



Funciones destacadas:

Haciendo uso del proceso estándar de Cross-Industry Standard Process for Data Mining (CRISP-DM) para la minería de datos, podrá evaluar los requisitos de la herramienta de minería de datos de su empresa.

Evaluación de una herramienta de minería de datos

Comience con una planificación de su evaluación

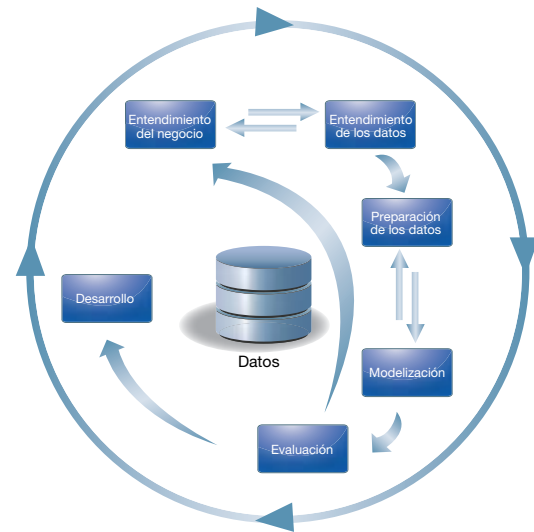
La minería de datos le ayuda a tomar mejores decisiones que permiten obtener resultados concretos y significativos, así como un aumento de los ingresos y procesos más eficaces. Aunque las capacidades de la minería de datos puedan ser espectaculares, algunos de los mitos difundidos sobre la minería de datos indican que pueden obtenerse increíbles resultados con el mínimo esfuerzo. Si se elige una herramienta de minería de datos con esta expectativa, puede que no se llegue al esperado retorno de la inversión.

Para obtener los resultados que espera de su proyecto de minería de datos, calcule la situación de su negocio y evalúe el rendimiento de las herramientas a lo largo de todo el proceso. Para ayudarle en su evaluación, se ha recopilado la siguiente lista de comprobación mediante CRISP-DM, Cross-Industry Standard Process for Data Mining.

CRISP-DM organiza el proceso de minería de datos en seis fases, tal y como se muestra en el diagrama de la página siguiente. La secuencia de las fases no es estricta, ya que siempre necesitará moverse entre las distintas fases. Tenga en cuenta este aspecto interactivo de la minería de datos en la evaluación de la herramienta. La herramienta que seleccione debe ser lo suficientemente flexible para hacer cambios en cualquier fase mientras trabaja en un proyecto.

Los siguientes criterios de evaluación se organizan en estas seis fases, con algunas consideraciones adicionales al final.





La fase 1 le ayuda a definir los objetivos de su proyecto y a crear un plan preliminar de minería de datos.

Fase 1 de CRISP-DM: Comprensión del sector comercial

La fase inicial se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva comercial y, a continuación, en la conversión de este conocimiento en la definición del problema de minería de datos y en un plan preliminar. Los elementos de la lista de comprobación indican si la herramienta tiene un historial con una gran variedad de problemas comerciales y cómo vincula la herramienta la comprensión comercial con los aspectos técnicos del proceso de minería de datos.

- ¿Tiene la herramienta un registro demostrado de soluciones de una gran variedad de problemas comerciales, incluyendo los problemas a los que nos enfrentamos?
 - ¿Se ha demostrado que la herramienta sea útil solucionando problemas en nuestro sector o con nuestra aplicación? ¿Tiene la herramienta un historial satisfactorio con las aplicaciones que queremos realizar?
- ¿Cómo vincula la herramienta la comprensión comercial con el aspecto técnico de la minería de datos?
 - ¿Pueden asignarse claramente los pasos para utilizar la herramienta a las necesidades de la minería de datos? ¿Se presentan de manera clara al usuario comercial los conceptos de la minería de datos? ¿Cómo se integra la herramienta con la gestión de proyectos u otras herramientas de planificación? ¿Requiere la herramienta aplicaciones escritas para unir este espacio?

En la fase 2 comparará cómo se integran las posibles herramientas de minería de datos con su entorno de tecnologías actual para habilitar las funciones de recopilación de datos y visualización.

En la fase 3 se evaluará cómo preparan los datos distintas herramientas de minería de datos para obtener un análisis óptimo.

Fase 2 de CRISP-DM: Comprensión de los datos

La fase de comprensión de los datos empieza con la recopilación de datos y continúa a través de actividades para familiarizarse con los datos. Los estándares abiertos para acceder a los datos y a las técnicas de visualización son consideraciones importantes a tener en cuenta en esta fase.

- ¿Cómo preserva la herramienta mi inversión en la infraestructura de TI?
 - ¿Funciona bien la herramienta de minería de datos con nuestros almacenes de datos existentes? ¿Funciona la herramienta con los estándares de datos abiertos o se necesita migrar los datos a un formato propio?
- ¿Cómo habilita la herramienta la visualización y la exploración interactiva de los datos?
 - ¿Ofrece la herramienta de minería de datos técnicas de visualización para ayudar a ver los patrones de los datos? ¿Puede realizarse la visualización de manera interactiva realizando cambios en los gráficos y creando nuevos gráficos según los diferentes aspectos de los datos?

Fase 3 de CRISP-DM: Preparación de los datos

La fase de preparación de los datos cubre todas las actividades para crear el conjunto de datos final a partir de los datos iniciales sin procesar. Los elementos de la lista de comprobación incluyen la eficacia y la facilidad de la preparación de los datos.

- ¿Cómo prepara la herramienta los datos?
 - ¿Se distribuye todo el proceso de minería de datos interactivo, incluyendo las actividades necesarias para preparar los datos sin procesar hasta que estos estén listos para crear modelos, para obtener un análisis en profundidad de los datos? ¿Presenta la herramienta pasos de preparación de datos de forma sencilla?
- ¿Puede la herramienta extraer datos automáticamente para su preparación?
 - ¿Puede la herramienta extraer datos automáticamente o se requiere un trabajo manual para escribir consultas SQL para las uniones, agregaciones, ordenaciones y otras operaciones de preparación de datos?

Una modelización eficaz es un componente clave de la minería de datos. La fase 4 le ayuda a examinar qué herramientas de minería de datos funcionan mejor con distintas técnicas de modelización.

Fase 4 de CRISP-DM: Modelización

En esta fase, se seleccionan y aplican varias técnicas de modelización y se calibran los parámetros hasta valores óptimos. A menudo, los analistas vuelven a la fase de preparación de datos para cumplir los requisitos de distintos tipos de modelos. Debido a que distintos modelos dirigen el mismo tipo de problemas de minería de datos, los elementos clave de la lista de comprobación incluyen el nivel de competencia de la herramienta mediante la aplicación y comparación de técnicas diferentes.

- ¿Cómo fomenta la herramienta la productividad de los analistas?
 - ¿Permite la herramienta que los analistas desarrollen modelos eficaces rápidamente? ¿Con qué facilidad pueden los usuarios probar distintos modelos para lograr la mejor solución? ¿Con qué facilidad puede realizarse la preparación de los datos para reunir las necesidades específicas del modelo?
- ¿Ofrece la herramienta una amplia gama de técnicas?
 - ¿Ofrece la herramienta técnicas o algoritmos para la visualización, clasificación, conglomerado, asociación y regresión?
- ¿Se pueden combinar técnicas con la herramienta?
 - ¿Se pueden combinar fácilmente distintas técnicas para obtener mejores resultados? ¿Se pueden incorporar los resultados de los algoritmos al conjunto de datos para el posterior procesamiento y el análisis?
- ¿Preserva la herramienta mi inversión existente en tecnología al igual que los algoritmos y otras herramientas?
 - ¿Puede funcionar la herramienta de minería de datos con los algoritmos existentes? ¿Cómo funciona la herramienta de minería de datos con las demás herramientas de análisis de datos?

En la fase número 5, la fase de evaluación, se evaluarán los distintos modelos de datos para garantizar que se han tratado todos los problemas comerciales.

Fase 5 de CRISP-DM: Evaluación

La fase de evaluación consiste en una evaluación a fondo del modelo o modelos antes de su distribución. Un objetivo clave es determinar si existe un problema comercial importante que no se haya considerado lo suficiente. Los elementos de la lista de comprobación para la fase de evaluación hacen referencia a la forma en la que se han incorporado las entradas de los usuarios comerciales al modelo y a cómo se reflejan los resultados a los destinatarios deseados.

- ¿Logra la herramienta de manera constante altos resultados?
 - ¿Crea su herramienta de minería de datos soluciones que funcionan constantemente a un alto nivel o sólo proporciona buenos resultados en determinados casos o con algunos tipos de datos? ¿Reflejan sus resultados de manera precisa todos los problemas comerciales y, por tanto, funcionan bien en los datos de prueba?
- ¿Ofrece la herramienta resultados fáciles de comprender?
 - ¿Comprenden fácilmente los usuarios comerciales sus resultados? Si no es así, ¿cuáles son los pasos necesarios a seguir para que los resultados sean convincentes? ¿Ha fomentado la herramienta la aportación de comentarios procedentes de expertos comerciales a través de todo el proceso de minería de datos?
- ¿Se puede aplicar todo el ámbito de visualización para validar los resultados de un modelo?
 - ¿Se pueden analizar fácilmente las puntuaciones y las predicciones del modelo, entre otros resultados, para validar el rendimiento de este?

La fase 6 de la lista de comprobación de CRISP-DM le ayuda a evaluar cómo se pueden distribuir de manera sencilla y eficaz las soluciones de minería de datos.

Fase 6 de CRISP-DM: Distribución

En la fase de distribución se utiliza el proceso de minería de datos, ya se ofrezca simplemente una perspectiva de un problema comercial o se implemente el proceso en una aplicación para ofrecer un conocimiento actualizado para los usuarios de información. A menudo, la distribución requiere amplios servicios, de ahí que el elemento clave de la lista de comprobación trate con la capacidad que tiene la herramienta para ayudar con esta tarea.

- ¿Cómo puedo distribuir mis soluciones de minería de datos (tanto ahora como en el futuro)?
 - ¿Cómo se pueden integrar mis soluciones de minería de datos en aplicaciones operativas? ¿Puede realizarse la integración de manera rentable o requerirá una inversión considerable tanto en coste como en tiempo? ¿Con qué facilidad pueden actualizarse mis soluciones? Si las soluciones no pueden actualizarse fácilmente, ¿en qué cantidad de trabajo y costes se incurrirá?

Consideraciones adicionales: Coste de la propiedad

Además de las fases de la minería de datos del modelo del proceso de CRISP-DM, necesitará realizar un análisis del retorno de la inversión.

- ¿Cuál es el coste de propiedad?
 - Cuantifique los costes de propiedad sobre la vida del producto y la oferta del servicio, incluyendo los sistemas complementarios necesarios. Cuantifique el retorno anticipado. ¿Cuándo se espera un retorno positivo de la inversión? ¿Cumple este resultado los objetivos comerciales?
- ¿Qué es el tiempo de implementación?
 - ¿Cuánto tardará en implementar su herramienta de minería de datos? ¿Requiere otras herramientas o hardware? ¿Qué cantidad de formación, asesoría y personalización se necesita para obtener resultados del proyecto?
- ¿Se corresponden las habilidades de nuestros usuarios (ahora y en el futuro) con las habilidades necesarias para utilizar la herramienta?
 - ¿Qué habilidades se necesitan para obtener resultados útiles con la herramienta de minería de datos? ¿Es una herramienta diseñada para usuarios con más experiencia técnica, para principiantes en minería de datos o puede utilizarse por ambos grupos y usuarios con habilidades intermedias? ¿Qué costes de formación tendré que soportar para que todo el mundo alcance una gran productividad? No olvide considerar los conjuntos de habilidades que tendrán los usuarios potenciales en el futuro.
- ¿Pueden nuestros usuarios personalizar la herramienta?
 - ¿Cómo se puede personalizar la herramienta para hacer diferentes usos de ella? ¿Pueden guardarse los procesos comunes para su reutilización? ¿Puede la herramienta automatizar tareas? ¿Están disponibles los servicios para personalizar interfaces u ofrecer otro tipo de ayuda?

Al determinar la eficacia de una herramienta de minería de datos, tenga en cuenta el coste total de propiedad, el retorno de la inversión y las oportunidades de asociación con futuros vendedores.

Consideraciones adicionales: Vendedor

Por último, considere la importancia del vendedor en la evaluación de su herramienta de minería de datos. La mayoría de las veces, la adquisición de una solución de minería de datos es una inversión de futuro y necesitará un compañero de confianza para el camino que le espera por delante.

- ¿Ofrece el vendedor otras herramientas y servicios que puedan ayudar a solucionar problemas similares?
 - ¿Ofrece el vendedor otras herramientas de minería o de análisis de datos? ¿Están la asesoría, la formación, la asistencia técnica, así como otros servicios disponibles? ¿Están disponibles estos servicios en una base mundial?
- ¿Será el vendedor una fuente de confianza que pueda actualizar el software y los servicios que cumplan con mis necesidades?
 - ¿Es el vendedor un líder en la oferta de soluciones de minería de datos? ¿Cuenta el vendedor con los recursos necesarios para continuar ofreciendo un servicio de alto nivel en el futuro?

Acerca de IBM Business Analytics

El software IBM Business Analytics ofrece información completa, coherente y precisa en la que los órganos de toma de decisiones confían para mejorar el rendimiento comercial. Un conjunto integral de inteligencia empresarial, análisis avanzado, rendimiento comercial y gestión de estrategias, así como de aplicaciones de análisis predictivo le ofrece una perspectiva clara, inmediata e interactiva del rendimiento actual y la capacidad para predecir resultados futuros.

Como parte de este portafolio, el software de Análisis Predictivo de IBM SPSS ayuda a las organizaciones a predecir eventos futuros y actuar proactivamente según esa información para llegar a mejores resultados de negocio. Clientes del entorno comercial, gubernamental y académico de todo el mundo confían en la tecnología de IBM SPSS como una ventaja competitiva para atraer, retener y aumentar los beneficios con sus clientes, a la vez que reducen el fraude y mitigan los riesgos. Al incorporar el software de IBM SPSS en sus operaciones diarias, las organizaciones se convierten en empresas predictivas – capaces de direccionar y automatizar decisiones para cumplir con los objetivos de negocio y conseguir una ventaja competitiva apreciable. Para más información, visite www.ibm.com/spss/es.



© Copyright IBM Corporation 2010

IBM Corporation
Route 100
Somers, NY 10589

Derechos restringidos de usuarios del Gobierno de EE.UU. - El uso o duplicación de revelación está restringido por el GSA ADP Schedule Contract con IBM Corp.

Creado en Estados Unidos de América
Mayo de 2010
Reservados todos los derechos

IBM y el logotipo de IBM, ibm.com, WebSphere, InfoSphere y Cognos son marcas comerciales de International Business Machines Corporation en EE.UU., otros países o ambos. Si estos u otros términos con marca comercial de IBM están indicados en su primera aparición en esta información con un símbolo de marca comercial (® o TM), dichos símbolos indican una marca comercial registrada en EE. UU. propiedad de IBM en el momento de la publicación de esta información. Dichas marcas comerciales también pueden estar registradas en otros países. Existe una lista actualizada de marcas comerciales de IBM bajo "Copyright and trademark information" en www.ibm.com/legal/copytrade.shtml.

SPSS es una marca comercial de SPSS, Inc., an IBM Company, registrada en muchas jurisdicciones de todo el mundo.

Otros nombres de empresas, productos y servicios pueden ser marcas comerciales o marcas de servicio de terceros.



Por Favor Recicle
