Ana Sofía Aponte Barriga

Universidad Nacional de Colombia – Sede Bogotá

Ingeniería Mecatrónica

Mayo de 2023

Desarrollo de un modelo de inteligencia artificial usando la nube de IBM para predecir fallas de un equipo industrial

TablA DE ContenIDOs

[I. Introducción 1](#_Toc132824391)

[II. GUÍA DE CONSTRUCCIÓN DE UN MODELO DE INTELIGENCIA ARTIFICIAL USANDO IBM AutoAI 2](#_Toc132824392)

[A. RECOLECCIÓN DE DATOS 2](#_Toc132824393)

[B. LIMPIEZA DE DATOS 3](#_Toc132824394)

[C. ANÁLISIS EXPLORATORIO DE DATOS 18](#_Toc132824395)

[D. CONSTRUCCIÓN DEL MODELO 25](#_Toc132824396)

[E. DESPLIEGUE DEL MODELO 37](#_Toc132824397)

[III. REFERENCIAS 52](#_Toc132824398)

# Introducción

El uso de la inteligencia artificial como herramienta de predicción permite optimizar gran cantidad de procesos, costos y tiempos. Por ello, se presenta una guía de construcción de un modelo de Inteligencia Artificial, el cual se construirá usando la herramienta AutoAI de IBM Cloud. El uso de esta herramienta permite un mayor acercamiento a lo que es la inteligencia artificial y al ciclo de vida de un problema en la ciencia de datos.

En este caso, y teniendo en cuenta el contexto industrial en que se desempeñan los ingenieros mecatrónicos, se seleccionó un caso de uso referente a la optimización del uso de máquinas de producción. Esto se logra a partir de un conjunto de datos que incluyen variables como el tamaño de la máquina, velocidades de rotación, torques y temperaturas. Sin embargo, el proceso presentado es aplicable en innumerables casos de uso en los cuales se tenga acceso a datos estructurados.

# GUÍA DE CONSTRUCCIÓN DE UN MODELO DE INTELIGENCIA ARTIFICIAL USANDO IBM AutoAI

A continuación, se presenta una guía que permitirá construir un modelo de inteligencia artificial usando las tecnologías Watson AutoAI y Watson Machine Learning de IBM Cloud.

Para la construcción del modelo se usará una aproximación basada en el ciclo de vida de proyectos de data science, el cual contiene las etapas de exploración, limpieza, análisis exploratorio de datos, construcción del modelo, y despliegue del modelo.

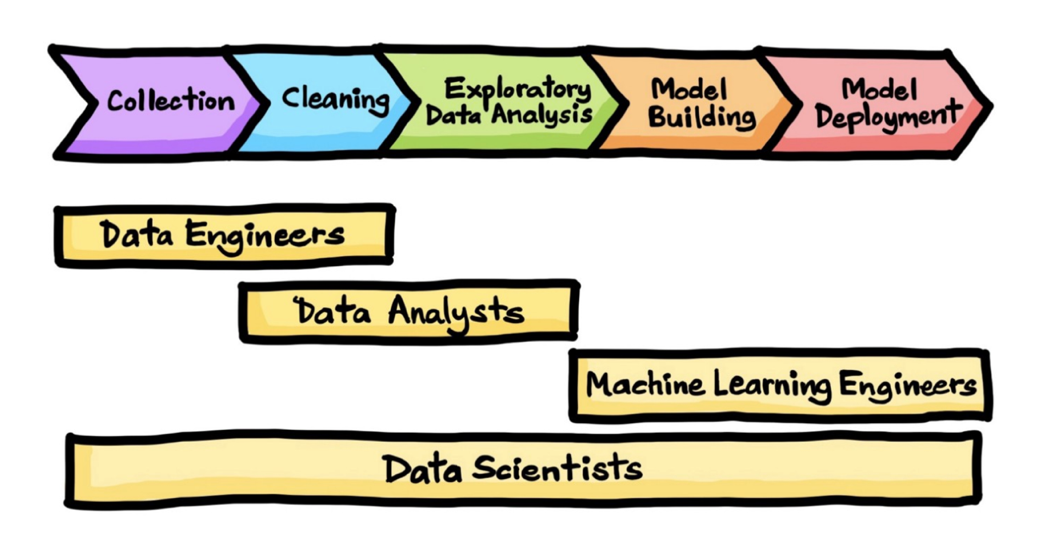


Figura 1 Ciclo de vida de un proyecto de data science. Fuente: [springboard.com](https://www.springboard.com/blog/data-science/data-science-process/#:~:text=comprising%20each%20one.-,What%20Is%20the%20Data%20Science%20Process%3F,presenting%20the%20solution%20to%20stakeholders.)

## RECOLECCIÓN DE DATOS

Para poder iniciar el proceso, se requiere tener un conjunto de datos estructurado, esto es, una tabla de datos. Para ello, en este caso, se usará un conjunto de datos obtenido de la plataforma de data science “Kaggle”, disponible en: <https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification?resource=download>

Sin embargo, cualquier tabla con variables definidas es adecuada para generar un modelo. Dicho modelo será mejor si hay una mayor cantidad de filas.

El conjunto de datos seleccionado contiene 10.000 filas, organizadas en las siguientes columnas (variables):

UDI: numeración del registro en la tabla. Contiene valores únicos

Product ID: Código interno del registro. Contiene valores únicos

Type: tipo de máquina. Clasificadas en H, L, M

Air Temperature [K]: Temperatura del aire. Valores entre 295.3 y 304.5

Process Temperature [K]: Temperatura del proceso. Valores entre 305.7 y 313.8

Rotational Speed [rpm]: Velocidad rotacional. Valores entre 1168 y 2886

Torque [Nm]: Torque. Valores entre 3.8 y 76.6

Tool wear [min]: Desgaste en la herramienta de corte. Valores entre 0 y 253

Target: Indica si hubo falla o no.

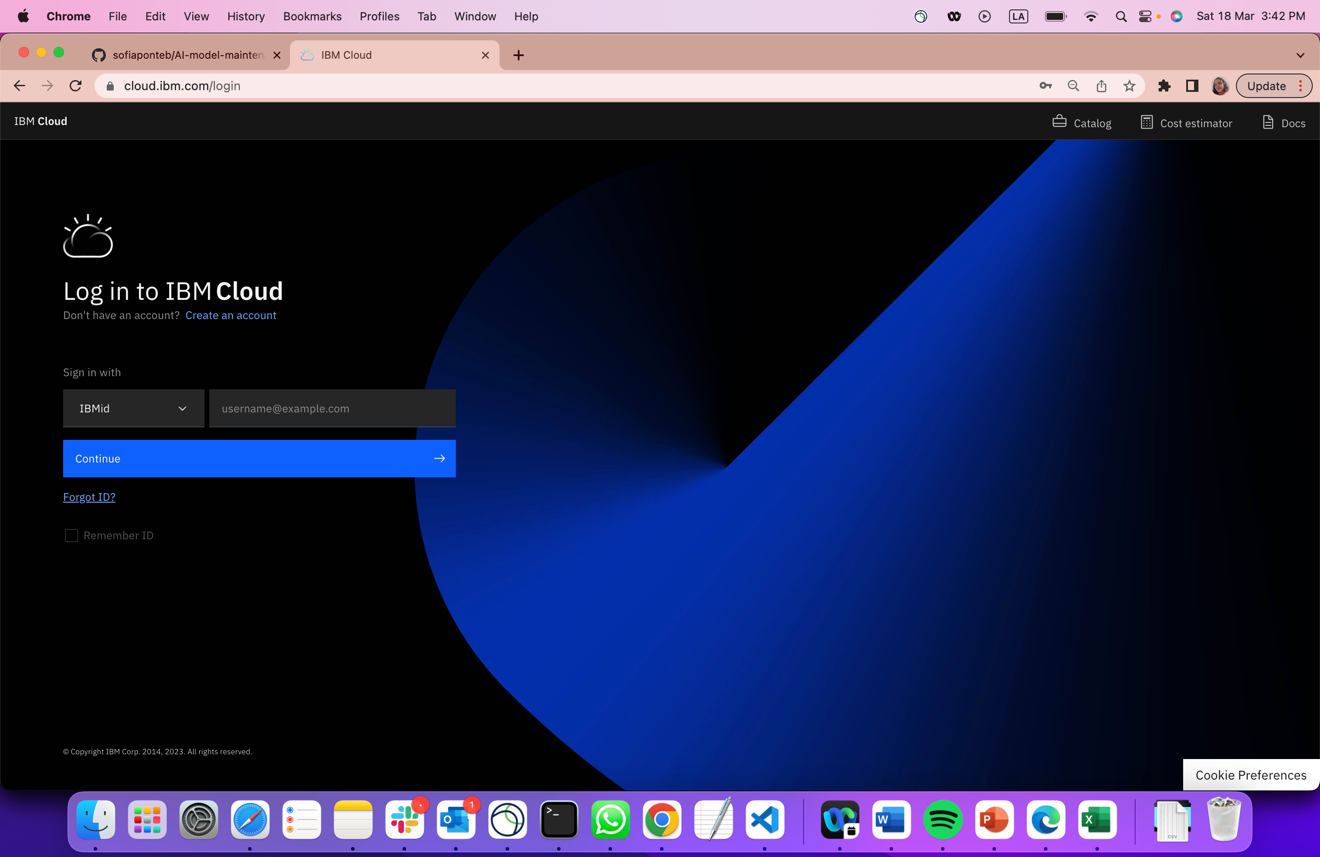
Failure type: Indica el tipo de falla.

### Para esta etapa del proceso el único paso a realizar es ingresar al link: <https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification?resource=download> y descargar el archivo .csv con los datos

## LIMPIEZA DE DATOS

La primera fase para generar el modelo es limpiar los datos, esto es eliminar aquellas columnas que no aportan al modelo y modificar las que se requiera. Adicionalmente, si hay alguna fila con valores en blanco, se debe eliminar.

### Ingresa a [cloud.ibm.com](http://cloud.ibm.com/) e inicia sesión con tu usuario y contraseña



### En la parte superior, selecciona la pestaña **Catalog**

### Graphical user interface, application Description automatically generated

### En el menú de la izquierda, ingresa a la sección **AI/Machine Learning**

### Graphical user interface, application Description automatically generated

### Selecciona el producto **Watson Studio**

### Graphical user interface, application Description automatically generated

### En la página de aprovisionamiento, elige una ubicación para tu instancia. En este caso seleccionamos **Dallas** por la baja latencia que genera en Colombia. Selecciona el plan de pago **Lite**, el cual es gratis y contiene capacidades de cómputo limitadas que serán suficientes para el desarrollo del modelo.

### Graphical user interface, application Description automatically generated

### Asígnale un nombre a tu instancia, en este caso dejamos un nombre asociado al modelo a realizar: **Watson Studio – machine maintenance.** Finalmente, da click para aceptar las condiciones y posteriormente en **Create** para crear la instancia.

### Graphical user interface, application Description automatically generated

### Luego de la creación, deberás esperar a que en el panel principal de Watson Studio aparezca el anuncio en verde en la parte superior que dice **Active**, el cual indica que ya se puede hacer uso de la instancia. Da click en **Launch in IBM Cloud Pak for Data**

### Graphical user interface, application Description automatically generated

### Se abrirá una nueva pestaña con tu instancia de Watson Studio, en la cual te sugerirá aprovisionar Watson Machine Learning. Da click en **Next**

### Graphical user interface Description automatically generated

### Selecciona la misma ubicación que elegiste para la instancia de Watson Studio. En este caso, **Dallas**. Nuevamente selecciona el plan **Lite**.

### Graphical user interface, application Description automatically generated

### Asigna un nombre a tu instancia de Watson Machine Learning. En este caso se asignó el nombre **Watson Machine Learning – machine maintenance**. Da click en **Create**

### Graphical user interface Description automatically generated

### Al volver a la página de Watson Studio, te sugerirá la creación de un projecto. Selecciona **New Project** y da click en **Next**

### Graphical user interface, application Description automatically generated

### Luego de esto deberás esperar un momento mientras se realiza la configuración de Watson Studio

### 

### Te aparecerá el panel de creación de proyecto nuevo. En este caso se le asignó el nombre **maintenance-prediction**. En la sección de la derecha, selecciona la opción **add**, para crear un servicio de almacenamiento para el proyecto.

### Graphical user interface, application Description automatically generated

### Se abrirá una nueva pestaña para confgurar la instancia de Cloud Object Storage. Selecciona el plan **Lite** y da click en **Create**. Si así lo deseas, puedes cambiar el nombre de la instancia, en este caso se dejó el nombre predefinido.

### Graphical user interface, application Description automatically generated

### Cuando vuelvas al panel de creación del proyecto, da click en **Refresh** para que aparezca la instancia de almacenamiento creada.

### Graphical user interface, application Description automatically generated

### Luego de que aparezca la instancia de almacenamiento, podrás dar click en **Create**

### Graphical user interface, application Description automatically generated

### Luego de crear el proyecto, la página te llevará a la página de inicio de tu proyecto, el cual se encuentra vacío

### Graphical user interface, text, application Description automatically generated

### Lo primero a realizar en el proyecto, es incluir el archivo CSV que contiene la tabla de datos a utilizar. Para ello, da click en la parte superior derecha de tu pantalla, en el botón esto abrirá un panel desplegable que te permite dar click y subir el archivo CSV que descargaste en la sección anterior [(Recolección de Datos)](#_Para_esta_etapa)

### Graphical user interface, application Description automatically generated

### Luego de seleccionar tu archivo, en la página de inicio del Proyecto te debe aparecer un único elemento en la lista de assets, que corresponde al CSV subido. Da click en el nombre de tu archivo para abrirlo.

### Graphical user interface, text, application Description automatically generated

### Podrás ver una vista previa de los datos almacenados en el CSV, da click en **Prepare Data** para iniciar la limpieza de datos.

### Graphical user interface, application, table Description automatically generated

### En la página de inicio de Data Refinery, se te sugerirá dar un recorrido inicial, da click en **Maybe Later**

### Graphical user interface Description automatically generated

### Podrás ver que Watson automáticamente detectó los tipos de variables presentes en el conjunto de datos, y agregó un paso que incluye dicha asignación de variables. En la tabla puedes verificar que todas hayan sido asignadas correctamente, y en caso de que algún tipo de variable haya sido asignada de forma errónea, podrás corregirlo dando click en los 3 puntos que aparecen al lado del nombre de la variable. Luego de realizar esta revisión, da click en la pestaña **Profile**, que se encuentra justo sobre la tabla.

### Graphical user interface, application, table, Excel Description automatically generated

### En esta visualización podrás ver una caracterización de los datos en diagramas de barras. Identificaremos las columnas que tienen 10000 valores únicos:

### UDI

### Product ID

### Graphical user interface, application Description automatically generated

### Verifica que las demás columnas no estén compuestas solo por valores únicos

### Graphical user interface Description automatically generated

### También eliminaremos la columna **Target**, que indica si hubo falla o no. Ya que nos enfocaremos en predecir el tipo de falla y esta columna no hace parte de los datos a ingresar en la predicción.

### Graphical user interface, application Description automatically generated

### Ahora eliminaremos las 3 columnas: **UDI, Product ID, Target**

### Para eliminar estas tres columnas repite los siguientes pasos con cada una:

### Da click en el nombre de la columna

### Luego en la parte inferior izquierda en **New Step**

### Graphical user interface, application, table, Excel Description automatically generated

### Selecciona la opción **Remove Column**

### Graphical user interface, application, table, Word Description automatically generated

### Da click en el botón azul **Apply**

### Graphical user interface, application, Word Description automatically generated

### Con esto damos por terminada la sección de limpieza de datos, puedes visualizar al lado izquierdo los 4 pasos añadidos para limpiar el conjunto de datos.

### 

## ANÁLISIS EXPLORATORIO DE DATOS

En esta etapa del proceso exploraremos las gráficas que se pueden generar con Data Refinery de Watson Studio. Esto nos permite tener un mejor contexto de qué es lo que hay en el conjunto de datos.

### Da click en la pestaña **Visualizations**, allí podrás ver diferentes tipos de gráficas que describen el conjunto de datos. En este caso, por defecto aparece Scatter Plot, en el menú de la izquierda podrás seleccionar las variables para los ejes.

### Graphical user interface, application, table, Excel Description automatically generated

### Da click en el menú desplegable en chart type, y selecciona **scatter matrix chat**. Podrás seleccionar variables a comparar para ver su correlación. En este caso se seleccionó velocidad rotacional vs. Temperatura del proceso, podemos ver que tienen una correlación de 0.036.

### Graphical user interface, application Description automatically generated

### Si seleccionas variables que se relacionan, como temperatura del aire y temperatura del proceso, verás que efectivamente se puede identificar una correlación fuerte usando Watson.

### Graphical user interface, application Description automatically generated

### Podemos también en un diagrama circular observar la proporción de procesos sin falla vs. Las fallas clasificadas por tipo.

### Graphical user interface, application Description automatically generated

### También encontrarás otros tipos de gráficas no tan comunes, como lo es el circle packing chart. Acá podemos ver la relación de tipos de falla agrupadas por tipo de máquina. A su vez, si lo deseas puedes dar click en la flecha del lado derecho para descargar un archivo con la imagen generada de la gráfica.

### Graphical user interface Description automatically generated

### Luego de haber visualizado los datos, vamos a guardar las modificaciones que se realizaron al archivo CSV en la sección anterior. Para ello, da click en el ícono de guardar en la parte superior. Asigna un nombre al proceso de limpieza de datos, en este caso se le puso **refine-dataset**. Da click en **next**

### 

### En la siguiente ventana se configura el tipo de ambiente, déjalo como está por defecto y da click en **next**.

### Graphical user interface Description automatically generated

### Da click en **next** sin hacer ninguna selección.

### Graphical user interface, text, application Description automatically generated

### Nuevamente da click en **next** sin hacer ninguna selección.

### Graphical user interface, application Description automatically generated

### Finalmente, en la página de resumen verifica el nombre del archivo CSV que se generará con las modificaciones seleccionadas. En este caso es **predictive\_maintenance\_csv\_shaped**. Da click en **Create and run**.

### Graphical user interface, application Description automatically generated

### Te aparecerá la ventana de creación, donde podrás ver el progreso en la columna status.

### Graphical user interface, application Description automatically generated

### Cuando aparezca el status **Completed**, habrás terminado esta sección y podrás iniciar con la construcción del modelo.

### Graphical user interface, application Description automatically generated

## CONSTRUCCIÓN DEL MODELO

En esta sección, utilizaremos la herramienta AutoAI de Watson Studio para crear el modelo de inteligencia artificial que predecirá si cierta configuración generará falla, y si es así, el tipo de falla generada.

### Iniciaremos en el panel de resumen del proyecto, allí da click en el botón azul **New Asset**

### Graphical user interface, text, application Description automatically generated

### Entre las opciones presentadas, selecciona **AutoAI**

### Graphical user interface, application Description automatically generated

### Asigna un nombre a tu experimento de AutoAI, podría ser **AutoAI-maintenance-prediction**, los demás ítems los dejaremos tal como están. Da click en **Create**.

### Graphical user interface Description automatically generated

### Ahora necesitarás el conjunto de datos, para ello da click en el botón azul **Select data from Project**.

### Graphical user interface, text, application Description automatically generated

### Como recordarás del final de la sección anterior, el conjunto de datos ya modificado fue almacenado en el archivo **predictive\_maintenance\_csv\_shaped**. Selecciona el conjunto de datos que se modificó en la sección de limpieza de datos y da click en el botón azul **Select asset**.

### Graphical user interface, text, application Description automatically generated

### En la siguiente ventana podrás configurar la creación de tu modelo. En la sección “Create a times series forecast” selecciona la opción **no.** Posteriormente, podrás seleccionar una columna a predecir, allí seleccionaremos **Failure type**. Al tener un servicio lite, la capacidad de procesamiento disponible es limitada, por lo cual te aparecerá cuánto almacenamiento te queda en el mes, la cantidad máxima está definida en 20 CUH. Da click en el botón azul **Run experiment** para iniciar la construcción de tu modelo.

### Graphical user interface, application Description automatically generated

### En la siguiente ventana te aparecerá el avance del experimento.

### Graphical user interface, text Description automatically generated

### Podrás ver el paso a paso del proceso. Inicialmente, te mostrará el porcentaje de datos seleccionados para entrenar y para probar el modelo.

### Graphical user interface, text Description automatically generated

### Posteriormente te mostrará la etapa de selección de algoritmos para el modelo

### Graphical user interface Description automatically generated

### Podrás ver que cada que se selecciona un algoritmo, al lado derecho aparecerá qué algoritmo es y el mapa de progreso en la generación del modelo.

### Graphical user interface, chart Description automatically generated

### Podrás ver también la creación de diferentes pipelines con los mejores algoritmos

### Graphical user interface, chart Description automatically generated

### Cuando la creación del modelo termine, te aparecerá un anuncio para ver los resultados. Da click en **Next**.

### Graphical user interface, application Description automatically generated

### La herramienta te dará la opción de dar un tour por los resultados obtenidos, podrás darle click en next o en la X para saltar el tour.

### Graphical user interface Description automatically generated

### Inicialmente, podrás ver el mapa de relaciones donde aparecen los algoritmos probados, de los cuales los mejores fueron seleccionados para la creación de pipelines, donde las mejores son seleccionadas para ser optimizadas.

### Graphical user interface Description automatically generated

### En la siguiente tabla podrás ver cada uno de los modelos generados, incluyendo para cada uno el algoritmo que usa, la exactitud obtenida, las optimizaciones implementadas y el tiempo que tardó en ser generado. Da click en el nombre del primer modelo (en este caso **Pipeline 4**) para ver sus detalles.

### Graphical user interface Description automatically generated

### En la ventana emergente, podrás ver una gráfica de evaluación del modelo, donde se clasifica por el tipo de falla.

### Graphical user interface, chart, line chart Description automatically generated

### En el menú de la izquierda, selecciona la matriz de confusión, allí podrás ver una comparación entre los datos observados y predichos.

### Graphical user interface, application Description automatically generated

### En el menú de la izquierda, puedes seleccionar la opción **Feature Summary**, donde se presentan las variables usadas por el modelo para realizar la predicción. En este caso podemos ver que las features más importantes en este modelo son productos de variables entregadas en el conjunto de datos.

### Graphical user interface, application Description automatically generated

### Finalmente, en la pestaña de **Model Information** podrás ver un resumen general del modelo generado. Ya que este modelo fue el de mejor desempeño, será el que guardaremos para realizar el despliegue. Para ello da click en el botón azul **Save as**

### Graphical user interface, text, application Description automatically generated

### Primero guardaremos el modelo, para ello da click en **Model** y luego en **create**.

### Graphical user interface, application Description automatically generated

### De la misma forma, puedes dar click en **Notebook** para obtener el código fuente del modelo generado, y posteriormente en **Create**.

### Graphical user interface, application Description automatically generated

### En la página de resultados del experimento, también podrás guardar el código de todos los modelos generados, dando click en **Save Experiment Code**.

### Graphical user interface Description automatically generated

### Para guardar el código del experimento, podrás elegir un nombre o dejar el que aparece por defecto, y luego dar click en **Create**.

### Graphical user interface, application Description automatically generated

### Finalmente, si quieres explorar un poco más el modelo generado, puedes dar click en la pestaña **Pipeline comparison**, donde aparece un comparativo por categorías entre todos los modelos.

### Graphical user interface, chart Description automatically generated

## DESPLIEGUE DEL MODELO

La última etapa del proceso de ciencia de datos es realizar el despliegue y evaluación del modelo generado. En este caso, realizaremos pruebas al modelo usando la interfaz web de IBM Cloud. Sin embargo, también es posible desplegar el modelo por medio de APIs embebidas en aplicaciones.

### Iniciamos volviendo a la pantalla inicial del proyecto:

### Graphical user interface, text, application Description automatically generated

### En el menú de la izquierda, selecciona la opción **Models**, así te aparecerá el modelo generado. Da click en el menú de hamburguesa de la esquina superior izquierda.

### Graphical user interface, text, application Description automatically generated

### Selecciona la opción **Deployments**

### Graphical user interface, text, application Description automatically generated

### Da click en el botón azul **New Deployment space**

### Graphical user interface, text, application Description automatically generated

### Ingresa un nombre para tu espacio de despliegue, en este caso se usó **Deployment-space-predictive-maintenance**. Las instancias de almacenamiento y machine learning deben estar seleccionadas por defecto. Da click en **Create**

### Graphical user interface, text, application Description automatically generated

### Aparecerá un aviso mientras se crea el espacio de despliegue.

### Graphical user interface, application Description automatically generated

### Cuando termine la creación del espacio, da click en **View new space**

### Graphical user interface, application Description automatically generated

### Podrás ver un resumen del espacio creado. Ahora volvemos a la pantalla inicial del proyecto. Da click en el menú de hamburguesa en la esquina superior izquierda.

### Graphical user interface, application Description automatically generated

### Selecciona la opción **View all projects**

### Graphical user interface, application Description automatically generated

### Aparecerá el proyecto en que hemos estado trabajando. Da click en el nombre del proyecto

### 

### Selecciona el menú **Models** y da click en el nombre del modelo creado.

### Graphical user interface, text, application, Word Description automatically generated

### Da click en el botón azul **Promote to deployment space**

### 

### Aparecerá seleccionado el espacio de despliegue que recién creaste, da click en la casilla **Go to the model in the space after promoting it**, y posteriormente da click en **Promote**

### Graphical user interface, application Description automatically generated

### Da click en **New deployment**

### 

### Selecciona la opción **online**, y asigna un nombre a tu despliegue, por ejemplo **deployment-predictive-maintenance**. Da click en **Create**

### Graphical user interface, application Description automatically generated

### En la pantalla se mostrará el estado del proceso de creación del despliegue.

Graphical user interface, text, application

Description automatically generated

### Cuando se complete la creación del despliegue, da click en el nombre para abrirlo.

### Graphical user interface, text, application Description automatically generated

### Te aparecerá la página de referencia API, donde se encuentran los códigos para incluir el modelo en aplicaciones de diferentes lenguajes de programación. En este caso usaremos la herramienta online, así que da click en la pestaña **Test**, justo debajo del título de la página.

### Graphical user interface, text, application Description automatically generated

### Inicialmente probaremos el modelo con una sola entrada de datos. Para ello, ingresa un valor por variable en cada ítem de la primera fila y da click en **Predict**

### Graphical user interface, text, application, table Description automatically generated

### Aparecerá una ventana emergente con el resultado de falla según los datos ingresados. En este caso se predice una falla por disipación de calor.

### Graphical user interface, application, table, Excel Description automatically generated

### Otra forma de probar el modelo generado es incluyendo un archivo CSV con datos de prueba. Da click en el botón **Download CSV template**, esto descargará un archivo CSV con los encabezados de la tabla de datos. Allí podrás incluir todas las filas de datos que desees. Para incluir el archivo da click en **Browse local files** y selecciona tu CSV con los datos de prueba.

### Table Description automatically generated

### Luego de dar click en **Predict**, te aparecerá una ventana emergente con los resultados de predicción en visualización de tabla y de diagrama circular. Asímismo se incluye un nivel de confianza de cada predicción en la vista de tabla. Estos resultados los puedes descargar dando click en **Download**.

### Graphical user interface, application Description automatically generated

### Finalmente, descargaremos y visualizaremos el código del modelo generado. Para ello regresa a la pantalla principal de IBM Cloud [cloud.ibm.com/resources](http://cloud.ibm.com/resources). Da click en **Storage** para ver la instancia creada de Cloud Object Storage, da click sobre el nombre de tu instancia.

### Graphical user interface, application Description automatically generated

### Dentro del Object Storage tendrás dos buckets, uno que fue el creado para almacenar modelos y el otro que es usado por Machine Learning para guardar información propia del servicio. Se distinguen porque el creado por Machine Learning es una cadena de caracteres aleatorios. Da click en el bucket creado para almacenar los resultados de Watson Studio, en este caso es **maintenance-prediction-donotdelete**

### Graphical user interface, text, application Description automatically generated

### Te aparecerá una lista con todos los archivos almacenados. Da click en el ícono de carpeta en la parte superior derecha para tener una mejor visualización.

### Graphical user interface, text, application Description automatically generated

### Selecciona la carpeta **notebook**

### 

### Selecciona el notebook almacenado con la última hora de modificación, y da click en **Download object**

### Graphical user interface, application Description automatically generated

### Abre el archivo con un editor de texto, como lo es Visual Studio Code, allí podrás ver el código completo del modelo generado, incluyendo explicaciones del contenido de cada sección. Con esto damos por finalizada la creación y prueba del modelo de inteligencia artificial.

### Graphical user interface, text Description automatically generated

# REFERENCIAS

### [springboard.com](https://www.springboard.com/blog/data-science/data-science-process/#:~:text=comprising%20each%20one.-,What%20Is%20the%20Data%20Science%20Process%3F,presenting%20the%20solution%20to%20stakeholders.)

### <https://www.kaggle.com/datasets/shivamb/machine-predictive-maintenance-classification?resource=download>

### <https://www.ibm.com/docs/en/cloud-paks/cp-data/4.0?topic=autoai-tutorial>

### <https://www.ibm.com/docs/en/cloud-paks/cp-data/4.6.x?topic=data-refining>

### <https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-overview.html>