

MLA702 Gestión de datos 1: Introducción a la gestión de datos

El contenido de este curso es propiedad de Marine Learning Alliance (MLA) y NO DEBE compartirlo con nadie por ningún medio, sin el permiso previo de MLA. El contenido provisto en este módulo se proporciona ÚNICAMENTE PARA SU USO INDIVIDUAL. Todas las imágenes pictóricas y diagramas utilizados en esta conferencia han sido creados por MLA o están libres de derechos de autor, a menos que se indique lo contrario. Cuando MLA no haya creado una imagen, se otorga el reconocimiento apropiado al pie de este documento.

Pantalla 1 – Título

Bienvenido a la primera de una serie de conferencias que exploran el tema de la gestión de datos

Pantalla 2 – Resultados del aprendizaje

Explicar qué se constituye como datos; reconocer que toman muchas formas

Definir qué es la administración de datos y por qué es necesaria

Reconocer qué datos son buenos y cuáles son datos incorrectos, y cómo las operaciones se ven afectadas por ambos

Pantalla 3 – Introducción y esquema

Cuando hablamos de datos, ¿qué queremos decir realmente y cómo se supone que debemos gestionarlos con éxito? Como era de esperar, ambos temas recibirán una gran atención, no solo en esta conferencia sino en todo momento. Por cierto, esta conferencia establece el escenario de las conferencias restantes de la serie, así que no se preocupe si está interesado en descubrir más sobre un tema dado presentado aquí, ya que destacaré los lugares donde se puede esperar encontrar una mayor cantidad de detalles en los lugares apropiados.

Lo que vamos a ver aquí es nuestra capacidad para detectar cuándo estamos analizando los datos correctos y las ocasiones en que no lo estamos, lo que es de importancia considerable y cuándo necesita confiar en los datos para tomar decisiones operativas informadas. En definitiva, este es nuestro objetivo aquí. Necesitamos una nueva apreciación de la importancia de los datos para nuestra vida profesional y tal vez personal. Como veremos a continuación, los datos deben considerarse como el producto más importante en su negocio. Muy poco se puede lograr sin ellos, pero se puede perder mucho si no tomamos una precaución adecuada al administrarlos.

Pantalla 4 – Datos definidos

Solo como recordatorio, es posible que no supiera que la palabra *datos* (*data*) en inglés, aunque parece singular, es plural. Tenga cuidado con esto en la escritura científica, donde verá frases escritas como 'los datos **son** buenos' en oposición a 'los datos **es** bueno'. Cuando se conversa en inglés, es común hablar de datos en forma singular y, a medida que el lenguaje evoluciona, esto es aceptable, pero es probable que se corrija al escribir datos si no usa la forma plural. Hablando

de los tiempos modernos, como sin duda sabe, hemos experimentado una especie de explosión en el crecimiento de la tecnología en los últimos tiempos. Los datos rodean casi todas las actividades.

Pero, ¿qué son exactamente los datos? Bueno, es posible que haya oído hablar del término bit y byte. Un bit es la unidad fundamental de datos, y contiene dos piezas de información, por ejemplo, sí y no. Un byte consta de ocho bits, por lo que estamos hablando de un total de 256 elementos de información, ya que 2^8 equivale a 256. Una oración o párrafo corto puede equipararse a esto; y puede verlo por sí mismo: quizás escriba algún texto en un editor de texto y ve cómo el tamaño del archivo cambia con la cantidad de texto que incluye. Examinaremos más a fondo los otros términos utilizados para describir los datos más adelante en el curso, pero esencialmente habrá encontrado los prefijos, sin duda familiares, de kilo, mega, giga, etc., antes del byte a lo largo de sus actividades diarias.

Habiendo dicho esto, no debemos olvidar que los registros en papel son igualmente valiosos, y tan válidos en términos de cualquier descripción de datos. Muchos aspectos de las actividades operativas siguen siendo dependientes y el humilde trozo de papel y bolígrafo, especialmente cuando se trata de registrar el contexto en el que se recopilan los datos y por qué, también se conoce como metadatos. En esencia, la gestión de todo tipo de datos es, probablemente, algo que emprendemos como parte de nuestra vida cotidiana, desde la banca personal hasta el llenado de formularios para acceder a los servicios locales. Lo que estamos tratando de hacer aquí es tomar más conciencia de la importancia de la gestión de los datos, con el objetivo de maximizar la eficiencia y la precisión de los sistemas utilizados para recopilar, almacenar y analizar.

Pantalla 5 – Datos buenos/datos incorrectos

La precisión de los datos es algo así como una fiesta móvil, y a menudo se puede llegar a un compromiso entre la calidad y la eficiencia. Esto puede parecer un poco contra-intuitivo, ya que podría pensar que se deben hacer todos los esfuerzos para que todos los datos sean lo más precisos posible. Este no es el caso para mucha instrumentación, por ejemplo, los instrumentos que registran el nivel de la marea o las corrientes de medida pueden tener configuraciones variables que compensan la demanda de energía con el número de mediciones realizadas para un período de tiempo determinado.

La recopilación de datos también puede afectar la calidad. La instrumentación es un aspecto, pero cuando esos datos se graban manualmente, a menudo pueden surgir problemas. Un ejemplo típico de esto en el mar puede ser la observación de mamíferos marinos. Por más hábiles que sean los observadores, con frecuencia tienen que soportar largos turnos y el cansancio suele ser un problema. En estas condiciones, no es raro que los períodos de reloj individual se eliminen o se descuenten debido a la falta de fiabilidad.

Muchos datos están disponibles de forma gratuita y tienen código abierto. A pesar de lo útil que es, no siempre podemos estar seguros de la fiabilidad de estos datos, ni de que se recopilaron de forma adecuada. Se debe tener mucho cuidado, incluso si los metadatos parecen ser sólidos como una roca.

En última instancia, nuestra capacidad para analizar los datos de manera eficiente puede generar un mayor compromiso sobre la calidad. Por mucho que nos gustaría que todos nuestros datos sean de la más alta calidad todo el tiempo, siempre que seamos transparentes, honestos acerca de cómo se recopilaron y procesaron, y permitiendo a cualquier usuario final realizar una evaluación educada de los mismos, entonces esto en gran medida será suficiente. Realmente dependerá de las circunstancias específicas de la situación a la que se enfrente.

Pantalla 6 – Uso y procesamiento de datos

A menudo, la verdadera imagen de la calidad de los datos no se materializará hasta que se complete todo lo que se denomina el postprocesamiento, y usted comience a analizar las señales de interés más de cerca. Quizás solo entonces determine que los datos recopilados pueden ayudar a informar de esa decisión operativa. Se requiere una cuidadosa inspección de los resultados de sus esfuerzos aquí; se requiere sabiduría para garantizar que los resultados no se sobreinterpreten. Por ejemplo, esta imagen resulta ser algunos datos de velocidad actuales para una ubicación en el canal de inglés occidental. Aquellos de ustedes con ojos penetrantes verán que es por un período de tiempo muy específico. Lo que no podemos hacer es interpretar en exceso esta información y decir que se aplica a todo el canal en inglés, todo el tiempo.

El postprocesamiento generalmente se refiere a la etapa de análisis de datos que sigue al tratamiento de los datos brutos. Es el momento en que obtienes una imagen más clara y realmente entiendes lo que está sucediendo para cualquier situación que se proponga estudiar. Los datos como estos que se muestran aquí son insuficientes para describir la actividad a largo plazo en esta ubicación, pero aún podrían validar un modelo de velocidad actual aquí, lo que tal vez contribuya a una mejor capacidad para pronosticar condiciones o ayudar a determinar algún otro proceso importante localmente. El punto clave es usar su juicio sabiamente y cuestionar todos los aspectos de lo que se muestra, en lugar de distraerse con lo que de otra manera podría parecer sólidamente fiable.

Pantalla 7 – Gestión del flujo

Hay un equilibrio entre el volumen de datos recopilados y los recursos disponibles para procesarlos y analizarlos. Por supuesto, esto depende de cuál sea su objetivo particular. Si tiene sistemas automatizados que procesan los datos sin mayor esfuerzo, entonces tal vez más es mejor. Por ejemplo, si tuviera que realizar un análisis de las corrientes de la marea en un lugar determinado, tal vez para afinar un modelo para predecir la velocidad actual, entonces realmente más es mejor. El mayor nivel de detalle o más atrás en el historial que toman sus mediciones siempre le ayudará en esta situación. Pero a veces, hay una necesidad de interpretar los resultados de los datos de una manera específica. Quizás esté buscando establecer la forma más eficiente en que los barcos entran y salen del puerto. Es poco probable que un sistema totalmente automatizado pueda gestionar esto, sin alguna orientación o aporte específico de un usuario que aprecie el contexto en el que la computadora llega a una decisión. En resumen, tenemos que pensar en gestionar el flujo de información, lo que esperamos quede claro en el proceso de planificación para la recopilación de datos.

Como era de esperar, la gestión de grandes cantidades de datos puede ser una ocupación a tiempo completo. Desde la perspectiva del científico oceanográfico del Reino Unido, este papel recae en un organismo financiado con fondos públicos llamado British Oceanographic Data Center (BODC).

Pantalla 8 – Servicio de limpieza y metadatos

Debemos tener en cuenta que los datos que adquiramos no se pueden usar de inmediato, ni siquiera para el fin para el que se recopilamos, y puede transcurrir un período de tiempo significativo entre la recopilación y el uso. Aquí es donde entran los metadatos, y estrictamente es el primer paso y el más importante en la buena gestión de datos. Los metadatos dan ese contexto tan importante a las circunstancias dentro de las cuales se recopilamos los datos originales. Sin ellos, los futuros usuarios de los datos están ciegos, y no pueden aplicar ningún pensamiento nuevo a los datos antiguos si falta este componente tan importante...

Por el momento, es suficiente decir que los pequeños detalles que rodean la recopilación de datos pueden ser increíblemente importantes. Como sugiere la imagen, los metadatos son realmente una nota de amor para los futuros usuarios, ya que no podemos saber de antemano cómo otros pueden querer utilizar los datos en los años siguientes. Los metadatos pueden ser cualquier cosa relacionada con la adquisición de los datos, incluyendo cosas simples como la fecha de recopilación, las coordenadas de posición, las condiciones climáticas, el tiempo de muestreo, la configuración del instrumento, etc. Para estar seguro, informe de todo esto y mucho más. El próximo usuario de los datos le querrá por ello.

Pantalla 9 – Soluciones de almacenamiento

Suponiendo que hayamos hecho un gran trabajo recopilando los datos, registrando los metadatos apropiados y comenzando el análisis, ¿qué hacemos con el almacenamiento de datos? De hecho, es más complejo de lo que se podría pensar y, si bien las opciones son abundantes, la solución debe ser adecuada para las necesidades del sector en el que trabaja. ¿Podemos permitirnos operar un gran almacén de datos? ¿Tenemos los recursos necesarios? ¿Necesitamos acceder a los datos al instante, desde cualquier parte del mundo?

Si es así, entonces tal vez la computación en la nube podría ayudarnos con esto y con todos los demás problemas. Pero, ¿y si trabajamos en alta mar y exigimos una conexión estable con mucho ancho de banda? Claramente, no existe una solución única para todos, pero tenemos muchas opciones aquí, así que volveremos sobre este tema más adelante en la serie. Una cosa que deberíamos hacer todos los días es hacer una copia de seguridad de nuestros datos. Esto incluye datos personales y profesionales. En mi trabajo como científico investigador, no es raro realizar varias copias de seguridad al día, ya que se puede desperdiciar un esfuerzo considerable en encontrar soluciones a problemas que serían completamente ineficaces repetir. La seguridad, por supuesto, es igualmente importante, especialmente si trabaja en entornos comercialmente sensibles o si tiene que procesar datos que están protegidos por legislación de cualquier tipo.

Pantalla 10 – Toma de decisiones operativas

El objetivo de todo esto es administrar de manera efectiva nuestros datos para informar mejor de nuestra toma de decisiones operativas. Si parece que hay mucho que pensar aquí, ¡es porque lo hay! Todos trabajamos en sectores complejos de una industria fuertemente influida por los datos y, si esperamos negociar con éxito escenarios complejos, este tema no puede pasarse por alto ni subestimarse.

En última instancia, para comprender mejor todos los aspectos del entorno en el que trabaja, la gestión de datos exitosa es fundamental. Hay otros beneficios asociados con las mejores prácticas, por supuesto, en el sentido de que aquellos que están altamente cualificados en esta área tienen una gran demanda. Existe una escasez mundial de personas con profundos conocimientos analíticos de procesamiento y gestión de datos. Un informe del Mckinsey Global Institute (2011) sugirió recientemente que la escasez mundial de administradores con conocimiento de datos se encontraba en torno a los 1,5 millones. No hay excusa para no darle a este aspecto de sus estudios la máxima atención.

Pantalla 11 – Responsabilidad

¿Quién debe asumir la responsabilidad de garantizar que se sigan las mejores prácticas en la gestión de datos? Francamente, nos corresponde a todos asegurarnos de que se sigan estos pasos básicos. No podemos dejarlo en manos de otros y asumir que todos nuestros datos valiosos se almacenaron y procesaron correctamente, y todos los metadatos se escribieron adecuadamente y se adjuntaron a los datos brutos.

Debemos comenzar a apreciar que los datos son un bien importante y valioso que podría influir si nosotros y las organizaciones que representamos tienen éxito. Mientras que algo de esto puede parecer evidente, estoy seguro de que todos nos hemos encontrado en más de una ocasión con que nosotros o alguien que conocemos ha sufrido innecesariamente con los datos como resultado de no seguir lo que puede parecer, al menos superficialmente, tener un buen y anticuado sentido común. ¡Puedo recordar a más de un estudiante que no presentó su trabajo de clase a tiempo porque había tenido un problema relacionado con los datos de su proyecto! La pelota está en nuestro tejado. Reconozca la importancia de una buena administración de datos, implemente algunas de sus sugerencias en la forma en que trabaja con los datos y obtendrá los beneficios para el resto de su carrera.

Pantalla 12 – Resultados de aprendizaje revisados

Explicar qué se constituye como datos; reconocer que toma muchas formas

Definir qué es la administración de datos y por qué es necesaria

Reconocer qué datos son buenos y cuáles son incorrectos, y cómo las operaciones se ven afectadas por ambos

Agradecimientos y referencias

McKinsey & Company (2011) Big data: La próxima frontera para la innovación, la competencia y la productividad. McKinsey Global Institute, Cambridge

Pantalla 3:

<https://www.flickr.com/photos/malfet/6870465199>

Pantalla 7:

https://upload.wikimedia.org/wikipedia/commons/5/53/USACE_Gavins_Point_Dam.jpg

Por Robert Etzel, Cuerpo de Ingenieros del Ejército de EE. UU. [Dominio público], a través de Wikimedia Commons

Pantalla 8:

<https://www.flickr.com/photos/centralasian/8071729256>

Pantalla 9:

https://upload.wikimedia.org/wikipedia/commons/1/19/Interior_of_StorageTek_tape_library_at_NERSC_%282%29.jpg

Por Derrick Coetzee de Berkeley, CA, EE. UU. [CC0], a través de Wikimedia Commons

Pantalla 11:

<https://www.flickr.com/photos/smemon/6032417950/in/photolist-ac4HG3-au8uNg-q1mcsn-91iRws-9mq7H9-eNgdFU-nuUQM7-pkNGFp-pakM17-7YzBq9-kR9oZt-5sTuyk-dJvTjJ-5eLK2r-9fBHYt-nuURRo-8YV4K5-oeYpCH-kRawJq-kKPBSS-7FGCDx-9Vk8NS-qrEzjs-e7M5Qx-dDwovw-dBd8Rf-7cSb1j-cwujQN-fdgXiR-9RWUDD-5Tj7MG-8YUWaE-getGqG-kR9kG6-8YVJ4b-4CgAaf-nTGrou-bCbzWz-bVJFVV-5o2hnx-iNK4nX-getFvs-d4ouN3-pYfF75-9YoGVQ-7YzDyE-8YV91s-nuUSuA-pMdVmL-pembtL>