CRISP DM – Modelamiento de reservas de seguros comerciales de autos con una metodología distinta a Chain ladder

Introducción

Existen un tipo de compañías que trabajan con la incertidumbre diaria que genera la propia vida o naturaleza, específicamente trabajan con eventos aleatorios inesperados que normalmente pueden ser funestos y toman como nombre de **siniestros**, estas compañías se ocupan de ofrecer el producto del seguro, el cual consiste en dar un respaldo económico a las personas, familias o entidades si algún siniestro suceda, por ejemplo, la existencia de un posible terremoto que genere daños y pérdidas es un evento inesperado y que contiene incertidumbre o aleatoriedad debido a que no se sabe en qué momento ocurre el siniestro, igualmente, también es aleatorio el valor del daño causado; para este tipo de eventos las empresas de seguros ofrecen "seguros" que protegen a las personas de estas posibles pérdidas cuando acontece el siniestro.

Específicamente, un seguro es un pago que hace un cliente a la aseguradora de acuerdo a unos intervalos de tiempo, esto a cambio de que cuando ocurra el siniestro la compañía de seguros cubra los daños causados por el evento inesperado.

Existen muchos tipos de seguros, dependiendo si son a corto o largo plazo, que siniestro cubre, como vida, accidentes automovilísticos, viajes, desempleo, entre otros, o dependiendo del cliente al que se cubra.

Otra característica que tienen los seguros es que a la hora de vender sus productos tienen en cuenta probabilidades de riesgo de cada cliente para dar un servicio muchas más específico, es decir, normalmente las aseguradoras miden características de sus clientes para estimar o cuantificar el riesgo de un evento y así poder asignarle un producto acorde a las cualidades del consumidor.

Luego de que las aseguradoras calculan estos riesgos el otro calculo que realizan es el de la reserva, la reserva es la cantidad de dinero que deben guardar las aseguradoras para poder responder por los diferentes siniestros de sus clientes que pueden ocurrir, comúnmente esta reserva se calcula de manera anual, es decir se calcula cuánto dinero se debe guardar para responder por los siniestros que pasen en el siguiente año, por otro lado, la forma en cómo se calcula esta reserva es por medio del método Chain-Ladder, el cual es un método cuyo objetivo es calcular las reservas de los siniestros futuros por medio de los históricos y bajo la suposición de que los patrones del pasado seguirán pasando en el futuro.

El cálculo de la reserva es un tema de extremo cuidado puesto que una mala estimación puede traer fuertes consecuencias para la compañía, en primer lugar si el cálculo de la reserva subestima los gastos la aseguradora presentara falta de liquidez y se tendrá que endeudar para cumplir sus obligaciones, en segundo lugar, si el cálculo de la reserva sobreestima los gastos entonces la aseguradora pierden la oportunidad de invertir ese dinero y obtener nuevas ganancias.

Así pues, el tema central de este trabajo es plantear un nuevo cálculo de la reservas para mejorar la estimación futura teniendo en cuenta los históricos de la empresa y nuevos modelos teóricos estadísticos o de machine Learning, este trabajo se realizara para una asegurador cuyo producto en un seguro sobre responsabilidad comercial de autos

Los seguros de responsabilidad comercial de autos se ofrecen a empresas o individuos que usan los automóviles en sus operaciones comerciales, como empresas construcción, transporte, entre otros. El seguro cubre daños a terceros que pueden suceder en el momento en que las personas o las empresas laboran, estos daños que cubre el seguro pueden ser lesiones corporales o daños a estructuras.

Comprensión del Negocio

1. Descripción general de la comprensión empresarial

Para lograr un descripción general de la comprensión empresarial exitosa se deben plantear diferentes pasos y fases que se debe argumentar como definir las personas clave de la organización, los diferentes objetivos, definir el personal, recursos, riesgos, continencias, costos, beneficios, entre otros pasos para luego finalizar con un plan del proyecto que define los tiempos de todo el proyecto de minería de datos.

2. Determinación de objetivos empresariales

A medida que van surgiendo nuevos métodos para la estimación y predicción de diferentes problemas por medio de modelos estadísticos y de Machine Learning las compañías de seguros se van interesando cada vez más en estas nuevas técnicas debido a los buenos resultados que ha tenido el aprendizaje de máquinas, Así mismo, las compañías que venden seguros comerciales automovilísticos desean utilizar estos nuevos métodos para poder estimar la reserva y minimizar los errores o la distancia entre la reserva estimada y la real, lo cual se traduce en una mayor eficiencia de recursos, mayores ganancias, confianza del cliente, estabilidad económica, cumplimiento normativo y atracción de inversionistas.

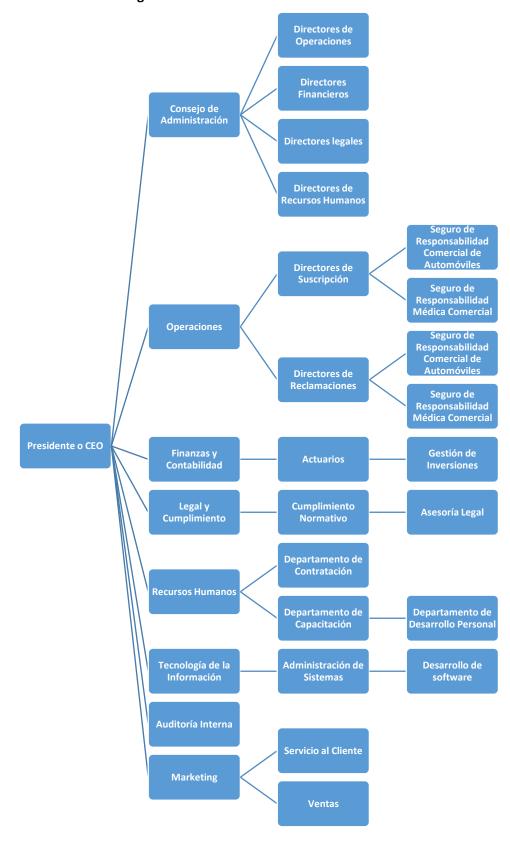
Para lograr el objetivo de minimizar los errores de la estimación de la reserva se procede a realizar o aplicar toda una metodología completa o proyecto de minería de datos y cuyo éxito será valorado si las estimaciones de la reserva con las nuevas metodologías mejoran considerablemente, concretamente el desarrollo del proyecto se considera un éxito si el rendimiento en la estimación es mejor que con la técnica de Chain-Ladder.

1. Compilación de antecedentes comerciales

La aseguradora tiene un buen historial respecto a los pagos de sus seguros a sus clientes y en el pago de sus deudas, lo cual lo ha llevado a tener buenas relaciones comerciales con otras compañías del sector bancario, además la estructura empresarial ha cambiado pocas veces y su solvencia es buena aunque se quiere mejorar con la estimación de las reservas de cada año.

En particular, la compañía de seguros ofrece una cobertura nacional de sus productos, es reconocida por la atención al cliente y la rápida respuesta cuando ocurre un siniestro lo cual lo lleva a ser una empresa con alta experiencia en el mercado y que tiene la reputación de cumplir con las regulaciones y requisitos gubernamentales.

• Determinar la estructura organizacional



Personas Clave en la organización:

Para esta empresa de seguros las personas clave dentro de la organización para hacer el nuevo cálculo de reservas son:

- Ceo
- Consejo de Administración
- Gerencia de Finanzas y Contabilidad
- Gerencia de Tecnología y la Información

Patrocinadores Internos para apoyo financiero o con experiencia para el Proyecto de minería de datos

- Ceo o Gerente General es un patrocinador financiero del proyecto
- Gerente de Finanzas y Contabilidad es un patrocinador financiero y con experiencia en el tema de la estimación de la reservas
 - Grupo de actuarios ofrecen su experiencia en el cálculo de la reserva para la aseguradora
- Gerente de Tecnología y la Información ofrece su experiencia
 - Grupo de informáticos ofrecen experiencia en sistemas, programación y modelos estadísticos y de machine Learning

Unidades de negocio que se verán afectadas por el proyecto de minería de datos

Las unidades de negocios más afectadas son la gerencia de Finanzas y Contabilidad puesto que se hará un cambio en el cálculo de la estimación de la reserva y la Gerencia General debido a que el nuevo cálculo de la reserva puede generar mejores estimaciones y mayor solvencia económica.

Describir el área problemática

El problema principal es mejorar la estimación de la reserva para los años posteriores por medio de un proyecto de minería de datos, cuya área encargada de este problema es la gerencia de Finanzas y Contabilidad que tiene a cargo el grupo de Actuaría de la organización.

Para el proyecto de minería de datos que ayudará a solucionar la problemática se necesita previamente una preparación académica y aplicada sobre las metodologías de la minería de datos en los profesionales de actuaría de la organización, pues son estos los que se encargaran de mejorar la estimación de la reserva.

La empresa ya antes ha utilizado proyecto de minería de datos, más que todo en el área de marketing para desarrollar modelos de Churn o tasa de cancelación de clientes, por lo tanto la aseguradora aprueba y apoya el proyecto de minería de datos para el problema en cuestión.

Describir la solución actual

La solución actual a la estimación del cálculo de la reserva para años posteriores se hace por medio del método Chain-Ladder que se basa en datos históricos de los siniestros para organizarlos en un elemento llamado triángulo de siniestros que luego se le aplica el método de la cadena que se basa en el supuesto de que la relación entre las reservas de un período y las de otro período consecutivo

se mantiene constante a medida que se avanza en el tiempo para último extrapolar los resultados de las reservas futuras en la cartera de seguros.

Normalmente esta técnica es muy aceptada dentro de la organización debido a la facilidad de su uso pero tiene la desventaja de que a la hora de estimar reservas se basa en suposiciones que pueden no ser ciertos, por eso en la literatura se recomienda aplicar este método en conjunto con otras herramientas, por tal motivo, el proyecto de minería de datos desea crear un mejor método para la estimación de la reserva que no dependa de supuesto tan fuertes y sea más robusto

2. Definición de objetivos comerciales

- Mejorar la solvencia económica de la aseguradora
- Mejorar la competitividad en el mercado
- Maximizar la satisfacción del cliente
- Innovación dentro de los procesos de la empresa

3. Criterios de éxito empresarial

Objetivo

Mejorar la estimación de la reserva por medio de nuevas metodologías de la minería de datos en comparación con el método de Chain-Ladder y que produzca un rendimiento mejor a largo de los años.

Subjetivo

Mejorar la solvencia económica de la aseguradora para poder gestionar mucho mejor los recursos de la compañía.

3. Evaluación de la situación

Para este proyecto la aseguradora tiene a su disposición todo el histórico de los datos sobre los siniestros y sobre seguros de autos comerciales como valores de las primas, valores de los siniestros, fecha del accidente, pérdidas acumuladas pagadas y gastos asignados, entre otros.

Actualmente la empresa aseguradora tiene a su disposición los profesionales y personal necesario para completar el proyecto de minería de datos puesto que cuenta con profesionales expertos en el cálculo de reservas y personal calificado en metodologías de minería de datos

Los mayores riesgos que tiene el proyecto es implementar un método que resulte peor que el método tradicional y que la estimación de la reserva no sea mejor, lo cual conlleva a que la empresa pueda caer en serios problemas de solvencia económica que a la larga puede llevar a la aseguradora a tiempos difíciles complicados de superar.

El plan de contingencia consiste es desistir en el modelo de minería de datos que se construya si los objetivos no se cumplen anualmente y volver rápidamente a utilizar el método tradicional de Chain-Ladder.

1. Inventario de Recursos

¿Qué hardware necesita soportar?

Para el proyecto de minería de datos se necesitan los siguientes hardware

- Computadoras y Servidores
- CPU (Unidad Central de Procesamiento)
- GPU (Unidad de Procesamiento Gráfico)
- Buena capacidad de memoria RAM
- Unidades de estado sólido (SSD)
- Almacenamiento en la Nube
- Buena red o conexión a internet
- Cluster de Servidores
- Copia de Seguridad

• Identificar fuentes de datos y almacenes de conocimiento

Los datos en la compañía estan almacenados en los servideros de los que dispone la empresa y cuyo acceso para los profesionales que trabajarán en el proyecto es simple, pues tienen acceso en vivo y disponen del permiso de seguridad necesaria puesto que en el proyecto estan los personas del área de informática. Por otro lado, la compañía dispone de recursos para comprar o adquirir bases de datos externas.

• Identificar recursos de personal

El personal encargada del Proyecto son expertos en el negocio de los seguros, cálculos actuariales, programación y datos, además se cuenta con profesionales en informática que tiene conocimientos en administración de bases de datos, ETL, modelado y análisis de datos.

2. Requerimientos, supuestos y restricciones

Requisitos

Respecto a los requisitos legales, se tiene que las personas encargadas del plan para construir el nuevo método para el cálculo de reservas deben cumplir con ciertas normas a la hora de manipular los datos y estar alineados con los requisitos, estos requisitos son:

- No debe acceder a datos sin autorización
- No debe compartir datos sin consentimiento
- No debe recopilar datos innecesarios
- No debe retener datos más tiempo del necesario
- No debe usar datos para fines no autorizados
- No debe dejar datos desprotegidos
- No debe ignorar las solicitudes de los titulares de datos
- No debe falsificar datos
- No debe compartir contraseñas o credenciales
- No debe dejar dispositivos sin protección
- No debe usar datos para discriminación
- No debe evadir la notificación de violaciones de datos

• Aclarar suposiciones

El Proyecto no cuenta con ninguna competencia dentro de la organización, sin embargo no se descarta que existan otras compañías que buscan el mismo objetivo, así mismo no contiene factores

económicos que puedan afectar su ejecución. Por otro lado, se consideran que los datos son de calidad si son completos, correctos, coherentes, precisos y actualizados.

Por último, respecto a los patrocinadores del proyecto, el proceso y resultados del proyecto serán mostrados a estos, enfocándose mucho más en la visualización de los resultados por medio de software de visualización de datos dinámica.

Verificas restricciones

El personal dedicado al proyecto cuenta con todas las credenciales y contraseñas necesarias para llevar a cabo el desarrollo del nuevo método.

No existen restricciones legales para utilizar los datos de los clientes y en general de la compañía gracias a las autorizaciones de tratamiento de datos a la cual se acogen los clientes, pero siempre recordando las buenas practicas sobre los datos mencionadas anteriormente.

Así mismo, el proyecto cuenta con el apoyo financiero para desarrollarse con éxito.

3. Riesgos y contingencias

A continuación se presenta una lista de riesgos posibles con su respectivo plan de contingencias

- Programación: el proyecto puede demorarse más de lo previsto, para este caso se planea redoblar esfuerzos para terminar de forma temprana, otra idea es alargar un poco la terminación del proyecto.
- Financiero: aunque es difícil que pase existe la posibilidad de que el proyecto quede sin recursos, para este caso se planea refinanciar el proyecto o en el peor de los casos terminarlo con los recursos disponibles.
- Mala calidad de los datos: Los datos pueden tener errores, para este caso se cuenta con un equipo especializado y con experiencia en la limpieza e imputación de datos.
- Resultados: si los resultados iniciales del proyecto sobre el nuevo método para calcular la reserva de la aseguradora no son tan impactantes esto reafirma el hecho de que el método tradicional es igualmente eficiente y eficaz que el nuevo y por lo tanto no hay motivo para cambiarlo o se puede pensar en otra metodología de minería de datos que pueda tener realmente mejores resultados.

4. Terminología

Para que las personas que trabajan en el proyecto minería de datos hablen el mismo idioma se dispone de un glosario con las palabras/frases más importantes o desconocidas para un proyecto de minería de datos para la estimación de la reserva en una empresa de seguros que ofrece seguros de automóviles comerciales. El siguiente glosario también puede ser encontrado en la intranet de la empresa.

- Minería de Datos (Data Mining): El proceso de descubrir patrones, tendencias y conocimientos ocultos en grandes conjuntos de datos utilizando técnicas estadísticas y de aprendizaje automático.
- Reserva de Siniestros: Una estimación de la cantidad de dinero que una compañía de seguros debe reservar para cubrir reclamaciones de seguros pendientes y futuras.

- Datos de Siniestros: Información detallada sobre los siniestros reportados, que incluye fechas, descripciones, costos estimados y pagos realizados.
- Desarrollo de Siniestros: El proceso por el cual los costos de un siniestro aumentan o disminuyen con el tiempo a medida que se investigan, se procesan y se resuelven las reclamaciones.
- Triángulo de Siniestros: Una representación tabular de los siniestros a lo largo del tiempo, que muestra cuándo se reportaron, cuánto se pagó en cada período y cuánto queda pendiente de pago.
- Tasa de Desarrollo: La tasa promedio a la que los costos de los siniestros se incrementan o disminuyen con el tiempo en función del análisis de datos históricos.
- Modelo de Reservas: Un modelo matemático o estadístico que se utiliza para prever los costos futuros de siniestros y, en última instancia, calcular las reservas necesarias.
- Ajuste de Reservas: Los cambios que se realizan en las estimaciones de reserva a medida que se obtienen más datos o se actualiza el modelo de reserva.
- Análisis de Pérdida Triangular: Una técnica que se utiliza para estimar las reservas basadas en la información contenida en el triángulo de siniestros.
- Exceso de Pérdida (Excess Loss): La cantidad de dinero que una aseguradora está dispuesta a pagar por encima de un cierto límite antes de que se active la cobertura de reaseguro.
- Reaseguro (Reinsurance): Un acuerdo en el que una compañía de seguros transfiere parte de sus riesgos a otra compañía de seguros o reaseguradora para limitar sus pérdidas potenciales.
- Cobertura de Responsabilidad Comercial: Un tipo de seguro que proporciona protección contra reclamaciones por lesiones corporales o daños a la propiedad que puedan surgir en el curso de las operaciones comerciales.
- Modelo de Aprendizaje Automático: Un enfoque que utiliza algoritmos y técnicas de aprendizaje automático para analizar datos históricos y hacer predicciones sobre futuros siniestros y costos.
- Validación Cruzada (Cross-Validation): Una técnica que se utiliza para evaluar la precisión y la eficacia de un modelo de aprendizaje automático mediante la división de los datos en conjuntos de entrenamiento y prueba.
- Clasificación de Riesgo: El proceso de categorizar diferentes tipos de riesgos y evaluar su probabilidad y gravedad.
- Primas de Seguro: Los pagos periódicos que los asegurados realizan a la compañía de seguros a cambio de la cobertura de seguro.
- Seguro de Automóviles Comerciales: Un tipo de seguro que proporciona cobertura para vehículos utilizados con fines comerciales, como camiones, furgonetas y flotas de vehículos
- Póliza: Un contrato legal que establece los términos y condiciones de la cobertura de seguro, incluyendo los riesgos cubiertos, las primas y otros detalles.
- Primas: Pagos regulares realizados por el asegurado a la compañía de seguros a cambio de la cobertura de seguro.
- Riesgo: La probabilidad de que ocurra un evento adverso que pueda dar lugar a una reclamación de seguro.
- Reclamación: Una solicitud presentada por un asegurado para recibir compensación por un evento cubierto por la póliza de seguro.

- Estadísticas: El análisis de datos numéricos y la aplicación de métodos estadísticos para obtener información sobre patrones y tendencias.
- Segmentación: La división de un conjunto de datos en grupos más pequeños o segmentos para un análisis más detallado.
- Ajuste: La modificación de las estimaciones de reservas de seguros en función de nuevos datos o información actualizada.
- Cobertura: El alcance y los términos de protección proporcionados por una póliza de seguro.
- Fraude: La presentación de información falsa o engañosa con el propósito de obtener beneficios indebidos del seguro.
- Exceso: El monto que una compañía de seguros no cubrirá y que debe ser asumido por el asegurado o por otra forma de seguro.
- Estimación: Una aproximación calculada o proyectada de un valor o cantidad, como la estimación de las reservas de siniestros.
- Modelo: Un conjunto de algoritmos y reglas matemáticas utilizado para predecir o analizar datos en función de patrones históricos.
- Experiencia: El historial de siniestros y reclamaciones de una compañía de seguros, utilizado para hacer estimaciones futuras.
- Aprendizaje: La capacidad de un sistema informático para mejorar su rendimiento a través de la experiencia y la adaptación a nuevos datos.
- Validación: El proceso de confirmar la precisión y eficacia de un modelo o método de estimación mediante la comparación con datos reales.
- Regulaciones: Las leyes y normativas gubernamentales que rigen la industria de seguros y establecen estándares para la conducta y las prácticas.
- Reserva: La cantidad de dinero que una compañía de seguros establece para cubrir futuras reclamaciones y obligaciones.
- Siniestro: Un incidente o evento adverso que da lugar a una reclamación de seguro.
- Historial: Un registro de eventos pasados y datos relacionados, como el historial de siniestros de un asegurado.
- Pérdida: La cantidad de dinero que una compañía de seguros paga como resultado de una reclamación realizada por un asegurado.

5. Análisis costo/beneficio

• Costo: Estimación de costos en Dólares

- 1. Costo por Recopilación de Datos:
 - Contratación de personal o servicios externos para recopilar datos históricos de seguros: \$10,000.
 - Adquisición de bases de datos externas relevantes para el análisis: \$20,000.
 - Gastos relacionados con la limpieza y preprocesamiento de datos: \$10,000.

2. Despliegue de Resultados:

- Desarrollo de un sistema o plataforma para implementar el modelo de cálculo de triángulos y reservas en un entorno de producción: \$5,000.
- Gastos asociados a la implementación de software y hardware necesarios para el despliegue: \$1,000.
- Capacitación de personal en el uso del sistema y la interpretación de los resultados: \$10,000.

3. Costos de Operación:

- Gastos recurrentes de mantenimiento y actualización del sistema de cálculo de triángulos y reservas: \$15,000 al año.
- Costos de almacenamiento de datos a largo plazo: \$5,000 al año.
- Servicios de consultoría o asesoría para ajustar el modelo de minería de datos: \$10,000 al año
- Licencias de software y herramientas utilizadas en el proyecto: \$5,000 al año.

4. Costos Laborales:

- Salarios y beneficios para los científicos de datos, ingenieros de datos, actuarios y analistas involucrados en el proyecto: \$15,000 al año.
- Posibles bonificaciones o incentivos para el personal en función de los resultados del proyecto: \$5,000 al año.
- Gastos de contratación si se necesita personal adicional: \$10,000.

Estimación total:

1. Costo por Recopilación de Datos: \$40,000

2. Despliegue de Resultados: \$11,500

3. Costos de Operación: \$35,000

4. Costos Laborales: \$30,000

Costos totales: \$116,500 para todo el proyecto de minería de datos

• Beneficios:

- Se cumple el objetivo del proyecto lo cual traerá mejoras en el cálculo de la reserva y a largo plazo mejores rentabilidades para la empresa
- Se logra una mayor organización de los datos para el cálculo de reservas
- Se evidencian nueva información y comprensión de los datos gracias a la exploración de los datos
- Se toman de Decisiones Basada en Datos
- Ventaja Competitiva
- Satisfacción del Cliente
- Cumplimiento Regulatorio
- Innovación en Productos y Servicios

4. Determinación de los objetivos de la minería de datos

1. Objetivos de minería de datos

Utilizando datos históricos de las reservas de la aseguradora se generará un modelo de machine Learning o estadístico para hacer cálculo y estimación de triángulos para cálculo de reservas anuales; las estimaciones se validarán con las reservas reales de cada año.

2. Criterios de éxito de la minería de datos

Se desea obtener un modelo que tenga un mejor rendimiento que el método usual de Chain ladder y tenga bueno valores en las diferentes medidas de rendimiento, este rendimiento será calculado con diferentes métricas.

- Error Cuadrático Medio
- Error Absoluto Medio

- Coeficiente de Determinación
- Error porcentual absoluto medio

Estas métricas se utilizan porque se van a pronosticar reservas, es decir valores de números continuos.

5. Plan del Proyecto

Fase	Tiempo	Recursos	Riesgos
Comprensión Empresarial	1 semana	Todos los involucrados en el proyecto pertenecientes a la compañía	Cambios en la dirección de la compañía, cambio en los productos relacionados con el cálculo de la reserva que se desea estimar.
Comprensión de los datos	1 semana	Todos los involucrados en el proyecto pertenecientes a la compañía, especialmente analistas de datos, mineros de datos, ingenieros de datos y actuarios. Lenguaje de programación Python, adquisición de bases de datos y almacenamiento.	Problemas de limpieza de datos, origen de la información, problemas de hardware o software, licencias, versiones de librerías del lenguaje de programación.
Preparación de los datos	2 semana	Analistas de datos, mineros de datos, ingenieros de datos y actuarios. Lenguaje de programación Python, adquisición de bases de datos y almacenamiento.	Problemas de hardware o software, licencias, versiones de librerías del lenguaje de programación, definir y crear la unidad de análisis.
Modelado	3 semanas	Analistas de datos, mineros de datos, ingenieros de datos y actuarios. Lenguaje de programación Python, adquisición de bases de datos y almacenamiento.	Incapacidad para escoger el modelo adecuado, problemas de hardware o software, licencias, versiones de librerías del lenguaje de programación
Evaluación	1 semana	Analistas de datos, mineros de datos, ingenieros de datos, actuarios y personas de la compañía que estan involucradas y les interesa el proyecto. Lenguaje de programación Python.	Cambios en la dirección de la compañía, incapacidad para implementar los resultados, mal escogencia de evaluación.
Despliegue	1 semana	Analistas de datos, mineros de datos, ingenieros de datos, actuarios y personas de la compañía que estan involucradas y les interesa el proyecto.	incapacidad para implementar los resultados porque no se tienen los recursos necesarios

1. Evaluación de herramientas y técnicas

La herramienta a utilizar para lograr el éxito en la minería de datos es el lenguaje de programación Python por medio de Notebook de Jupyter el cual es idóneo para presentar algoritmos, texto y lenguaje matemático sobre un modelo de minería de datos. El Lenguaje de programación Python tienen diferentes ventajas.

Legibilidad y Simplicidad

- Amplia Comunidad y Soporte
- Multiplataforma
- Librerías y Ecosistema
- Aprendizaje Automático y Ciencia de Datos
- Desarrollo Web
- Integración
- Automatización
- Comunidad Activa
- Gratuito y de Código Abierto