

Cómo comenzar con Machine Learning: consejos de expertos de vanguardia

Las compañías audaces saben que la inteligencia artificial (AI) y machine learning (ML) tienen el potencial de transformar los negocios, pero muchas no están seguras con respecto al punto de partida. Por ello, aquí les ofrecemos la guía de un experto basada en 10 de las preguntas más candentes sobre lo que se necesita para lanzar una iniciativa de machine learning.



Larry Pizette
Gerente senior, Laboratorio de soluciones de aprendizaje automático Amazon Web Services

INTRODUCCIÓN

Para muchas compañías, machine learning (ML) es, a menudo, el catalizador que convierte los datos comerciales en predicciones precisas e información procesable. Sin embargo, las barreras de entrada parecen desalentadoras:

- Se deben encontrar, organizar y estandarizar los datos.
- Se deben obtener los recursos de informática para manejar las cargas de trabajo de ML.
- Científicos de datos deben crear y formar los modelos.
- Los productos deben integrarse en los procesos del negocio.

Cuando se enfrentan a requisitos que son difíciles de alcanzar y objetivos que requieren imaginación, las compañías necesitan ayuda de expertos. Larry Pizette, gerente senior y líder del laboratorio de soluciones de machine learning (ML) en Amazon Web Services, sabe mucho sobre cómo superar las barreras de ML. Se reúne con clientes de diversas industrias, ayudándoles a planificar e implementar sus iniciativas de ML, en una amplia gama de casos de uso.

Tuvimos la oportunidad de sentarnos con Larry y hacerle 10 preguntas, y así obtener información que ayudará a los líderes empresariales que desean comprender las tendencias de ML, cómo iniciar un proyecto y cómo pasar de la prueba de concepto a la producción.

ENFOQUE HACIA ML

Trim. 1 | ¿Qué motiva el interés en el aprendizaje automático hoy en día?

Estamos en un punto de inflexión, donde el concepto de machine learning es conocido por todos. Se hace referencia a la inteligencia artificial y a machine learning en televisión, revistas, sitios web y anuncios de productos; están en todas partes. Los clientes buscan llevar a cabo procesos de ML por diferentes razones, pero éstas tienden a estar en torno a los mismos objetivos: mejores experiencias y resultados para sus clientes, mejores operaciones y, para los clientes comerciales, mejores resultados financieros. Los dueños de las compañías quieren aprovechar todo el potencial de sus datos. Es difícil no prestar atención a ML.

Trim. 2 | ¿Existen industrias o funciones específicas que sean fuertes candidatos para implementar ML?

Tenemos clientes de diferentes áreas de negocios: petróleo y gas, salud, ciencias biológicas, seguros, servicios financieros, restaurantes, deportes profesionales, transporte, gobierno e instituciones educativas. Los clientes acuden al laboratorio de soluciones de ML porque desean comenzar con las capacidades de ML en sus organizaciones y les gustaría entender el arte de lo posible.

AWS TECH

ML Plug-and-Play

Muchos líderes piensan que ML es tecnología compleja y que requiere mucha mano de obra, pero no tiene por qué ser así. AWS puede ayudarlo a comenzar hoy mismo, sin que necesite experiencia en esta tecnología, mediante varios servicios de ML que se pueden agregar fácilmente a sus aplicaciones:

- Analice sus imágenes y videos con Amazon Rekognition.
- Cree chatbots de voz y texto para su negocio con Amazon Lex, que funcionan con la misma tecnología que Amazon Alexa.
- Cree aplicaciones que hablen convirtiendo el texto en voz natural, con Amazon Polly.
- Descubra relaciones e información en texto, como los emails de servicio al cliente o transcripciones, con Amazon Comprehend.
- Traduzca el contenido como sus sitios web y aplicaciones, de manera rápida y rentable, con Amazon Translate.

"Comenzar implica solo seguir tres pasos: aprender sobre ML, crear ideas basadas en los casos de uso y los enfoques para validar las ideas, e implementar una o dos pruebas de conceptos de alto valor".

Por ejemplo, las compañías de seguros quieren ayudar a evaluar mejor el riesgo, poner un mejor precio de sus productos y mantener su rentabilidad. Los líderes empresariales, gubernamentales y educativos están empezando a pensar en cómo ML puede afectar a sus organizaciones. Ellos saben que necesitan comenzar. Simplemente no están 100 % seguros sobre cuál es el punto de partida. Saben que Amazon tiene una extensa historia en machine learning, por lo que acuden a nosotros para ayudarlos a seguir adelante.

IMPLEMENTAR ML

Trim. 3 | ¿Qué pasos deben seguir las organizaciones para comenzar con ML?

Comenzar es realmente una cuestión de tres pasos: aprender sobre ML, crear conceptos basados en casos de uso y enfoques para validar las ideas, e implementar una o dos pruebas de conceptos (POC) de alto valor.

Si su organización aún no tiene una base sólida de conocimiento sobre ML, le recomendamos que se tome el tiempo para educar a sus líderes empresariales y operativos. Cuando los clientes nos piden esta información, primero los hacemos participar en una sesión de descubrimiento sobre el aprendizaje automático, sentados con nuestra organización de desarrollo de negocios especialmente formada. Como parte de ese paso, nos aseguramos de que los clientes comprendan, a nivel comercial, cómo funciona el ML y la importancia de los datos para sus organizaciones.

El siguiente paso suele ser tener una sesión creativa e identificar la POC. Nos gusta incluir tanto a los líderes de negocios (los dueños de compañías que buscan los resultados que desean) como a la gente de TI que puede ayudarnos a identificar los datos para satisfacer las necesidades del negocio. Este es el punto donde comenzamos a identificar los detalles de una POC y los resultados del negocio.

También detallamos la importancia de comprender los datos y, en muchos casos, de tener los datos comentados, es decir, tener datos que están etiquetados con información sobre los datos en sí. Un ejemplo sería una imagen y una descripción de la imagen. Este paso es significativo, ya que muchas organizaciones no conocen la información que tienen o cómo acceder a ella, y los datos son necesarios para formar los modelos de ML.

DEMOSTRAR EL VALOR DE ML

Trim. 4 | Eso nos lleva al tercer paso. ¿Cómo es el aspecto de una prueba de concepto exitosa?

Descubrimos que los clientes que están dispuestos a identificar casos de negocios impactantes, algo importante que realmente puede hacer una enorme diferencia en su organización, son los que pueden demostrar resultados rápidamente con ML.

Es realmente importante que los clientes estén dispuestos a hacer prototipos y a hacer pruebas de concepto para ponerse en marcha. Si intentan hacer lo imposible y tener un producto que dure varios años, es mucho más difícil demostrar el progreso.

He visto que las pruebas de concepto van realmente rápido cuando los clientes tienen los datos y están muy motivados. Con esos clientes, obtienen los datos; asignan recursos para colaborar con nosotros. Nos permiten trabajar con acceso a los datos y un objetivo claro; y podemos mostrar los resultados iniciales, del modelado de la prueba de concepto, en tan solo tres semanas desde el momento en que accedemos a los datos.

Trim. 5 | Una vez que finaliza la etapa de la prueba de concepto, ¿cuánto tiempo lleva implementar una solución de ML en producción?

Una vez que haya probado el potencial de ML, el siguiente paso es trasladar la capacidad a la

¿Sus datos están listos para ML?

Los datos son oro para aquellos líderes que buscan alterar sus industrias con ML. La pregunta es, ¿sus datos están listos? Muchas empresas solo comienzan a trabajar para organizar los datos cuando planifican una prueba de concepto, pero no es necesario esperar. Siga la guía a continuación para posicionar mejor a su organización con el fin de implementar y escalar ML.

1. Localice los datos

Realice un viaje de descubrimiento para detectar todos los orígenes de datos en su negocio. Explore diferentes funciones, divisiones, sistemas y procesos para capturar y categorizar los diferentes tipos de datos que posee. Todo tiene relevancia, desde datos estructurados como los de la facturación y de CRM, hasta datos no estructurados, registros web y feeds de redes sociales.

2. Evalúe sus datos

Solo los datos de calidad permitirán que se beneficie de ML, tal como usted quiere. Antes de ir demasiado lejos, es importante determinar qué parte de sus datos se pueden utilizar realmente. Las bases de datos a menudo están incompletas, difieren en su organización o taxonomía o tienen contenido duplicado. Tómese el tiempo para evaluar los datos e identificar las correcciones hoy mismo y ahorrará tiempo y gastos considerables en el futuro.

3. Depure los datos

Ahora es el momento de darle forma a los datos para que puedan utilizarse con el fin de impulsar sus objetivos de ML. Elimine la información duplicada, complete los campos o los detalles que faltan, corrija cualquier error y organícelos. Sin duda, tendrá datos que no se pueden normalizar a la perfección para que cumplan con los modelos o flujos de trabajo. Solo márkuelos y sepárelos para evitar que perjudiquen el desarrollo de los algoritmos.

4. Centralice sus datos

Mantenga y proteja los nuevos datos de alta calidad, almacenándolos y administrándolos de manera centralizada. Éste es también el momento de agregar anotaciones o etiquetas necesarias, para que los sistemas de machine learning analicen con precisión los datos y transformen la descripción de lo que sucedió en el pasado, en potentes predicciones y acciones para su negocio.



"Recomendamos focalizarse en cómo se puede ofrecer una mejor experiencia para sus clientes e identificar los resultados comerciales y operativos deseados, y luego aprovechar ML para satisfacer esas necesidades".

producción, lo que puede incluir la integración de la capacidad de ML en un sistema de TI mayor.

Cuando un cliente traslada una carga de trabajo a producción, suele durar más que la prueba de concepto. Realmente depende de la complejidad, porque estamos incorporando machine learning en todo el sistema. Si es un tipo de capacidad independiente, puede entrar en producción muy rápidamente. Si tiene que integrarse con un ecosistema mucho mayor de sistemas de TI y procesos empresariales, puede demorar varios meses.

Pero lo bueno de esto es que, una vez que se ponga en marcha, puede seguir actualizando sus modelos. Está recopilando datos y resultados, y está mejorando sus modelos con el correr del tiempo. A medida que reúne más datos e información sobre cómo debe ser el "ideal", es fundamental repetir el proceso de formación y reevaluación de sus modelos. Esto lo ayudará a asegurarse de que esté obteniendo los mejores resultados posibles.

Como existe variabilidad en cada caso de uso personalizado, nuestro laboratorio de soluciones de ML puede ayudar a nuestros clientes a delimitar el alcance de las pruebas de concepto y la implementación de producción.

Trim. 6 | ¿Puede dar algunos ejemplos de cosas de las que está muy orgulloso?

Estoy particularmente orgulloso del excelente trabajo que hemos hecho con ligas deportivas profesionales como la de béisbol (MLB). Por ejemplo, ideamos junto con ellos una función convincente para la audiencia televisiva. Una vez que decidimos cuál sería el modelo que identifica la posibilidad de éxito de un corredor para robar la segunda base, nos tomó tres semanas, a partir del 1.º de

septiembre de 2018, preparar los datos, formar, evaluar e implementar el modelo.

Aprovechamos Amazon SageMaker como parte clave de esta iniciativa. Incluso, la implementación en la producción en la nube. SageMaker proporcionó el tiempo de respuesta inferior a un segundo que se requiere para hacer predicciones (llamadas inferencias) y, al usar un grupo de autoescalado, proporcionó tanto la alta disponibilidad como el alto rendimiento que MLB necesitaba. Podrá ver esto en acción en futuros juegos de MLB, incluso en la Serie Mundial 2018.

ML EN LA PRÁCTICA

Trim. 7 | ¿Qué consejo puede dar a las empresas que buscan adoptar ML?

Recomendamos centrarse en cómo puede ofrecer una mejor experiencia para sus clientes e identificar los resultados comerciales y operativos deseados, y luego aprovechar ML para satisfacer esas necesidades.

Para ayudar a desarrollar este potencial, culturalmente, debe haber una aceptación de que machine learning es una parte importante del negocio y de las operaciones. Con frecuencia, las empresas tienen diferentes grupos de TI, de los cuales, cada uno tiene su propia área de experiencia o unidades de negocio a las que brindan soporte. Algunas iniciativas de ML pueden requerir información de estos dominios, por lo que es importante entender la estrategia de datos y dónde encajará machine learning.

Trim. 8 | ¿Qué papel juega en ML la informática en la nube?

La nube ha beneficiado realmente al machine learning, ya que es posible probar las capacidades. Ahora puede hacer pruebas de conceptos. Si tiene un problema, no se

AWS TECH

Amazon SageMaker: una poderosa plataforma de ML

Cuando esté listo para probar e implementar sus propias soluciones de AI, AWS le ofrece un conjunto incomparable de infraestructura y servicios de machine learning. De hecho, en AWS se compila más machine learning que en cualquier otro lugar. Amazon SageMaker es un excelente punto de partida. Es una plataforma completamente administrada que permite a los desarrolladores y a los científicos de datos compilar, formar e implementar rápida y fácilmente modelos de machine learning a cualquier escala. Amazon SageMaker elimina todas las barreras que normalmente demoran a los desarrolladores que desean utilizar machine learning proporcionándoles una variedad de algoritmos precompilados para funciones comerciales comunes, y es compatible con prácticamente todos los marcos de ML conocidos, como Apache MxNet, TensorFlow, PyTorch, Cafe2 y Chainer.

preocupe. Acaba de ponerlo a prueba. No hay hardware, y puede escalar vertical u horizontalmente. Usted tiene los datos allí, así que puede probar otros enfoques si uno no funciona. Debido a que puede escalar rápidamente cantidades significativas de capacidad con las instancias de Informática Acelerada adaptadas a ML, es posible formar

"Estamos tratando de poner machine learning a disposición de cada organización y de desarrollar servicios para poder ayudar con el trabajo pesado, no diferenciado, de machine learning".

grandes cantidades de datos en minutos u horas, lo cual, en el pasado hubiera llevado días.

Trim. 9 | ¿Qué hace que ML de AWS sea una tecnología única?

Estamos tratando de poner machine learning a disposición de todas las organizaciones y desarrollar servicios para poder ayudar con el trabajo pesado, no diferenciado, de machine learning.

Si observa nuestra pila de servicios de ML, puede ver que hay muchos niveles diferentes en los que puede emplear los servicios de AWS. En el nivel más alto, tenemos servicios de machine learning basados en API como Rekognition, Lex, Polly, Comprehend, Translate y Transcribe. Estos servicios son muy fáciles de usar y no requieren un científico de datos.

Para los clientes que desean profundizar en una capa, contamos con servicios de infraestructura, como SageMaker, que facilitan la creación, la formación y la implementación de capacidades en la nube. SageMaker es un increíble servicio de infraestructura que hace que sea excepcionalmente fácil ejecutar un software de marco ampliamente utilizado como Apache MxNet, TensorFlow, PyTorch, Caffe2 y Chainer. Además, SageMaker incluye algoritmos incorporados que se han optimizado y pueden mejorar el rendimiento 10 veces más, en comparación con lo que puede ejecutar en otro lugar.

Trabajamos en estrecha colaboración con empresas como Intel para asegurarnos que el hardware que respalda esa infraestructura se diseñe a la medida y se optimice para lograr el mejor rendimiento con los marcos más conocidos. Además, SageMaker hace que sea más fácil para nuestros clientes implementar capacidades de ML en producción. Con un clic en el botón implementar en entornos de

zonas de múltiple disponibilidad y escalables, hemos logrado eliminar la complejidad de TI durante la implementación en producción.

Trim.10 | Si estuviera en el lugar de un líder empresarial que recién comienza, ¿cuáles serían sus próximos pasos?

Primero, buscaría formas de ayudar a los clientes basándome en los datos, optimizando las operaciones para que sean más eficientes y maximizando los resultados financieros. En segundo lugar, me aseguraría de estar desarrollando las capacidades de las personas dentro de nuestra organización para aprovechar el ML, incluso contrataría o formaría científicos de datos donde sean necesarios. En tercer lugar, buscaría identificar o establecer el enfoque de nuestra organización con respecto a los datos, incluso al control y los repositorios, y determinaría el alcance de TI para respaldar el ML en la empresa. No hay manera de que las compañías ignoren una tecnología innovadora como ML, y estoy muy orgulloso de ser parte de un equipo que ayuda a responder preguntas y a que avancen estas iniciativas.

Para obtener más información sobre ML de AWS e Intel, visite ml.aws

Innovación de Intel en ML



Intel y Amazon han estado colaborando estrechamente para impulsar la innovación en AI y ML, brindándole acceso a una infraestructura de alto rendimiento, además de herramientas de formación de modelos de machine learning y motores de inferencia optimizados para ofrecer velocidad y eficiencia.

Por ejemplo, las instancias C5 de AWS, con la tecnología de los procesadores escalables Intel Xeon personalizados, están diseñadas específicamente para manejar las cargas de trabajo de ML más complejas. Las implementaciones a gran escala pueden ejecutarse utilizando servicios de plataformas como Amazon SageMaker, que están optimizados para mejorar el rendimiento en instancias EC2 basadas en Intel. Los motores de aprendizaje profundo de AWS, con la última tecnología de Intel, están aumentando el rendimiento de machine learning en más de 100 veces.