren minería de datos para abordar esta problemática.

## Tarea 3: Determinación de la solución actual al problema

Actualmente, se utiliza la metodología de triángulos run-off, cuyo objetivo es estimar de manera predictiva los montos de dinero pagados por la compañía aseguradora en el futuro, gráficamente, se espera completar un triángulo de información conocida hasta obtener un rectángulo con montos estimados de dinero en pago de siniestros, esto se ve de la siguiente manera:

Accident year	Development year								
	$0$	$1$	$\dots$	$k$	$\dots$	$n-i$	$\dots$	$n-1$	$n$
$0$	$Y_{0,0}$	$Y_{0,1}$	$\dots$	$Y_{0,k}$	$\dots$	$Y_{0,n-i}$	$\dots$	$Y_{0,n-1}$	$Y_{0,n}$
$1$	$Y_{1,0}$	$Y_{1,1}$	$\dots$	$Y_{1,k}$	$\dots$	$Y_{1,n-i}$	$\dots$	$Y_{1,n-1}$	$Y_{1,n}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$			
$i$	$Y_{i,0}$	$Y_{i,1}$	$\dots$	$Y_{i,k}$	$\dots$	$Y_{i,n-i}$	$\dots$	$Y_{i,n-1}$	$Y_{i,n}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$					
$n-k$	$Y_{n-k,0}$	$Y_{n-k,1}$	$\dots$	$Y_{n-k,k}$	$\dots$	$Y_{n-k,n-i}$	$\dots$	$Y_{n-k,n-1}$	$Y_{n-k,n}$
$\vdots$	$\vdots$	$\vdots$							
$n-1$	$Y_{n-1,0}$	$Y_{n-1,1}$	$\dots$	$Y_{n-1,k}$	$\dots$	$Y_{n-1,n-i}$	$\dots$	$Y_{n-1,n-1}$	$Y_{n-1,n}$
$n$	$Y_{n,0}$	$Y_{n,1}$	$\dots$	$Y_{n,k}$	$\dots$	$Y_{n,n-i}$	$\dots$	$Y_{n,n-1}$	$Y_{n,n}$

$Y_{i,n-i}$  ~ cumulative losses of the present calendar year  $n$  or current (cumulative) losses.

$Y_{i,n}$  ~ ultimate (cumulative) losses.

La reserva IBNR se obtiene a partir del rectángulo mencionado hace unos momentos, particularmente interesa la última columna, normalmente llamada “Ultimate” y la diagonal principal, es decir la que contiene aquellos montos pagados por siniestros a la última fecha de información conocida, es decir, la que divide la información conocida de la proyectada. Estos dos elementos se restan y el resultado obtenido es la reserva IBNR de un ramo determinado de seguros.

Para obtener el triángulo run-off de información conocida se parte de un conjunto de datos de siniestros ocurridos y pagados de un determinado ramo de seguros, como puede ser Autos o Responsabilidad Civil, entre otros. Posteriormente se constituyen triángulos acumulados de pagos por periodicidad mensual. Después entra en juego la metodología Chain-Ladder para la estimación de los pagos futuros. El problema que la compañía enfrenta es que la metodología es relativamente sencilla de aplicar y usar, por lo tanto, es de esperar que los resultados no sean los deseados en muchos casos.

Como se mencionó, la metodología parece ser insuficiente, por lo que la compañía se ve en la necesidad de combinar la metodología tradicional con algunas adicionales que requieren insumos como el monto de primas cobradas por la compañía a unas fechas determinadas para la estimación de la reserva IBNR, tales como Bornhuetter-Ferguson, Loss Ratio, entre otras. Esta mezcla de metodologías es aceptada por la compañía y el regulador ya que es efectiva, sin embargo, esto implica manualidades adicionales en el proceso, lo cual puede representar riesgos para la solvencia de la compañía y retrasos en la presentación de los informes mensuales a la junta directiva de la compañía, esto repercute en otras áreas como la de contabilidad, que debe generar los estados financieros basándose en los montos establecidos de la reserva IBNR.

### **Definición de objetivos de negocio**

Con base en la información anterior, se puede identificar que el problema a resolver en términos generales es lograr estimar la reserva IBNR utilizando metodologías robustas de data mining que permitan lograr resultados acertados basados en ejercicios de comparación con las metodologías tradicionales, que satisfagan los requisitos del regulador y los intereses económicos de la compañía. Como objetivo adicional, sería deseable que el ejercicio diera como resultado un menor monto de reservas constituidas, lo cuál se traduciría en un menor impacto en fondos propios. Para el proceso, es posible que con la información actual no se logren los mejores resultados, de manera que un objetivo intermedio es lograr completitud de información.

### **Criterio de éxito de negocio**

El criterio principal de éxito será obtener resultados razonables de reserva IBNR comparando contra las metodologías tradicionales que son aceptadas por el regulador y la compañía. Este objetivo se verá acompañado por uno adicional que determine que se desea reducir la reserva constituida, lo cual permita disminuir el impacto en fondos propios e incrementar las ganancias de los accionistas.

## **Evaluación de la situación**

### **Inventario de recursos**

#### **Tarea 1: Investigación de recursos de Hardware**

Se estima que el software a utilizar no es muy complejo, pero se requerirá manejo de grandes volúmenes de información.

#### **Tarea 2: Identificación de fuentes de datos**

La compañía aseguradora cuenta con fuentes de información relativas a la ocurrencia de siniestros, es decir, se cuenta con distinción por ramos, fechas de pago, fechas de reporte, montos pagados, montos asignados al reasegurador, entre otros. También hay información relevante con respecto al cobro de las primas, como las fechas de pago, montos correspondientes al intermediario de seguros, vigencia de las pólizas, entre otros. La aseguradora también cuenta con información estándar como rendimientos de las inversiones, gastos incurridos por la compañía, entre otros.

En esta compañía, la información se almacena en una bodega de datos con históricos “útiles” que se ha aprovisionado cumplidamente desde el año 2011 con base en información comercial y de las áreas de la compañía, sin embargo, la información no es perfecta y hay problemas de completitud. La compañía tiene los recursos necesarios para comprar fuentes de información que permitan llevar a cabo el proyecto de minería de datos.

#### **Tarea 3: Identificación de recursos personales**

La compañía cuenta con limitados expertos de modelación, si bien existen los suficientes actuarios que conocen la información, no se cuenta con muchos científicos de datos relevantes en la etapa de modelación. Los administradores de las bases de datos tienen gran conocimiento de la información y tienen disponibilidad para ayudar al proyecto.

### **Requerimientos, supuestos y restricciones**

#### **Tarea 1: Determinación de requerimientos**

No existen problemas de seguridad y todos están alineados con los requisitos del proyecto, sin embargo, es muy importante que la solución final cumpla con los requisitos del regulador.

#### **Tarea 2: Aclaración de supuestos**

Las regulaciones que dispongan los entes encargados juegan un papel importante en el proyecto, se deben contemplar en el desarrollo del proyecto. Es necesario incurrir en costos adicionales para contar con un especialista en la modelación de la solución y en minería de datos. La dirección desea tomar un rol ejecutivo, de manera que está interesada en que los resultados representen un beneficio para la compañía y cumplan los resultados del regulador, no tienen interés en los tecnicismos del asunto.

### **Tarea 3: Verificación de restricciones**

Las restricciones económicas se cumplen, el presupuesto es amplio y permite un nivel de maniobra aceptable.

#### **Riesgos y contingencias**

Principalmente hay riesgos asociados a los requisitos cambiantes del regulador, financieros, si la tarea de modelación implica aumentos en costos por la necesidad de un especialista de datos se afectaría el presupuesto. La contingencia para el primer riesgo incluye desarrollos adicionales en la etapa post-proyecto. El segundo riesgo se mitigaría con la contratación temporal de firmas de consultoría que ayuden a complementar la labor en secciones específicas. El tercer riesgo se puede mitigar.

Nota: Los análisis de costo beneficio y terminología o se incluyen en este documento

### **Determinación de metas de la minería de datos**

#### **Metas de la minería de datos**

Como se ha podido evidenciar, el tipo de problema de minería de datos es predictivo, pues se desea estimar el monto de reserva IBNR a partir de ciertos insumos históricos. Se desea obtener un modelo que obtenga resultados consistentes con el negocio y ayude a reducir la intervención humana y los errores en este proceso.

#### **Criterios de éxito de la minería de datos**

Los criterios iniciales a tener en cuenta, son los propios de los modelos estadísticos como pruebas de significancia estadística, de residuos, entre otros. Adicionalmente, se establecerá una métrica de éxito comparando los modelos tradicionales con resultados aceptados y determinado las diferencias y explicando los motivos. Los errores no deben superar el 5%, aunque con pruebas razonables se pueden aceptar.

### Plan de proyecto preliminar

Fase	Tiempo	Recursos	Riesgos
Entendimiento del negocio	4 días	Todas las áreas	Cambios dispuestos por el regulador
Entendimiento de los datos	5 días	Todas las áreas	Problemas de calidad de la información y aprovisionamiento
Preparación de los datos	7 días	Actuaría, Siniestros, I.T, científico de datos	Falta de científicos de datos que satisfagan las necesidades del proyecto
Modelación	5 días	Actuaría, Siniestros, I.T, científico de datos	Falta de científicos de datos que satisfagan las necesidades del proyecto
Evaluación	3 días	Todas las áreas	Cambios dispuestos por el regulador
Lanzamiento	3 días	Actuaría, Siniestros, I.T, científico de datos	Cambios dispuestos por el regulador