

› White Paper



Gestionando del ciclo de vida analítico para la toma de decisiones a gran escala:

Los pasos a seguir para pasar de los datos a las decisiones de la forma más rápida posible

Contents

Introducción	1
Decisiones a gran escala	2
¿Cuáles son los retos?	2
SAS® Analytical Life Cycle: Mejores prácticas para mejorar los resultados de la modelización predictiva	4
La fase de exploración del ciclo de vida analítico	4
La fase de implementación del ciclo de vida analítico	6
Pasar del concepto a la acción: cómo crear un entorno analítico eficaz	7
SASCómo puede contribuir SAS® al ciclo de vida analítico al ciclo de vida analítico	8
Preparación y exploración de los datos: Un enfoque sistemático	8
Desarrollo de modelos mediante un entorno analítico innovador	9
Implementación de modelos desde un solo entorno fácil de usar	9
¡En marcha! Pon en práctica los resultados de la modelización lo más rápido posible	10
Gestión y evaluación de modelos: Un proceso continuo	10
Casos prácticos	11
Institución financiera del Reino Unido: Modernizando su ciclo de vida analítico	11
Orlando Magic: La magia al descubierto	11
Visa: Mil millones de decisiones operacionales al año mejoran la experiencia de los clientes y reducen el fraude	11
Recursos adicionales	12

Introducción

Considera el supuesto que se expone a continuación.

Una organización cuenta con cientos de modelos analíticos integrados en producción que respaldan la toma de decisiones en marketing, la fijación de precios, los riesgos de crédito, los riesgos operativos, el fraude y la gestión financiera.

Los analistas de las diferentes unidades de negocio desarrollan sus modelos sin seguir ningún proceso formalizado o estándar de almacenamiento, implementación y gestión. Algunos modelos no disponen de documentación que indique el responsable del modelo, el objetivo empresarial, las instrucciones de uso u otro tipo de información necesaria para gestionarlo o presentarlo a las autoridades reguladoras.

Los responsables de la toma de decisiones obtienen resultados de los modelos, pero con requisitos y controles limitados. Además, los resultados son incoherentes debido al uso de variables y conjuntos de datos distintos en su creación. Los procesos de validación y las pruebas backtesting no son suficientes para asegurar el ajuste de los modelos.

Las decisiones se adoptan en función de los resultados obtenidos por los modelos, y todos esperamos lograr los mejores resultados

A muchas organizaciones les suena mucho este tipo de problemas con sus modelos. En un entorno de modelización variado y poco gestionado, a veces es muy difícil responder a preguntas muy importantes en relación a los modelos predictivos. Entre estas cuestiones, podemos encontrar las siguientes:

- ¿Quién creó los modelos y por qué?
- ¿Qué variables de entrada se usan para hacer predicciones y, en última instancia, tomar decisiones?
- ¿Cómo se usan los modelos?
- ¿Cuál es el rendimiento de estos modelos y cuándo se actualizaron por última vez?
- ¿Dónde se encuentra la documentación complementaria?
- ¿Por qué se tarda tanto en poner en marcha modelos nuevos o actualizados?

Las organizaciones que no pueden contestar con confianza a estas preguntas no pueden garantizar que sus modelos analíticos verdaderamente ofrecen buenos resultados.

Los modelos analíticos son fundamentales para las decisiones empresariales importantes. Gracias a ellos, puedes identificar nuevas oportunidades, forjar nuevas o mejores relaciones con clientes y gestionar la incertidumbre y los riesgos. Por estos y otros muchos motivos, los modelos se deben crear y considerar como activos organizativos de gran valor. Pero ¿cómo?

Para empezar, necesitas un software potente y fácil de usar que te ayude a dar forma a tus datos y crear rápidamente muchos modelos predictivos precisos. A continuación, debes recurrir a procesos potentes e integrados para gestionar tus modelos analíticos, de modo que consigas un rendimiento óptimo

Uso de modelos para tomar decisiones de forma automática

La toma de decisiones operativas es un proceso muy repetitivo que, a menudo, ocurre muchas veces al día. En la toma de decisiones, pueden participar personas, por ejemplo, empleados de call centers que realizan ventas incrementales o complementarias llamando por teléfono a clientes, o bien pueden ser decisiones completamente automatizadas, que simulan la lógica humana, por ejemplo, si se rechaza una compra con tarjeta de crédito. Sin embargo, una decisión operativa no tiene por qué afectar a un solo cliente, sino que puede hacer referencia al mejor grupo de clientes al que dirigir una determinada oferta. Este tipo de decisiones se meten en los sistemas de gestión de campañas cada semana o mes. Cuando los modelos analíticos predictivos están integrados en los sistemas de producción y cuando negocio usa los resultados para ofrecer respuestas al instante, has implementado eficazmente estos modelos analíticos.

durante todo el ciclo de vida. Los equipos analíticos y de TI necesitan procesos eficaces y sistemáticos, así como una arquitectura fiable para gestionar datos y realizar el seguimiento de los modelos analíticos predictivos durante el ciclo de implementación.

Y lo que es más importante: la clave para conseguir el éxito en materia de análisis es convertir rápidamente los datos en conocimiento y este, a su vez, en acciones, lo que implica que debes integrar con eficacia modelos predictivos precisos en sistemas de producción que impulsen la toma de decisiones de forma automatizada.

La complejidad creciente de gestionar cientos o miles de modelos inestables hace que las organizaciones se encuentren al borde de una revolución de la información. El enfoque artesanal, que resulta ineficaz y ha quedado obsoleto, debe evolucionar y convertirse en un proceso automatizado más eficaz.

“Las decisiones operacionales son las que convierten tu estrategia de negocio en una realidad y garantizan que tu organización trabaje de manera eficaz”.

James Taylor y Neil Raden,
Smart (Enough) Systems

Decisiones a gran escala

¿Cuántas decisiones que afectan a las operaciones se toman en tu organización cada día? Probablemente más de lo que te imaginas; por ejemplo, elegir una entidad financiera. ¿Cuántas transacciones con tarjeta de crédito se procesan cada hora? (En Visa, la tasa puede alcanzar las 14 000 por segundo. Consulta la página 3). Cada una de ellas representa una decisión en materia de operaciones: permitir o prohibir que la transacción se procese teniendo en cuenta la valoración de riesgo de fraude. Aunque cada transacción o decisión de este tipo pueda tener un riesgo individual bajo, el gran número de estas decisiones que se toman cada hora o cada día incrementa mucho el riesgo asociado.

Por ello, la capacidad de tomar rápidamente las decisiones operativas más adecuadas, al mismo tiempo que se aprovechan los crecientes volúmenes de información, puede traducirse en que se produzca un fraude o no, es decir, en fracaso o éxito empresarial.

Por lo tanto, ¿qué hay que hacer para tomar rápidamente muchas decisiones adecuadas y alineadas con la estrategia? Algunos requisitos son:

1. Aplicaciones operacionales que se basen en los datos para ofrecer respuestas a personas (o sistemas) con lo que garantizar una toma de decisiones optimizada
2. Modelos analíticos actualizados y basados en un conocimiento profundo que la empresa pueda emplear como material fiable a la hora de tomar decisiones acertadas en el momento adecuado
3. Integración de reglas empresariales y análisis predictivos en los procesos de toma de decisiones operativas que proporcionen la orientación necesaria para tomar decisiones basadas en datos reales
4. Un método para gestionar y supervisar los modelos analíticos con el fin de garantizar que son eficaces y siguen ofreciendo las respuestas acertadas
5. Una arquitectura y unos procesos que pueden crecer para satisfacer nuevas necesidades, como difundir datos y crear modelos predictivos más detallados de la forma más rápida posible

¿Cuáles son los retos?

Desafortunadamente, a pesar de que los datos abundan y sepamos qué necesitamos, no resulta sencillo convertir grandes cantidades de datos estructurados y desestructurados en información útil que permita tomar mejores decisiones de forma automatizada. A continuación, figuran algunos retos:

- **Retrasos:** Debido a procesos que, a menudo, son manuales y ad hoc, la implementación de un modelo en los sistemas de producción puede tardar meses. Puesto que las fases de desarrollo y prueba de los modelos lleva demasiado tiempo, pueden quedarse obsoletos en el momento en el que alcanzan la fase de producción. O puede suceder que nunca lleguen a implementarse. Los problemas de cumplimiento externo e interno pueden plantear aún más retos.

- **Dificultades para identificar los desencadenantes** La fase que implica convertir las respuestas de los modelos analíticos en decisiones operativas requiere reglas de negocio acordadas y claras. Estas reglas de negocio tienen que formar parte del entorno controlado, ya que determinan cómo se usan los resultados de los modelos. Por ejemplo, un modelo de detección de fraude podría identificar una calificación de riesgo de fraude como un número comprendido entre 100 y 1000 (similar a la calificación crediticia FICO). La empresa tendrá que decidir qué nivel de riesgo implica llevar a cabo una acción. Si el nivel de un aviso de fraude se establece en un nivel demasiado alto, el fraude podría pasar inadvertido. En cambio, si se establece en un nivel demasiado bajo, los avisos señalan demasiados falsos positivos. Ambas metodologías restan el valor aportado por estos modelos y disminuyen la confianza en los resultados.
- **Resultados deficientes:** Con demasiada frecuencia, los modelos de bajo rendimiento se quedan en la fase de producción, incluso si ofrecen resultados imprecisos que conllevan decisiones empresariales desacertadas. Los resultados de los modelos cambiarán a medida que los datos se adaptan a las nuevas condiciones y comportamientos. Las principales razones son la falta de un repositorio central de modelos y la inexistencia de métricas coherentes para determinar si un modelo necesita actualizarse o sustituirse.
- **Confusión:** Las organizaciones actúan de forma reactiva, es decir, reaccionan con prisas cuando el regulador marca el momento de la entrega. Dado que cada departamento tiene una manera diferente de manejar y validar un modelo, las revisiones y ajustes de los modelos son complicados lo que dificulta enormemente el trabajo de TI. No se sabe con certeza por qué se seleccionó el modelo definitivo, cómo se calculó un determinado scoring ni por qué criterios se rigen las reglas empresariales que activan el modelo.
- **Falta de transparencia:** Hay poca visibilidad en las fases en las que se desarrollan los modelos o se desconoce quién está involucrado en ellos a medida que avanzan por su ciclo de vida. Por lo tanto, surgen presunciones conflictivas. Se debe recurrir a revisores imparciales para validar los modelos a medida que pasan por cada grupo, lo que supone una gran despilfarro de recursos.
- **Pérdida de información importante acerca de los modelos:** Si la información acerca de los modelos no está correctamente documentada, cuando los responsables se van se pierde la información. La propiedad intelectual también es un aspecto importante a tener en cuenta.

Visa: Mil millones de decisiones operacionales al año mejoran la experiencia de los clientes y reducen el fraude

Visa, que es una marca reconocida a nivel mundial, facilita las transferencias electrónicas de fondos a través de productos de marca de terceros comercializados por sus miles de entidades financieras asociadas. La empresa ha tramitado 64 900 millones de transacciones en 2014, y se realizaron compras con un valor de 4,7 billones de dólares por medio de tarjetas Visa durante ese mismo año.

Visa cuenta con una capacidad de cálculo que es capaz de tramitar 56 000 mensajes de transacción por segundo, que supera cuatro veces la tasa de transacción máxima real hasta la fecha. Visa no solo tramita y calcula, sino que realiza análisis constantemente para proporcionar información estratégica y operativa a sus entidades asociadas y les ayuda a mejorar el rendimiento.

El objetivo de negocio está respaldado por un sistema de gestión de datos muy potente. Visa también ayuda a sus clientes a mejorar el rendimiento y les proporciona información analítica exhaustiva. "Examinamos los patrones de comportamiento mediante una agrupación y segmentación muy detalla, y proporcionamos esta información a nuestros clientes", afirma Nathan Falkenberg, responsable de Visa Performance en el norte de Asia.

Hace muy poco tiempo Visa hizo una prueba de concepto con una solución de alto rendimiento de SAS que incluía almacenamiento en memoria para potenciar los algoritmos estadísticos y de machine-learning para posteriormente presentar dicha información de manera visual. Falkenberg declara: "El reto que se nos presenta, tal y como ocurre en cualquier empresa que pretende gestionar y usar grandes cantidades de datos, es cómo sacarle el mayor partido a todos esos datos para que nos ayuden a solucionar nuestros retos de negocio, ya sea mejorar los modelos de prevención del fraude, o bien ofrecer asistencia a nuestros clientes para comunicarse de forma más eficaz con sus propios clientes".

[Más información](#)

- **Escasez de habilidades analíticas:** Aunque haya un número creciente de científicos de datos que entran en el mercado, la escasez de habilidades analíticas necesarias para la creación e implementación de modelos supone aún un gran reto para muchas organizaciones.

SAS® Analytical Life Cycle: Mejores prácticas para mejorar los resultados de la modelización predictiva

Las organizaciones líderes reconocen que los modelos analíticos son activos empresariales fundamentales que generan respuestas que ayudan a mejorar las relaciones con los clientes y las operaciones, además de aumentar los ingresos y reducir los riesgos. Por lo tanto, es evidente que tratan de crear los mejores modelos posibles.

Sin embargo, son pocas las entidades que saben lidiar plenamente con las complejidades que surgen en todo el ciclo de vida del modelo analítico, ya que se trata de una tarea de carácter polifacético.

En SAS, hemos desarrollado un **ciclo de vida analítico** sistemático que te guía durante cada fase del proceso durante el cual los datos se transforman en decisiones. Empezaremos diciendo que hay dos fases en el ciclo de vida analítico: detección e implementación. A continuación, describiremos las tareas de cada fase y cómo interactúan para crear un proceso sistemático que puedas usar para producir resultados predictivos, fiables y constantes.

La fase de exploración del ciclo de vida analítico

- **Preguntas que se quieren responder:** El proceso de exploración se rige por el planteamiento de cuestiones empresariales que dan lugar a la innovación. Por lo tanto, el primer paso implica determinar lo que la empresa quiere averiguar.

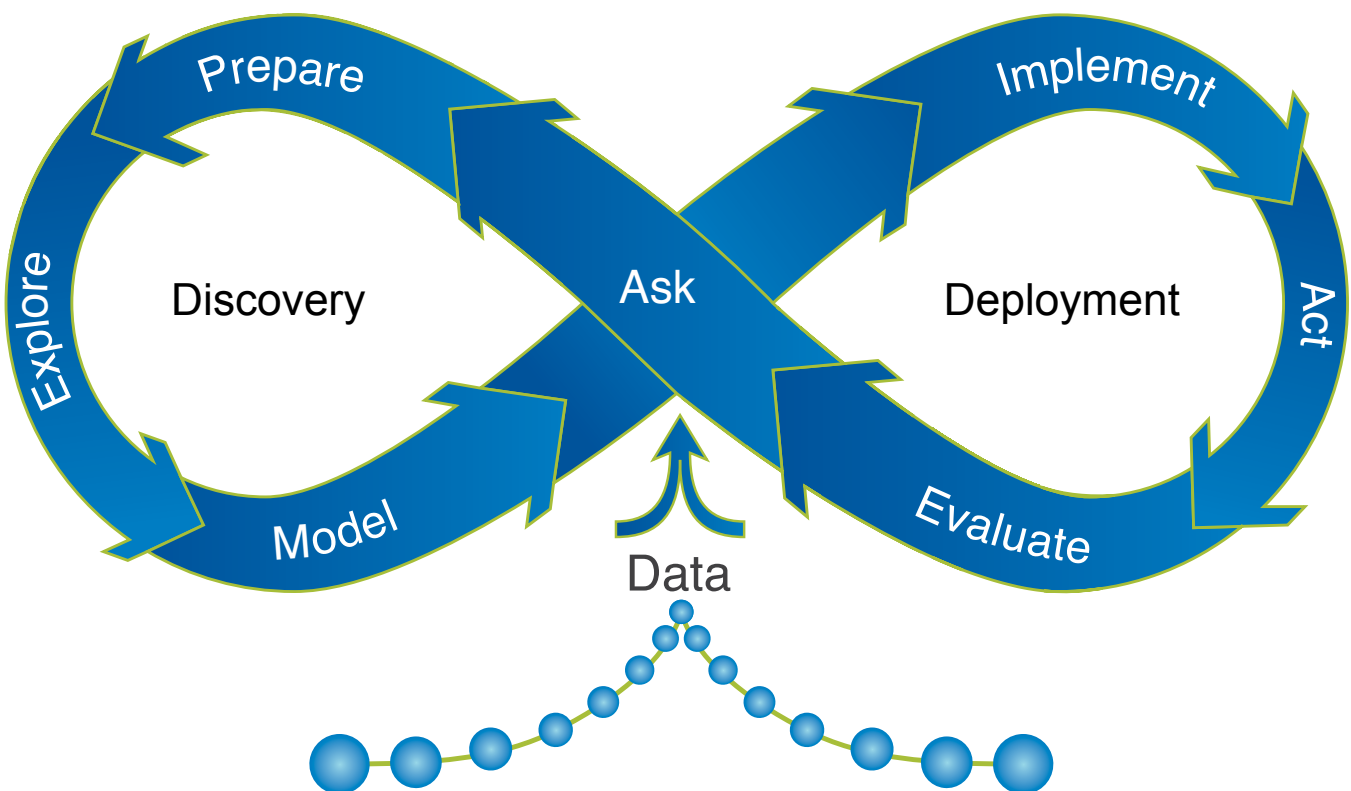


Ilustración 1: El ciclo de vida analítico de SAS

A continuación, esa pregunta se traduce en una representación matemática del problema, que se puede solucionar con análisis predictivos. Las áreas de negocio también tienen que especificar las necesidades, el alcance, las condiciones del mercado y los objetivos relacionados con la preguntas que están tratando de responder, lo que ayuda a seleccionar las técnicas de modelización más apropiadas.

- **Preparación de los datos:** Las tecnologías, como Hadoop, y los ordenadores cada vez más rápidos y baratos permiten almacenar y usar una cantidad y tipos de datos mayor que nunca. No obstante, esto solo ha aumentado la necesidad de combinar datos en diferentes formatos y de diferentes orígenes, así como de transformar datos en bruto, de modo que se puedan usar en la modelización predictiva. Debido a los nuevos tipos de datos que provienen de dispositivos conectados, de sensores o los registros web de interacciones on line, la fase de preparación de datos plantea aún más retos. Muchas organizaciones aún declaran que emplean una cantidad de tiempo excesiva, a veces, hasta un 80 %, para lidiar con las tareas de preparación de datos. La mayor parte del tiempo durante la fase de exploración se debe emplear en estudiar los datos y crear modelos idóneos, en lugar de preparar los datos.
- **Exploración de los datos:** Las herramientas de visualización self-service interactivas deben satisfacer las necesidades de una amplia variedad de usuarios (desde los analistas de negocio que no poseen conocimientos estadísticos hasta los científicos de datos que son expertos en análisis), de modo que puedan buscar fácilmente relaciones, tendencias y patrones para comprender mejor los datos. Durante esta fase, se pule tanto la pregunta como el enfoque que surgieron en la fase inicial de planteamiento del proyecto. Asimismo, se desarrollan y se ponen a prueba ideas sobre cómo afrontar el reto de negocio desde una perspectiva analítica. Durante la exploración de los datos, es posible que te des cuenta de que debes añadir, eliminar o combinar variables para crear modelos mucho más concretos. Soluciones ágiles e interactivas contribuyen a que este proceso sea sistemático, que es esencial para identificar las mejores preguntas y respuestas.
- **Modelización de los datos:** Durante esta fase, se aplican numerosos algoritmos de modelización analítica y machine-learning a los datos para encontrar la mejor representación de las relaciones entre ellos que ayudará a resolver la pregunta de negocio. Las herramientas analíticas buscan una combinación de técnicas de modelización y datos que pronostican el mejor resultado. Sin embargo, no existe un solo algoritmo que valga para todo, sino que el "más idóneo" para resolver las preguntas hechas se establece en función de los datos. La experimentación es un aspecto

clave para encontrar la respuesta más fiable, y la creación de modelos automatizados puede ayudar a minimizar el tiempo en el que se obtienen resultados, así como impulsar la productividad de los equipos analíticos. Con las herramientas de creación manual de modelos los responsables de la minería de datos y los científicos de datos solo podían crear varios modelos en una semana o un mes. Las nuevas soluciones y los ordenadores más rápidos han acelerado este proceso, de modo que, hoy en día, se pueden crear cientos o miles de modelos en el mismo plazo. Sin embargo, con ello, ha surgido otro problema: ¿cómo identificar de la forma más rápida y fiable el modelo (entre muchos) que mejor funciona? La comparación entre modelos sirve para elegir el que ofrezca los mejores resultados para un conjunto de datos específico. El hecho de que existan modelos machine-learning automatizados y métricas claramente definidas se consigue el identificar el mejor modelo. Los analistas y científicos de datos pueden dedicar tiempo a centrarse en el análisis y cuestiones más estratégicas.

La fase de implementación del ciclo de vida analítico

- **Implementación de los modelos:** En este punto, recopilas la información generada en la fase de detección y la incorporas utilizando procesos automatizados y sistemáticos. En muchas organizaciones, el proceso de modelización analítica se ralentiza considerablemente en esta fase, dado que no se ha definido una etapa de transición entre la detección y la implementación, ni tampoco existe una colaboración entre los desarrolladores de modelos y los arquitectos de implementación del departamento de TI, ni mucho menos una automatización optimizada. En la mayoría de las organizaciones, el entorno de implementación es muy diferente del entorno de detección, particularmente si los modelos predictivos fomentan la toma de decisiones que afectan a las operaciones. A menudo, el departamento de TI tiene que aplicar políticas de control estrictas en este entorno para garantizar los acuerdos de nivel de servicio con la empresa. Si integras las fases de detección e implementación, puedes crear una transición sistemática, flexible y automatizada que mejora las decisiones operativas. Asimismo, un proceso controlado y transparente resulta muy importante para todos, especialmente para los auditores. Una vez creado, el modelo se registra, se somete a pruebas o se valida, se aprueba y, por último, se informa de que está listo para usarse junto con los datos en producción, que están integrados en los sistemas operativos.
- **Trabajando con la nueva información:** Existen dos tipos de decisiones que se pueden tomar en función de los resultados analíticos. Por un lado, los encargados de examinar los resultados y actuar toman decisiones estratégicas, normalmente de cara al futuro. Por otro, las decisiones operativas se toman de forma automatizada, como las calificaciones crediticias o las mejores ofertas recomendadas. Estas decisiones no involucran a personas, dado que las reglas que estas últimas podrían aplicar se pueden codificar en sistemas de producción. Cada vez más organizaciones tratan de automatizar las decisiones que atañen a las operaciones y proporcionar resultados en tiempo real para reducir las demoras a la hora de tomar decisiones. Si estas decisiones operativas se basan en las respuestas que dan los modelos analíticos, adquieren objetividad, coherencia, regularidad y un valor cuantificable. La integración de los modelos con herramientas de gestión de decisiones ayuda a que las organizaciones creen flujos completos y exhaustivos para tomar decisiones operativas. Estos flujos combinan modelos analíticos con indicadores basados en reglas de negocio para tomar las mejores decisiones de forma automatizada. Y dado que estos aspectos están formalmente definidos en la

herramienta de gestión de decisiones, las actualizaciones y mejoras que sirven para cambiar las condiciones resultan sencillas y mejoran el control y la agilidad de los negocios. Tras aprobar el modelo para la fase de producción, la herramienta de gestión de decisiones lo aplica a los nuevos datos operativos, generando así la información predictiva necesaria para actuar de la mejor forma posible.

- **Evaluación de los resultados:** El siguiente paso, y quizás el más importante, es evaluar los resultados de las acciones que se han tomado siguiendo las indicaciones del modelo. ¿Han dado lugar tus modelos a las predicciones correctas? ¿Se obtuvieron resultados tangibles, como el aumento de los ingresos o la reducción de los costes? Es imprescindible una supervisión y medición continuas del rendimiento de los modelos a partir de estadísticas estandarizadas. Los resultados de esta evaluación, a su vez, pueden incorporarse a la siguiente iteración del modelo, lo que da lugar a un bucle continuo de aprendizaje automático. Si detectas un deterioro de ciertos modelos analíticos, puedes definir la estrategia óptima de actualización para que sigan produciendo los resultados deseados. Dado el incremento del número de los modelos analíticos, la automatización se impone como un elemento necesario para identificar con rapidez aquellos modelos que reclaman una mayor atención, e incluso realizar un ajuste automatizado.
- **Nuevo planteamiento:** Los modelos predictivos no son eternos. Los factores que permiten elaborar predicciones en el marco de un modelo tienen un carácter dinámico: los clientes van cambiando; algunos competidores abandonan el mercado, mientras que otros nuevos se incorporan, y nuevos datos pasan a estar disponibles. En consecuencia, incluso los modelos más precisos tienen que actualizarse, lo que obliga a las organizaciones a repetir las fases de exploración e implementación. Se trata de un proceso constante y cambiante. Si un modelo se deteriora, se somete, bien a un ajuste, mediante el cambio de sus coeficientes, bien a una reformulación, en la que se conservan algunas características y se añaden otras nuevas. Cuando el modelo deja de satisfacer una necesidad empresarial, se retira.

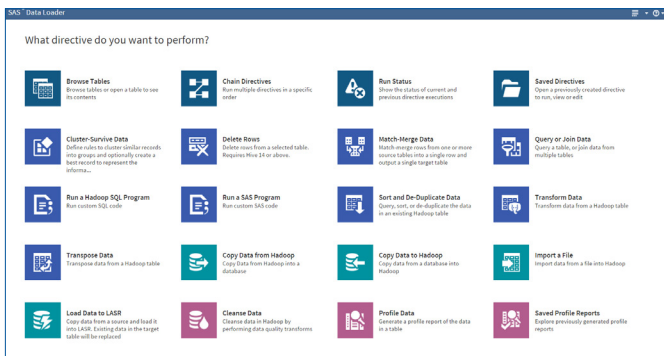
Resulta fácil imaginar las muchas formas en las que puede fallar este proceso. Con frecuencia, las organizaciones tardan meses, o incluso años, en recorrer este trayecto. Se pueden mencionar numerosos que hacen que ocurra esto:

- Las fuentes de datos necesarias pueden estar dispersas en diversos puntos de la organización.
- Es posible que los datos deban integrarse y limpiarse varias veces para satisfacer distintos requisitos analíticos.
- La traducción manual de los modelos a distintos lenguajes de programación para su integración con sistemas

operativos importantes (tanto en sistema de lotes como en tiempo real) puede prolongarse considerablemente.

- Las organizaciones pueden tardar en reconocer la necesidad de cambiar un modelo, por lo que avanzan adoptando decisiones inadecuadas en función de resultados obsoletos.
- Muchos de los pasos incluidos en el ciclo de vida analítico son repetitivos por naturaleza, por lo que pueden exigir la vuelta a un paso anterior del ciclo para incorporar o actualizar datos.
- Los distintos perfiles añaden complejidad al proceso, lo que confiere una importancia capital a las tareas de colaboración y documentación. En muchas organizaciones, el departamento de IT se encarga de preparar los datos durante la fase de detección, mientras que las labores de exploración de datos y de desarrollo de modelos suelen recaer en los analistas de negocios y los científicos de datos. El departamento de IT también es responsable de la implementación (en especial, cuando incluye la integración en procesos operativos). No obstante, el equipo específico del departamento encargado de esta tarea puede diferir del grupo de gestión de datos.

Todo esto tiene como efecto neto que los modelos, que supuestamente deben generar información empresarial útil y sólida, conducen a la adopción de decisiones subóptimas, a la pérdida de oportunidades y a la implantación de medidas inadecuadas. Pero existe una solución a este problema.



Pasar del concepto a la acción: cómo crear un entorno analítico eficaz

En un entorno analítico eficaz, que cuente con un buen almacenamiento y acceso a los datos y con una estructura correcta para su exploración y para su desarrollo, los modelos se efectúan con gran rapidez. Los modelos se generan y prueban con celeridad, y se implementan en un entorno de producción sin apenas demora. Los modelos en producción no tardan en generar resultados fiables y su rendimiento se supervisa constantemente para sustituir de inmediato los que muestran un funcionamiento deficiente por otros más actualizados.

En resumen, una estrategia analítica correcta entraña algo más que la mera creación de un modelo de gran capacidad predictiva: comprende la gestión holística de cada una de las fases del ciclo de vida tanto para cada uno de los modelos como para el conjunto completo de estos. Y no se trata de una tarea sencilla.

No olvidemos que los analistas y los científicos de datos no se limitan a desarrollar un único modelo para solucionar un problema empresarial. Elaboran un conjunto de modelos enfrentados y usan diversas técnicas para abordar problemas complejos. Contarán con varios de ellos en distintas fases del

Ilustración 2: SAS Data Loader incluye instrucciones para minimizar las necesidades de formación. Este ejemplo muestra la norma que permite unir y combinar datos, mediante la que puedes combinar dos tablas en una, y que se ejecuta en Hadoop para conseguir un mayor rendimiento.

Ilustración 3: Analiza los datos desde un entorno de visualización de datos interactivo.

desarrollo, además de con otros adaptados a líneas de productos y problemas empresariales específicos. De este modo, la organización puede llegar con facilidad a gestionar miles de modelos.

Y, más aún, el entorno de los modelos no tiene nada de estático: estos se actualizan continuamente conforme se prueban y nuevos datos pasan a estar disponibles. El objetivo consiste en generar los mejores modelos predictivos posibles con los datos óptimos disponibles.

Los modelos predictivos constituyen activos muy valiosos de las organizaciones, cuyo éxito requiere algo más que el mero uso de la tecnología. Las organizaciones también deben tener muy en cuenta tanto a las personas como a los procesos. Por ejemplo, es importante actualizar constantemente las habilidades analíticas de carácter empresarial y técnico para que puedan identificarse los problemas de negocio más acuciantes, además de aplicarse la información analítica pertinente a los procesos operativos.

El ciclo de vida analítico es sistemático y colaborativo por naturaleza. El personal, que posee distintos perfiles y habilidades, participe en varias etapas del proceso. Los responsables de negocio tienen que detectar el problema o la pregunta que quieren responder y, a continuación, tomar la decisión apropiada, además de supervisar los ingresos obtenidos a partir de dicha decisión. Los analistas de negocios llevan a cabo la visualización y el estudio de los datos y trabajan para identificar variables claves que influyan en los resultados. Los equipos de gestión de datos y del departamento de TI facilitan la preparación de los datos, además de la implementación y supervisión de los modelos. Los científicos de datos o responsables de la extracción de datos llevan a cabo tareas más complejas de análisis exploratorios, segmentación descriptiva y modelización predictiva.

Para obtener los mejores resultados analíticos, las organizaciones necesitan contratar personal con la cualificación adecuada y facilitarles el trabajo colaborativo para realizar sus funciones.

Cómo puede contribuir SAS al ciclo de vida analítico

SAS utiliza componentes integrados para reducir el tiempo de elaboración de los modelos, desde los datos hasta la gestión de las decisiones. Los procesos y las tecnologías consistentes que se emplean en el desarrollo y la implementación de los modelos reducen los riesgos inherentes a la modelización y, al mismo tiempo, fomentan la colaboración y el control entre los principales interesados de negocio y del departamento de TI.

Preparación y exploración de los datos: Un enfoque sistemático

- Preparación de los datos:** SAS® Data Management te permite definir perfiles y limpiar los datos, así como crear rutinas con procesos ELT (extraer, cargar y transformar) que generan los data marts analíticos, en los que se utilizan solo los datos necesarios. Los datos se almacenan en la base de datos para que se carguen de forma rápida, y se les de una estructura que facilite la creación de modelos y se simplifiquen para crear variables derivadas. Estos procesos se pueden automatizar y planificar en lotes, o bien se pueden ejecutar en tiempo real y ad hoc, en función de la fase del ciclo de vida analítico. Las herramientas de manipulación y preparación de datos self-service, como SAS Data Loader para Hadoop, ayudan a los analistas y a los científicos de datos a facilitar el acceso, combinando y limpiando datos sin que suponga una carga para el departamento de TI. El procesamiento secuencial de eventos de SAS implica que se

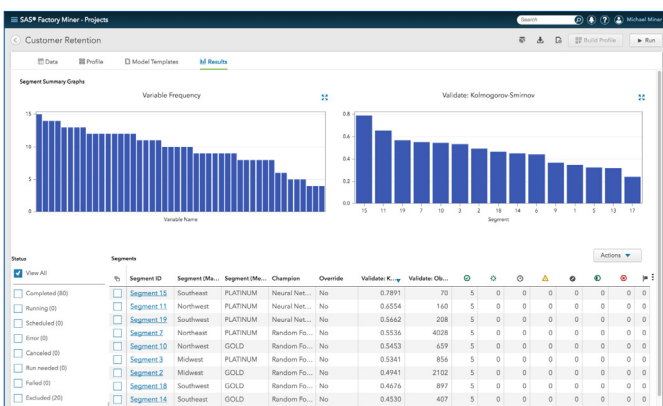


Ilustración 4: Técnicas de evaluación personalizables de SAS Factory Miner que te permiten generar los modelos ganadores para todos los segmentos

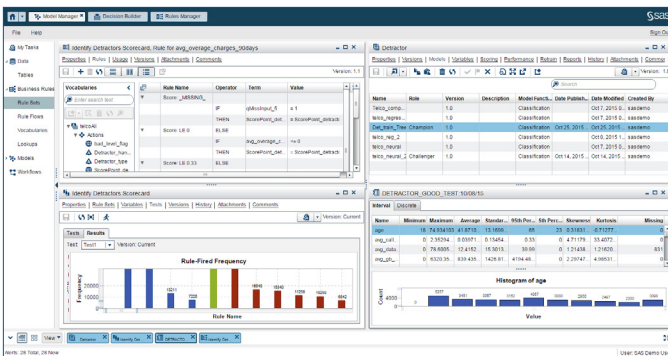


Ilustración 5: SAS Decision Manager agiliza proceso de implementación de los modelos. Integra la automatización del desarrollo de los modelos. SAS Factory Miner acelera las tareas manuales habituales, como la fijación de reglas de negocio y la terminología, que se genera automáticamente

puedan gestionar con la mayor eficiencia cientos de millones de eventos por segundo. Además, te ayuda a identificar qué aspectos requieren atención inmediata, cuales se pueden ignorar o cuales deben almacenarse. Finalmente, el procesamiento in data-based se usa para reducir el movimiento de datos y mejorar el rendimiento.

- **Exploración de datos:** SAS Visual Analytics permite que los analistas detecten con facilidad relaciones importantes entre datos y se centren rápidamente en ámbitos de interés o que presentan oportunidades, descubrir patrones inesperados, examinar distribuciones de datos, averiguar la frecuencia de valores extremos e identificar variables importantes para incorporar en el proceso de desarrollo de modelos.

Desarrollo de modelos mediante un entorno analítico innovador

Los analistas pueden crear modelos predictivos utilizando una variedad de herramientas de SAS que incluyen un amplio conjunto de algoritmos para analizar datos estructurados, así como no estructurados.

- **Comparaciones automáticas de los modelos machine-learning.** Gracias a SAS Factory Miner, consigues un entorno de modelización predictiva interactivo que facilita crear, modificar y evaluar cientos, o incluso miles, de modelos de forma muy rápida. Con unos cuantos clics, puedes acceder, modificar y transformar tu datos, elegir qué técnicas de machine-learning quieres aplicar y ejecutar los modelos en un entorno de comparación automatizada de modelos para identificar rápidamente el que mejor funciona.

Las plantillas pueden crearse directamente usando best practices que se pueden compartir con otros usuarios sin tener que escribir ninguna línea de código.

- **Modelización predictiva y minería de datos.** SAS® Enterprise Miner™ optimiza el proceso de minería de datos, de modo que puedas crear rápidamente modelos descriptivos y predictivos optimizados basados en grandes volúmenes de datos. Un entorno de trabajo interactivo y los flujos de proyecto pueden servir para documentar el trabajo y reducir el tiempo de desarrollo de los modelos. Los responsables del desarrollo de los modelos pueden usar funciones de evaluación visual y estadísticas de validación, así como comparar distintos modelos a la vez.
- **Minería de textos:** SAS Text Analytics proporciona una amplia gama de herramientas para detectar y extraer conceptos e información de las fuentes textuales, entre las que se incluyen la web, notas de call centers, libros, etc. Además, puedes añadir a tus modelos analíticos información de fuentes textuales para lograr una mayor capacidad predictiva.
- **Análisis en memoria:** El procesamiento de análisis en memoria distribuido en las soluciones de SAS aprovecha una potentes, fiable y escalable infraestructura analítica, que incluye bases de datos, como Pivotal Greenplum, Teradata, Oracle y SAP HANA, así como el hardware básico que usa Hadoop. Este procesamiento tan rápido permite que los modelos machine-learning y los modelos estadísticos se optimicen rápidamente bebiendo de bases de datos de gran tamaño.

Implementación de modelos desde un solo entorno fácil de usar

SAS Enterprise Decision Manager es la clave para mejorar la fase de implementación del ciclo de vida analítico. Proporciona un entorno compartido basado en web que permite gestionar el ciclo de vida y el control de tus activos de modelización y, además, se ajusta a las necesidades de los usuarios de negocio y del departamento de TI. Los analistas seleccionan datos y modelos de un repositorio central y, al mismo tiempo, pueden determinar las reglas de negocio que se aplican en sus modelos. Este marco de trabajo facilita el seguimiento de las actividades de modelización y las actualizaciones y las en un entorno único.

- **Registro de modelos:** Tras finalizar el desarrollo del modelo, los analistas registran un paquete que contiene el modelo, que incluye todas las transformaciones o imputaciones de datos etc., además de todos los resultados y documentos

relacionados. Este paquete garantiza que se han adoptado las medidas apropiadas y que el modelo lanzado a producción es potente y adecuado. También ayuda a las organizaciones a estandarizar el proceso para crear, gestionar, implementar y supervisar modelos analíticos.

- **Governance:** La gestión del riesgo de los modelos incluye el seguimiento de las métricas, el control de las diferentes versiones y quién ha cambiado qué cuando el control pasa de un departamento a otro, entre otros aspectos. Un repositorio centralizado de modelos, las plantillas de ciclos de vida y los controles de versión hacen que los procesos analíticos sean más transparentes y garantizan que se pueden auditar para cumplir con las normativas externas y de control interno.
- **Repositorio de modelos:** Los repositorios centrales sirven para almacenar documentación completa sobre el modelo, su scoring y los metadatos relacionados. Los responsables de la modelización pueden colaborar fácilmente y volver a usar el código de los modelos, gracias a las actividades que se quedan registradas a través de la autenticación de usuarios o grupos, los controles de versión y los controles de auditoría.
- **Calificación:** Después de revisar un modelo, aprobarlo e informar de que está listo para producción, este adquiere el estado de mejor modelo. Con un solo clic, todo el flujo de trabajo de tu modelo ganador se puede convertir en código de scoring que se puede implementar en SAS, bases de datos externas, Hadoop y servicios web.
- **Validación:** Los criterios de calificación se validan antes de elaborar los modelos, utilizando una plantilla y un proceso sistemáticos que permiten registrar todas las pruebas que el motor de calificación ha superado, para garantizar que los criterios integrados en el modelo ganador son adecuados.

¡En marcha! Pon en práctica los resultados de la modelización lo más rápido posible

Gracias a SAS, puedes elegir entre varias opciones de implementación para integrar los mejores modelos en los sistemas en producción. SAS Scoring Accelerator permite una implementación integrada y automatizada que puede impulsar el rendimiento de tus modelos de varias formas.

- **Implementación más rápida:** Implementa información y procesos analíticos con una infraestructura y un coste mínimos. Todo el código de scoring analítico se convierte de forma automática en servicios web ligeros o en lenguajes nativos para realizar el procesamiento en bases de datos. Por ello, desaparecen los procesos manuales, tediosos y propensos a errores con los que se traduce y se valida el código de scoring de los modelos para distintos entornos.

- **Implementación flexible:** La misma solución no se ajusta a todas las necesidades si nos referimos a los distintos sistemas operativos. Si no se tienen que implementar cambios, los mismos modelos se pueden implementar en sistemas batch y de tiempo real. Por ello, los analistas se desprenden de la pesada carga que supone adaptar cada entorno con distintos parámetros, códigos de scoring y códigos cifrados personalizados.
- **Procesos de scoring controlados mejores y más rápidos:** Puesto que el modelo se califica directamente en la base de datos, la tarea de ejecución de los modelos aprovecha la velocidad de procesamiento y escalabilidad de la base de datos. Las tareas que solían durar días y horas se pueden finalizar en cuestión de minutos o segundos. El movimiento de datos se minimiza dado que el scoring analítico puede llevarse a cabo donde se guardan los datos. Como los datos no tienen que extraerse del entorno de datos sumamente controlado y seguro. Incluso en entornos muchos menos gestionados, como Hadoop, si se elimina el movimiento y la duplicación de datos, las ventajas en el rendimiento pueden ser considerables.

Gestión y evaluación de modelos: Un proceso continuo

Una vez que un modelo se encuentra en un entorno de producción y se ejecuta para ofrecer respuestas, el modelo ganador se supervisa de forma centralizada a través de una serie de informes basados en indicadores clave de rendimiento. Si el rendimiento del modelo empieza a disminuir por debajo del nivel de aceptación, dicho modelo se puede sustituir rápidamente por uno nuevo o volver a ajustarlo.

- **Flujo de trabajo automatizado:** Las consolas de flujo de trabajo basadas en web fomentan que el proceso de gestión de modelos sea más automatizado, sistemático, colaborativo y controlado. Tanto la organización como los auditores pueden realizar el seguimiento de todas las fases de un proyecto de modelización, desde el planteamiento del problema hasta las etapas de desarrollo, implementación y retirada.
- **Gestión del ciclo de vida general:** Todas las fases del ciclo de vida de un modelo se coordinan de manera integral mediante plantillas prediseñadas y personalizadas por el cliente que se ajustan a los procesos de negocio de la organización.
- **Supervisión automatizada de los modelos:** Las tareas de control de rendimiento de los modelos se pueden planificar de forma periódica y los resultados se pueden supervisar de manera regular, tanto en el caso de un modelo como en todo el inventario de modelos. Estos resultados se pueden

publicar en dashboards, y fijar los umbrales mínimos de rendimiento. Si estos umbrales se incumplen, los analistas reciben avisos sobre los modelos que requieren su atención, ahorrando así dinero y tiempo.

Gracias a un marco de trabajo formal de gestión de modelos, los mejores modelos alcanzan la fase de producción de forma más rápida para empezar a satisfacer las necesidades de la empresa mucho antes. La compañía puede generar más y mejores modelos, con una amplia variedad de métodos analíticos y con menos recursos. Los modelos analíticos se supervisan y se mejoran continuamente para que estén actualizados y gocen de precisión. Todo el proceso de exploración e implementación se vuelve más transparente y mejor controlado, por lo tanto, resulta sencillo explicar las decisiones basadas en los análisis a las autoridades reguladoras y a los responsables de negocio.

Casos prácticos

Gracias a un enfoque de ciclo de vida analítico predictivo, el escenario “de después” y la forma de actuar cambia bastante, obteniéndose una ventaja competitiva considerable.

Institución financiera del Reino Unido: Modernizando su ciclo de vida analítico

Una de las principales instituciones financieras del Reino Unido se dio cuenta que la duración del ciclo desde la iniciación del modelo hasta su implementación no cumpliría con las expectativas del siglo XXI. El proceso era manual, propenso a errores y exigía numerosos recursos. Además, contaba con poca o ninguna supervisión a la hora de identificar el deterioro de los modelos.

Al trabajar con SAS y Teradata, la organización diseñó una plataforma de análisis predictiva y flexible en la que integró funciones para gestionar datos, desarrollar e implementar modelos utilizando tecnología en bases de datos. Esta plataforma aprovecha la escalabilidad del entorno de Teradata para calificar los modelos y utiliza la potencia de SAS Analytics para crearlos.

Gracias a la nueva plataforma, más de 55 millones de registros se pueden calificar en Teradata muchas veces durante el día; un resultado que jamás se habría logrado con el proceso anterior. El tiempo necesario para facilitar

que un modelo alcanzase el entorno de producción bajó de los tres meses a cuestión de días. La preparación de datos se redujo un 40 % y, por otro lado, la productividad de los analistas aumentó un 50 %.

Orlando Magic: La magia al descubierto

Gracias a SAS Analytics y SAS Data Management, Orlando Magic es uno de los equipos de la NBA que más ingresos genera.

El equipo lo consiguió analizando el mercado de reventa de entradas y fijando menores precios, previendo el número de titulares de abonos de temporada con riesgo de cancelar su suscripción (y para recuperarles de nuevo) y analizando las ventas de productos con el fin de asegurarse de que la organización pone a disposición de los aficionados todo lo que ellos quieren cada vez que entran al estadio. El club ha utilizado SAS incluso para ayudar a los entrenadores a reunir la mejor alineación.

Orlando Magic no tiene una bola de cristal, pero cuentan con SAS Enterprise Miner, que les permitió entender mejor los datos y desarrollar modelos analíticos que pronostiquen las renovaciones que llevarán a cabo los titulares de abonos de temporada. Gracias a las herramientas de minería de datos, el equipo generó scorings más precisos que contribuyeron a previsiones más exactas sobre su mercado y marcaron una diferencia con respecto a la forma de abordar la retención de clientes y el marketing.

Visa: Mil millones de decisiones operacionales al año mejoran la experiencia de los clientes y reducen el fraude

Visa, que es una marca reconocida a nivel mundial, facilita las transferencias electrónicas de fondos a través de productos de marca de terceros comercializados por sus miles de entidades financieras asociadas. La empresa ha tramitado 64 900 millones de transacciones en 2014, y se realizaron compras con un valor de 4,7 billones de dólares por medio de tarjetas Visa durante ese mismo año.

Visa cuenta con una capacidad de cálculo que es capaz de tramitar 56 000 mensajes de transacción por segundo, que supera cuatro veces la tasa de transacción máxima real hasta la fecha. Visa no solo tramita y calcula, sino que realiza análisis constantemente para proporcionar información estratégica y operativa a sus entidades asociadas y les ayuda a mejorar el rendimiento.

El objetivo de negocio está respaldado por un sistema de gestión de datos muy potente. Visa también ayuda a sus clientes a mejorar el rendimiento y les proporciona información analítica exhaustiva. “Examinamos los patrones de comportamiento mediante una agrupación y segmentación muy detallada, y proporcionamos esta información a nuestros clientes”, afirma Nathan Falkenberg, responsable de Visa Performance en el norte de Asia.

Hace muy poco tiempo Visa hizo una prueba de concepto con una solución de alto rendimiento de SAS que incluía almacenamiento en memoria para potenciar los algoritmos estadísticos y de machine-learning para posteriormente presentar dicha información de manera visual. Falkenberg declara: “El reto que se nos presenta, tal y como ocurre en cualquier empresa que pretende gestionar y usar grandes cantidades de datos, es cómo sacarle el mayor partido a todos esos datos para que nos ayuden a solucionar nuestros retos de negocio, ya sea mejorar los modelos de prevención del fraude, o bien ofrecer asistencia a nuestros clientes para comunicarse de forma más eficaz con sus propios clientes”.

[Más información](#)

Recursos adicionales

Para obtener más información sobre la fase de detección del ciclo de vida analítico, lea el informe técnico [La extracción de datos de principio a fin: Cómo conseguir información y generar mejores oportunidades](#).

Para obtener más información sobre la fase de implementación, lea el informe técnico [De los datos a la decisión: La automatización de las decisiones referentes a las operaciones con SAS Decision Manager](#).

Aspectos diferenciales de SAS*

- Preparación de datos para el análisis, exploración y Discovery interactivo, gestión, control y desarrollo de modelos; todo ello, en una plataforma integrada.
- Integración de distintas disciplinas analíticas (por ejemplo, minería de datos, machine learning, forecasting, análisis y optimización de textos) que permite lograr una mayor agilidad en el desarrollo de modelos y solucionar problemas empresariales más complejos.
- Automatización inteligente, así como plantillas de machine learning y de modelización predictiva con mejores prácticas recomendadas integradas, que proporcionan herramientas de productividad y machine learning a un amplio grupo de usuarios.
- Con SAS, puedes gestionar un gran número de modelos complejos con un número de variables prácticamente ilimitado y bases de datos de gran tamaño.
- El entorno de comparación de modelos automatizado y basado en la web fomenta la productividad al permitir que los responsables de la modelización prueben de forma rápida y sencilla muchos enfoques a la vez utilizando los algoritmos estadísticos y de machine learning.
- Gracias a las funciones de flujo de trabajo basadas en la web, los usuarios pueden fácilmente determinar procesos personalizados, gestionarlos hasta que finalicen, fomentar la colaboración mediante notificaciones y establecer estándares empresariales.
- El seguimiento de la modelización desde la fuente de datos hasta los resultados analíticos proporciona un control esencial, que resulta clave para cumplir las normativas o los estrictos requisitos de informes.
- Dashboards gráficos e intuitivos facilitan el seguimiento del rendimiento de los modelos en distintos proyectos, de modo que los equipos puedan centrarse en aquellos que requieren su atención de manera inmediata y evitar el deterioro de modelos.
- La interoperabilidad con herramientas de modelización externas permite a las organizaciones importar, gestionar y supervisar en un repositorio central los activos de modelización creados por SAS y otras herramientas (por ejemplo, modelos PMML, R).
- Las bases de datos más comunes, como Teradata, Aster Data, EMC Greenplum, IBM Netezza, IBM DB2, Oracle y Teradata, permiten incorporar funciones de calificación en bases de datos.
- Los metadatos de SAS engloban el análisis y la gestión de datos para que las transformaciones de datos utilizadas en la fase de exploración se puedan volver a utilizar en la de implementación.
- SAS proporciona conocimientos técnicos y empresariales muy amplios durante la asistencia anterior y posterior a la venta que te ayuda a reducir el tiempo a la hora de sacar más partido al retorno de la inversión y mejorarlo.

To contact your local SAS office, please visit: sas.com/offices