

La Minería de Datos de la A a la Z:

Cómo Descubrir Conocimientos y Crear Mejores Oportunidades

Contenido

Introducción	1
El Ciclo de Vida Analítico de SAS®: Combinando Datos, el Descubrimiento y la Implementación.....	2
¿Qué Puede Ayudarle a Descubrir la Minería de Datos?	3
Un Vistazo al ROI de la Minería de Datos en el Proceso de Descubrimiento	5
Paso 1: Convierta una Pregunta de Negocio en una Hipótesis Analítica.....	5
Paso 2: Prepare los Datos para la Minería de Datos	5
Paso 3: Explore los Datos	5
Paso 4: Modele los Datos	6
Soluciones de Minería de Datos de SAS®	6
Usando SAS® Enterprise Miner™ para la Minería de Datos y Machine Learning	6
Usando SAS® Factory Miner para la Automatización de la Minería de Datos.....	8
Escalando su Proceso de Descubrimiento para Manejar Big Data y los Problemas Complejos .	8
La Integración Facilita la Implementación de Modelos, el Monitoreo y la Gestión	10
Conclusión.....	11
Más Información.....	11

Introducción

Hay tantos datos y una gran cantidad de decisiones que tomar. Las organizaciones de todo el mundo se están enfrentando a este dilema. Los datos están creciendo, pero ¿y su capacidad para tomar decisiones de acuerdo con esos enormes volúmenes de datos? ¿También están creciendo? Para muchos, desafortunadamente, la respuesta es 'NO'.

Los datos fluyen a velocidades y volúmenes nunca antes vistos, y de todas partes. Pero tomar decisiones basadas en hechos no depende de la cantidad de datos que uno tenga. De hecho, tener tantos datos puede ser un obstáculo. ¿Por dónde comenzar? Su éxito dependerá de lo rápido que pueda descubrir conocimientos en todos esos datos y utilizar dichos conocimientos para llevar a cabo mejores acciones dentro de su organización.

Es ahí donde la analítica predictiva, la minería de datos, el machine learning y la gestión de decisiones entran en acción. La analítica predictiva ayuda a evaluar lo que sucederá en el futuro. La minería de datos (Data Mining) busca los patrones ocultos en los datos que pueden utilizarse para predecir el comportamiento futuro. Las empresas, los científicos y los gobiernos han utilizado este enfoque por años para transformar los datos en conocimientos proactivos. La gestión de decisiones convierte esos conocimientos en acciones que se utilizan en sus procesos operativos. De modo que mientras se puedan seguir aplicando hoy los mismos enfoques, necesitan suceder más rápidamente y a una mayor escala, utilizando las técnicas más modernas disponibles actualmente.

Las organizaciones innovadoras utilizan la minería de datos y la analítica predictiva para detectar los fraudes y los problemas de seguridad cibernética, para gestionar los riesgos, anticipar las demandas de recursos, aumentar las tasas de respuesta de las campañas de marketing, generar las próximas mejores ofertas, reducir el abandono de los clientes e identificar efectos adversos que tendrían los medicamentos durante las pruebas clínicas, entre otras muchas cosas.

Debido a que pueden producir conocimientos predictivos a partir de datos diversos y de gran volumen, las metodologías de minería de datos, el machine learning y el modelado analítico avanzado son esenciales para identificar los factores que pueden mejorar el desempeño organizacional y, cuando se automatiza en las decisiones de todos los días, crear una ventaja competitiva. Y hoy que hay mucho más de todo (datos, poder de cómputo, preguntas de negocio, riesgos y consumidores), la capacidad de aumentar su poder analítico es esencial para adelantarse a su competencia.

Implementar rápidamente los conocimientos analíticos asegura que la conveniencia de sus modelos analíticos no se pierda debido a procesos lentos como son reescribir el código para cada entorno, revalidar los modelos reescritos y otros procesos manuales. Si usted puede implementar rápidamente sus modelos analíticos, el contexto y la relevancia de los modelos no se pierde, usted puede conservar su ventaja competitiva.

Por lo tanto, ¿cómo crea usted un entorno que pueda ayudarle a su organización a lidiar con todos los datos que se están recolectando, con todos los modelos que se están creando, y con todas las decisiones que necesitan tomarse, todo a una mayor escala? La respuesta es un ciclo de vida analítico iterativo que reúna:

- **Datos** - el fundamento de las decisiones.
- **Descubrimiento** - el proceso de identificar nuevos conocimientos en los datos.
- **Implementación** - el proceso de utilizar conocimientos recientemente encontrados para impulsar acciones mejoradas.

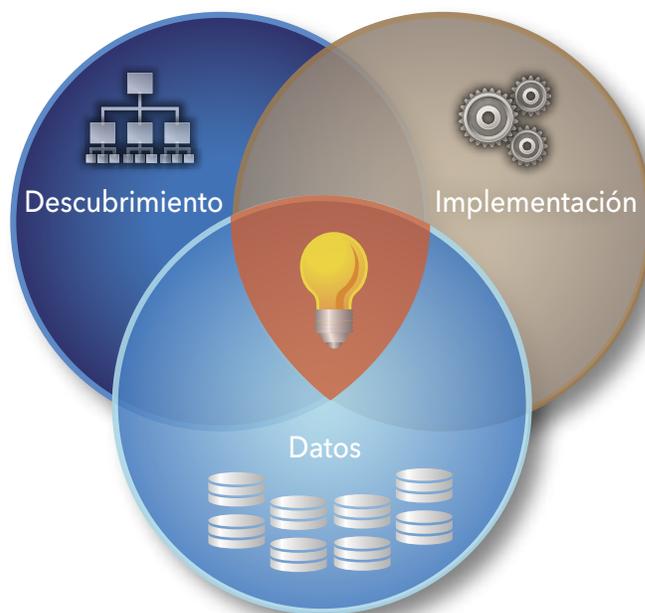


Figura 1: Se necesita una combinación integrada de datos, descubrimiento e implementación para impulsar y poner en acción los conocimientos necesarios para tomar decisiones escalables en poco tiempo.

El Ciclo de Vida Analítico de SAS®: Combinando Datos, el Descubrimiento y la Implementación

Si bien gran parte de este documento se enfoca en utilizar la minería de datos para descubrir conocimientos, echemos un vistazo al ciclo de vida analítico iterativo completo, porque eso es lo que hace posible el descubrimiento predictivo y le da más valor a las acciones que de él se derivan.

- **Haga una pregunta de negocio.** Todo comienza aquí. El proceso de descubrimiento se da al hacer preguntas de negocio que impulsen la innovación. El primer paso se enfoca en explorar lo que usted necesita saber, cómo puede aplicar la analítica predictiva a sus datos para resolver un problema o mejorar un proceso.
- **Prepáre los datos.** Actualmente recolectar datos no es un problema - provienen de todos lados. Tecnologías como Hadoop y computadoras más rápidas y más accesibles han permitido almacenar y utilizar más datos, y más tipos de datos, que nunca antes. Pero aún persiste el problema de unir los datos en diferentes formas de diferentes fuentes y la necesidad de transformar los datos brutos en datos que puedan utilizarse como una entrada a la minería de datos. Los científicos de datos aún pasan mucho de su tiempo lidiando con estas tareas.
- **Explorar los datos.** Las herramientas de visualización interactivas necesitan servir a una amplia gama de usuarios en una organización (desde el analista de negocio sin conocimientos analíticos hasta un científico de datos) para permitir hacer búsquedas de relaciones, tendencias y patrones para entender más a fondo la información capturada por variables en los datos. En este paso, la hipótesis formulada en la fase inicial del proyecto se refinará y las ideas de cómo afrontar el problema de negocio desde una perspectiva analítica se desarrollan y se prueban. Mientras examina sus datos, puede surgir la necesidad de crear, seleccionar y transformar algunos datos para crear modelos enfocados con mayor precisión. Las herramientas interactivas veloces ayudan a hacer de éste un proceso interactivo, el cual es crucial para identificar las mejores preguntas y respuestas.
- **Modele los datos.** En esta etapa, el científico de datos aplica varios algoritmos de modelado analítico a los datos para identificar representaciones robustas de las relaciones de los datos que ayuden a responder a la pregunta de negocio. Las herramientas analíticas buscan una combinación de datos y de técnicas de modelado que predigan de manera confiable un resultado deseado. Para encontrar la respuesta más confiable, la clave es la experimentación, y crear un modelo automático puede ayudar a reducir el tiempo para obtener los resultados y aumentar la productividad de los equipos analíticos. En el pasado, con las herramientas manuales para crear modelos, los mineros de datos y los científicos de datos podían crear varios modelos en una semana o en un mes. Actualmente, pueden crear cientos o incluso miles. Pero ¿cómo pueden encontrar de manera rápida y confiable un modelo (de muchos) que tenga el mejor desempeño? Con torneos automatizados de algoritmos de machine-learning y un modelo campeón claramente definido, esto se ha convertido en un proceso sencillo. Los analistas y los científicos de datos ahora pueden invertir su tiempo en preguntas e investigaciones más estratégicas.
- **Implemente modelos.** Aquí pasamos de la fase de descubrimiento a la implementación - tomando los conocimientos aprendidos y poniéndolos en acción usando procesos automáticos y repetibles. En muchas organizaciones este es el punto en el que a menudo el proceso se hace considerablemente lento debido a que no hay un choque de manos definido entre el mundo del descubrimiento y el mundo de la implementación, mucho menos con la automatización. Unir estos mundos para crear una transición integrada ayuda a reducir el tiempo para obtener valor para la analítica predictiva. Cuanto más rápido su empresa pueda utilizar las respuestas generadas por la analítica predictiva para tomar mejores decisiones, se generará más valor. Y, un proceso transparente es importante para todos - especialmente para los auditores.
- **Utilice la nueva información.** Existen dos tipos de decisiones que pueden tomarse de acuerdo con los resultados analíticos. Las decisiones estratégicas son tomadas por los humanos que observan los resultados y actúan. Las decisiones operativas son automáticas - como las calificaciones de crédito o las mejores ofertas recomendadas - y no requieren de la intervención humana. Son cada vez más las organizaciones que buscan automatizar las decisiones operativas y brindar respuestas y resultados en tiempo real para reducir las latencias de las decisiones. Basar las decisiones operativas en las respuestas de los modelos analíticos también hace a las decisiones objetivas, repetibles y cuantificables. La integración con herramientas de gestión de decisiones empresariales permite a las organizaciones crear flujos de decisiones operativas completos que combinen la analítica impulsada por los datos y las reglas de negocio para decisiones automáticas óptimas.
- **Evalúe los resultados.** El siguiente paso -y tal vez el más importante- es evaluar el resultado de las acciones que produjo el modelo analítico. ¿Sus modelos predictivos produjeron resultados tangibles, como un aumento de los ingresos o una reducción de los costos? Con el monitoreo y medición continuos del desempeño de los modelos, usted puede evaluar el éxito de estos recursos y asegurarse de seguir produciendo los resultados deseados.

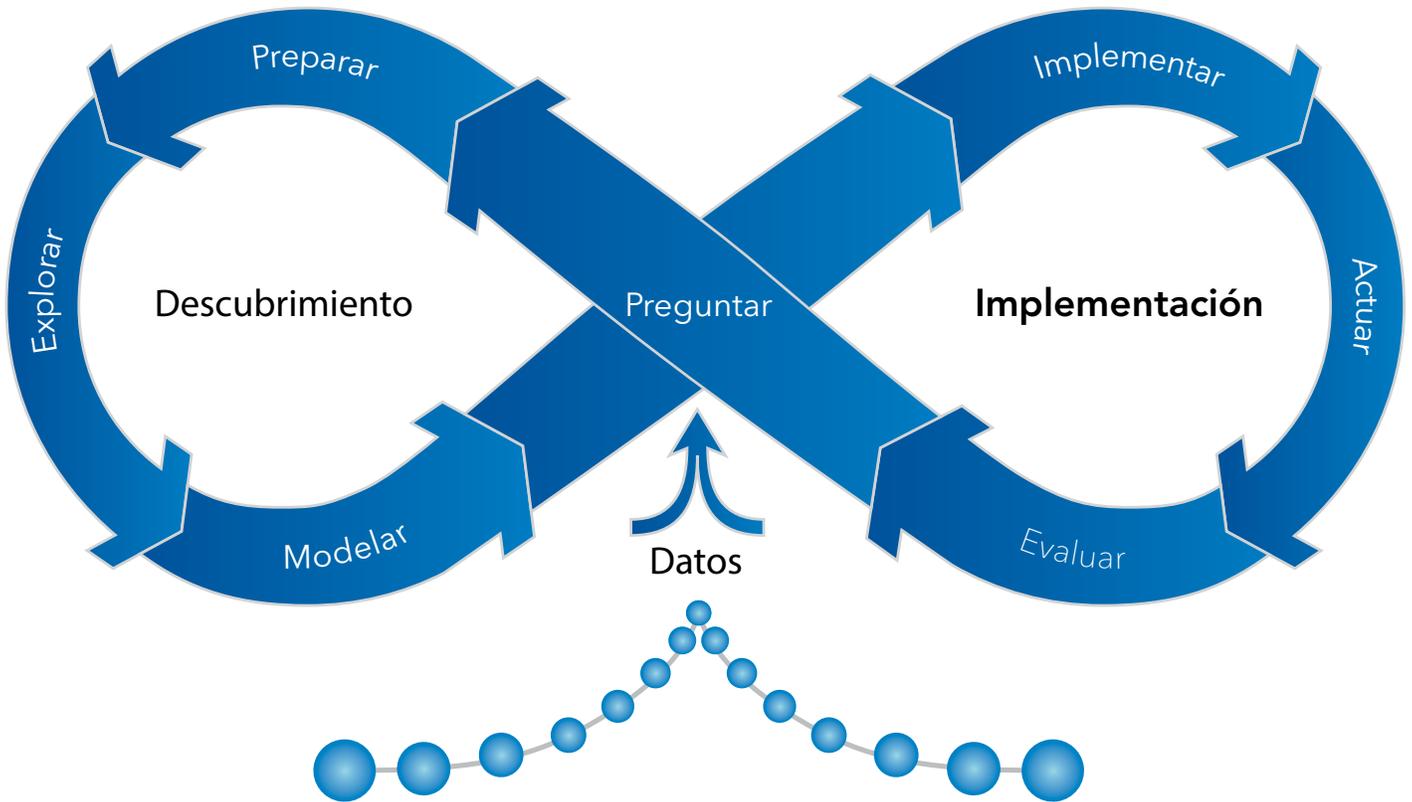


Figura 2: El ciclo de vida analítico es un proceso iterativo de hacer descubrimientos en sus datos y aplicar nuevos conocimientos para mejorar continuamente los modelos predictivos y sus resultados.

- **Pregunte una vez más.** Debido a que los datos están creciendo y cambiando continuamente, las relaciones de los datos que utilizan sus modelos para las predicciones también cambian con el tiempo. La evaluación constante de sus resultados analíticos identificará la degradación de la precisión de los modelos. Incluso los modelos más precisos tendrán que renovarse con el tiempo, y las organizaciones necesitarán pasar por los pasos de descubrimiento e implementación una vez más. Es un proceso constante y en evolución.

SAS ofrece una plataforma analítica completa e integrada que maneja cada paso del ciclo de vida analítico iterativo. Este es recordatorio de que el documento se enfocará en la parte del descubrimiento de datos del ciclo de vida - y las herramientas de minería de datos que usted necesitará para crear rápidamente los modelos predictivos más precisos posible.

¿Qué Puede Ayudarle a Descubrir la Minería de Datos?

La minería de datos ofrece una serie de tecnologías que ayudan a las organizaciones a anticipar resultados futuros, descubrir nuevas oportunidades y mejorar el desempeño del negocio. Puede aplicarse a una variedad de problemas de los clientes en todas las industrias - desde la segmentación de clientes y la detección de fraudes y la calificación de riesgo de crédito, hasta identificar los efectos adversos de un medicamento durante las pruebas clínicas.

Un uso común de la minería de datos y de las técnicas de machine-learning es la segmentación automática de los clientes por comportamiento, demografías o actitudes- para entender mejor las necesidades de grupos específicos y atenderlos de una manera más dirigida. Esta segmentación analítica, o modelado sin supervisión, ayuda a identificar a grupos de clientes que son similares y que podrían reaccionar a ciertas ofertas o actividades de manera parecida.

Utilizando estos segmentos, usted puede crear modelos para cada grupo para predecir la siguiente mejor oferta o actividad a la cual probablemente responderán. Para asegurar que usted sólo llega a los clientes que quiere, puede complementar el modelo de adquisición con un modelo de calificación de riesgos para averiguar quién es un buen riesgo de crédito y que vale la inversión para adquirirlo o retenerlo.

Otro uso importante para la minería de datos y machine-learning es ayudar a detectar los fraudes, lo que es importante a medida que los defraudadores desarrollan tácticas más sofisticadas. Pueden construirse modelos para cruzar datos de una amplia variedad de fuentes, correlacionar variables no obvias con características conocidas para identificar nuevos patrones de actividades fraudulentas.

Debido a su potencial de producir conocimientos predictivos precisos de enormes volúmenes de datos diversos, la minería de datos ha demostrado ser un componente valioso de muchas iniciativas analíticas. La minería de datos y el machine learning pueden ayudarle a:

- Descubrir automáticamente patrones, tendencias y relaciones representadas en los datos
- Desarrollar modelos para entender y describir mejor las características y actividades basadas en estos patrones.
- Utilizar esos conocimientos para ayudar a evaluar las opciones futuras y tomar decisiones basadas en hechos.
- Crear código de calificación que exprese los cálculos hechos para acciones adecuadas oportunas.

Aplicaciones Comunes para la Minería de Datos en las Industrias

Pregunta de Negocio	Aplicación	¿Qué se predice?
¿Cómo dirigir mejor las ofertas de productos/servicios?	Perfiles y segmentación.	Los comportamientos de los clientes y sus necesidades por segmento.
¿Qué producto/servicio recomendar?	Venta cruzada y venta directa.	Compras probables de los clientes.
¿Cómo aumentar los clientes valiosos y retenerlos?	Adquisición y retención.	Preferencias de los clientes y patrones de compra.
¿Cómo dirigir la oferta correcta a la persona correcta en el momento correcto?	Gestión de campañas.	El éxito de las comunicaciones con los clientes.
¿En qué clientes invertir y cómo atraerlos mejor?	Rentabilidad y valor de vida.	Factores de valor futuro (margen y retención).

Aplicaciones de Minería de Datos para Industrias Específicas

Pregunta de Negocio	Aplicación	¿Qué se predice?
¿Cómo evaluar y controlar el riesgo dentro del portafolio de consumo actual (o nuevo)?	Calificación de crédito (banca).	Solvencia de grupos de clientes nuevos y existentes.
¿Cómo incrementar las ventas con ventas directas/cruzadas, programas de lealtad y promociones?	Sistemas de recomendación (retail en línea).	Productos que probablemente se compren la siguiente vez.
¿Cómo reducir las interrupciones operativas y los costos de mantenimiento?	Mantenimiento de activos (servicios, manufactura, petróleo y gas).	Factores reales de falla de activos o equipo.
¿Cómo reducir los costos de salud y satisfacer a los pacientes?	Gestión de salud y condiciones (seguros de salud).	Pacientes en riesgo de sufrir enfermedades crónicas, tratables/prevenibles.
¿Cómo reducir las pérdidas por fraudes y disminuir los positivos falsos?	Gestión de fraudes y seguridad informática (gobierno, seguros, bancos).	Casos de fraude desconocidos y riesgos futuros.
¿Cómo llevar rápidamente los medicamentos al mercado y de forma efectiva?	Descubrimiento de medicinas (ciencias biológicas).	Compuestos que tienen los efectos deseados.

Un Vistazo al ROI de la Minería de Datos en el Proceso de Descubrimiento

La minería de datos y machine learning se encuentran en el centro del proceso de descubrimiento. Pero hay más por descubrir que sólo crear un modelo analítico. Usted obtiene mejores resultados si adopta un enfoque holístico iterativo.

Paso 1: Convierta una Pregunta de Negocio en una Hipótesis Analítica

El primer paso del proceso de descubrimiento es hacer una pregunta de negocio (ver las tablas de la página 4). Normalmente, una organización tiene una idea general de lo que quiere lograr – algo como, “Queremos reducir el abandono de nuestros clientes valiosos”. Para enfrentar estos problemas con la analítica, las preguntas deben ser detalladas o transformarse en hipótesis analíticas. Por ejemplo, cada modelo predictivo requiere de un resultado bien definido, una etiqueta o un objetivo. Si usted quiere predecir el abandono de los clientes, necesita definirlo como un resultado para el modelo. Sin embargo, probablemente el abandono es definido de manera distinta en diferentes organizaciones. ¿Se refiere a alguien que cancela activamente un contrato o a alguien que no tiene ninguna actividad? ¿Cuánto tiempo tiene que permanecer inactivo un cliente antes de ser clasificado como alguien que potencialmente se alejará de la compañía? ¿Qué es valioso? ¿Incluimos sólo el valor histórico o el valor futuro potencial (el valor de vida) de un cliente? El primer paso del proceso de descubrimiento es identificar un problema y traducirlo en una pregunta que pueda ser respondida con la analítica.

Paso 2: Prepare los Datos para la Minería de Datos

Para comenzar, usted debe determinar qué datos son necesarios para responder a la pregunta. De acuerdo con los detalles específicos de la pregunta de negocio, un analista evalúa los datos que están disponibles y decide si éstos tienen el potencial de responder a la pregunta. Si no, podrían necesitarse datos externos o que se recolecten datos nuevos. A menudo, los datos se encuentran en diferentes sistemas y es necesario tener acceso a ellos y convertirlos en un conjunto de datos que pueda usarse para la minería de datos y machine learning. Los modelos predictivos o supervisados requieren de un solo registro por entidad para modelarse. (Una tabla base analítica para hacer predicciones o para el análisis del mercado lucirá distinta a la tabla para el modelado predictivo o supervisado). Si usted desea modelar la probabilidad de que un cliente lo abandone, necesitará crear una sola tabla donde cada registro contenga todos los atributos de datos para un solo cliente. Esto a menudo requiere gran cantidad de agregación y transformación de datos. Una vez que se ha agregado una sola tabla base analítica para el análisis, los otros aspectos del ciclo de vida entran en acción. Debido a que es necesario experimentar con los datos, la etapa de preparación también es muy iterativa, y donde el analista prueba diferentes tipos de datos para obtener los resultados predictivos más precisos.

En algunos casos para acelerar los procesos de modelado, usted podría necesitar tomar una muestra de los datos – esto es, crear un subconjunto más pequeño de los datos que representan al conjunto de datos objetivo. La minería de datos solamente puede descubrir los patrones que ya están presentes en los datos, por lo que la muestra debe ser representativa y lo suficientemente grande para contener la información importante. La tabla base de analítica también se divide generalmente en por lo menos dos conjuntos: el de capacitación y el conjunto de prueba. El primero se utiliza para capacitar a el algoritmo (s) de minería de datos y de machine-learning, al tiempo que el segundo se utiliza para verificar la precisión de los patrones encontrados.

Paso 3: Explore los Datos

Enseguida, usted querrá explorar los datos y buscar relaciones anticipadas, tendencias no anticipadas y anomalías para entender la información con la que está trabajando y refinar más las ideas y las preguntas. Asimismo, la exploración de datos puede ayudar a identificar los problemas de calidad de los datos como los errores, valores faltantes o distribuciones de datos que necesitan transformarse para la etapa de modelado. Además, usted puede usar varios otros tipos de técnicas para detectar patrones en los datos que puedan ayudarle a crear modelos predictivos más precisos o ayudarle a crear datos de entrada adicionales para su modelo predictivo.

- **Clustering (o modelado no supervisado)** identifica grupos o estructuras en los datos que sean similares, más allá de las estructuras visibles en los datos.
- **El aprendizaje por asociación** de reglas busca relaciones entre variables, como productos que se compran juntos con frecuencia (conocido como análisis de la canasta del mercado), que puede llevar a mayores recomendaciones de compra.
- **La analítica de textos** puede ayudarle a crear nueva información estructurada a partir de datos de texto electrónicos. Estos nuevos datos pueden ayudar a mejorar la precisión de sus modelos. Por ejemplo, integrar los comentarios de los clientes sobre sus productos o servicios de las anotaciones hechas en el centro de atención telefónica (call center) o de las revisiones en los foros de medios sociales a menudo produce modelos de predicción de abandono más precisos.
- **La visualización interactiva** de datos presenta los resultados de forma gráfica y le permite a los usuarios interactuar con éstas gráficas para identificar más fácilmente los patrones importantes o las anomalías con los datos que podrían tener un impacto en la fase de creación de modelos.

Con frecuencia usted necesitará modificar sus datos antes de hacer el modelado, de modo que debe planear un paso para crear, seleccionar y transformar variables para enfocarse en su proceso de selección de modelos. De acuerdo con sus descubrimientos en la fase de exploración, usted podría necesitar manipular sus datos para introducir nuevas variables, llenar los valores faltantes o buscar valores atípicos de modo que usted puede reducir el número de variables a sólo los más importantes.

Paso 4: Modele los Datos

Después de explorar y preparar cuidadosamente sus datos de entrada, usted está listo para crear modelos predictivos o supervisados para buscar una combinación de los datos que predigan de modo confiable un resultado deseado. Dependiendo de los datos y los problemas que se tengan, usted puede escoger de una variedad de técnicas modernas estadísticas y de machine-learning para resolver su problema - incluyendo la clasificación, la regresión, las redes neuronales, los bosques aleatorios, la máquina de vectores de soporte, la respuesta gradual o la minería de datos por series de tiempo - así como técnicas de industrias específicas como la calificación de crédito en banca o primas para seguros.

La selección de las técnicas más adecuadas depende de varios factores: ¿Es importante tener un modelo que prediga su resultado deseado con la mayor precisión o es también (o incluso más) importante tener transparencia en las relaciones de datos que impulsan las predicciones? Las técnicas de machine-learning automática a menudo son demasiado complejas para permitir la exploración de los factores de negocio a partir de los resultados de los modelos, mientras que otras técnicas estadísticas como la regresión y los árboles de decisiones son más transparentes y las industrias reguladas las prefieren.

Para obtener al valor máximo de sus modelos predictivos, usted querrá evaluar constantemente la utilidad y la confiabilidad de los hallazgos de sus procesos de minería de datos. No todos los patrones encontrados por los algoritmos de la minería de datos serán válidos. Los algoritmos podrían encontrar patrones en el conjunto de datos de capacitación que no están presentes en el conjunto de datos general. (A esto se le llama sobreajuste -overfitting). Con el fin de solucionar esta preocupación, los patrones son validados con un conjunto de datos de prueba. Los patrones aprendidos en los datos de capacitación se aplicarán al conjunto de prueba, y el resultado se compara con el resultado deseado (o conocido).

Por ejemplo, un algoritmo de minería de datos que ha sido capacitado para distinguir transacciones con tarjetas de crédito fraudulentas de las legítimas se aplicaría al conjunto de prueba de transacciones en las que no había sido capacitado. La precisión de los patrones puede entonces medirse a partir de cuántas transacciones con tarjetas de crédito se clasificaron correctamente. Si los patrones aprendidos no cumplen con los estándares deseados, se hacen modificaciones a las técnicas de pre-procesamiento y de minería de datos hasta que el resultado sea satisfactorio y los patrones aprendidos puedan aplicarse exitosamente a los sistemas operativos.

Los científicos de datos y los mineros de datos necesitan experimentar con varios algoritmos de modelado predictivo y de machine-learning con el fin de encontrar uno que funcione mejor para su problema específico. Los torneos de modelado automático donde los usuarios

pueden experimentar para identificar rápidamente la estrategia de modelado ganadora pueden ayudar a ahorrar mucho tiempo. Cuando usted está satisfecho con los resultados de sus esfuerzos de modelado, entonces comienza el proceso de implementación. Pero ya que es un proceso completamente iterativo, hay exámenes y ajustes constantes. Como se dijo anteriormente, el proceso de desarrollo sigue varios pasos (vea la sección Ciclo de Vida Analítico de SAS en la página 2). Para consultar más información sobre el proceso de implementación, lea [Desde los Datos hasta la Decisión: Cómo SAS® Decision Manager Automatiza las Decisiones Operativas](#). Para saber más sobre minería de datos y el descubrimiento, ¡siga leyendo!

Soluciones SAS® de Minería de Datos

La minería de datos y el machine learning le permiten descubrir conocimientos que mejoran la toma de decisiones. Con las soluciones de minería de datos de SAS, usted puede agilizar el proceso de descubrimiento para desarrollar modelos más rápidamente para que pueda entender las relaciones clave y encontrar los patrones que más importan.

Usando SAS® Enterprise Miner™ para la Minería de Datos y Machine Learning

SAS Enterprise Miner es un banco de trabajo gráfico completo para la minería de datos. Esta extensa plataforma ampliamente aclamada ofrece capacidades para preparar los datos para la analítica predictiva, identificar las variables más importantes, desarrollar modelos usando los algoritmos más modernos de minería de datos y machine-learning, validar fácilmente la precisión y aptitud del modelo(s), y generar activos que permiten una implementación sencilla de los modelos analíticos en sus aplicaciones operativas para la toma automática de decisiones.

Las poderosas herramientas de preparación de datos solucionan los problemas de calidad, como valores faltantes y valores atípicos, y le ayudan a desarrollar reglas de segmentación. La exploración interactiva de datos les permite a los usuarios crear gráficos dinámicos vinculados para identificar relaciones dentro de los datos. SAS Enterprise Miner ofrece docenas de algoritmos estadísticos avanzados y de machine-learning para el modelado descriptivo y predictivo, incluyendo clustering, análisis de vínculos y de la canasta de mercado, análisis de los principales componentes, árboles de decisiones, potenciación y empaquetado, redes bayesianas, redes neuronales, bosques aleatorios, regresión lineal, regresión logística, máquina de vectores de apoyo, minería de datos de series de tiempo y mucho más.

Al final de la línea de desarrollo de modelos, se entrega código de calificación completo y optimizado para la implementación sencilla de modelos no supervisados y supervisados en SAS, C, Java y PMML

para calificar datos en SAS así como en otros entornos. El código de calificación también se entrega automáticamente como una función en la base de datos para calificar dentro de Hadoop así como en bases de datos líderes de la industria como Teradata, IBM, Oracle, Pivotal, Aster Data, SAP HANA, etc., para cada interacción con aplicaciones de negocio y rápidos resultados operativos.

Además de generar código de calificación en diferentes lenguajes y formatos, SAS Enterprise Miner también genera muchos activos que permiten la implementación, gestión y monitoreo sencillos de modelos predictivos como parte de los procesos de negocio operativos. Todos estos activos son soportados por metadatos para brindar la documentación importante alrededor del proceso completo.

El proceso de minería de datos de SAS Enterprise Miner es guiado por un diagrama de flujo de procesos que usted puede modificar, guardar y compartir. La GUI de arrastrar y colocar le permite a los analistas con poca experiencia estadística navegar por el proceso de minería de datos, mientras que el experto cuantitativo puede afinar los modelos analíticos en segundo plano.

Con SAS Enterprise Miner, usted puede:

- Crear series de datos de capacitación y prueba de muestra con un alto valor predictivo.
- Explorar interactivamente las relaciones y anomalías de los datos.
- Crear, transformar y seleccionar las variables más adecuadas para el análisis.
- Aplicar un rango de técnicas de modelado para identificar los patrones en los datos.
- Validar la utilidad y confiabilidad de los hallazgos del proceso de minería de datos.
- Crear todos los activos necesarios para la implementación, monitoreo y gestión de los modelos.

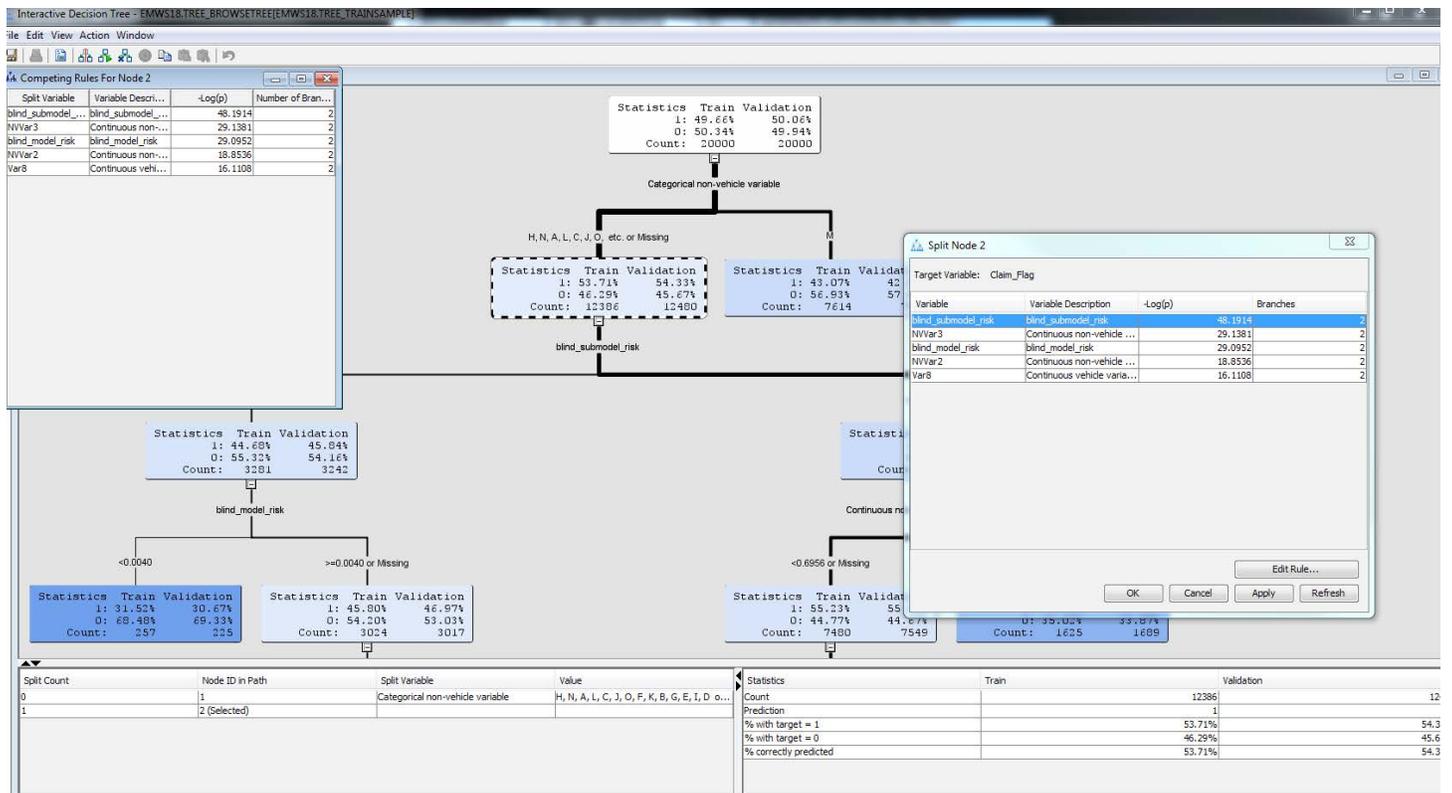


Figura 3: Los árboles de decisiones son sólo una de las muchas técnicas de modelado que se incluyen con SAS Enterprise Miner. Pueden desarrollarse interactivamente o en un modo en lote. Varias gráficas de evaluación ayudan a medir la estabilidad general de los árboles.

Usando SAS® Factory Miner para la Automatización de la Minería de Datos

A medida que las organizaciones aplican analíticas más dirigidas a sus crecientes segmentos de clientes y de negocio, hay una necesidad de crear modelos aún más predictivos a niveles más granulares. Por ejemplo, en lugar de desarrollar un modelo para la base de clientes completa, los departamentos de marketing quieren crear modelos específicos para muchos segmentos de clientes. Una tienda tal vez quiera desarrollar modelos de ventas cruzadas para un gran número de categorías de productos. O una compañía de transporte querrá crear modelos de mantenimiento predictivos para diferentes componentes de los vehículos que tiene en operación. Y si bien esto hace necesario crear muchos más modelos, la mayoría de los analistas y científicos de datos no tienen mucho tiempo para eso.

Con SAS Factory Miner, usted tiene un entorno de modelado predictivo interactivo que hace extremadamente sencillo crear, modificar y evaluar cientos, o incluso miles, de modelos muy rápidamente. Con solo unos clics, usted puede acceder, modificar y transformar sus datos, elegir qué técnicas de machine-learning quiere aplicar y ejecutar los modelos en un entorno de torneo automático de modelos para identificar rápidamente al mejor para cada segmento. Las técnicas de modelado incluidas en SAS Factory Miner son:

- Redes bayesianas.
- Árboles de decisiones.
- Potenciación de gradientes.
- Redes neuronales.
- Bosques aleatorios.
- Máquinas de vectores de soporte.
- Modelos lineales generalizados
- Regresión Logística

Los usuarios pueden identificar fácilmente las excepciones de modelado (segmentos en los que el enfoque automático no genera modelos que cumplan con los criterios de aceptación). El diseño de caja blanca de SAS Factory Miner permite que los usuarios modifiquen fácilmente las líneas de modelado predictivo y afinen los parámetros de los componentes de las líneas para mejores resultados cuando es necesario. Pueden incluso crear sus propias líneas de modelado a la medida para sus proyectos analíticos favoritos, incluyendo la preparación de datos, ingeniería de características y algoritmos de selección y de aprendizaje, y compartílos con otros usuarios para crear un repositorio de mejores prácticas organizacionales. Esta colaboración en toda la organización puede ayudar a ampliar la reserva de talento académico en su organización. El científico de datos actúa como el productor de las líneas de modelado de mejores prácticas organizacionales para

Con SAS Factory Miner, usted puede:

- Potenciar la productividad del descubrimiento.
- Automatizar el desarrollo de modelos.
- Explorar nuevas ideas más rápidamente.
- Colaborar con sus colegas analíticos en su organización.
- Ampliar su reserva de talento analítico a través del machine-learning de autoservicio automático.
- Poner grandes portafolios de modelos predictivos en producción de forma más eficiente y gestionarlos con facilidad.

diferentes proyectos y otros usuarios del entorno consumen estas mejores prácticas bajo un modelo de autoservicio para obtener resultados óptimos.

Y SAS Factory Miner no se detiene con la identificación de un modelo campeón para cada segmento. El código completo se crea automáticamente para toda la línea de calificación (incluyendo las transformaciones de los datos) de cada modelo para implementarse en SAS o en otros entornos, como las bases de datos o Hadoop.

Además, todos los activos de desarrollo de modelos y de calificación pueden registrarse en SAS Decision Manager, un entorno centralizado basado en la Web para gestionar el ciclo de vida y el gobierno de sus activos de modelado de SAS o de otros proveedores, incluyendo la analítica de código abierto.

La automatización, la facilidad de uso, la escalabilidad y las capacidades de colaboración de SAS Factory Miner aumentan su poder para crear modelos predictivos, incrementar la productividad de su personal analítico, permitir la colaboración de los equipos analíticos dispersos, así como ampliar su reserva de talento analítico a través de la democratización de las técnicas de machine-learning.

Escalando su Proceso de Descubrimiento para Manejar Big Data y los Problemas Complejos

Big data y los problemas complejos exigen grandes soluciones analíticas. En SAS, ampliamos su poder de descubrimiento con analítica en-memoria distribuida. La idea es sencilla pero poderosa. Divida sus datos en piezas más pequeñas y distribuya el volumen de los datos y la complejidad de los problemas entre sus motores de cómputo, ya sea en una sola máquina con varios núcleos de procesamiento (CPUs) o en una red de computadoras, como

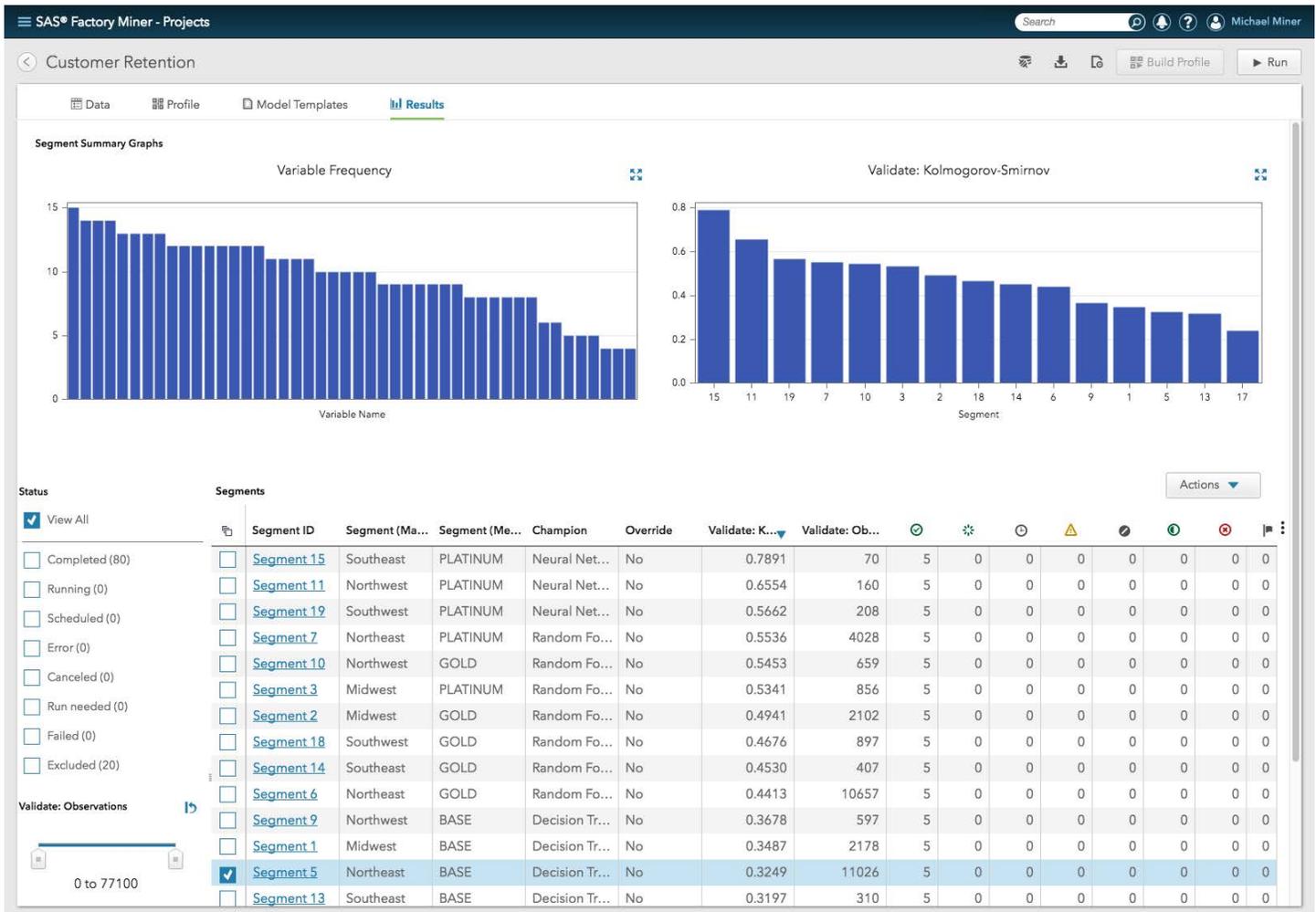


Figura 4: Las técnicas de evaluación personalizables de SAS Factory Miner le permiten generar modelos campeones para cada segmento de sus datos.

un cluster Hadoop. El proceso se realiza completamente en la memoria cada vez que es posible, incluyendo la comunicación entre las unidades de procesamiento (CPUs), lo que hace a este proceso realmente rápido.

El proceso analítico en-memoria distribuido de SAS aprovecha una infraestructura analítica altamente escalable y confiable - incluyendo dispositivos de bases de datos como Pivotal Gremlin, Teradata, Oracle y SAP HANA - y hardware estándar que utilice Hadoop de código abierto, o Hadoop Cloudera y distribuciones de Hortonworks. Para los usuarios, no cambia mucho. Pueden trabajar desde la misma interfaz conocida para sus proyectos de minería de datos, de analítica predictiva y de machine-learning, mientras que SAS In-Memory Analytics se encarga de la distribución óptima de cargas de trabajo en el sistema disponible.

SAS High-Performance Data Mining le permite analizar grandes volúmenes de datos diversos usando una interfaz de arrastrar y

colocar, así como poderosos métodos descriptivos, predictivos y de machine-learning. Una variedad de técnicas de modelado - incluyendo bosques aleatorios, máquinas de vectores de soporte, redes neuronales y clustering - se combina con la preparación de datos, la exploración de datos y las capacidades de calificación. Gracias a que usted puede crear y ejecutar más modelos más rápidamente, puede hacer más preguntas y traer nuevas ideas a su proceso de minería de datos. SAS High-Performance Text Mining le permite obtener conocimientos rápidos de los grandes conjuntos de datos no estructurados que implican millones de documentos, correos electrónicos, notas, fragmentos de reportes, fuentes de medios sociales, etc. Se incluye el soporte para el análisis, la extracción de entidades, derivación automática y detección de sinónimos, descubrimiento de temas y la descomposición de valores singulares (SVD). Los resultados de la minería de datos pueden utilizarse como entradas en la minería de datos de alto desempeño para mejorar el poder del modelado predictivo.

La Integración Facilita la Implementación de Modelos, el Monitoreo y la Gestión

Si bien este documento se enfoca en la minería de datos y en el proceso de descubrimiento analítico, usted realmente no puede terminar una conversación sobre minería de datos y machine learning para las aplicaciones de negocio sin abordar lo que sucede después de que se crean los modelos predictivos y se eligen los modelos campeones. Entonces, ¿qué sucede? Usted pasa a la fase de implementación (vea la Figura 2).

Después de que se ha seleccionado el modelo campeón, tiene que implementarse en el entorno de producción correcto. Las organizaciones utilizan modelos predictivos de diferentes maneras. Por ejemplo, podrían utilizarse para seleccionar a clientes para las campañas de marketing al ejecutar un proceso de calificación en lote y entregar a marketing una lista de los clientes seleccionados. Un número cada vez mayor de organizaciones están recurriendo a procesos más integrados y automáticos para hacer que los resultados de los modelos predictivos estén disponibles para tomar decisiones operativas. En lugar de hacer que el proceso de calificación se ejecute en lote, les gustaría que el modelo dé las respuestas bajo demanda como parte de una aplicación de negocio. Las organizaciones también podrían querer respuestas en tiempo real del flujo de datos (por ejemplo, para la detección automática de fraudes o el mantenimiento predictivo).

Además, se utilizan reglas de negocio junto con los modelos analíticos para tomar decisiones más flexibles y ágiles. Con SAS Decision Manager, las reglas de negocio ayudan a definir las acciones de acuerdo con las condiciones específicas de los procesos de negocio.

Anteriormente, TI era el encargado de realizar manualmente la implementación de un modelo predictivo en el entorno de producción, lo que a menudo provocaba grandes retrasos antes de que el modelo pudiera utilizarse. Con las condiciones del mercado en constante cambio y la llegada continua de nuevos datos, es posible que los modelos se vuelvan obsoletos antes de que siquiera sean implementados. Con la integración perfecta de las fases de descubrimiento y de implementación del ciclo de vida analítico, SAS permite que las organizaciones automaticen este proceso. SAS Decision Manager ofrece una interfaz ágil para implementar modelos en los entornos de ejecución en tiempo real o en lote sin volver a codificar los modelos para diferentes entornos. Esto maximiza la inversión en la analítica mediante la reutilización de los activos analíticos en los entornos y reduce los riesgos al eliminar la necesidad de realizar la recodificación manual y la posterior revalidación: desarrolle una vez, implemente varias veces.

Las organizaciones innovadoras están encontrando nuevas formas de ser más eficientes y tomar mejores decisiones automáticas. SAS Decision Manager ofrece las funcionalidades que las organizaciones necesitan para la implementación de modelos más rápida y más sencilla en las situaciones de producción.

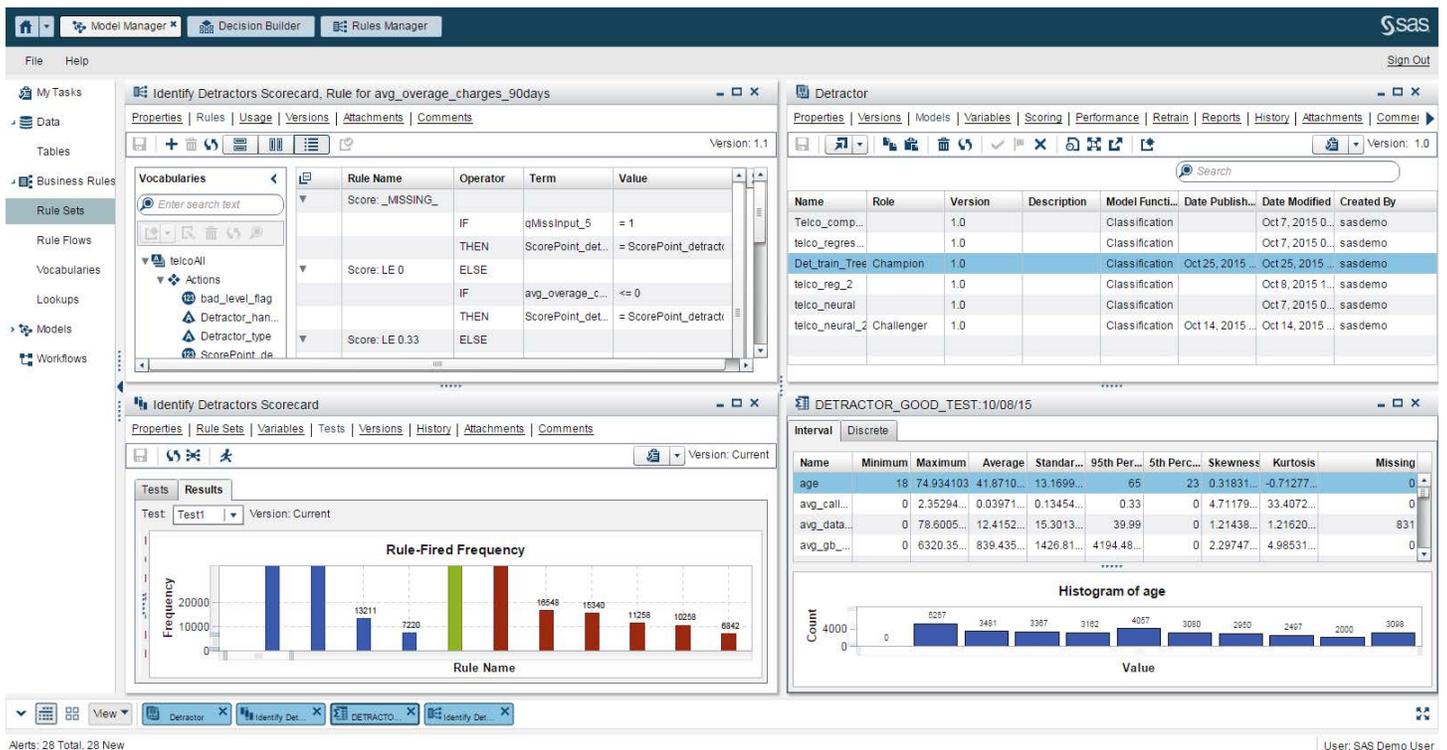


Figura 5: SAS Decision Manager ayuda a acelerar el proceso de implementación de modelos. Integra la automatización del desarrollo de modelos con SAS Factory Miner y acelera las tareas manuales comunes, como la definición de reglas de negocio y la generación automática de vocabularios.

Conclusión

Actualmente, más organizaciones están reconociendo el valor de los resultados de la analítica predictiva. Y eso es bueno porque si usted está recolectando y almacenando datos, debería estar usándolos para obtener el conocimiento que le dará una ventaja competitiva.

¡En especial si su organización le está pagando a la gente para crear modelos analíticos! Pero el truco siempre ha sido tener todas las diferentes piezas y las partes que se mueven juntas para extraer el valor máximo de todos sus datos. SAS ofrece un proceso completo del ciclo de vida analítico que ayuda a las organizaciones a ir de los datos a las decisiones a una escala muy grande, de una forma muy confiable.

Comienza con el acceso a los datos y su preparación (los volúmenes de datos no importan), pasa por el proceso de descubrimiento de datos y de modelado analítico para producir conocimientos predictivos, y continúa con la implementación y la gestión de los resultados - todo en un entorno integrado.

Aunque este documento presentó todas las fases del ciclo de vida analítico, su principal enfoque fue la parte del descubrimiento. Y en SAS, el descubrimiento significa usar la analítica predictiva para encontrar conocimientos nuevos y confiables de los datos de una manera rápida y sencilla. Con el software de minería de datos reconocido por la industria como SAS Enterprise Miner, la nueva solución SAS Factory Miner, las tecnologías en memoria y las capacidades de gestión de modelos empresariales, las organizaciones son capaces de enfrentar cualquier problema de analítica de big data.

- SAS Factory Miner ofrece una solución automática basada en la Web para crear y retener modelos predictivos en múltiples segmentos. Potencia la productividad al permitir que los modeladores prueben rápida y fácilmente varios enfoques simultáneamente usando algoritmos de machine learning y estadísticos.
- En situaciones en las que el modelado automático no funciona, SAS Enterprise Miner puede utilizarse para hacer a la medida modelos predictivos avanzados estratégicos.
- El cómputo en-memoria distribuido mantiene al procesamiento moviéndose a máximas velocidades.
- SAS Decision Manager agiliza la implementación de modelos analíticos - todo desde una sola interfaz.

Estas soluciones agilizan el proceso de descubrimiento de datos/ minería de datos, permitiéndole crear modelos predictivos y descriptivos altamente precisos basados en el análisis de datos en su empresa.

Más Información

Visite sas.com/datamining para consultar más información sobre nuestras soluciones de minería de datos y de descubrimiento de datos.

Únase a la [SAS Data Mining Community](#), donde los usuarios y los empleados de SAS comparten consejos y otra información.

Para tener un panorama completo del ciclo de vida analítico, lea [Gestione el Ciclo de Vida Analítico para la Innovación Continua](#).

Para consultar más información sobre la fase de implementación, lea [Desde los Datos hasta la Decisión: Cómo SAS® Decision Manager Automatiza las Decisiones Operativas](#).

Para contactar a SAS México, visite por favor: sas.com/mexico

SAS y el resto de los nombres de productos y servicios de SAS Institute Inc. son marcas registradas o marcas comerciales de SAS Institute Inc. en Estados Unidos y otros países. ® indica el registro en Estados Unidos. Otros nombres de marcas y productos son marcas registradas de sus respectivas compañías. Copyright© 2015, SAS Institute Inc. Todos los derechos reservados. 104937_S149733.0116

