pitney bowes

Spectrum™ Technology Platform

Version 12.0

Guía de Machine Learning

Contents

1 - Introducción

In this section

Módulo Machine Learning (Vista previa de la tecnología)

El equipo de Spectrum se complace en compartir con usted una vista previa de nuestro potente nuevo módulo: Machine Learning. Esta vista previa de la tecnología es una implementación anticipada de Machine Learning que contiene las funciones esenciales que consideramos las más útiles para usted. Queremos agregar nuevas funciones importantes en versiones futuras, por lo que ponemos a su disposición esta vista previa de la tecnología para que podamos asegurarnos de agregar las funciones que desea. Sus comentarios nos permitirán orientar la evolución del módulo Machine Learning. A medida que explora Machine Learning, tenga en cuenta lo siguiente:

- Esta vista previa de la tecnología contiene un conjunto limitado de funciones. Cuando descubra alguna función que le gustaría tener disponible, llene una solicitud de mejora mediante la asistencia técnica para informarnos. Sus sugerencias determinarán qué funciones agregaremos en las versiones futuras. Para obtener información acerca de cómo comunicarse con la asistencia técnica, consulte www.pitneybowes.com/us/contact-dcs.html.
- Incluso los mejores softwares tienen algunos errores. Si encuentra un error, envié un informe de errores a la asistencia técnica para informarnos. Como se trata de una vista previa de la tecnología, no podemos garantizarle que resolvamos su problema específico de inmediato. Para obtener información acerca de cómo comunicarse con la asistencia técnica, consulte www.pitneybowes.com/us/contact-dcs.html.
- No dude en utilizar esta vista previa de la tecnología en su entorno de producción. Tenga en cuenta que no podemos respetar los acuerdos de nivel de servicio (SLA) normales para las vistas previas de la tecnología.
- Esperamos obtener comentarios imprevistos e interesantes que tengan una repercusión drástica en la próxima versión de Machine Learning, por lo que no podemos garantizar que podrá conservar todo el trabajo que realiza con Machine Learning cuando actualice a versiones futuras.
- Use su criterio en el momento de tomar decisiones empresariales basadas en análisis generados con esta vista previa de la tecnología. No podemos respetar los acuerdos a nivel de servicio (SLA) normales en el caso de las funciones que se encuentran en la vista previa de la tecnología.

Esperamos que disfrute experimentar con el módulo Machine Learning y nos envíe sus comentarios.

Un primer vistazo a Machine Learning

El módulo Machine Learning de Spectrum™ Technology Platform ofrece la capacidad de ajustar modelos de machine learning supervisados y no supervisados.

Nota: El módulo Machine Learning solo se admite en sistemas operativos Windows y Linux.

Binning

Binning divide los registros en grupos (contenedores) para una variable continua sin considerar la información de objetivos. Puede ejecutar binning no supervisado de una de estas dos maneras: usando contenedores del mismo ancho o contenedores de la misma frecuencia.

K-Means Clustering

K-Means Clustering permite crear modelos según agrupaciones en clústeres analíticas, lo cual segmenta un conjunto de registros en clústeres de registros similares a partir de valores de datos.

Logistic Regression

Logistic Regression crea modelos a partir de conjuntos de datos que usan objetivos binarios con variables de entrada.

Java Model Scoring

Esta característica evalúa nuevos datos usando la fórmula creada cuando ajusta un modelo de machine learning.

Administración de modelo de aprendizaje automático

Machine Learning Model Management le permite administrar todos los modelos de machine learning en su servidor Spectrum™ Technology Platform. Puede exponer, no exponer o eliminar modelos. Además, puede ver información detallada de cada modelo y comparar cualquier par de modelos del mismo tipo.

> **Nota:** El módulo Machine Learning utiliza una biblioteca H2O.ai subyacente para modelado de algoritmos en K-Means Clustering, Logistic Regression y Java Model Scoring.

Un flujo de trabajo de Machine Learning

Un flujo de trabajo de Machine Learning típico implica los siguientes pasos que se deben realizar en uno o más flujos de trabajo:

1. Acceda a los datos usando otros módulos de Spectrum, como Data Integration.
2. Prepare los datos usando etapas de otros módulos de Spectrum, como las de Data Integration, Data Quality y Core.
3. Ajuste un modelo de Machine Learning, ejecute el flujo de datos y, luego, revise el contenido de la pestaña Salida del modelo en la etapa del modelo. Después, puede retocar el modelo, si es necesario, y volver a ejecutar el flujo de datos. A continuación, debe revisar el conjunto completo de datos de salida de evaluación del modelo en la herramienta Gestión de modelos de Machine Learning. Puede revisar un modelo a la vez o comparar dos modelos.

4. (Opcional) Si usará el modelo para evaluar datos, exponga el modelo en la herramienta Gestión de modelos de Machine Learning, que pone el modelo a disposición de la etapa Java Model Scoring.
   a. Cree un flujo de datos de Spectrum™ Technology Platform siguiendo los pasos 1 y 2 anteriores y luego, reemplace el paso 3 por la etapa Java Model Scoring. Configure este flujo de datos para ejecutarlo en modo de lote a fin de completar un archivo con evaluaciones de modelo aplicadas a datos actualizados (los campos utilizados como X o datos de entrada se actualizan en el paso 1-2 como una parte natural de hacer negocios).
   b. De manera alternativa, use un servicio web en Spectrum™ Technology Platform para evaluar datos según demanda. Por ejemplo, acceda al sitio web, obtenga la ID de cliente y los datos de entrada del modelo, evalúelos y devuelva la evaluación a un proceso que personaliza el contenido web para su cliente.
5. (Opcional) También puede implementar evaluaciones de modelo en una base de datos de gráficos Data Hub como una propiedad de entidad, en mapas o en aplicaciones CES.

2 - Binning

In this section

Introducción a Binning

La etapa Binning realiza lo que se conoce como binning supervisado, que divide una variable continua en grupos (contenedores) sin considerar la información de objetivos. Los datos capturados incluyen rangos, cantidades y porcentaje de valores dentro de cada rango.

Las ventajas de ejecutar binning incluyen:

- Permite incluir registros con datos faltantes en el modelo.
- Controla o mitiga el impacto de valores atípicos en el modelo.
- Soluciona el problema de tener escalas diferentes entre las características, permitiendo que las ponderaciones de los coeficientes en el modelo final se puedan comparar.

En binning no supervisado de Spectrum™ Technology Platform, puede usar contenedores del mismo ancho, donde los datos se dividen en contenedores de igual tamaño, o contenedores de la misma frecuencia, donde los datos se dividen en grupos que contienen aproximadamente el mismo número de registros. En la etapa Binning, los contenedores que tienen el mismo ancho se denominan contenedores de Rango igual y los contenedores que tienen la misma frecuencia se denominan contenedores de Completación igual.

Configuración de opciones de binning

1. Seleccione si desea realizar un **estilo de binning** de rango o de población equivalente.
2. En **Intervalo de valor nulo**, seleccione cómo desea manejar los campos bin vacíos, que representan valores desconocidos debido a datos faltantes. Seleccione **El más alto** para asignar valores nulos al bin más alto y seleccione **El más bajo** para asignar valores nulos al bin más bajo. El bin más bajo siempre será el bin 1.
3. Haga clic en **Bines internos objetivo** e ingrese la cantidad de bines que desea completar entre los bines finales. Si realiza binning de rango equivalente, puede seleccionar este tipo de procesamiento o el **Ancho de bin**, pero no ambos. Si realiza binning de población equivalente, solo pueden realizar el procesamiento de bin interno.
4. Si realizar binning de rango equivalente y desea seleccionar este tipo de procesamiento en lugar del procesamiento de bin interno, haga clic en **Ancho de bin** e ingrese la cantidad de unidades que desea en cada bin.
5. Haga clic en **Incluir** para cada campo cuyos datos desea agregar al binning. Tenga en cuenta que solo aparecerán campos numéricos en esta lista.
6. Haga clic en **Aceptar** para guardar la configuración.

Salida Binning

La etapa Binning tiene dos puertos de salida. El primer puerto enviará todos los campos de entrada más un campo de bin por cada campo de entrada seleccionado. Por ejemplo, si la entrada contiene los campos Nombre, Edad e Ingresos y se realiza binning en los campos Edad e Ingresos, la salida del primer puerto contendrá los siguientes campos:

- Nombre
- Edad
- Binned_Age
- Ingresos
- Binned_Income

El segundo puerto envía cuatro tipos de información por cada campo de entrada seleccionado. Por ejemplo, si realiza binning en el campo Edad, la salida del segundo puerto contendrá los siguientes campos:

- Age_Bins
- Age_BinValue
- Age_Count
- Age_Percentage

3 - K-Means Clustering

In this section

Introducción a K-Means Clustering

K-Means Clustering permite crear modelos según agrupaciones en clústeres analíticas, lo cual segmenta un conjunto de registros en clústeres de registros similares a partir de valores de datos.

Para crear su modelo, primero debe completar la ficha Propiedades del modelo. Las fichas Opciones básicas y Opciones avanzadas ofrecen configuraciones predeterminadas suficientes para completar un trabajo, pero puede cambiarlas de acuerdo con sus necesidades. Luego se puede ejecutar el trabajo y una versión limitada de los detalles de salida de modelo resultantes aparece en la pestaña Salida de modelo; el modelo es almacenado en el servidor de Spectrum™ Technology Platform y la salida completa se encuentra disponible en la herramienta de gestión de modelos de Machine Learning.

Definición de las propiedades del modelo

1. En **Etapas principales**/**Etapas implementadas**/**Machine Learning**, haga clic en la etapa **K-Means Clustering** y arrástrela hasta el lienzo, colóquela donde desee en el flujo de datos y conéctela con otras etapas. Tenga en cuenta que la etapa de entrada debe ser el origen de datos que contiene los campos de variables de entrada de su modelo. No se requiere una etapa de salida, a menos que seleccione la opción Calificar datos de entrada en la pestaña Opciones básicas. También puede conectar una etapa de salida si desea capturar su salida, independiente de la herramienta de gestión de modelo Machine Learning.
2. Haga doble clic en la etapa K-Means para que aparezca el cuadro de diálogo **Opciones de K-Means Clustering**.
3. Ingrese un **Nombre de modelo** si no desea utilizar el nombre predeterminado.
4. Opcional: marque la casilla **Sobrescribir** para sobrescribir el modelo existente con datos nuevos.
5. Ingrese el **Número de agrupamiento** que desea en su modelo si no quiere el número predeterminado (5).
6. Opcional: Ingrese una **Descripción** del modelo.
7. Haga clic en **Incluir** para cada campo cuyos datos desea agregar al modelo.
8. Utilice la lista desplegable **Tipo de datos de modelo** para especificar si el campo de entrada se debe utilizar como un campo numérico, categórico, o de fecha y hora.
9. Haga clic en **Aceptar** para guardar el modelo y la configuración, o continúe a la ficha siguiente.

Configuración de opciones básicas

1. Deje marcada la opción **Estandarizar campos de entrada** para estandarizar las columnas numéricas a fin de que la variación media y por unidad sea igual a cero.

   Si no utiliza la estandarización, los resultados podrían incluir componentes dominados por variables que aparentarán tener variaciones mayores en relación con otros atributos como una cuestión de escala y no como una contribución verdadera.
2. Revise el **Número estimado de agrupamiento** para hacer que el algoritmo de K-Means intente determinar el número de agrupamiento que contendrá el modelo. Aunque designe el número de agrupamiento deseado en la pestaña Propiedades del modelo, la rutina podría descubrir durante su procesamiento que un número de agrupamiento diferente resulta más apropiado en vista de los datos.
3. Especifique un valor entre 1 y 100 como **Porcentaje para datos de capacitación** cuando los datos de entrada se dividen aleatoriamente en muestras de datos de capacitación y de prueba.
4. Ingrese el valor de 100 menos la cantidad que ingresó en el Paso 5 como **Porcentaje para datos de prueba**.
5. Ingrese un número en **Propagar para muestras** para garantizar que cuando los datos se dividan en datos de prueba y de capacitación, esto ocurra siempre de la misma manera cada vez que ejecute el flujo de datos. Deje “0” en este campo para obtener una división aleatoria cada vez que ejecuta el flujo.
6. Haga clic en **Aceptar** para guardar el modelo y la configuración, o continúe a la ficha siguiente.

Configuración de opciones avanzadas

1. Deje marcada la opción **Ignorar campos constantes** para omitir campos que tienen el mismo valor para cada registro.
2. Seleccione el modo de inicialización correcto en el menú desplegable **Inic**.

   Más lejano — Inicializa el primer centroide al azar, pero luego inicializa el segundo centroide para que sea el punto de datos más lejano de él. Inicializa los centroides para que queden bien separados entre sí.

   Plus-Plus — Inicializa los centros de clúster antes de proceder con las iteraciones de optimización *k*-means estándar. Con la inicialización *k*-means++, se garantiza que el algoritmo encuentre una solución que es O (registro *k*) competitiva con la solución *k*-means óptima.

Aleatorio	Opción predeterminada. Elige clústeres K del conjunto de observaciones N en forma aleatoria, de manera que cada observación tenga la misma posibilidad de ser elegida.

3. Deje marcada la opción **Propagar para N iteraciones** e ingrese el número de propagación para garantizar que cuando los datos se dividan en datos de prueba y de capacitación, esto ocurra siempre de la misma manera cada vez que ejecute el flujo de datos. Deje “0” en este campo para obtener una división aleatoria cada vez que ejecuta el flujo.
4. Marque la opción **N iteraciones** e ingrese la cantidad de iteraciones si va a realizar una validación cruzada.
5. Marque la opción **Asignación de iteración** y seleccione la lista despegable si está ejecutando una validación cruzada. Este campo solo se aplica si ingresó un valor en **N iteraciones**.

AUTO	Opción predeterminada. Permite que el algoritmo seleccione automáticamente una opción; actualmente utiliza Aleatorio.
Módulo	Divide de manera uniforme el conjunto de datos en las iteraciones y no depende de la raíz.
Aleatorio	Divide de manera aleatoria los datos en piezas de n iteraciones; es ideal para grandes conjuntos de datos.
Estratificado	Estratifica las iteraciones en función de la variable de respuesta para los problemas de clasificación. Distribuye de manera uniforme las observaciones de las diferentes clases en todos los conjuntos mediante la división de un conjunto de datos en datos de capacitación y de prueba. Puede resultar útil si hay muchas clases y el conjunto de datos es relativamente pequeño.

6. Marque la opción **Iteraciones máximas** e ingrese el número de iteraciones de capacitación que deben ocurrir.
7. Haga clic en **Aceptar** para guardar el modelo y la configuración, o continúe a la ficha siguiente.

Salida de modelo

Esta pestaña muestra las métricas que está usando para evaluar el modelo ajustado. No puede editar estos campos. La columna Capacitación siempre va a contener datos. Si seleccionó la división capacitación/prueba en la pestaña Opciones básicas, la columna Prueba también se completará, a menos que haya seleccionado una validación de N iteraciones en la pestaña Opciones avanzadas, en cuyo caso se completará la columna N iteraciones. Haga clic en el botón **Salida** para regenerar los datos de salida y, luego, en **Para obtener detalles, haga clic aquí** para ver los datos de salida completos en la herramienta Machine Learning Model Management.

4 - Logistic Regression

In this section

Introducción a Logistic Regression

Logistic Regression le permite realizar aprendizaje de máquina mediante la creación de modelos a partir de conjuntos de datos que usan objetivos binarios con variables de entrada.

Para crear su modelo, primero debe completar la ficha Propiedades del modelo. Las fichas Opciones básicas y Opciones avanzadas ofrecen configuraciones predeterminadas suficientes para completar un trabajo, pero puede cambiarlas de acuerdo con sus necesidades. Luego se puede ejecutar el trabajo y una versión limitada del modelo resultante aparece en la pestaña Salida de modelo; la salida completa se encuentra disponible en la herramienta de gestión de modelos de Machine Learning.

Definición de las propiedades del modelo

1. En **Etapas principales**/**Etapas implementadas**/**Machine Learning**, haga clic en la etapa **Logistic Regression** y arrástrela hasta el lienzo, colóquela donde desee en el flujo de datos y conéctela con otras etapas. Tenga en cuenta que la etapa de entrada debe ser el origen de datos que contiene los campos de variables objetivo y de entrada de su modelo. No se requiere una etapa de salida, a menos que seleccione la opción Calificar datos de entrada en la pestaña Opciones básicas. También puede conectar una etapa de salida si desea capturar su salida, independiente de la herramienta de gestión de modelo Machine Learning.
2. Haga doble clic en la etapa Logistic Regression para que aparezca el cuadro de diálogo **Opciones de Logistic Regression**.
3. Ingrese un **Nombre de modelo** si no desea utilizar el nombre predeterminado.
4. Opcional: marque la casilla **Sobrescribir** para sobrescribir el modelo existente con datos nuevos.
5. Haga clic en la opción desplegable **Campo objetivo** y seleccione "Categórico".
6. Opcional: Ingrese una **Descripción** del modelo.
7. Haga clic en **Incluir** para cada campo cuyos datos desea agregar al modelo.
8. Utilice la lista desplegable **Tipo de datos de modelo** para especificar si el campo de entrada se debe utilizar como un campo numérico, categórico, o de fecha y hora.
9. Haga clic en **Aceptar** para guardar el modelo y la configuración, o continúe a la ficha siguiente.

Configuración de opciones básicas

1. Deje marcada la opción **Estandarizar campos de entrada** para estandarizar las columnas numéricas a fin de que la variación media y por unidad sea igual a cero.

   Si no utiliza la estandarización, los resultados podrían incluir componentes dominados por variables que aparentarán tener variaciones mayores en relación con otros atributos como una cuestión de escala y no como una contribución verdadera.
2. Marque la opción **Calificar datos de entrada** para agregar una columna para la predicción del modelo (calificación) a los datos de entrada.
3. Marque **Anterior** si se tomaron muestras de los datos y la media de respuesta no refleja la realidad; luego, ingrese la probabilidad anterior para p(y==1) en el campo de texto.
4. Para especificar cómo manejar los datos faltantes, marque **Omitir** o **Imputar medios**, que agregará el valor medio para cualquier dato faltante.
5. Especifique un valor entre 1 y 100 como **Porcentaje para datos de capacitación** cuando los datos de entrada se dividen aleatoriamente en muestras de datos de capacitación y de prueba.
6. Ingrese el valor de 100 menos la cantidad que ingresó en el Paso 5 como **Porcentaje para datos de prueba**.
7. Ingrese un número en **Propagar para muestras** para garantizar que cuando los datos se dividan en datos de prueba y de capacitación, esto ocurra siempre de la misma manera cada vez que ejecute el flujo de datos. Deje “0” en este campo para obtener una división aleatoria cada vez que ejecuta el flujo.
8. Haga clic en **Aceptar** para guardar el modelo y la configuración, o continúe a la ficha siguiente.

Configuración de opciones avanzadas

1. Deje marcada la opción **Ignorar campos constantes** para omitir campos que tienen el mismo valor para cada registro.
2. Deje marcada la opción **Calcular valores de p** para calcular valores de p para las estimaciones de parámetros.
3. Deje marcada la opción **Quitar columna alineada** para quitar automáticamente las columnas alineadas durante la construcción del modelo. Esto dará como resultado un coeficiente de 0 en el modelo devuelto.

   Esta opción debe estar marcada si la opción **Calcular valores de p** también está marcada.
4. Deje marcada la opción **Incluir término constante (interceptar)** para incluir un término constante (interceptar) en el modelo.

Debe marcar este campo si también marca la opción **Quitar columna alineada**.

5. Seleccione un **Solucionador** de la lista desplegable. Tenga en cuenta que COORDINATE_DESCENT y COORDINATE_DESCENT_NAIVE se encuentran en etapa experimental.

AUTO	El solucionador se determinará en función de los datos y parámetros de entrada.
COORDINATE_DESCENT	IRLSM con la versión de actualizaciones de covarianza del descenso cíclico por coordenadas en el bucle interior.
COORDINATE_DESCENT_NAIVE	IRLSM con la versión de actualizaciones naive del descenso cíclico por coordenadas en el bucle interior.
IRLSM	Ideal para problemas con una pequeña cantidad de predictores o para búsquedas Lambda con penalidad L1.
L_BFGS	Ideal para conjuntos de datos con muchas columnas.

6. Deje marcada la opción **Propagar para N iteraciones** e ingrese el número de propagación para garantizar que cuando los datos se dividan en datos de prueba y de capacitación, esto ocurra siempre de la misma manera cada vez que ejecute el flujo de datos. Deje “0” en este campo para obtener una división aleatoria cada vez que ejecuta el flujo.
7. Marque la opción **N iteraciones** e ingrese la cantidad de iteraciones si va a realizar una validación cruzada.
8. Marque la opción **Asignación de iteración** y seleccione la lista despegable si está ejecutando una validación cruzada. Este campo solo se aplica si ingresó un valor en **N iteraciones** y no se especificó el **Campo de iteración**.

AUTO	Permite que el algoritmo seleccione automáticamente una opción; actualmente utiliza Aleatorio.
Módulo	Divide de manera uniforme el conjunto de datos en las iteraciones y no depende de la raíz.
Aleatorio	Divide de manera aleatoria los datos en piezas de n iteraciones; es ideal para grandes conjuntos de datos.
Estratificado	Estratifica las iteraciones en función de la variable de respuesta para los problemas de clasificación. Distribuye de manera uniforme las observaciones de las diferentes clases en todos los conjuntos mediante la división de un conjunto de datos en datos de capacitación y de prueba. Puede resultar útil si hay muchas clases y el conjunto de datos es relativamente pequeño.

9. Si está ejecutando una validación cruzada, marque la opción **Campo de iteración** y seleccione el campo que contiene la asignación del índice de iteración de validación cruzada en la lista desplegable.

   Este campo solo se aplica si no ingresó un valor en **N iteraciones** y **Asignación de iteración**.

10. Marque la opción **Iteración máxima** e ingrese el número de iteraciones de capacitación que deben ocurrir.
11. Marque la opción **Objetivo épsilon** e ingrese el umbral de convergencia; este debe ser un valor entre 0 y 1. Si el valor objetivo es menor que este umbral, el modelo se convergerá.
12. Marque la opción **Beta épsilon** e ingrese el umbral de convergencia; este debe ser un valor entre 0 y 1. Si el valor objetivo es menor que este umbral, el modelo se convergerá. Si la normalización L1 del cambio beta actual está por debajo de este umbral, considere el uso de la convergencia.
13. Haga clic en **Aceptar** para guardar el modelo y la configuración, o continúe a la ficha siguiente.

Salida de modelo

Esta pestaña muestra las métricas que está usando para evaluar el modelo ajustado. No puede editar estos campos. La columna Capacitación siempre va a contener datos. Si seleccionó una división capacitación/prueba en la pestaña Opciones básicas, también se completará la columna Prueba, a menos que haya seleccionado una validación de N iteraciones en la pestaña Opciones avanzadas, en cuyo caso se completará la columna N iteraciones.

Después de ejecutar su trabajo, el modelo resultante se guarda en el servidor Spectrum™ Technology Platform. Haga clic en el botón **Salida** para regenerar los datos de salida y luego, en **Para obtener detalles, haga clic aquí** para ver los datos de salida completos en la herramienta Machine Learning Model Management.

5 - Java Model Scoring

In this section

Introducción a Java Model Scoring

Java Model Scoring le permite evaluar nuevos datos usando la fórmula creada cuando ajusta un modelo de machine learning.

> **Nota:** Primero debe exponer los modelos a través de Machine Learning Model Management antes de que queden disponibles en la etapa Java Model Scoring. Consulte **Introducción a Machine Learning Model Management** en la página 23 para obtener más información.

Para evaluar sus datos, debe completar las dos pestañas del cuadro de diálogo **Opciones de Java Model Scoring**. Primero, identifique el modelo y su tipo y, después, asegúrese de que los campos del modelo se asignen correctamente a los campos de Spectrum™ Technology Platform. Luego, configure la salida seleccionando los campos que desea incluir y ejecute su trabajo. La pestaña **Salida de modelo** contiene mapas para tipos de datos para Spectrum™ Technology Platform y su modelo.

Si su trabajo contiene una etapa que captura los datos de salida en un archivo o una tabla, puede usar esos datos de salida en un flujo de datos o servicio web subsiguiente.

Definición de las propiedades del modelo

1. En **Etapas principales**/**Etapas implementadas**/**Advanced Analytics**, haga clic en la etapa **Java Model Scoring** y arrástrela hasta el lienzo, colóquela donde desee en el flujo de datos y conéctela con las etapas de entrada y salida. Observe que la etapa de entrada debe ser la fuente de datos que contiene los campos de variables objetivo y de entrada para su modelo. Si está ejecutando su trabajo en modo de lote, también necesitará una etapa de salida para capturar calificaciones de modelo; en caso contrario, usará un servicio web de Spectrum™ Technology Platform para calificar los datos en tiempo real.
2. Haga doble clic en la etapa Java Model Scoring para que aparezca el cuadro de diálogo **Opciones de Model Scoring**.
3. Opcional: seleccione el tipo de modelo que está evaluando en la lista desplegable **Filtro de tipo**.
4. Seleccione el **Filtro de tipo** que está usando para evaluar el modelo.
5. Seleccione el **Nombre de modelo** de la lista desplegable.
6. Ingrese el tipo de modelo que está evaluando en el campo **Tipo de modelo**.
7. Opcional: Ingrese una **Descripción** del modelo.
8. La tabla **Entradas** muestra información de los campos de entrada del modelo. Estos campos y sus tipos de datos se asignan automáticamente a campos y tipos de datos de Spectrum.
9. Haga clic en **Aceptar** para guardar estas opciones o continúe a la ficha siguiente.

Salida de modelo

La tabla **Salidas** muestra información de los campos de salida del modelo. Estos campos y sus tipos de datos se asignan automáticamente a campos y tipos de datos de Spectrum.

1. Haga clic en **Incluir** para cada campo cuyos datos desea agregar a la salida del modelo.
2. Haga clic en **Aceptar** para guardar el modelo.

6 - Administración de modelo de aprendizaje automático

In this section

Introducción a Machine Learning Model Management

La pestaña Análisis de modelo en Machine Learning Model Management muestra una lista de todos los modelos de Machine Learning en su servidor Spectrum™ Technology Platform. Puede filtrar esta lista ingresando una cadena en el cuadro de texto; se buscará esa cadena en cada campo de la tabla.

Puede realizar varias operaciones en estos modelos. Puede exponer, no exponer o eliminar modelos. Los modelos expuestos se usan en la etapa Java Model Scoring para evaluar nuevos datos usando fórmulas creadas cuando ajusta los modelos de Machine Learning. Además, puede ver información detallada de cada modelo; los detalles devueltos dependen del tipo de modelo cuyos datos está visualizando. Para terminar, puede comparar cualquier par de modelos del mismo tipo. Esta comparación muestra, lado a lado, la misma información que aparece en la pestaña Detalle de modelo de cada uno de los modelos que está comparando.

Acceder al análisis de modelo de la gestión de modelos de Machine Learning

Hay tres maneras de acceder a la gestión de modelos de Machine Learning:

- Use la página de bienvenida de Spectrum™ Technology Platform:
  - Abra un navegador web y acceda a la página de bienvenida de Spectrum™ Technology Platform:

    http://<nombre del servidor>:<puerto>

    Por ejemplo, si instaló Spectrum™ Technology Platform en una computadora denominada "MiPlataformaSpectrum" y utiliza el puerto HTTP predeterminado 8080, accederá a:

    http://MiPlataformaSpectrum:8080

  - Haga clic en **Spectrum Machine Learning**.
  - Haga clic en **Abrir repositorio de Machine Learning**.
- Haga clic en **Para obtener más detalles haga clic aquí** desde una de las etapas de la generación de modelos.
- Use un navegador web:
  - Abra un navegador web y vaya a la página de gestión de modelos de Machine Learning de Spectrum™ Technology Platform:

    http://<servername>:<port>/machinelearning

    Por ejemplo, si instaló Spectrum™ Technology Platform en una computadora denominada "MiPlataformaSpectrum" y utiliza el puerto HTTP predeterminado 8080, accederá a:

    http://myspectrumplatform:8080/machinelearning

- Ingrese un nombre de usuario y contraseña de Spectrum™ Technology Platform válidos.
- Cuando se abra la herramienta, haga clic en la pestaña **Análisis de modelo**.

Operaciones de análisis de modelo en Gestión de modelo

Realice estas operaciones seleccionando un modelo y haciendo clic en el botón correspondiente:

	Exponga el modelo para ponerlo a disposición de la etapa Java Model Scoring. Si no expone el modelo, no lo puede usar para evaluación.
	Anule la exposición del modelo.
	Elimine el modelo. **Nota:** No puede eliminar un modelo expuesto; sin embargo, en este momento, no existe seguridad inherente que impida a un usuario eliminar los modelos de otro usuario.
	Vea los detalles de salida del modelo. También puede acceder a esta información desde las etapas K-Means Clustering y Logistic Regression haciendo clic en "Para obtener más detalles del modelo, haga clic aquí" en la pestaña Salida del modelo.
	Compare los modelos.

Ficha Detalles de modelo

La pantalla Detalle de modelo muestra la siguiente información para todos los modelos:

- **Nombre de modelo**: el nombre del modelo
- **Tipo de modelo**: el tipo de modelo de Machine Learning
- **Usuario**: el nombre de usuario de la persona que creó el modelo
- **Descripción**: la descripción del modelo en caso de que se haya proporcionado una cuando se creó
- **Estado**: si el modelo se expuso o si se anuló la exposición
- **Nombre de flujo de datos**: el nombre del flujo de datos que produce el modelo
- **Tiempo de creación**: la fecha y la hora en que se creó el modelo

Se proporcionan detalles adicionales en función del tipo de modelo.

Detalles de K-Means Clustering

La pantalla Detalle de modelo incluye la siguiente información para modelos K-Means Clustering:

Resumen de modelo

- Número de filas
- Número de clústeres
- Número de columnas categóricas
- Número de iteraciones
- Suma de cuadrados dentro del clúster
- Suma total de cuadrados
- Suma de cuadrados entre el clúster

Métricas

Proporciona datos de capacitación, prueba y N subconjuntos para lo siguiente:

- Suma total de cuadrados dentro del clúster
- Suma total de cuadrados
- Suma de cuadrados entre el clúster

Estadísticas de centroide

Proporciona datos de capacitación, prueba y N subconjuntos para cada centroide:

- Tamaño
- Suma de cuadrados dentro del clúster

Agrupamiento de medias

Proporciona información detallada de cada centroide. El contenido varía según los datos de entrada. Un clúster es un grupo de observaciones de un conjunto de datos identificado como similar según un algoritmo de agrupamiento específico

Agrupamiento estandarizado de medias

Proporciona información estandarizada de cada centroide. El contenido varía según los datos de entrada.

Detalles de Logistic Regression

La pantalla Detalle de modelo incluye la siguiente información para modelos Logistic Regression:

Métricas

Proporciona datos de capacitación, prueba y N subconjuntos para lo siguiente:

- Error cuadrático medio (MSE)
- Error cuadrático medio de raíz (RMSE)
- Número de observaciones
- R-cuadrado (R2)
- Pérdida logarítmica (Logloss)
- Área bajo la curva (AUC)
- Coeficiente Gini
- Error medio por clase
- AIC
- Desviación residual
- Desviación nula
- Grado de libertad nulo
- Grado de libertad residual

Umbral de métricas máximas

Proporciona el Umbral de métricas máximas de capacitación para datos de capacitación, prueba, N subconjuntos usando las métricas siguientes:

- max f1
- max f2
- max f0point5
- max accuracy
- max precision
- max recall
- max specificity
- max absolute_mcc
- max min_per_class_accuracy
- max mean_per_class_accuracy

Matriz de confusión

Ilustra el rendimiento de un modelo en un conjunto de datos de capacitación, prueba y N subconjuntos para los que se conocen los valores verdaderos.

Gráfico de coeficiente estándar

Muestra los predictores más importantes proporcionando el valor relativo de los coeficientes, lo que indica cuánto cambia el objetivo por un cambio en la entrada.

Coeficientes de GLM

Coeficientes para un modelo lineal generalizado, que estiman los modelos de regresión para resultados que siguen distribuciones exponenciales.

Curvas AUC

Área bajo la curva; determina cuál de los modelos usados predice las clases que mejor usan los datos de capacitación, prueba y N subconjuntos.

Curvas de ganancia/elevación

Evalúan la capacidad de predicción de un modelo de clasificación binaria usando datos de capacitación, prueba y N subconjuntos.

Notices

Avisos de USPS®

Pitney Bowes Inc. posee una licencia no exclusiva para publicar y vender bases de datos ZIP + 4® en medios magnéticos y ópticos. Las siguientes marcas comerciales son propiedad del Servicio Postal de los Estados Unidos: CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS^{Link}, NCOA^{Link}, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite^{Link} , United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code, y ZIP + 4. Esta lista no es exhaustiva de todas las marcas comerciales que pertenecen al servicio postal.

Pitney Bowes Inc. es titular de una licencia no exclusiva de USPS® para el procesamiento NCOA^{Link}®.

Los precios de los productos, las opciones y los servicios del software de Pitney Bowes no los establece, controla ni aprueba USPS® o el gobierno de Estados Unidos. Al utilizar los datos RDI™ para determinar los costos del envío de paquetes, la decisión comercial sobre qué empresa de entrega de paquetes se va a usar, no la toma USPS® ni el gobierno de Estados Unidos.

Proveedor de datos y avisos relacionados

Los productos de datos que se incluyen en este medio y que se usan en las aplicaciones del software de Pitney Bowes Software, están protegidas mediante distintas marcas comerciales, además de un o más de los siguientes derechos de autor:

© Derechos de autor, Servicio Postal de los Estados Unidos. Todos los derechos reservados.

© 2014 TomTom. Todos los derechos reservados. TomTom y el logotipo de TomTom son marcas comerciales registradas de TomTom N.V.

© 2016 HERE

Fuente: INEGI (Instituto Nacional de Estadística y Geografía)

Basado en los datos electrónicos de © National Land Survey Sweden.

© Derechos de autor Oficina del Censo de los Estados Unidos

© Derechos de autor Nova Marketing Group, Inc.

Algunas partes de este programa tienen © Derechos de autor 1993-2007 de Nova Marketing Group Inc. Todos los derechos reservados

© Copyright Second Decimal, LLC

© Derechos de autor Servicio de correo de Canadá

Este CD-ROM contiene datos de una compilación cuyos derechos de autor son propiedad del servicio de correo de Canadá.

© 2007 Claritas, Inc.

El conjunto de datos Geocode Address World contiene datos con licencia de GeoNames Project (**www.geonames.org**) suministrados en virtud de la licencia de atribución de Creative Commons (la “Licencia de atribución”) que se encuentra en **http://creativecommons.org/licenses/by/3.0/legalcode**. El uso de los datos de GeoNames (según se describe en el manual de usuario de Spectrum™ Technology Platform) se rige por los términos de la Licencia de atribución. Todo conflicto entre el acuerdo establecido con Pitney Bowes Software, Inc. y la Licencia de atribución se resolverá a favor de la Licencia de atribución exclusivamente en cuanto a lo relacionado con el uso de los datos de GeoNames.

3001 Summer Street
Stamford CT 06926-0700
USA

www.pitneybowes.com