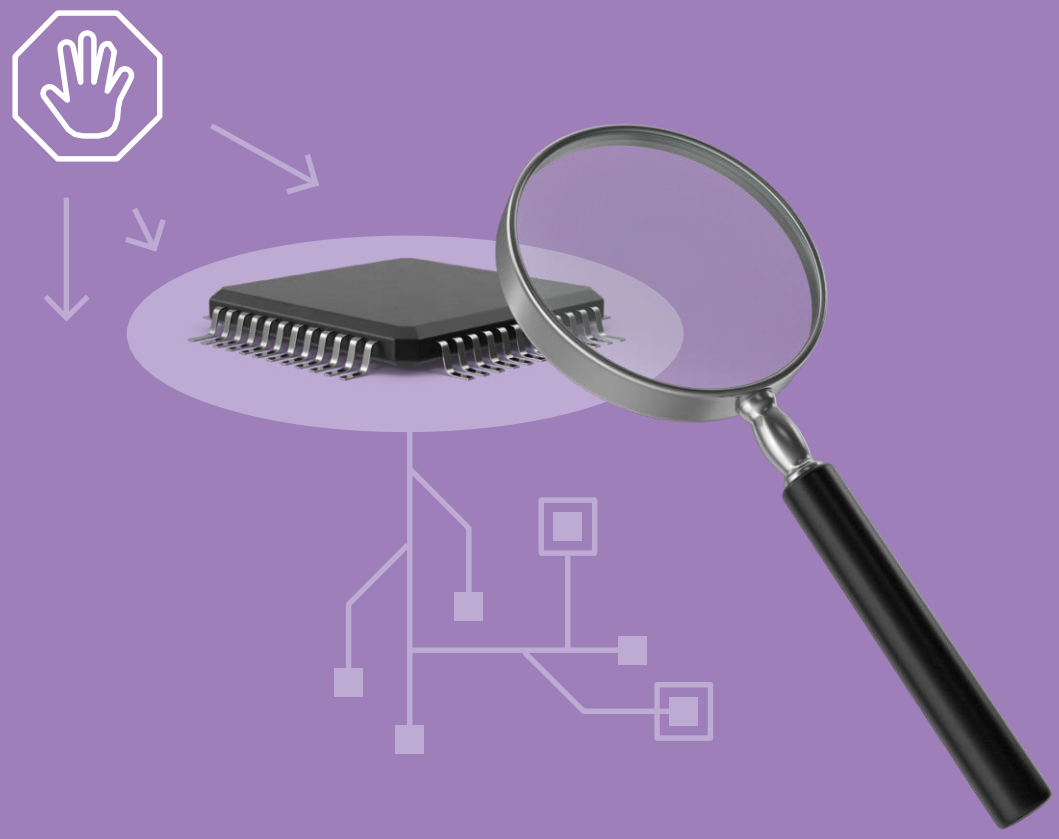
GUIA

¿Cómo Ejecutar detección de anomalías a escala?

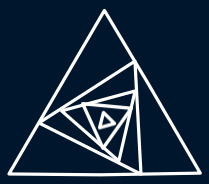
Para descubrir fraudes, monitoreo de red, cuidado de la salud, descubrimiento de tendencias emergentes y más.



SOLEX

Acerca de esta Guía

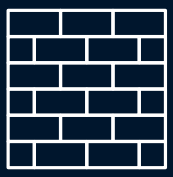
Esta guía pretende proporcionar una descripción general de alto nivel de **qué es la detección de anomalías**, cómo funciona y cuándo se puede aplicar. Al final, los lectores deben tener una comprensión de:



Qué es la detección de anomalías y en qué se diferencia de otros sistemas similares.

Los tres tipos de anomalías que se pueden detectar.

Casos de uso y ejemplos de dónde se emplea la detección de anomalías.



Cómo construir un sistema básico de detección de anomalías.

Para aquellos que no están familiarizados con la ciencia de datos en el contexto de la detección de anomalías, esta guía proporcionará una breve introducción al tema y un recorrido por los aspectos principales.

Qué es?

La detección de anomalías se trata de encontrar patrones de interés (valores atípicos, excepciones, peculiaridades, etc.) que se desvían del comportamiento esperado dentro de los conjuntos de datos. Dada esta definición, vale la pena señalar que la detección de anomalías es, por lo tanto, muy similar a la eliminación de ruido y la detección de novedades. Aunque los patrones detectados con la detección de anomalías son en realidad de interés, la detección de ruido puede ser ligeramente diferente porque el único propósito de la detección es eliminar esas anomalías, o ruido, de los datos.

Al igual que con la mayoría de los proyectos de ciencia de datos, el objetivo final o el resultado final de la detección de anomalías no es solo un algoritmo o un modelo de trabajo. En cambio, se trata del valor de la información que proporcionan los valores atípicos. Es decir, para una empresa es el dinero ahorrado al prevenir daños en el equipo, es dinero perdido en transacciones fraudulentas, etc. En el cuidado de la salud, puede significar una detección más temprana o un tratamiento más fácil.

LA DETECCIÓN DE ANOMALÍAS REQUIERE UN SISTEMA ÁGIL Y EN CONSTANTE APRENDIZAJE PORQUE:

La naturaleza misma de los casos de uso para la detección de anomalías (específicamente en los sectores de TI y finanzas) significa que muchas veces los estafadores intentan específica y deliberadamente producir entradas que no parezcan valores atípicos. Adaptarse y aprender de esta realidad es fundamental.

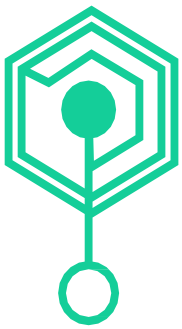
Además de las intenciones maliciosas, los conjuntos de datos generalmente tienden a cambiar con el tiempo a medida que cambian los usuarios, por lo que un sistema debe evolucionar junto con esos usuarios. Las anomalías, por su naturaleza, son inesperadas, por lo que es importante que los métodos utilizados se adapten a los datos subyacentes.

Muchos casos de uso son extremadamente sensibles al tiempo y las empresas (o pacientes, clientes, etc., de esas empresas) no pueden darse el lujo de esperar. La detección temprana de patrones en función de una combinación de puntos de datos puede ayudar a anticipar problemas antes de que sea demasiado tarde

Es importante tener en cuenta que, a pesar de que los casos de uso más comunes son la detección de fraude o la intrusión en el sistema, las anomalías no siempre son malas, es decir, no siempre tienen que indicar que algo anda mal.

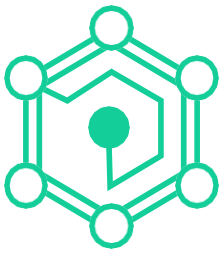
La detección de anomalías también se puede utilizar, por ejemplo, para detectar o predecir cambios leves en el comportamiento de los clientes o usuarios que luego pueden resultar en un cambio en la estrategia de ventas, desarrollo o marketing, lo que permite a las empresas mantenerse un paso por delante de las nuevas tendencias.

Hay tres tipos básicos de anomalías que se pueden detectar:



Anomalías de puntos:

Las anomalías de puntos son simplemente instancias únicas y anómalas dentro de un conjunto de datos más grande. Por ejemplo, una temperatura de 60 grados centígrados en un conjunto de datos sería una anomalía puntual, ya que sería la temperatura más alta jamás registrada en la Tierra. Los sistemas de detección de anomalías a menudo comienzan identificando anomalías puntuales, que se pueden utilizar para detectar anomalías contextuales o colectivas más sutiles.



Anomalías contextuales (o condicionales):

Estos son puntos que solo se consideran anómalos en cierto contexto. Un buen ejemplo es la temperatura nuevamente; mientras que se considera que 30 grados centígrados están dentro del rango de temperaturas posibles, dado el contexto de diciembre en la ciudad de Nueva York, este punto de datos es ciertamente una anomalía. Con los datos espaciales, la latitud y la longitud son el contexto, mientras que con los datos de series temporales, el tiempo es el contexto.

Anomalías colectivas:

Cuando conjuntos de datos relacionados o partes del mismo conjunto de datos son tomados como anómalos respecto al conjunto de datos completo (incluso cuando los conjuntos de datos individuales no contienen anomalías).

Por ejemplo, supongamos que hay datos de una compra con tarjeta de crédito en EE. UU., pero también un conjunto de datos que muestra que se retira dinero de cajeros automáticos en Francia al mismo tiempo. Una anomalía colectiva puede ocurrir si no ocurre una sola anomalía en ningún conjunto de datos, pero todos los conjuntos de datos que miden varios componentes en conjunto indican un problema.

Porqué?

La detección de anomalías es un enfoque que puede ser útil en una variedad de industrias y para una variedad de propósitos. Pero el factor unificador subyacente es la capacidad de detectar pequeños cambios o diferencias en un sistema que, de otro modo, podrían pasar desapercibidos. Descubrir anomalías mediante el aprendizaje automático permite que los humanos (u otros sistemas) tomen medidas en función de estos valores atípicos.

Específicamente, una mirada no exhaustiva a los casos de uso de los sistemas de detección de anomalías incluye:

INDUSTRIA	CASOS DE USO	RETO(S)
IT, DEVOPS	Detección de intrusiones (seguridad del sistema, malware), monitoreo del sistema de producción o monitoreo de subidas/bajadas de tráfico en la red.	Necesidad de una canalización en tiempo real para reaccionar; grandes volúmenes de datos más la falta de disponibilidad de datos etiquetados correspondientes a intrusiones que dificultan el entrenamiento/prueba; generalmente tienen que adoptar un enfoque semi-supervisado o no supervisado.
MANUFACTURA E INDUSTRIA, CONSTRUCCIÓN, AGRICULTURA Y MÁS	Mantenimiento predictivo, detección de fraude en el servicio.	Los sistemas industriales producen datos de diferentes sensores que varían enormemente: diferentes niveles de ruido, calidad, frecuencia de medición.
CUIDADO DE LA SALUD	Supervisión del estado, incluida la detección de convulsiones o tumores	Los costos de clasificar erróneamente las anomalías son muy altos; Además, los datos etiquetados con mayor frecuencia pertenecen a pacientes sanos, por lo que generalmente se debe adoptar un enfoque semisupervisado o no supervisado.

INDUSTRIA	CASOS DE USO	RETO(S)
FINANZAS & SEGUROS	Detección de fraude (tarjetas de crédito, seguros, etc.), análisis bursátil, detección temprana de uso de información privilegiada	La detección de anomalías financieras es de alto riesgo, por lo que debe realizarse en tiempo real para que pueda detenerse tan pronto como suceda. Además, quizás sea más importante que otros casos de uso ser cuidado con los falsos positivos que pueden perturbar la experiencia del usuario.
SECTOR PUBLICO	Detección de imágenes inusuales de vigilancia	Requiere técnicas de aprendizaje profundo, lo que encarece este tipo de detección de anomalías.

La detección de anomalías puede ser útil en una serie de otros campos e industrias donde los eventos raros son muy importantes o impactantes, pero son difíciles de encontrar dentro de los datos.

Debido a su amplia gama de aplicaciones, dominar la detección de anomalías desde la perspectiva de un científico de datos es un caso de uso increíblemente aplicable.

Cómo?

Si está familiarizado con los siete pasos fundamentales para crear un proyecto de datos, entonces ya conoce los conceptos básicos sobre cómo iniciar a usar la detección de anomalías para beneficiar a su equipo o negocio. Pero también hay varias particularidades a tener en cuenta a la hora de trabajar con la detección de anomalías:

1 Entender el negocio

El primer paso para una detección de anomalías exitosa es comprender realmente qué tipo de sistema necesita el negocio y establecer un marco para los requisitos y objetivos de la detección de anomalías antes de sumergirse. Estas son discusiones preliminares importantes porque no todo el trabajo de detección de anomalías es el mismo. ; exactamente lo que califica como una anomalía y los procesos posteriores iniciados por la detección de anomalías varían enormemente según (e incluso entre) los casos de uso.

En particular, la naturaleza de los datos, del problema en cuestión y los objetivos del proyecto dictan necesariamente las técnicas empleadas para la detección de anomalías. Incluso dentro de una industria única y específica como la atención médica, diferentes proyectos tendrán diferentes definiciones de lo que hace que un punto de datos sea una anomalía. Las fluctuaciones muy pequeñas en un sistema de seguimiento de la temperatura corporal, por ejemplo, podrían considerarse anomalías, mientras que otros sistemas podrían tolerar una gama mucho más amplia de entradas. Por lo tanto, no es tan fácil aplicar universalmente un solo enfoque como lo es para otros tipos de proyectos de datos.

Para garantizar el éxito de un proyecto que involucre la detección de anomalías, los miembros del equipo de datos deberán trabajar directamente juntos, colaborando con otros equipos que no sean de datos (negocios, operaciones, legal, etc., según el dominio) para:

Definir y refinar continuamente lo que constituye una anomalía. Puede cambiar constantemente, lo que significa una reevaluación continua.

Determinar, una vez detectada una anomalía, qué hará el sistema a continuación. Por ejemplo, envíe anomalías a otro equipo para su posterior análisis y revisión, acciones automáticas en un activo/cuenta asociado, etc..

Definir objetivos y parámetros para el proyecto en general. Por ejemplo, el objetivo final probablemente no sea solo detectar anomalías, sino algo más grande que impacta en el negocio, como bloquear cobros fraudulentos, curar más pacientes al detectar condiciones de salud antes, aumentar los ingresos al anticipar tendencias futuras, etc. Tener metas más grandes le permitirá definir mejor el alcance del proyecto y el resultado esperado.

Desarrolle un plan para monitorear y evaluar el éxito del sistema en el futuro.

Identifique la frecuencia de detección de anomalías (en tiempo real o por lotes) es adecuada para el negocio y el caso de uso en cuestión.

2 Obtención de datos

Tener tantos datos para la detección de anomalías como sea posible permitirá modelos más precisos porque nunca se sabe qué características pueden ser indicativas de una anomalía. El uso de múltiples tipos y fuentes de datos es lo que permite a una empresa ir más allá de las anomalías puntuales para identificar anomalías contextuales o colectivas más sofisticadas. En otras palabras, la variedad es clave.

Por ejemplo, al observar la detección de fraudes, es posible que los datos de transacciones no sean anómalos porque el estafador se ha mantenido dentro del rango "normal" de los hábitos reales del usuario. Pero los datos del uso de cajeros automáticos o los weblogs de cuentas pueden revelar anomalías.

IR MÁS ALLA

En aras de la simplicidad, esta guía analizará un caso simple de detección de anomalías, más específicamente un ejemplo de detección de fraude donde el objetivo es predecir si una transacción de pago móvil es fraudulenta o no.

El conjunto de datos teórico contiene varios campos de información sobre las transacciones en sí, los destinatarios y los clientes que realizaron los pagos. El esquema quedaría así:

```
transaction_id | transaction_date | transaction_type | transaction_amount | recipient_
id | recipient_country | client_ip_address | client_card_mask | client_card_country |
client_email_address | is_fraudulent
```

En un contexto supervisado, la columna `is_fraudulent` representa la variable de destino, es decir, el estado real de la transacción (fraudulenta o legítima).

3 Explorar, limpiar y enriquecer los datos

Al realizar la detección de anomalías, esta etapa es incluso más importante que lo habitual, porque a menudo los datos contienen ruido (generalmente errores, ya sean humanos o no) que tienden a ser similares a las anomalías reales. Por lo tanto, es fundamental distinguir entre los dos y eliminar cualquier dato problemático que pueda producir falsos positivos.

En un mundo ideal, tendría una cantidad suficiente de datos etiquetados para comenzar; es decir, podrá enriquecer los conjuntos de datos que tiene con información sobre qué registros representan anomalías y cuáles son normales. Si es posible, comenzar con los datos que sabe que son anómalos o normales es la forma preferida de comenzar a construir un sistema de detección de anomalías porque será el camino más simple a seguir, permitiendo métodos supervisados con clasificación (a diferencia de los métodos de detección de anomalías no supervisados).

Para algunos de los casos de uso detallados anteriormente, es probable que esto sea muy factible. Específicamente en finanzas para la detección de fraudes o fabricación/industria para el mantenimiento predictivo porque existe un mecanismo claro para la retroalimentación sobre qué casos son anómalos (datos del administrador de relaciones con el cliente que detallan las quejas de fraude o registros de mantenimiento). Para otros casos de uso, como el monitoreo de condiciones en el cuidado de la salud o la detección de intrusiones, tener suficientes datos etiquetados para comenzar puede ser difícil, aunque aún es posible detectar anomalías sin datos etiquetados.

IR MÁS ALLA

En la mayoría de los casos de uso de ciencia de datos, y especialmente en la detección de anomalías, la parte de preparación de datos puede tener un impacto significativo en el resultado del modelo. La aplicación de los pasos de procesamiento correctos y las características relevantes de la ingeniería son, de hecho, muy buenas formas de caracterizar mejor los posibles valores atípicos.

En el caso de la detección de fraude y dado este conjunto de datos en particular, se pueden realizar varias operaciones útiles en el conjunto de datos inicial para crear información adicional sobre cada transacción, por ejemplo:

- Analizar `transaction_date` y extraer características de fecha (por ejemplo, día de la semana, semana del año).
- Derivar el país del cliente de `client_ip_address` via IP geolocalización.
- Extraiga el dominio de correo electrónico de `client_email_address`.

Las operaciones de ingeniería de características más avanzadas también resultarán útiles aquí. En el caso de la detección de fraudes, es común calcular las funciones de clasificación aprovechando las funciones de ventana (realizando un cálculo en un conjunto de filas de la tabla que de alguna manera están relacionadas con la fila actual).

Por ejemplo, en este caso de detección de fraude, es importante saber cuántas direcciones IP diferentes utilizó un cliente determinado (identificado por su dirección de correo electrónico) para sus compras. La consulta de PostgreSQL correspondiente es:

```
SELECT “transaction_id”,
              SUM(CASE WHEN “first_seen” THEN 1 ELSE 0 END)
            OVER (PARTITION BY “client_email_address” ORDER BY “transaction_date”)
AS “distinct_ip”
FROM(
  SELECT “transaction_id”,
                “client_email_address”,
                “transaction_date”,
              “transaction_date” = MIN(“transac_date”) OVER (PARTITION BY “client_
email_address”, “client_ip_addr”) AS ”first_seen” FROM ”transactions_dataset”
```

Otros ejemplos de funciones típicas de clasificación para la detección de fraude incluyen:

- ¿Cuántas veces un cliente realizó una transacción desde un país determinado?
- ¿Cuántas veces un usuario determinado y un beneficiario determinado estuvieron involucrados en una transacción común?

4 Predecir

Hay dos arquitecturas principales para construir sistemas de detección de anomalías:

Detección supervisada de anomalías, que puede usar si tiene un conjunto de datos etiquetado donde sabe si cada punto de datos es normal o no.

Detección de anomalías no supervisada, donde el conjunto de datos no está etiquetado (es decir, si cada punto de datos es o no una anomalía es poco confiable o desconocido).

Cuando utilice un enfoque supervisado, aplicará un algoritmo de clasificación binaria. Exactamente qué algoritmo es menos importante que asegurarse de tomar las medidas adecuadas con respecto al desequilibrio de clases (es decir, el hecho de que para la detección de anomalías, es muy probable que tenga muchos más casos "normales" que anómalos).

Cuando se utiliza un enfoque no supervisado, hay dos formas de entrenar sus algoritmos:

DETECCIÓN DE NOVEDADES

El conjunto de entrenamiento está hecho exclusivamente de inliers para que el algoritmo aprenda el concepto de "normalidad" (de ahí el prefijo "una clase" que se encuentra en algunos métodos). En el momento de la prueba, los datos también pueden contener valores atípicos. Esto también se conoce como detección semisupervisada.

DETECCIÓN DE OTROS:

El conjunto de entrenamiento ya está contaminado por valores atípicos. Se supone que la proporción de valores atípicos es lo suficientemente pequeña como para que se puedan utilizar algoritmos de detección de novedades. En consecuencia, se espera que esos algoritmos sean lo suficientemente robustos en el momento del entrenamiento para ignorar los valores atípicos y ajustarse solo a los valores internos.

IR MÁS ALLA

En un contexto supervisado, el `is_fraudulent` la etiqueta está disponible y, por lo tanto, el problema de predicción puede verse como una tarea de clasificación binaria que toma una matriz (X) que contiene todas las características en forma numérica y un vector objetivo (y) que representa las etiquetas.

Sin embargo, se deben tomar algunos pasos adicionales porque este caso trata con un problema de alto desequilibrio de clases (es decir, los valores atípicos están muy poco representados):

1. **Durante la fase de validación cruzada, asegúrese de que tanto los conjuntos de entrenamiento como los de prueba tengan la misma proporción de valores atípicos. Esto se puede hacer usando una división estratificada: aquí hay un ejemplo en Python usando la biblioteca Scikit-learn:**

```
from sklearn.model_selection import train_test_split  X_train, X_test,y_train,
y_test = train_test_split(X, y, test_size=0.35, stratify=y, random_state=42)
```

2. **Una opción adicional es utilizar técnicas de remuestreo, es decir, tomar solo un subconjunto de valores atípicos y mantener la población completa de valores atípicos.**

En un caso totalmente sin supervisión, no hay acceso a la `is_fraudulent` etiqueta. En consecuencia, es necesario recurrir a algoritmos especiales de detección de valores atípicos que se entrenan solo en la matriz de características (X) y devuelven una puntuación de anomalía para cada punto de datos en el momento de la evaluación.

Siguiendo el ejemplo de un caso no supervisado, primero entrene el algoritmo Isolation Forest en los datos de la transacción:

```
from sklearn.ensemble import IsolationForest dtr =
IsolationForest().fit(X)  df[“y_scored”] =
                  -dtr.decision_function(X)
```

Luego, calcule el rango de anomalía de cada transacción y agréguelo al conjunto de datos original (en formato pandas DataFrame). Al hacerlo, será posible clasificar las transacciones disminuyendo el nivel de anomalía para acceder directamente a las más sospechosas (según lo definido por el modelo) en la parte superior de la lista:

```
df[“anomaly_rank”] = df[“y_scored”].rank(method=”dense”, ascending=0)
df.sort_values(by=[“anomaly_rank”], ascending=True, inplace=True)
```

En algunos casos, dicho método de clasificación puede mejorarse agregando una columna de daño proyectado, es decir, la cantidad de dinero que se perdería en una transacción determinada si fuera fraudulenta.

5 Visualizar

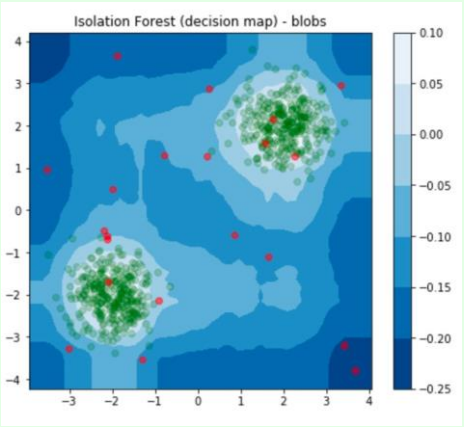
Las visualizaciones son especialmente útiles en el proceso de creación y prueba de modelos de detección de anomalías porque, a veces, son la forma más clara de ver valores atípicos, especialmente en conjuntos de datos muy grandes.

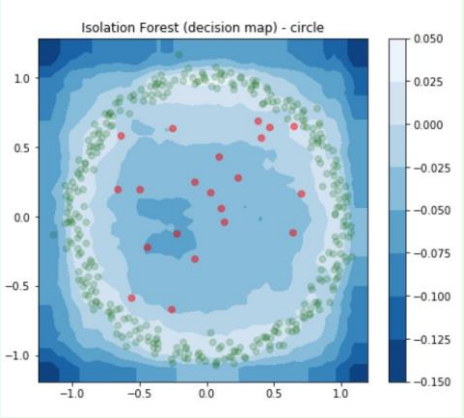
IR MÁS ALLA

En un contexto no supervisado, es posible construir un mapa de decisión donde la puntuación de anomalía se calcula en cada punto de una cuadrícula que se extiende sobre el espacio de características. En la práctica, esto permite la observación de las zonas donde es probable que los inliers se reagrupen de acuerdo con el modelo. Cualquier punto que se encuentre fuera de esas áreas tiene una mayor probabilidad de ser una anomalía.

En los siguientes ejemplos, los mapas de decisión se construyen después de entrenar un algoritmo de bosque de aislamiento en conjuntos de datos bidimensionales simples con varias formas de clúster para los valores internos (en verde) y una partición aleatoriamente uniforme para los valores atípicos (en rojo). Es más probable que los datos ubicados en un área con tonos de azul más oscuros sean una anomalía.

Tenga en cuenta que, en la práctica, los conjuntos de datos de la vida real suelen tener más de dos características, por lo que para poder aplicar una metodología de mapa de decisiones, existe el requisito previo de aplicar la reducción de características al conjunto de datos inicial.





6 Implementar e Iterar

Para tener un impacto real con un sistema de detección de anomalías, su modelo debe tener datos en tiempo real en producción. La detección de anomalías generalmente es sensible al tiempo, por lo que ir a producción para hacer predicciones sobre datos en vivo en lugar de retroactivamente sobre datos de prueba o obsoletos es más importante que nunca.

Pero poner un modelo en producción no es el final. La iteración y el monitoreo de los sistemas de detección de anomalías son fundamentales para garantizar que el modelo continúe aprendiendo y sea lo suficientemente ágil para continuar detectando anomalías incluso cuando cambien los comportamientos de los usuarios. Sin embargo, a diferencia de otros tipos de modelos de aprendizaje automático, la precisión no es una métrica viable para la detección de anomalías. Dado que la gran mayoría de los datos no se componen de anomalías (es decir, podría haber cientos de miles de puntos de datos "normales"), el sistema podría lograr una precisión muy alta, pero aún así no identificaría las anomalías con precisión.

IR MÁS ALLA

En lugar de precisión, los sistemas de detección de anomalías pueden basarse en los siguientes métodos de evaluación:

Recordar:

La proporción de anomalías detectadas correctamente a anomalías totales.

Tasa de falsos positivos

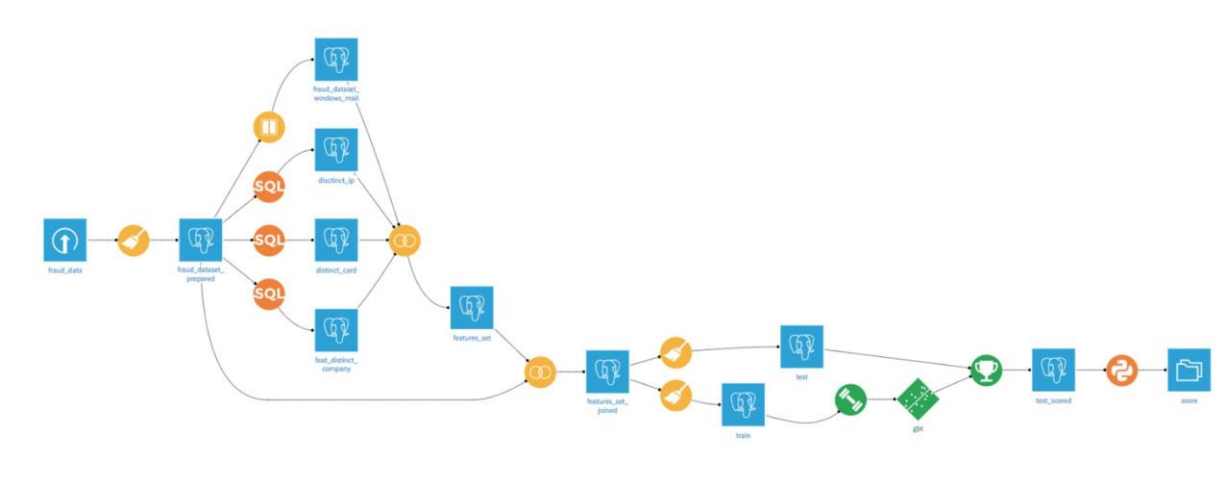
La proporción de anomalías mal clasificadas con respecto al total de registros.

Curva característica del operador del receptor(ROC)

Equilibrio entre detección y tasa de falsas alarmas.

En cuanto a la iteración, tenga en cuenta que, con mucho, el paso más laborioso cuando se trata de la detección de anomalías es la ingeniería de características. Continuar con la iteración hasta que se reduzcan los falsos positivos/negativos y el sistema sea efectivo pero ágil es una parte crítica del proceso que requiere mucho tiempo.

Lo que puede ser útil aquí es tener una representación visual de todo el proceso para que la iteración sea cada vez más simple y rápida, incluso una vez que el modelo está en producción; por ejemplo, aquí hay una descripción general de nuestro ejemplo de detección de fraude en el que trabajamos en Dataiku Data Science Studio (DSS), pero cualquier representación visual puede ser efectiva:



Cuáles son las fallas?

DEMASIADOS FALSOS NEGATIVOS/POSITIVOS:

La evaluación de la detección de anomalías es un equilibrio particularmente delicado. Los falsos negativos pueden ser perjudiciales, por supuesto, pero por otro lado, un sistema que identifica demasiados falsos positivos tampoco sirve. En un sistema en tiempo real, no hay lugar para una segunda revisión de posibles anomalías si el sistema está produciendo demasiado ruido. Y en muchos casos de uso, los falsos positivos en la detección de anomalías podrían destruir la reputación de la empresa y provocar la pérdida de negocios (piense en la frustración, por ejemplo, si su banco bloqueara constantemente sus fondos debido a la detección de falsos positivos de fraude).

Solución: dedicar tiempo por adelantado a la ingeniería de características para asegurarse de que no haya demasiados falsos negativos o positivos y continuar iterando, refinando y realizando mejoras incluso después de que el modelo esté en producción: ambos son críticos.

DATOS FALTANTES O NO CONFIABLES

Un sistema que no es lo suficientemente robusto no es la única causa de falsos positivos. Otro puede ser simplemente datos poco confiables, que desafortunadamente es un problema en muchas industrias diferentes, pero particularmente en el cuidado de la salud y la fabricación.

Solución: Invertir en la mejora de los sistemas y sensores para garantizar datos completos y confiables es un componente esencial de la detección precisa de anomalías. Porque no importa cuán bueno sea su modelo, no funcionará bien si se basa en datos de mala calidad.

FALTA DE AGILIDAD PARA MANEJAR CAMBIOS EN LA NORMATIVA:

En el mundo actual, el cambio es la única constante. Los seres humanos cambian con el tiempo, por lo que la idea de un comportamiento normal en el contexto de la detección de anomalías seguirá cambiando. Además, los sistemas también cambian con el tiempo, pero el cambio gradual no siempre equivale a un comportamiento anómalo.

Solución: cualquier sistema de detección de anomalías que construya, sin importar el caso de uso, debe ser lo suficientemente ágil para adaptarse a las normas cambiantes. De manera similar, debe existir un plan para monitorear y revisar continuamente el sistema para garantizar que aún funcione como se espera con el tiempo. Los sistemas también deben tener en cuenta cualquier estacionalidad u otros patrones; por ejemplo, un cliente en un banco generalmente puede realizar compras más grandes de lo normal durante las festividades, pero estas no deben necesariamente marcarse como fraudulentas.

FALTA DE VINCULO CON LOS OBJETIVOS DEL NEGOCIO:

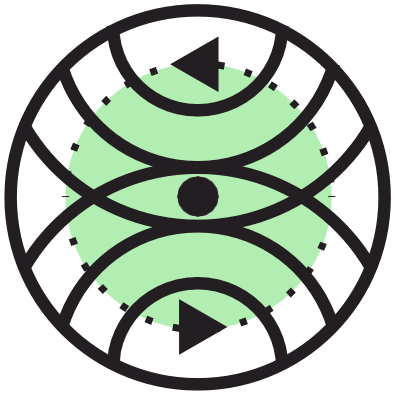
Al igual que muchos sistemas de inteligencia, es fácil desconectarse del lado comercial y desarrollar un sistema que no toma los siguientes pasos adecuados o no realiza un seguimiento en función de las anomalías detectadas.

Solución: ya sea enviar el caso para que lo revise un equipo de servicio al cliente, notificar a un miembro del personal en particular o innumerables otras acciones del siguiente paso, los sistemas de detección de anomalías deben tomar alguna acción al final para ser efectivos. Construir un sistema en el vacío de un equipo de ciencia de datos no servirá de nada.

MIRANDO HACIA EL FUTURO

DATOS NO ESTRUCTURADOS

Debido a la amplitud de su utilidad, especialmente a medida que la actividad fraudulenta y los ataques a los sistemas se generalizan, la detección de anomalías seguirá siendo más sofisticada. Ha habido, y seguirá habiendo, desarrollos en el ámbito de la detección de anomalías con datos no estructurados (como imágenes y texto). Los desarrollos en la detección de anomalías utilizando el aprendizaje profundo en imágenes serán particularmente influyentes en la atención médica.



PREVENCIÓN AUTOMATIZADA

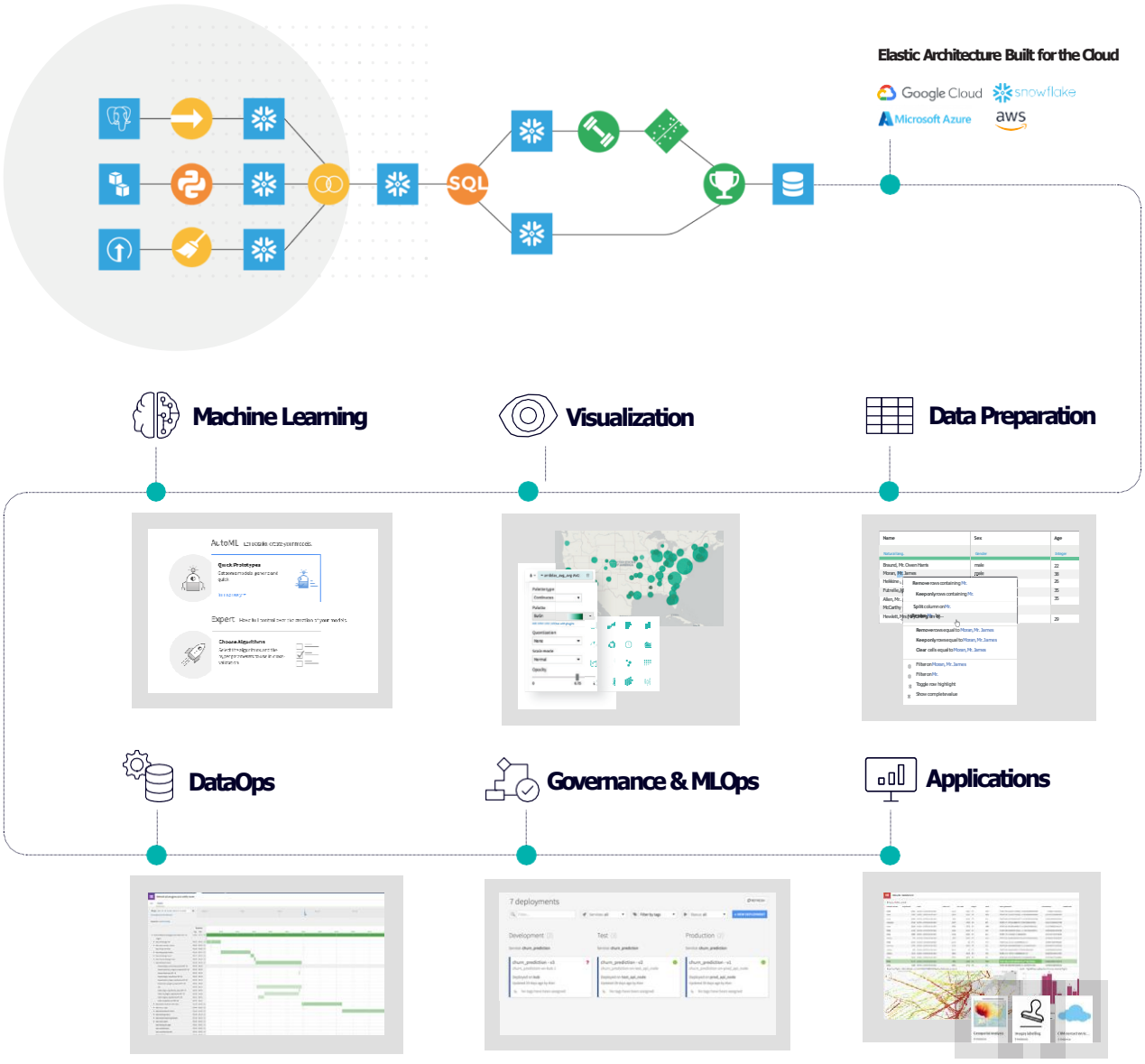
Además, los desarrollos irán desde la detección pura de anomalías hasta técnicas de prevención automatizadas. Ser capaz de detener el comportamiento anómalo potencialmente dañino antes de que suceda tiene el potencial de tener un amplio impacto en los próximos años, particularmente, nuevamente, cuando se trata de seguridad cibernética y detección de fraude.

PROCESOS MÁS ESCALABLES

Y finalmente, en los próximos años, a medida que más y más industrias dependan de la detección de anomalías, espere ver una simplificación general de los procesos de detección de anomalías a medida que las empresas continúan escalando. Esto significa que cada vez más empresas invierten en la arquitectura adecuada para recuperar datos críticos para el trabajo de detección de anomalías, los medios para procesarlos rápidamente y aplicar modelos para el impacto en la producción.

AI todos los días

Gente Extraordinaria



45,000+
USUARIOS ACTIVOS

450+
CLIENTES

Dataiku es la plataforma de AI, que sistematiza el uso de datos para obtener resultados comerciales excepcionales. Las organizaciones que usan Dataiku elevan a su gente armándolos con la capacidad de tomar mejores decisiones diarias con datos.

SOLEX

¿Tienes alguna necesidad en tu organización o iniciativa relacionada con Ciencia de Datos, Machine Learning o Gobierno de Datos? En Solex somos expertos en brindar soluciones confiables e innovadoras enfocados en Business Analytics. Comuníquese con nosotros a través del formulario de contacto o escríbanos al Whatsapp (+57 316 4576123). Más información en: www.solex.biz/dataiku

www.solex.biz/dataiku | info@solex.biz

Traducción cortesía Solex. Ver fuente del ebook https://pages.dataiku.com/anomaly-detection-at-scale-guidebook