

AutoML with SageMaker Autopilot

Use a SageMaker domain in the AWS Academy and in the Virginia Region to try AutoPilot

Canvas

Build a custom numeric or categorical prediction model

Numeric and categorical prediction models support both **Quick builds** and **Standard builds**.

To build a numeric or categorical prediction model, use the following procedure:

1. Open the SageMaker Canvas application.
2. In the left navigation pane, choose **My models**.
3. Choose **New model**.
4. In the **Create new model** dialog box, do the following:
 1. Enter a name in the **Model name** field.
 2. Select the **Predictive analysis** problem type.
 3. Choose **Create**.
5. For **Select dataset**, select your dataset from the list of datasets. If you haven't already imported your data, choose **Import** to be directed through the import data workflow.
6. When you're ready to begin building your model, choose **Select dataset**.
7. On the **Build** tab, for the **Target column** dropdown list, select the target for your model that you would like to predict.
8. For **Model type**, Canvas automatically chooses the problem type for you. If you want to change the type, choose **Change model type** and select your desired model type.
9. Select or deselect columns in your data to include or drop them from your build.

Note

If you make batch predictions with your model after building, Canvas adds dropped columns to your prediction results. However, Canvas does not add the dropped columns to your batch predictions for time series models.

10. (Optional) Use the visualization and analytics tools that Canvas provides to visualize your data and determine which features you might want to include in your model. For more information, see [Explore and analyze your data](#).
11. (Optional) Use data transformations to clean, transform, and prepare your data for model building. For more information, see [Prepare your data with advanced transformations](#). You can view and remove your transforms by choosing **Model recipe** to open the **Model recipe** side panel.

12. (Optional) For additional features such as previewing the accuracy of your model, validating your dataset, and changing the size of the random sample that Canvas takes from your dataset, see [Preview your model](#).

13. After reviewing your data and making any changes to your dataset, choose **Quick build** or **Standard build** to begin a build for your model. The following screenshot shows the **Build** page and the **Quick build** and **Standard build** options.

The image consists of three vertically stacked screenshots of the AWS S3 'Create bucket' wizard, showing the process of setting up a new bucket named 'nr-autopilot-canvas' in the 'US East (N. Virginia) us-east-1' region.

- General configuration:** Shows the 'Bucket name' field set to 'nr-autopilot-canvas'. The 'AWS Region' dropdown is set to 'US East (N. Virginia) us-east-1'. A note indicates that the bucket settings in the following configuration are copied from an existing bucket. A 'Choose bucket' button is present.
- Object Ownership:** A note states that control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). It offers two options:
 - ACLs disabled (recommended)**: All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.
 - ACLs enabled**: Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.A note at the bottom indicates 'Upcoming permission changes to disable ACLs'.
- Default encryption:** A note states that server-side encryption is automatically applied to new objects stored in this bucket. It shows the 'Encryption key type' dropdown set to 'Amazon S3 managed keys (SSE-S3)', with 'AWS Key Management Service key (SSE-KMS)' as an alternative. A note about 'Bucket Key' and 'AWS KMS' is present, with a link to learn more. The 'Disable' and 'Enable' buttons are shown.

At the bottom of each screen, there is a note: 'After creating the bucket you can upload files and folders to the bucket, and configure additional bucket settings.' Below this is a 'Cancel' button and a prominent 'Create bucket' button.

The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with tabs like Content, Class Collaborate, Inbox, Learner Lab, AWS Manager, S3 Management, Amazon SageMaker, and SageMaker Core. Below the navigation bar, the main title is "Amazon S3 > Buckets". On the left, there's a sidebar with sections for Buckets, Access Points, Object Lambda Access Points, Multi-Region Access Points, Batch Operations, IAM Access Analyzer for S3, Storage Lens, Dashboards, AWS Organizations settings, Feature spotlight, and AWS Marketplace for S3. The main content area displays an "Account snapshot" with a link to "View Storage Lens dashboard". Below that is a table titled "Buckets (2) info" with two entries:

Name	AWS Region	Access	Creation date
nr-autopilot-canvas	US East (N. Virginia) us-east-1	Bucket and objects not public	April 5, 2023, 17:08:53 (UTC-04:00)
sagemaker-studio-731623272175-4hg26lern9gg9	US East (N. Virginia) us-east-1	Objects can be public	April 5, 2023, 16:06:45 (UTC-04:00)

At the bottom of the page, there are links for CloudShell, Feedback, Language, and a footer with copyright information and a timestamp of 2023-04-05.

The screenshot shows the AWS S3 console interface for uploading files. The URL in the address bar is https://s3.console.aws.amazon.com/s3/upload/nr-autopilot-canvas?region=us-east-1. The main title is "Amazon S3 > Buckets > nr-autopilot-canvas > Upload". The left sidebar has the same structure as the previous screenshot. The main content area has a heading "Upload info" with instructions: "Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. Learn more [?]". Below this is a large input field with placeholder text "Drag and drop files and folders you want to upload here, or choose Add files, or Add folders." A table titled "Files and folders (2 Total, 54.5 KB)" lists the uploaded files:

Name	Folder	Type	Size
Data_Banknote_Authen	-	text/csv	9.1 KB
ication_Test.csv			
Data_Banknote_Authen	-	text/csv	45.4 KB
ication_Train.csv			

Below the table is a section titled "Destination" with a dropdown menu set to "s3://nr-autopilot-canvas". There are also "Destination details" and "Permissions" sections. At the bottom right are "Cancel" and "Upload" buttons. The footer includes CloudShell, Feedback, Language, and a timestamp of 2023-04-05.

The screenshot shows the AWS S3 Manager console with the following details:

- Upload status:** The upload was successful.
- Summary:** Destination is `s3://nr-autopilot-canvas`.
 - Succeeded:** 2 files, 54.5 KB (100%)
 - Failed:** 0 files, 0 B (0%)
- Files and folders:** (2 Total, 54.5 KB)
 - Data_Banknote_Authentication_Test.csv**: text/csv, 9.1 KB, Status: Succeeded
 - Data_Banknote_Authentication_Train.csv**: text/csv, 45.4 KB, Status: Succeeded

Amazon SageMaker Home Page:

The screenshot shows the Amazon SageMaker Home page with the following details:

- Domains:** A single domain named "onederdomain" is listed, which is InService and was created on April 05, 2023, at 20:06 UTC.
- JumpStart:** Foundation models (NEW), Computer vision models, Natural language processing models.
- Governance:** Ground Truth.

The screenshot shows the Amazon SageMaker console with the URL <https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/studio/d-rij6gb8emx5j>. The left sidebar includes links for Getting started, Studio, Studio Lab, Canvas, RStudio, Domains, SageMaker dashboard, Images, Lifecycle configurations, Search, JumpStart (Foundation models, Computer vision models, Natural language processing models), Governance, and Ground Truth. The main content area displays the 'Domain details' page for 'ondomain'. A warning message at the top states: '⚠️ No new Jupyter Lab 1 version apps can be created from March 30, 2023 onwards, with only Jupyter Lab 3 version app creation being supported. All existing apps running on Jupyter Lab 1 version will be removed on April 30, 2023.' Below this, the 'User profiles' tab is selected, showing a table with one row: 'nrawscloud' (Modified on Apr 05, 2023 20:14 UTC, Created on Apr 05, 2023 20:14 UTC). There is a 'Launch' button next to the row. The bottom of the screen shows the AWS navigation bar with CloudShell, Feedback, Language, and various icons.

The screenshot shows the Amazon SageMaker Canvas console with the URL <https://d-rij6gb8emx5j.studio.us-east-1.sagemaker.aws/canvas/default/readyToUseModels>. The left sidebar includes links for Amazon SageMaker Canvas, Ready-to-use models (selected), My models, Shared models, Datasets, Help, and Log out. The main content area displays the 'Ready-to-use models' section. A message at the top says: '💡 New launch! SageMaker Canvas now provides 9 ready-to-use models & support for text and image classification custom models. Learn more' with a link. Below this, a section titled 'Here are some ready-to-use models we've prepared for you to use.' lists four models: 'Sentiment analysis' (Detect sentiment in lines of text, which can be positive, negative, neutral, or mixed. Powered by Amazon Comprehend), 'Entities extraction' (Extract entities, which are real-world objects such as people, places, and commercial items, or units such as dates and quantities, from text. Powered by Amazon Comprehend), 'Language detection' (Determine the dominant language in text such as English, French or), and 'Personal information detection' (Detect personal information that could be used to identify an individual.). The bottom of the screen shows the AWS navigation bar with CloudShell, Feedback, Language, and various icons.

Screenshot of the AWS SageMaker Canvas Import interface:

Import

Data Source: Amazon S3

Choose files to import

Amazon S3 / nr-autopilot-canvas

Search Amazon S3

Name	Last updated
Data_Banknote_Authentication_Test.csv	04/05/2023 5:11 PM
Data_Banknote_Authentication_Train.csv	04/05/2023 5:11 PM

Cancel Import data

9°C Cloudy Search ENG IN 5:14 PM 2023-04-05

Screenshot of the AWS SageMaker Canvas Import interface showing an import preview:

Import

Data Source: Amazon S3

Import preview Previewing first 100 rows

Close preview Import data

Data_Banknote_Authentication_Test.csv

Use first row as header Delete

Variance	Skewness	Curtosis	Entropy	class
1.5691	6.3465	-0.1828	-2.4099	0
-0.27802	8.1881	-3.1338	-2.5276	0
0.051979	7.0521	-2.0541	-3.1508	0
-1.7559	11.9459	3.0946	-4.8978	0
2.4287	9.3821	-3.2477	-1.4543	0
4.6352	-3.0087	2.6773	1.212	0

Data_Banknote_Authentication_Train.csv

Use first row as header Delete

9°C Cloudy Search ENG IN 5:14 PM 2023-04-05

Screenshot of the Amazon SageMaker Canvas Datasets page.

The left sidebar shows navigation options: Amazon SageMaker Canvas, Ready-to-use models, My models, Shared models, Datasets (selected), Help, and Log out.

The main content area displays a list of datasets:

Name	Data type	Source	Cells (Columns x Rows)	Items	Create
Data_Banknote_Authentication_Train(1).csv	Tabular	S3	6,860 (5 x 1,372)	04/	
Data_Banknote_Authentication_Test.csv	Tabular	S3	1,375 (5 x 275)	04/	
canvas-sample-loans-part-2.csv	Tabular	S3	5,000 (5 x 1,000)	04/	
canvas-sample-loans-part-1.csv	Tabular	S3	19,000 (19 x 1,000)	04/	
canvas-sample-maintenance.csv	Tabular	S3	9,000 (9 x 1,000)	04/	
canvas-sample-product-descriptions.csv	Tabular	S3	600 (5 x 120)	04/	
canvas-sample-sales-forecasting.csv	Tabular	S3	5,000 (5 x 1,000)	04/	
canvas-sample-shipping-logs.csv	Tabular	S3	12,000 (12 x 1,000)	04/	
canvas-sample-housing.csv	Tabular	S3	10,000 (10 x 1,000)	04/	

Buttons at the top right include: Dataset search, Join data, Import, Create, and Create a model.

Screenshot of the Create new model dialog box.

The left sidebar is identical to the previous screenshot.

The dialog box has the following fields:

- Model name:** BNA-CanvasModel (Spelling correction: Open with the left-click (Alt+Down Arrow) characters.)
- Problem type:** Predictive analysis (selected)
- Predictive analysis description:** Build models using tabular datasets to predict single or multiple categories as well as regression and time-series forecast problems.
- Image analysis:** Build models using image datasets to predict single or multiple categories for image classification problems.
- Text analysis:** Build models using tabular datasets to predict single or multiple categories for text classification problems.

Buttons at the bottom right are: Cancel and Create.

Screenshot of the SageMaker Canvas interface showing the 'Build' tab for the 'BNA-CanvasModel'. The 'Validate your data' section indicates it might take several minutes depending on the dataset size. The 'Select a column to predict' section shows 'class' as the target column, with a 'Value distribution' chart showing 0 and 1. The 'Model type' section recommends '2 category prediction' for classifying 'class' into two categories. A 'Quick build' button is available. Below is a data visualization table for 'Data_Banknote_Authentication_Train(...)' with 1.4k rows:

Column name	Data type	Missing	Mismatched	Unique	Mean / Mode	Correlation to target
Variance	Numeric	0.00% (0)	0.00% (0)	1,338	0.57	-0.725
Skewness	Numeric	0.00% (0)	0.00% (0)	1,256	-4.46	-0.445
Entropy	Numeric	0.00% (0)	0.00% (0)	1,156	-0.99	-0.023
Curtosis	Numeric	0.00% (0)	0.00% (0)	1,270	1.24	0.156
class	Target	Binary	0.00% (0)	2	0	--

Total columns: 5, Total rows: 1,372, Total cells: 6,860, Show dropped columns.

Screenshot of the SageMaker Canvas interface showing the 'Build' tab for the 'BNA-CanvasModel'. The 'Validate your data' section indicates it might take several minutes depending on the dataset size. The 'Select a column to predict' section shows 'class' as the target column, with a 'Value distribution' chart showing 0 and 1. A 'Validate your data' modal is open, prompting to check for issues before training. The 'Model type' section recommends '2 category prediction' for classifying 'class' into two categories. A 'Quick build' button is available. Below is a data visualization table for 'Data_Banknote_Authentication_Train(...)' with 1.4k rows:

Column name	Data type	Missing	Mismatched	Unique	Mean / Mode	Correlation to target
Variance	Numeric	0.00% (0)	0.00% (0)	1,338	0.57	-0.725
Skewness	Numeric	0.00% (0)	0.00% (0)	1,256	-4.46	-0.445
Entropy	Numeric	0.00% (0)	0.00% (0)	1,156	-0.99	-0.023
Curtosis	Numeric	0.00% (0)	0.00% (0)	1,270	1.24	0.156
class	Target	Binary	0.00% (0)	2	0	--

Total columns: 5, Total rows: 1,372, Total cells: 6,860, Show dropped columns.

Screenshot of the BNA-CanvasModel page in the AWS SageMaker Canvas interface.

The page title is "BNA-CanvasModel". The navigation bar includes "Select", "Build", "Analyze" (which is underlined), and "Predict".

Model overview

Your model is being created. Quick build usually takes 2-15 minutes. You can now leave this view.

Expected build time: 2-15 minutes | Build type: Quick build | Detailed progress: Generating column impact

A small illustration of a person working at a desk with a laptop.

Bottom status bar: Data_Banknote_Authentication_Train(1).csv, Total columns: 5, Total rows: 1,372, Total cells: 6,860, class, 2 category prediction. Weather: 9°C Cloudy. System status: ENG IN. Date: 5:21 PM 2023-04-05.

Screenshot of the BNA-CanvasModel page in the AWS SageMaker Canvas interface, showing the model is now ready.

The page title is "BNA-CanvasModel". The navigation bar includes "Select", "Build", "Analyze" (underlined), and "Predict" (which is highlighted).

Model status

100% The model predicts the correct Class 100% of the time. ⓘ

Predict button

Overview and **Scoring** tabs are visible.

Column impact table:

Rank	Feature	Impact (%)
1	Variance	41.062%
2	Skewness	35.786%
3	Curtosis	19.343%
4	Entropy	3.809%

Impact of Variance on prediction of class scatter plot:

Scatter plot details: Impact on prediction vs Variance.

Bottom status bar: Data_Banknote_Authentication_Train(1).csv, Total columns: 5, Total rows: 1,372, Total cells: 6,860, class, 2 category prediction. Weather: 9°C Cloudy. System status: ENG IN. Date: 5:21 PM 2023-04-05.

BNA-CanvasModel

Select Build Analyze Predict

Model status

100%
The model predicts the correct Class 100% of the time. ⓘ

Overview Scoring

Predicted vs. Actual

	Predicted	Actual
Total	275	275
0	0	153
1	122	1

Model accuracy insights

If the model predicts **0**, it is correct **100% of the time**. ⓘ

For the values that are **0** in the dataset, the model predicted **100%** of them to be **0**. ⓘ

Data_Banknote_Authentication_Train(1).csv Total columns: 5 Total rows: 1,372 Total cells: 6,860 class 2 category prediction

Predict

This screenshot shows the SageMaker Canvas interface for the 'BNA-CanvasModel'. It displays a large '100%' indicating perfect accuracy. Below this, a 'Predicted vs. Actual' chart shows a 1x2 grid where all predictions for class 0 (purple) are correct (blue), and all predictions for class 1 (blue) are correct (purple). A sidebar on the right provides 'Model accuracy insights' stating that if the model predicts 0, it is correct 100% of the time.

Advanced metrics

Positive Class	F1 ⓘ	Accuracy ⓘ	Precision ⓘ	Recall ⓘ	AUC ⓘ
0	100%	100%	100%	100%	1

Model performance

	0	1
Actual values	<p>True positive (TP) 153 55.6% of predicted results</p> <p>False positive (FP) 0 0.0% of predicted results</p>	<p>False negative (FN) 0 0.0% of predicted results</p> <p>True negative (TN) 122 44.4% of predicted results</p>

Close Download

This screenshot shows the 'Advanced metrics' and 'Model performance' sections of the SageMaker Canvas interface. The advanced metrics table shows F1, Accuracy, Precision, Recall, and AUC all at 100%. The confusion matrix shows 153 True Positives (TP), 0 False Positives (FP), 0 False Negatives (FN), and 122 True Negatives (TN).

Select dataset for predictions

To make predictions on a dataset, select it or import it. The dataset that you select must have the same number of feature columns as the training dataset. [?](#)

+ Import

Search datasets in Canvas

Name	Columns	Rows	Cells	Created	Status
Data_Banknote_Authentication_Train(1).csv	5	1,372	6,860	04/05/2023 5:16 PM	Ready
Data_Banknote_Authentication_Test.csv	5	275	1,375	04/05/2023 5:14 PM	Ready
canvas-sample-loans-part-2.csv	5	1,000	5,000	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-loans-part-1.csv	19	1,000	19,000	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-maintenance.csv	9	1,000	9,000	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-product-descriptions.csv	5	120	600	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-sales-forecasting.csv	5	1,000	5,000	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-shipping-logs.csv	12	1,000	12,000	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-housing.csv	10	1,000	10,000	04/05/2023 5:13 PM	Incompatible ?
canvas-sample-diabetic-readmission.csv	16	1,000	16,000	04/05/2023 5:13 PM	Incompatible ?

Close Generate predictions

BNA-CanvasModel

Select Build Analyze Predict

Predict target values

Batch prediction Single prediction

Generate predictions for an entire dataset.

Select dataset

Predictions

Dataset Rows Created Status

batchInfer-BNA-CanvasModel-Data_Banknote_Authentication_Test.csv-1680729751	275	04/05/2023 5:22 PM	Ready
---	-----	--------------------	-------

batchInfer-BNA-CanvasModel-Data_Banknote_Authentication_Test.csv-1680729751 predictions ready View

9°C Cloudy 5:22 PM 2023-04-05

Screenshot of a web browser showing the Amazon SageMaker Canvas interface for a "BNA-CanvasModel".

The main panel displays a table titled "batchInfer-BNA-CanvasModel-Data_Banknote_Authentication_Test.csv-1680729751" with the following columns:

Prediction (class)	Probability	Variance	Skewness	Curtosis	Entropy
0	99.6%	1.5691	6.3465	-0.1828	-2.4099
0	95.8%	-0.27802	8.1881	-3.1338	-2.5276
0	97.1%	0.051979	7.0521	-2.0541	-3.1508
0	99.9%	-1.7559	11.9459	3.0946	-4.8978
0	99.3%	2.4287	9.3821	-3.2477	-1.4543
0	99.4%	4.6352	-3.0087	2.6773	1.212
0	99.8%	0.24835	7.6439	0.9685	-0.87371
0	99.3%	5.1731	3.9606	-1.983	0.40774
0	99.1%	4.0047	0.45937	1.3621	1.6181
0	99.8%	0.96441	5.8395	2.3235	0.066365
1	79.8%	1.5077	1.9596	-3.0584	-0.12243

A "Download CSV" button is located at the bottom right of the table.

The left sidebar shows navigation options: "Select", "Predict target values", "Batch prediction", "Generate predictions for", "Select dataset", "Predictions", "Dataset", and "batchInfer-BNA-CanvasModel".

The status bar at the bottom indicates "batchInfer-BNA-CanvasModel-Data_Banknote_Authentication_Test.csv-1680729751 predictions ready".

Autopilot

This guide shows how to create an Amazon SageMaker Autopilot experiment (that is, how to start an Autopilot job in SageMaker), so that you can explore, pre-process, and train various model candidates on a given dataset. This can help you get started with machine learning quickly.

You can use a user interface (Amazon SageMaker Studio UI) to help you populate the input, output, target, and parameters to run and evaluate an Autopilot experiment or use SageMaker API Reference. The UI has descriptions, toggle switches, dropdown menus, radio buttons, and more to help you navigate creating your model candidates. You can also view statistics while the experiment is running. After it runs, you can compare trials and delve into the details of the pre-processing steps, algorithms, and hyperparameter ranges of each model. You also have the option to download their [explainability](#) and [performance](#) reports. Use the provided [notebooks](#) to see the results of the automated data exploration or the candidate model definitions.

The following instructions show how to create an Amazon SageMaker Autopilot job as a pilot experiment using Studio UI or SageMaker API reference. You name your experiment, provide locations for the input and output data, and specify which target data to predict. Optionally, you can also specify the type of machine learning problem that you want to solve, choose your modeling strategy (*stacked ensembles* or *hyperparameters optimization*), select the list of algorithms used by the Autopilot job to train the data, and more.

Create an Autopilot experiment using Studio

To create an Amazon SageMaker Autopilot experiment using Studio

1. Sign in at <https://console.aws.amazon.com/sagemaker/>, select Studio from the left navigation pane, then choose Open Studio.
2. In Studio, choose the home icon () from the left navigation pane to view the Studio top-level navigation menu.
3. On the Home tab, choose the AutoML card. This opens a new AutoML tab.
4. Choose Create an AutoML experiment. This opens a new Create experiment tab.
5. In the Experiment and data details section, enter the following information:
 - a) Experiment name – Must be unique to your account in the current AWS Region and contain a maximum of 63 alphanumeric characters. Can include hyphens (-) but not spaces.
 - b) Input data – Provide the Amazon Simple Storage Service (Amazon S3) bucket location of your input data. This S3 bucket must be in your current AWS Region. The URL must be in an s3:// format where Amazon SageMaker has write permissions. The file must be in CSV or Parquet format and contain at least 500 rows. Select Browse to scroll through available paths and Preview to see a sample of your input data.
 - c) Is your S3 input a manifest file? – A manifest file includes metadata with your input data. The metadata specifies the location of your data in Amazon S3. It also specifies how the data is formatted and which attributes from the dataset to use when training your model. You can use a manifest file as an alternative to preprocessing when your labeled data is being streamed in Pipe mode.

- d) Auto split data? – Autopilot can split your data into an 80-20% split for training and validation data. If you prefer a custom split, you can choose the Specify split ratio. To use a custom dataset for validation, choose Provide a validation set.
 - e) Output data location (S3 bucket) – The name of the S3 bucket location where you want to store the output data. The URL for this bucket must be in an Amazon S3 format where Amazon SageMaker has write permissions. The S3 bucket must be in the current AWS Region. Autopilot can also create this for you in the same location as your input data.
6. Choose Next: Target and features. The Target and features tab opens.
 7. In the Target and features section, select a column to set as a target for model predictions. You can also select features for training and change their data type. The following data types are available: Text, Numerical, Categorical, Datetime, Sequence, and Auto. All features are selected by default.
 8. Choose Next: Training method. The Training method tab opens.
 9. In the Training method section, select your training option: Ensembling, Hyperparameter optimization (HPO), or Auto to let Autopilot choose the training method automatically based on the dataset size. Each training mode runs a pre-defined set of algorithms on your dataset to train model candidates. By default, Autopilot pre-select all the available algorithms for the given training mode. You can run an Amazon SageMaker Autopilot training experiment with all the algorithms or choose your own subset.

For more information on the training modes and the available algorithms, see the Autopilot training modes section in the [Training modes and algorithms](#) page.

10. Choose Next: Deployment and advanced settings to open the Deployment and advanced settings tab. Settings include auto display endpoint name, machine learning problem type, and additional choices for running your experiment.
 - a) Deployment settings – Autopilot can automatically create an endpoint and deploy your model for you.

To auto deploy to an automatically generated endpoint, or to provide an endpoint name for custom deployment, set the toggle to Yes under Auto deploy? If you are importing data from Amazon SageMaker Data Wrangler, you have additional options to auto deploy the best model with or without the transforms from Data Wrangler.

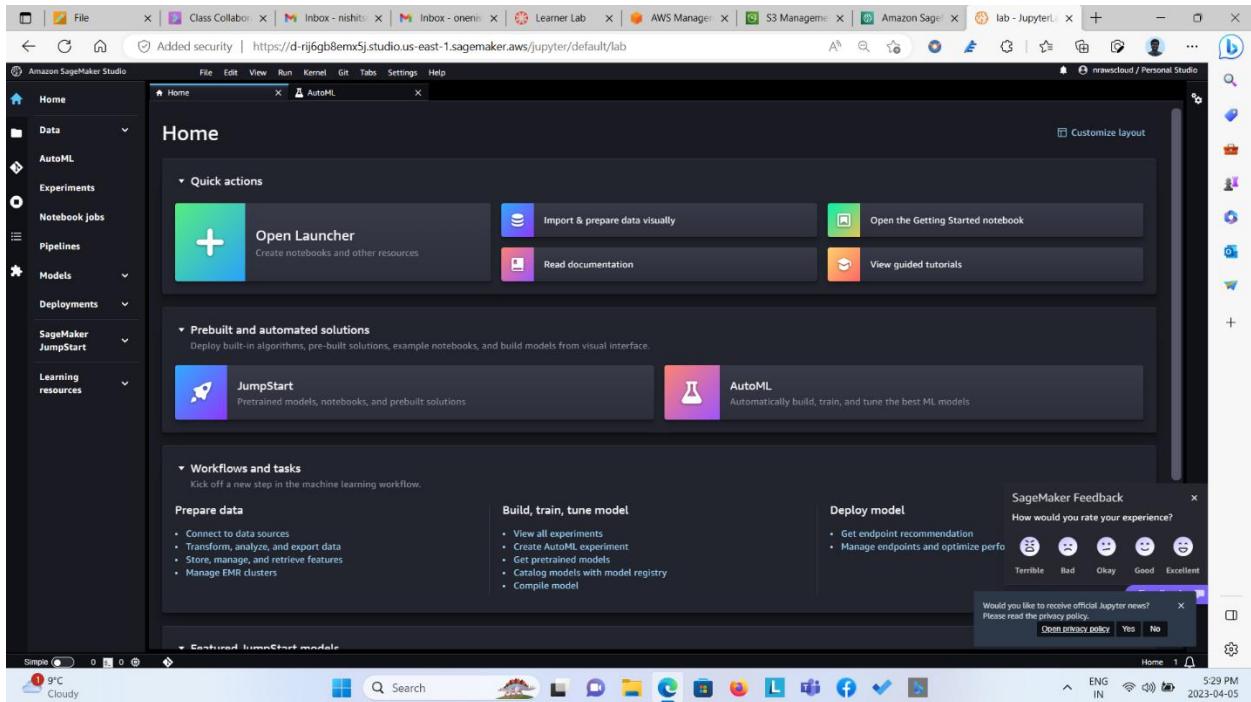
Note

If your Data Wrangler flow contains multi-row operations such as groupby, join, or concatenate, you won't be able to auto deploy with these transforms. For more information see [Automatically Train Models on Your Data Flow](#).

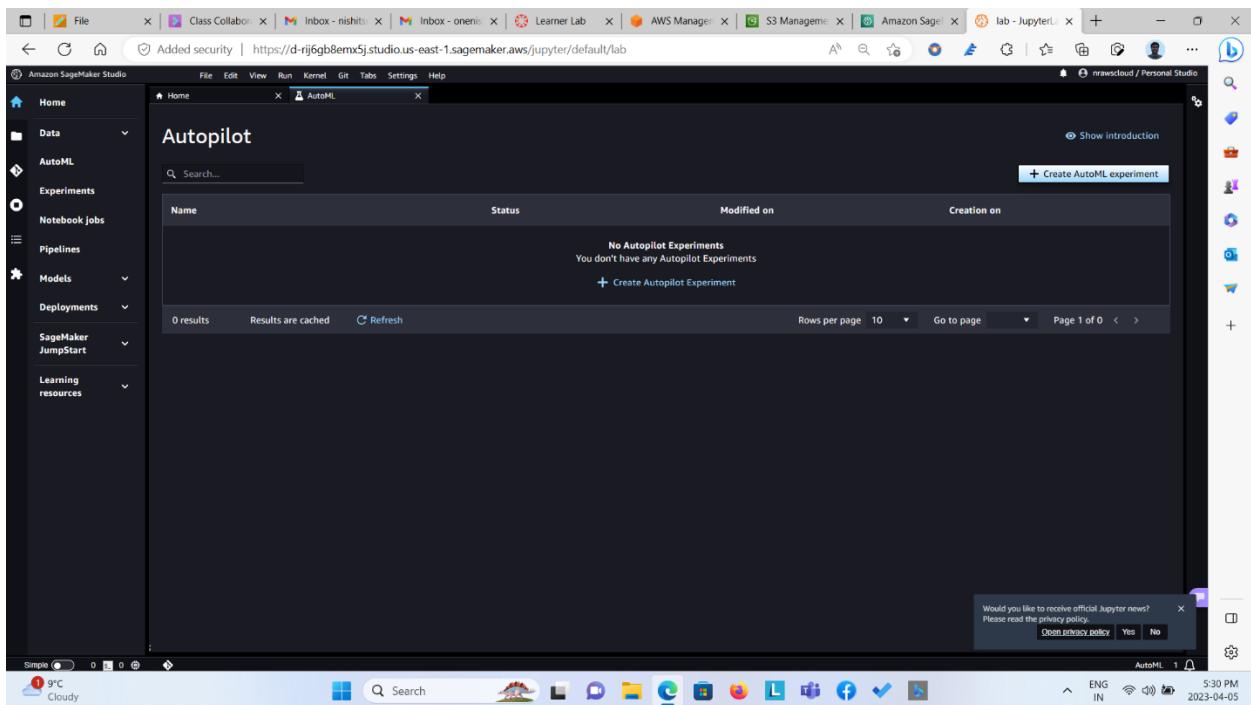
- a) Advanced settings (optional) – Autopilot provides additional controls to manually set experimental parameters such as defining your problem type, time constraints on your Autopilot job and trials, security, and encryption settings.
 - i. Machine learning problem type – Autopilot can automatically select the machine learning problem type. If you prefer to choose it manually, use the Select the machine learning problem type dropdown menu.

- A. Auto – Autopilot infers the problem type from the values of the attribute that you want to predict. In some cases, SageMaker is unable to infer accurately. When that happens, you must provide the value for the job to succeed.
 - B. Binary classification– Binary classification is a type of supervised learning that assigns an individual to one of two predefined and mutually exclusive classes, based on their attributes. For example, medical diagnosis based on results of diagnostic tests that determine if someone has a disease.
 - C. Regression – Regression estimates the values of a dependent target variable based on one or more variables or attributes that are correlated with it. For example, house prices based on features, such as square footage and number of bathrooms.
 - D. Multiclass classification – Multiclass classification is a type of supervised learning that assigns an individual to one of several classes based on their attributes. For example, the prediction of the topic most relevant to a text document, such as politics, finance, or philosophy.
- b) Choose Next: Review and create to get a summary of your Autopilot experiment before you create it.
11. Select Create experiment. The creation of the experiment starts an Autopilot job in SageMaker. Autopilot provides status on the course of the experiment, information on the data exploration process and model candidates in notebooks, a list of generated models and their reports, and the job profile used to create them.

For information on the notebooks generated by an Autopilot job, see [Amazon SageMaker Autopilot notebooks generated to manage AutoML tasks](#). For information on the details of each model candidate and their reports, see [Models generated by Amazon SageMaker Autopilot](#).



The screenshot shows the Amazon SageMaker Studio Home page. The left sidebar includes sections for Home, Data, AutoML, Experiments, Notebook jobs, Pipelines, Models, Deployments, SageMaker JumpStart, and Learning resources. The main content area features a "Quick actions" section with links to Open Launcher, Import & prepare data visually, Open the Getting Started notebook, Read documentation, and View guided tutorials. Below this is a "Prebuilt and automated solutions" section with links to JumpStart and AutoML. The "Workflows and tasks" section contains links for Prepare data, Build, train, tune model, and Deploy model. A "SageMaker Feedback" section at the bottom right allows users to rate their experience from Terrible to Excellent. A notification bar at the bottom asks if the user wants to receive official Jupyter news.



The screenshot shows the Amazon SageMaker Studio Autopilot page. The left sidebar is identical to the Home page. The main content area displays a table for Autopilot experiments, showing columns for Name, Status, Modified on, and Creation on. A message indicates "No Autopilot Experiments" and "You don't have any Autopilot Experiments". A "Create Autopilot Experiment" button is available. A "Show introduction" link is located in the top right. A notification bar at the bottom right asks if the user wants to receive official Jupyter news.

S3 URI path
Enter an S3 URI

Object name	Size	Last modified
Data_Banknote_Authentication_Test.csv	9.11KB	2023-04-05T21:11:04.000Z
Data_Banknote_Authentication_Train.csv	45.36KB	2023-04-05T21:11:04.000Z

Preview
Data_Banknote_Authentication_Train.csv (First 100 rows shown)

Variance	Skewness	Kurtosis	Entropy	class
3.6216	8.6661	-2.8073	-0.44699	0
4.5459	8.1674	-2.4585	-1.4621	0
3.866	-2.6383	1.9242	0.10645	0
3.4566	9.5228	-4.0112	-3.5944	0
0.32324	-4.4552	4.5718	-0.9888	0
4.3684	9.6718	-3.9606	-3.1625	0
3.5912	3.0129	0.72888	0.56421	0
2.0922	-6.81	8.4636	-0.60216	0
3.2032	5.7588	-0.75345	-0.61251	0

Cancel Select

Create an Autopilot experiment

When you create an Autopilot experiment, Amazon SageMaker analyzes your data and creates a notebook with candidate model definitions. This notebook provides visibility into how models are selected, trained, and tuned.

Experiment and data details > Target and features > Training method and algorithms > Deployment and advanced settings > Review and create

Experiment name: BNA-AutoML

Input data

Specify the input data location and choose how to split your data. Please ensure your input file is in CSV, parquet or manifest format and contains valid data.

S3 location: s3://nr-autopilot-canvas/Data_Banknote_Authentication_Train.csv

Is your S3 input a manifest file? No

Auto split data? Yes

Output data

Specify the output data location.

Auto create output data location? Yes

Cancel Next: Target and features

Screenshot of the Amazon SageMaker Studio interface showing the "Create an Autopilot experiment" wizard. The current step is "Target and features".

Target
Select the Target for the model to predict.
Target: class

Features
By default, all features are used for training. You can also select specific features to use for training.
4/5 columns selected
Column name Data type
class Target
Variance Auto
Skewness Auto
Curtosis Auto
Entropy Auto

Buttons: Cancel, Previous: Experiment and data details, Next: Training method.

Screenshot of the Amazon SageMaker Studio interface showing the "Create an Autopilot experiment" wizard. The current step is "Training method and algorithms".

Training method and algorithms
Select the training method for solving your machine learning problems.

Auto
Let Autopilot automatically decide the training method based on your dataset size.

Ensembling
Autopilot uses an AutoML algorithm that trains a multi-layer stack ensemble model to predict on regression and classification datasets directly from your data.

Hyperparameter optimization (HPO)
Autopilot finds the best version of a model by tuning hyperparameters and running training jobs on your data set.

Buttons: Cancel, Previous: Target and features, Next: Deployment and advanced settings.

The screenshot shows the 'Create an Autopilot experiment' wizard in Amazon SageMaker Studio. The current step is 'Deployment settings'. Key configuration includes:

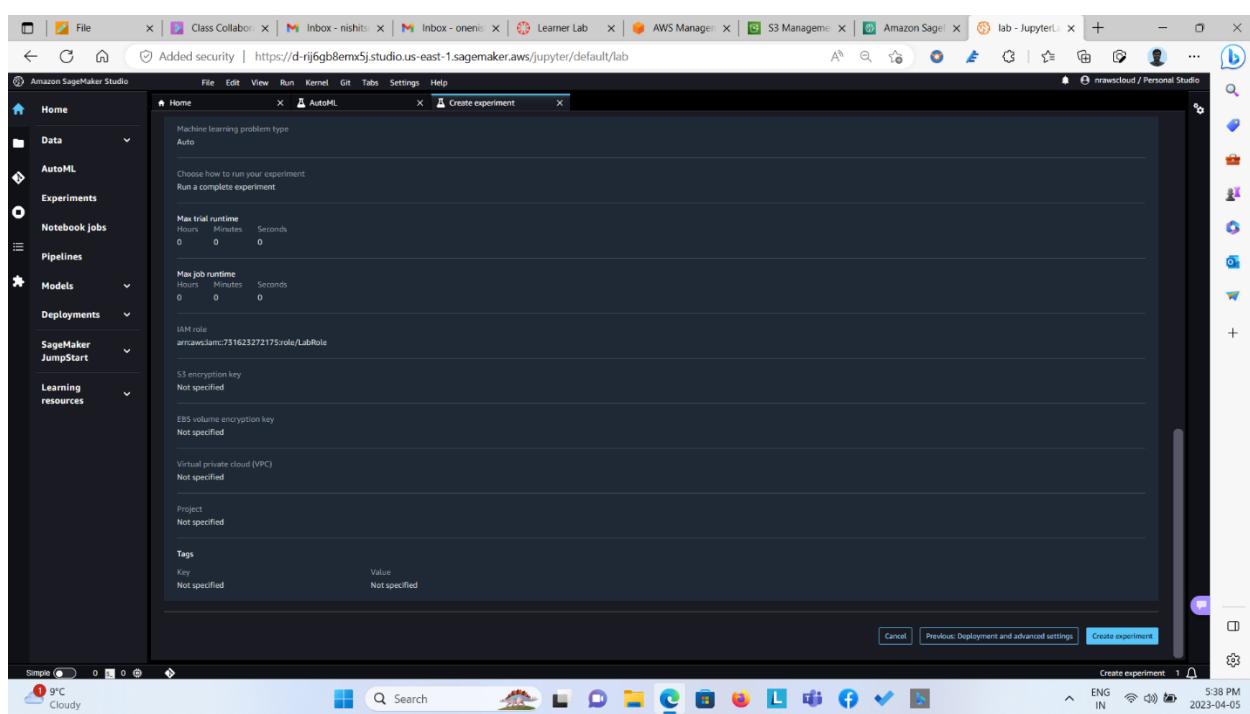
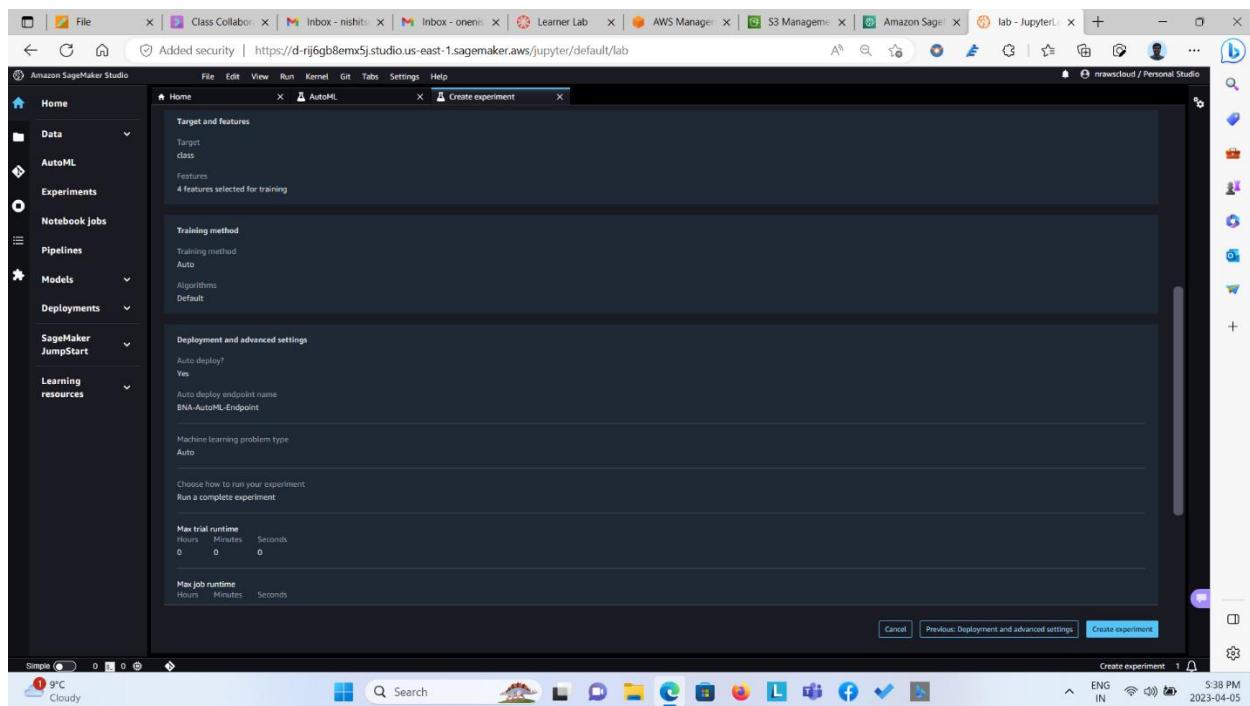
- Auto deploy:** Yes
- Auto deploy endpoint name:** BNA-AutoML-Endpoint
- Machine learning problem type:** Auto

The sidebar on the left shows the navigation menu for SageMaker Studio.

The screenshot shows the 'Create an Autopilot experiment' wizard in Amazon SageMaker Studio at the 'Review and create' step. Configuration details are as follows:

- Experiment name:** BNA-AutoML
- Input data location:** s3://nv-autopilot-canvay/Data_Banknote_Authentication_Train.csv
- Is your S3 input a manifest file?**: No
- Auto split data?**: Yes
- Auto create output data location?**: Yes

The sidebar on the left shows the navigation menu for SageMaker Studio.



The screenshot shows the Amazon SageMaker Studio interface. On the left, the navigation sidebar includes sections for Home, Data, AutoML, Experiments, Notebook jobs, Pipelines, Models, Deployments, SageMaker JumpStart, and Learning resources. The main area displays an 'AUTOPILOT JOB' titled 'BNA-AutoML'. It shows a progress bar indicating 'less than 20 seconds ago' and a status message: 'Your models will display here as they are generated.' A large circular icon with a spiral pattern is centered. To the right, a 'Training progress' section lists three steps: '1. Training Models', '2. Insights Report Generated', and '3. Deploying Model'. A small feedback dialog box is overlaid at the bottom right. The system tray at the bottom shows the date and time as 2023-04-05 5:38 PM.

The screenshot shows the AWS Management Console with the URL https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/processing-jobs. The left sidebar has a tree view with categories like JumpStart, Processing (selected), and Training. The main content area shows a table of 'Processing jobs' with columns for Name, ARN, Creation time, Duration, and Status. There are 12 entries listed, all in progress. The bottom of the page includes standard AWS footer links and the date and time as 2023-04-05 5:40 PM.

Name	ARN	Creation time	Duration	Status
BNA-AutoML-t7-1-3ce7ac6c522b46dfb82678790096ff6069b8652ffe33492	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t7-1-3ce7ac6c522b46dfb82678790096ff6069b8652ffe33492	4/5/2023, 5:39:31 PM	a few seconds	② InProgress
BNA-AutoML-t6-1-8f066be26cc4597e23472ff46a7b1c82d263de495	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t6-1-8f066be26cc4597e23472ff46a7b1c82d263de495	4/5/2023, 5:39:23 PM	a few seconds	② InProgress
BNA-AutoML-t2-1-b500375de9e14d1d9870fe938122916ebaf0bd6f1f6f45b	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t2-1-b500375de9e14d1d9870fe938122916ebaf0bd6f1f6f45b	4/5/2023, 5:39:16 PM	a few seconds	② InProgress
BNA-AutoML-t10-1-dd5146a0474c98823b2ff8833d570332e952e261746	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t10-1-dd5146a0474c98823b2ff8833d570332e952e261746	4/5/2023, 5:39:10 PM	a minute	② InProgress
BNA-AutoML-t1-1-c8f25a4dc1e141e38ba2848d0bbd3d2ea0e623ca832f4c8	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t1-1-c8f25a4dc1e141e38ba2848d0bbd3d2ea0e623ca832f4c8	4/5/2023, 5:39:03 PM	a minute	② InProgress
BNA-AutoML-t9-1-8a43a50562664d9eb310196afe0c5a79bca66414026415	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t9-1-8a43a50562664d9eb310196afe0c5a79bca66414026415	4/5/2023, 5:38:59 PM	a minute	② InProgress
BNA-AutoML-t5-1-f46f25fb8f841968d76b2e78b17e22d957b8061a4c415	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t5-1-f46f25fb8f841968d76b2e78b17e22d957b8061a4c415	4/5/2023, 5:38:53 PM	a minute	② InProgress
BNA-AutoML-t8-1-599bb4ee0ea84565fa3a8314a2fefa85d9d7caa7e8ab400	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t8-1-599bb4ee0ea84565fa3a8314a2fefa85d9d7caa7e8ab400	4/5/2023, 5:38:48 PM	a minute	② InProgress
BNA-AutoML-t3-1-dfffc148c0dc4491d8764baa2b704d16eedaa819245574c0	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t3-1-dfffc148c0dc4491d8764baa2b704d16eedaa819245574c0	4/5/2023, 5:38:42 PM	a minute	② InProgress
BNA-AutoML-t4-1-8475cf8353b47f88d6fb8fc324602d4fb8edae65594a9	arn:aws:sagemaker:us-east-1:731623272175:processing-job/bna-automl-t4-1-8475cf8353b47f88d6fb8fc324602d4fb8edae65594a9	4/5/2023, 5:38:34 PM	a minute	② InProgress

Screenshot of Amazon SageMaker Studio showing the AutoML interface. The left sidebar shows various notebooks and pipelines. The main panel displays an 'AUTOPilot JOB' titled 'BNA-AutoML' for 'BinaryClassification'. It lists several trials, each with metrics like Objective, Accuracy, AUC, Precision, Recall, Loss, and InferenceLatency. The best model is 'NeuralNetTorch-BAG-L1-FULL-i731623272175BNA-AutoML' with an Objective of 1, Accuracy of 1, AUC of 1, Precision of 1, Recall of 1, Loss of 0.001, and InferenceLatency of 0.089. The status for all trials is 'Completed'.

Screenshot of the AWS Management Console showing the 'Processing jobs' section under the 'Amazon SageMaker' service. The left sidebar includes sections for JumpStart, Processing, Training, Inference, Edge Manager, Augmented AI, and AWS Marketplace. The main table lists 14 completed processing jobs, each with a name, ARN, creation time (e.g., 4/5/2023, 5:39:31 PM), duration (e.g., 10 minutes), and status (e.g., Completed). The names of the jobs follow a consistent pattern involving 'BNA-AutoML' and various job IDs.

The screenshot shows the Amazon SageMaker Studio interface. On the left, a sidebar lists various services: Home, Data, AutoML, Experiments, Notebook jobs, Pipelines, Models, Deployments, SageMaker JumpStart, and Learning resources. The main content area displays the 'Model Details' for a job named 'BNA-AUTOML-AWS-AUTO-ML-JOB'. The status is 'COMPLETED'. The job details include:

- Job Name: BNA-AutoML.aws-auto-ml-job
- Model: NeuralNetTorch-BAG-L1-FULL-t5731623272175BNA-AutoML-aws-trial (Best model)
- Status: Completed
- Algorithm: NeuralNetTorch

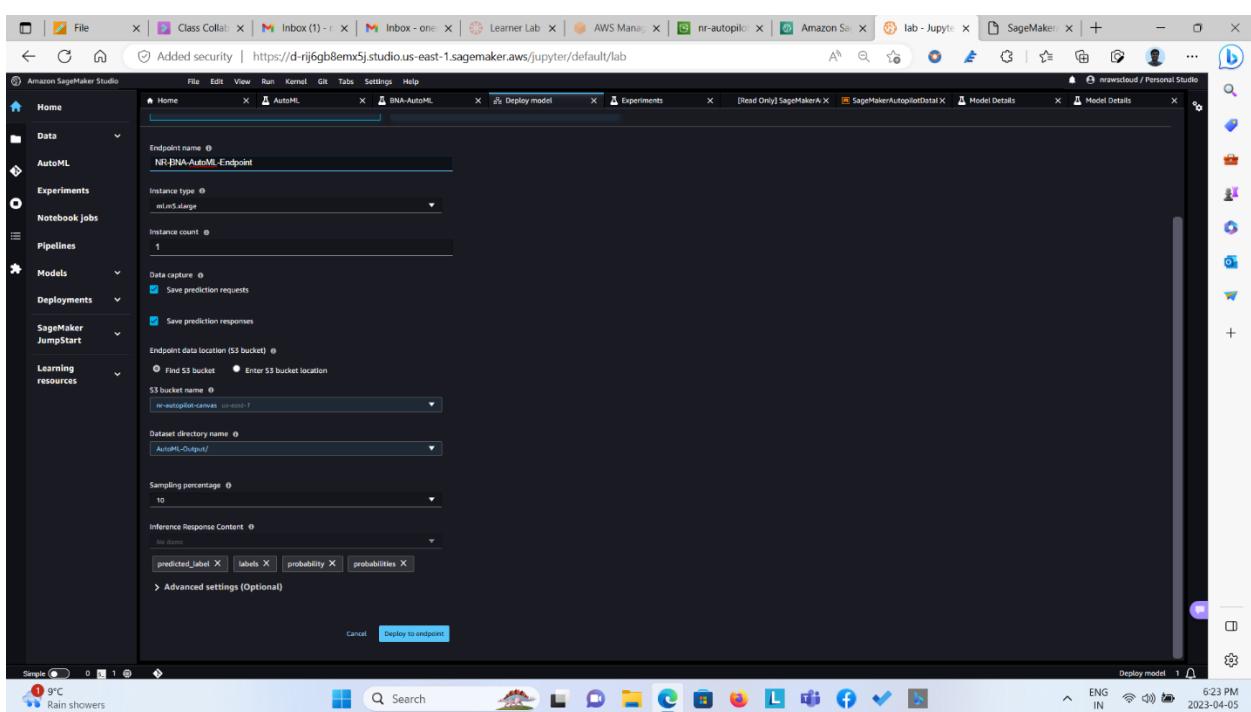
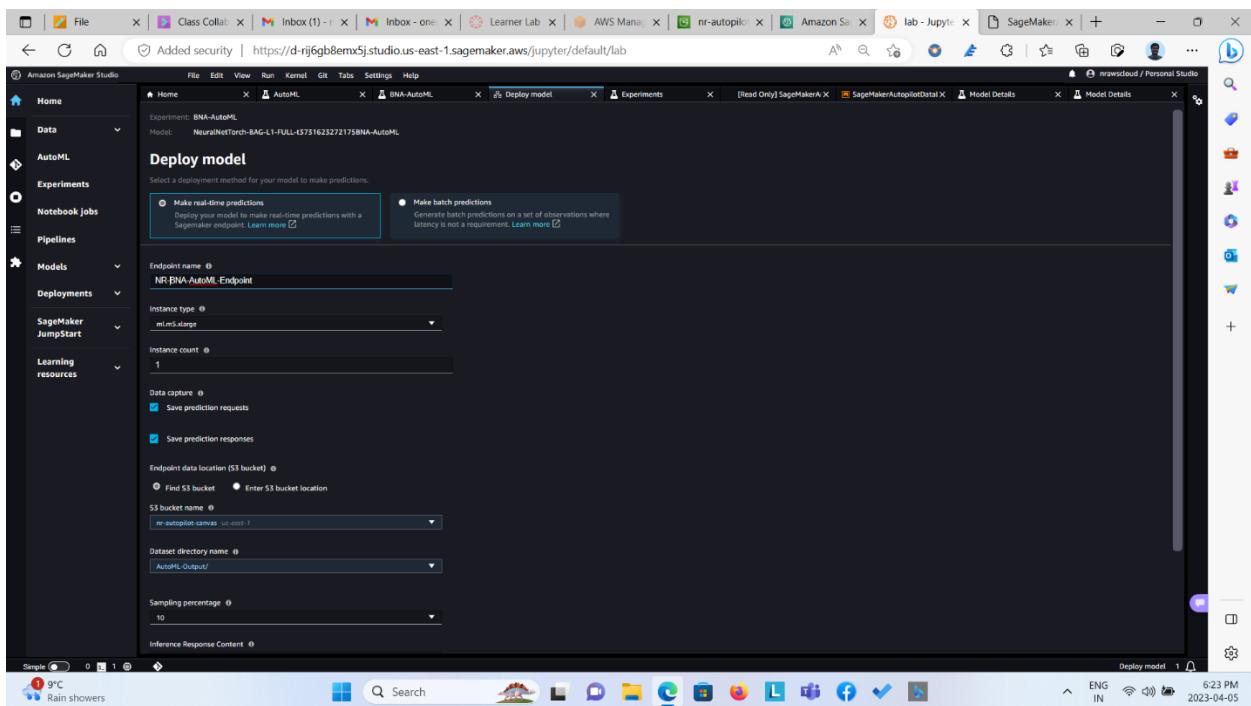
Below the job details, there are tabs for Explainability, Performance, Artifacts, and Network. Under the Artifacts tab, it shows the input data URL: https://us-east-1.console.aws.amazon.com/s3/object/nr-autopilot-canvas/Data_Banknote_Authentication_Train.csv. It also lists the shuffled training and validation splits URL: <https://console.aws.amazon.com/s3/buckets/nr-autopilot-canvas/BNA-AutoML/> and the algorithm model URL: <https://console.aws.amazon.com/s3/object/nr-autopilot-canvas/BNA-AutoML/candidates>.

This screenshot is identical to the one above, showing the 'Model Details' page for the 'BNA-AUTOML-AWS-AUTO-ML-JOB'. The job status is 'COMPLETED'. The job details and artifact URLs are the same. However, the 'Network' tab is selected, revealing two configuration options:

- Enable network isolation: True
- Enable inter-container traffic encryption: True

The screenshot shows the Amazon SageMaker Studio interface. On the left, a sidebar navigation bar includes Home, Data, AutoML, Experiments, Notebook jobs, Pipelines, Models, Deployments, SageMaker JumpStart, and Learning resources. The main content area displays the 'Experiments' dashboard with a title 'Experiments 1'. A table lists one experiment: 'BNA-AutoML-aws-auto-ml-job' (Type: Autopilot job, Description: --, Modified on: 04/05/23 5:52 PM GMT-4, Created on: 04/05/23 5:58 PM GMT-4, Created by: nrawscloud). Below the table are buttons for Refresh, Experiments per page (10), Go to page (1), and Page 1 of 1.

The screenshot shows the AWS S3 console. The left sidebar lists Buckets, Storage Lens, Dashboards, AWS Organizations settings, and AWS Marketplace for S3. The main area shows the path 'Amazon S3 > Buckets > nr-autopilot-canvas > Create folder'. A modal window titled 'Create folder' is open. It contains a note about using folders to group objects and a warning about bucket policies preventing folder creation if they block specific object types. The 'Folder name' field is populated with 'AutoML-Output'. Under 'Server-side encryption', it says 'The following settings apply only to the new folder object and not to the objects contained within it.' and shows 'Encryption key type' with 'Amazon S3 managed keys (SSE-S3)' selected. At the bottom right of the modal are 'Cancel' and 'Create folder' buttons.



Endpoint Details

Endpoint settings

Name	Endpoint type
NR-BNA-AutoML-Endpoint	Real-time

Data capture settings

Enable data capture	Sampling percentage (%)
Enabled	10

Endpoint runtime settings

Variant name	Current weight	Desired weight	Instance type	Elastic inference	Current instance count	Desir.
default-variant-name	1	1	ml.m5.xlarge	-	1	1

Endpoint configuration settings

Name	Encryption key
NR-BNA-AutoML-Endpoint-config-1680733406227	-

Model monitoring

Disabled

Sampling percentage (%)

10

Data capture options

- Prediction request
- Prediction response

Capture content type

- CSV/Text
- JSON

Production variants

default-variant-name

Model name: NR-BNA-AutoML-Endpoint-model-1680733406227

Training job: -

The screenshot shows the Amazon SageMaker Studio interface. The left sidebar navigation includes Home, Data, AutoML, Experiments, Notebook jobs, Pipelines, Models, Deployments, SageMaker JumpStart, and Learning resources. The main content area displays the 'Endpoint configuration settings' and 'Production variants' sections for a model named 'NR-BNA-AutoML-Endpoint-config-1680733406227'. Under 'Production variants', there is a section for 'default-variant-name' which includes 'Model name' (NR-BNA-AutoML-Endpoint-model-1680733406227), 'Training job' (with 'Instance details' and 'Lineage' options), and 'Tags' (with two entries: 'sagemaker-user-profile-arm' and 'sagemaker-domain-arm'). The status bar at the bottom shows the date and time as 2023-04-05 6:30 PM.

The screenshot shows the AWS Management Console with the URL https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/models. The left sidebar navigation includes Governance, Ground Truth, Notebook, Processing, Training, Inference, Edge Manager, Augmented AI, and AWS Marketplace. The main content area displays the 'Models' section under 'Amazon SageMaker > Models'. It shows a table with two rows:

Name	ARN	Creation time
NR-BNA-AutoML-Endpoint-model-1680733406227	arn:aws:sagemaker:us-east-1:731623272175:model/nr-bna-automl-endpoint-model-1680733406227	4/5/2023, 6:23:24 PM
BNA-AutoML-best-model	arn:aws:sagemaker:us-east-1:731623272175:model/bna-automl-best-model	4/5/2023, 5:52:08 PM

The status bar at the bottom shows the date and time as 2023-04-05 6:32 PM.

Screenshot of the AWS Management Console showing the Endpoint configuration page for Amazon SageMaker.

The left sidebar shows the navigation menu under the "Amazon SageMaker" section, including "Endpoint configuration".

The main content area displays the "Endpoint configuration" table:

Name	ARN	Creation time
NR-BNA-AutoML-Endpoint-config-1680733406227	arn:aws:sagemaker:us-east-1:731623272175:endpoint-config/nr-bna-automl-endpoint-config-1680733406227	4/5/2023, 6:23:25 PM

Buttons at the top right include "Apply to endpoint", "Clone", "Actions", and "Create endpoint configuration".

System tray icons and status bar information are visible at the bottom.

Screenshot of the AWS Management Console showing the Endpoints page for Amazon SageMaker.

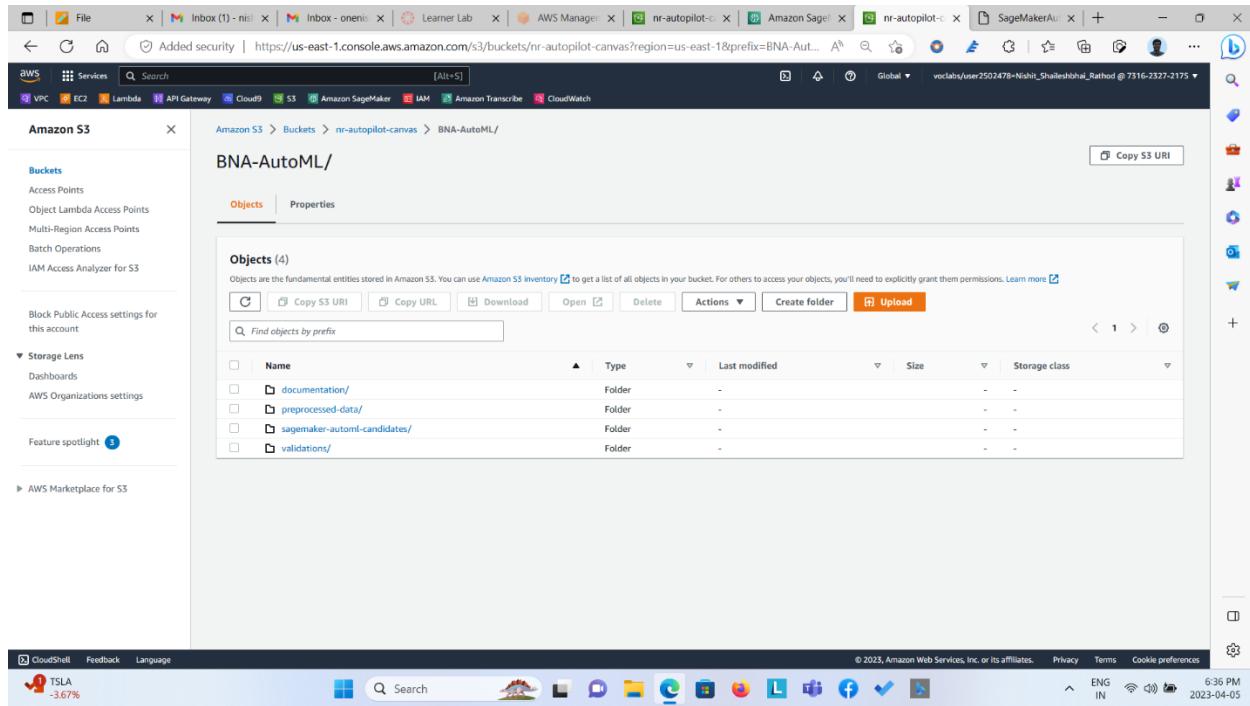
The left sidebar shows the navigation menu under the "Amazon SageMaker" section, including "Endpoints".

The main content area displays the "Endpoints" table:

Name	ARN	Creation time	Status	Last updated
NR-BNA-AutoML-Endpoint	arn:aws:sagemaker:us-east-1:731623272175:endpoint/nr-bna-automl-endpoint	4/5/2023, 6:23:25 PM	InService	4/5/2023, 6:25:40 PM

Buttons at the top right include "Update endpoint", "Actions", and "Create endpoint".

System tray icons and status bar information are visible at the bottom.



Assignment

- Select a data set and use Canvas to train a model based on that. Capture the results in a report (**0.5 mark**)
- Use AutoPilot to create a model based on the same dataset (find best mode) (**0.5 mark**)
- Compare the AutoPilot model quality with the model created by Canvas(**1 mark**)
- Select one item in AutoPilot and research on that. That could be different gained results by using different algorithms that Autopilot has used, or understanding the data exploration notebook that it has generated or understanding different reports that AutoPilot generates (like the report I have uploaded into BB). You create the result of research comprehensibly in a report (**3 marks**).

Observations

The Linear Learner Model trained on Canvas gave 100% accuracy in classifying banknote, whether the note is genuine or forged. The model evaluation metrics has identical results when compared to AutoML.

BNA-CanvasModel

Select Build Analyze Predict

Model status
100%
The model predicts the correct Class 100% of the time. ⓘ

Predict

Overview Scoring

Column Impact ⓘ

Column	Impact (%)
Variance	41.062%
Skewness	35.786%
Curtosis	19.343%
Entropy	3.805%

Impact of Variance on prediction of class

Data_Banknote_Authentication_Train(1).csv Total columns: 5 Total rows: 1,372 Total cells: 6,860 class 2 category prediction

9°C Cloudy 5:21 PM 2023-04-05

BNA-CanvasModel

Select Build Analyze Predict

Model status
100%
The model predicts the correct Class 100% of the time. ⓘ

Predict

Overview Scoring

Predicted vs. Actual

		Predicted	Actual
		0	1
Total	0 275	1 1	
Predicted	0 Correct 0 153	1 Correct 1 122	

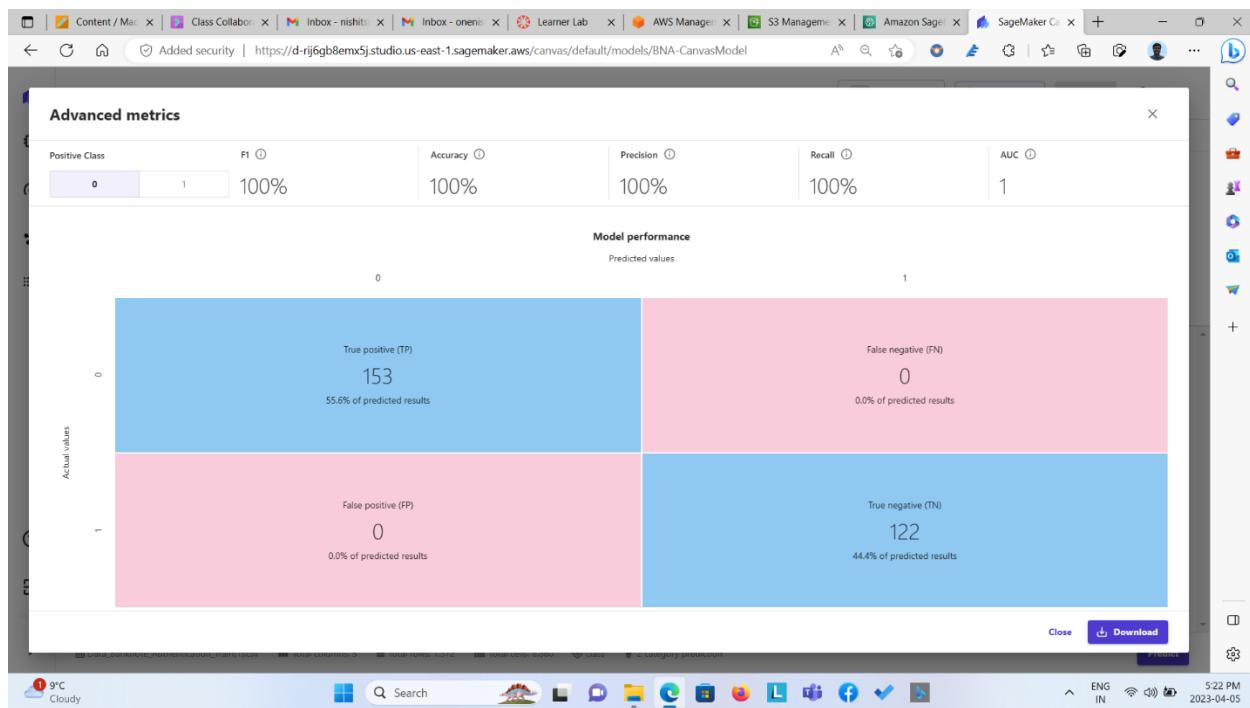
Model accuracy insights
If the model predicts 0, it is correct 100% of the time. ⓘ

Advanced metrics

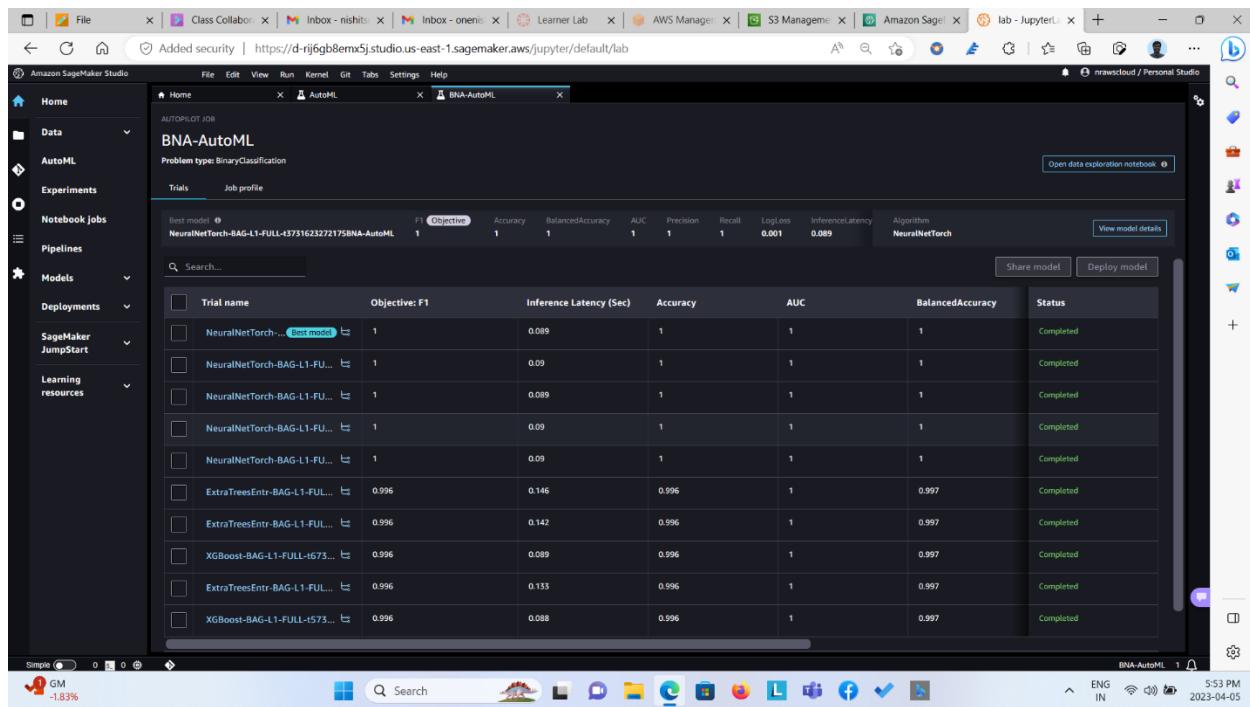
For the values that are 0 in the dataset, the model predicted 100% of them to be 0. ⓘ

Data_Banknote_Authentication_Train(1).csv Total columns: 5 Total rows: 1,372 Total cells: 6,860 class 2 category prediction

9°C Cloudy 5:21 PM 2023-04-05

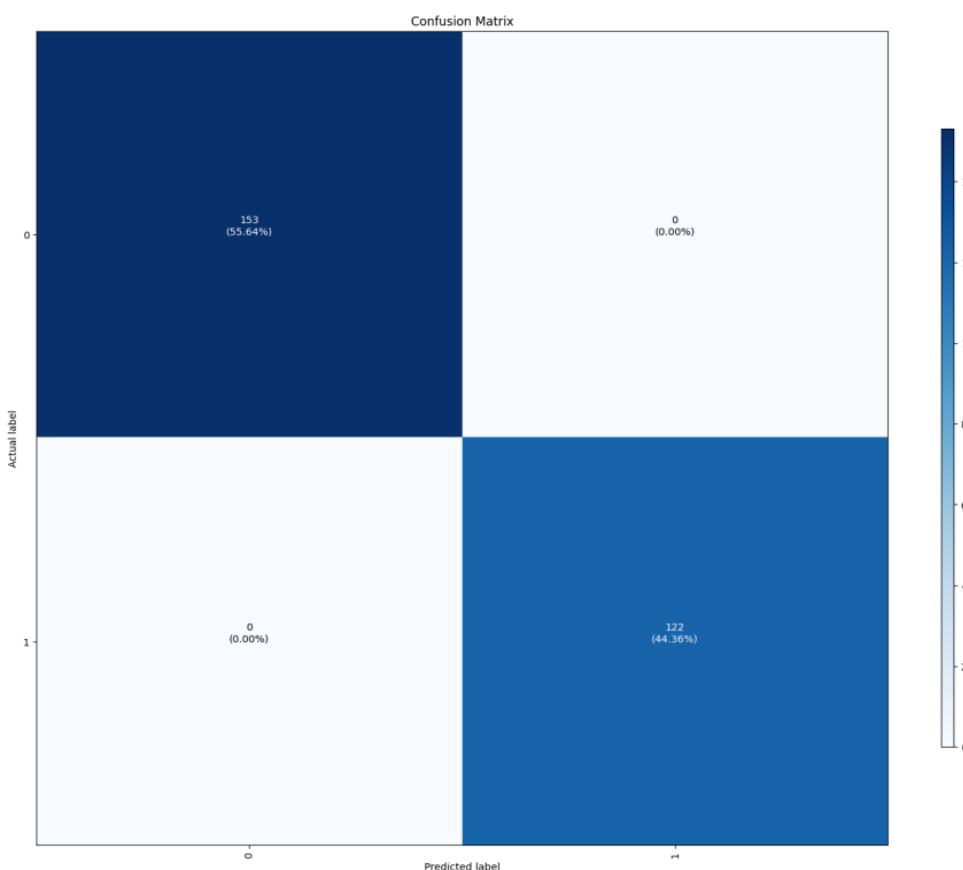


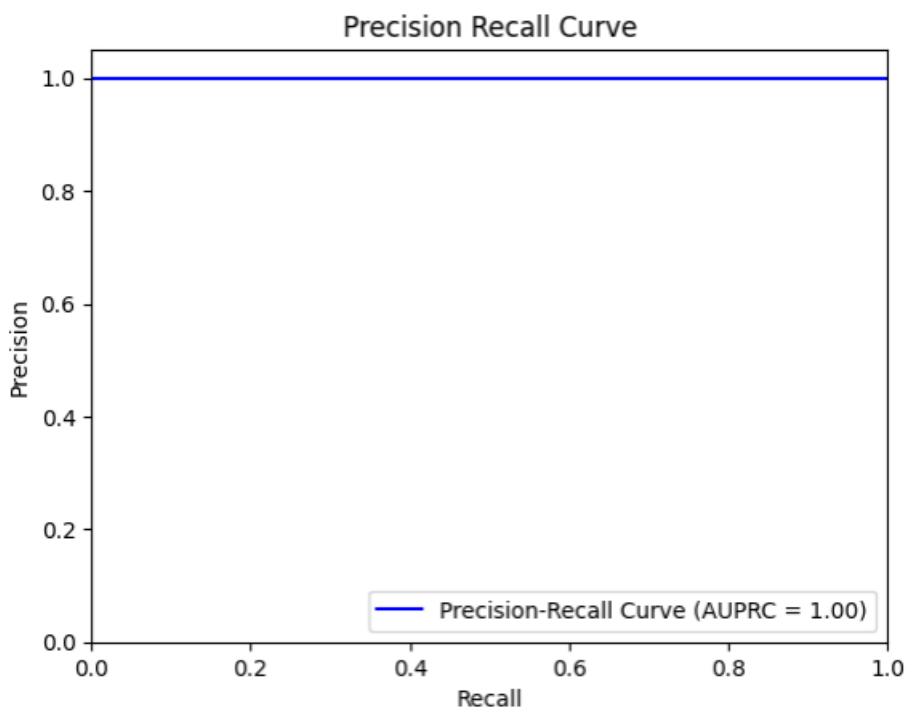
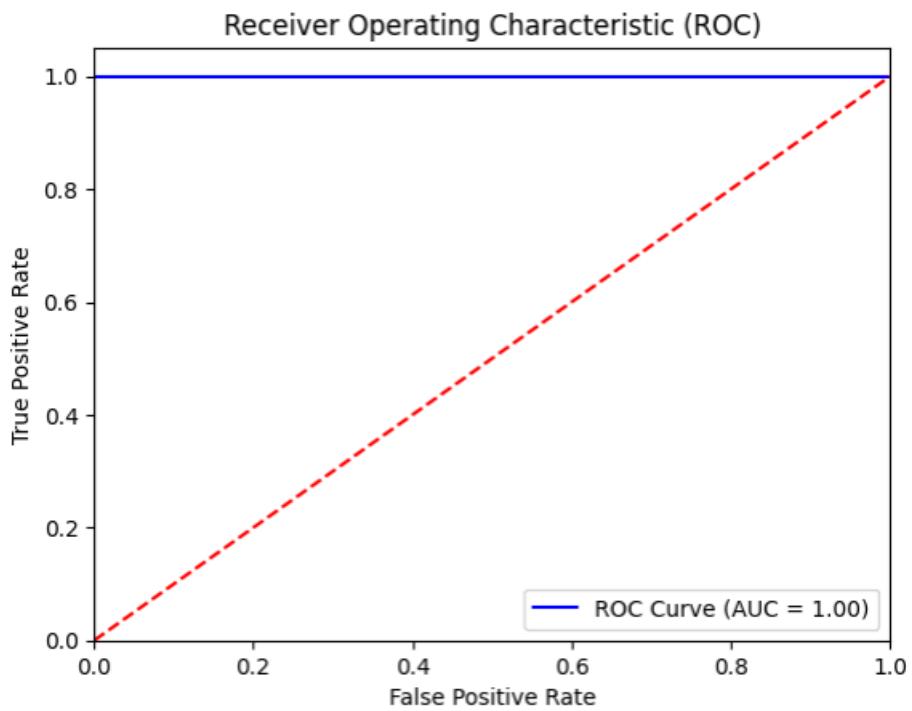
The models when implemented in Sagemaker Studio's via. AutoML also gave same accuracy of 100% in classifying Banknote. Among 10 models “Neural Net Torch BAG L1” gave the best classification results. The AutoML model evaluation metrics has identical results when compared to Canvas.



Metrics table

Metric Name	Value	Standard Deviation
recall	1.000000	0.000000
precision	1.000000	0.000000
accuracy	1.000000	0.000000
f0_5	1.000000	0.000000
f1	1.000000	0.000000
f2	1.000000	0.000000
recall_best_constant_classifier	0.000000	0.000000
precision_best_constant_classifier	0.000000	0.000000
accuracy_best_constant_classifier	0.556364	0.013502
f0_5_best_constant_classifier	0.000000	0.000000
f1_best_constant_classifier	0.000000	0.000000
f2_best_constant_classifier	0.000000	0.000000
true_positive_rate	1.000000	0.000000
true_negative_rate	1.000000	0.000000
false_positive_rate	0.000000	0.000000
false_negative_rate	0.000000	0.000000
auc	1.000000	0.000000
au_prc	1.000000	0.000000



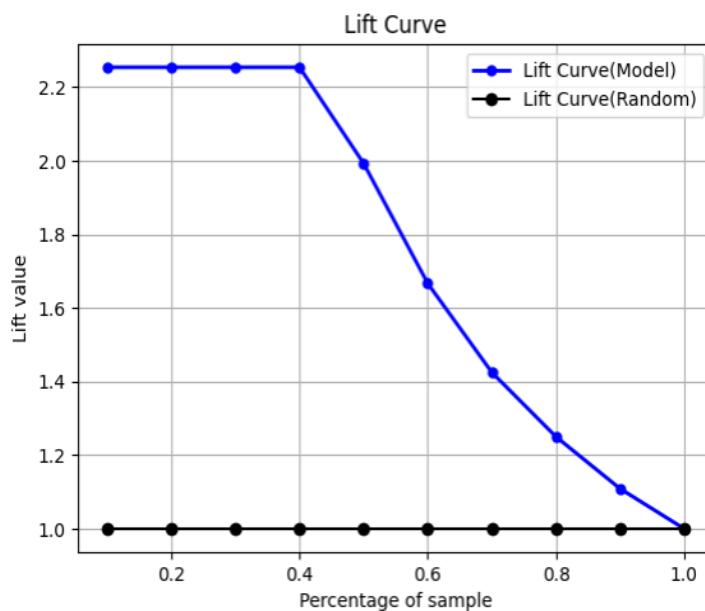
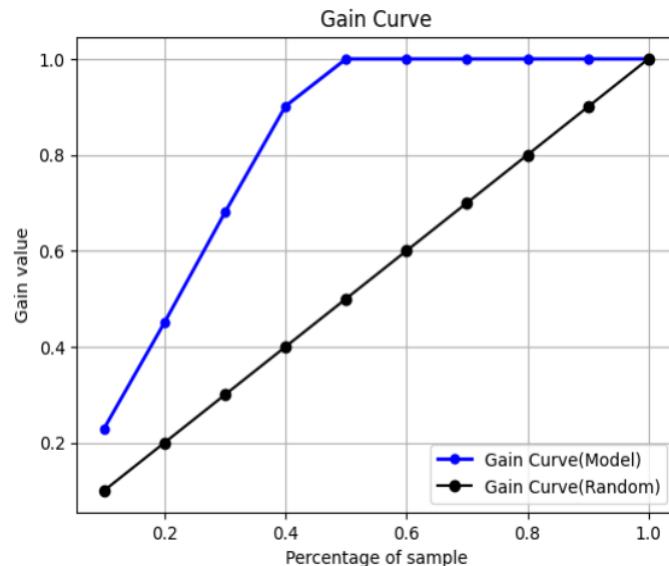


Question

Select one item in AutoPilot and research on that. That could be different gained results by using different algorithms that Autopilot has used, or understanding the data exploration notebook that it has generated or understanding different reports that AutoPilot generates (like the report I have uploaded into BB). You create the result of research comprehensibly in a report.

Answer

The thing on which, I had no knowledge was the Gain Curve and Lift curve which was derived and evaluated in the best model performance report i.e. on NEURAL NET TORCH BAG L1 FULL T3 model.



When researched more onto it I came to know that the Gain Curve and Lift Curve are two important evaluation metrics used in machine learning to measure the performance of binary classification models. Both of these metrics are based on the concept of ranking the predicted probabilities of a classifier and comparing them with the actual binary labels. Here, I will discuss in detail what Gain Curve and Lift Curve are, how they are calculated, and how they are used to evaluate machine learning models.

The gain curve and lift curve are two measures that are used for Measuring the benefits of using the model and are used in business contexts such as target marketing. It's not just restricted to marketing analysis. It can also be used in other domains such as risk modelling, supply chain analytics, etc. In other words, Gain and Lift charts are two approaches used while solving classification problems with imbalanced data sets.

Example: In target marketing or marketing campaigns, the customer responses to campaigns are usually very low (in many cases the customers who respond to marketing campaigns are less than 1%). The organization will raise the cost for each customer contact and hence would like to minimize the cost of the marketing campaign and at the same time achieve desired response level from the customers.

The gain chart and lift chart is the measures in logistic regression that will help organizations to understand the benefits of using that model. So that better and more efficient output carry out.

The gain and lift chart is obtained using the following steps:

1. Predict the probability $Y = 1$ (positive) using the LR model and arrange the observation in the **decreasing order of predicted probability** [i.e., $P(Y = 1)$].
2. Divide the data sets into deciles. Calculate the number of positives ($Y = 1$) in each decile and the cumulative number of positives up to a decile.
3. **Gain** is the ratio between the cumulative number of positive observations up to a decile to the total number of positive observations in the data. The gain chart is a chart drawn between the gain on the vertical axis and the decile on the horizontal axis.

$$Gain = \frac{\text{Cumulative number of positive observations upto decile } i}{\text{Total number of positive observations in the data}}$$

4. **Lift** is the ratio of the number of positive observations up to decile i using the model to the expected number of positives up to that decile i based on a random model. Lift chart is the chart between the lift on the vertical axis and the corresponding decile on the horizontal axis.

$$Lift = \frac{\text{Cumulative number of positive observations upto decile } i \text{ using ML model}}{\text{Cumulative number of positive observations upto decile } i \text{ using random model}}$$

Gain Chart Calculation:

Input

Decile	Number of Observations divided into 10 sections	Positive Response using Model [X]	Cumulative Positive Response using Model	Gain Score of individual decile [Xi/Total]	Gain	Gain(%)
1	500	223	223	0.4280	0.4280	42.80%
2	500	122	345	0.2341	0.6622	66.22%
3	500	74	419	0.1420	0.8042	80.42%
4	500	38	457	0.072	0.8772	87.72%
5	500	27	484	0.052	0.9290	92.90%
6	500	11	495	0.021	0.9501	95.01%
7	500	18	513	0.034	0.9846	98.46%
8	500	3	516	0.005	0.9904	99.04%
9	500	4	520	0.007	0.9981	99.81%
10	500	1	521	0.001	1.000	100.00%

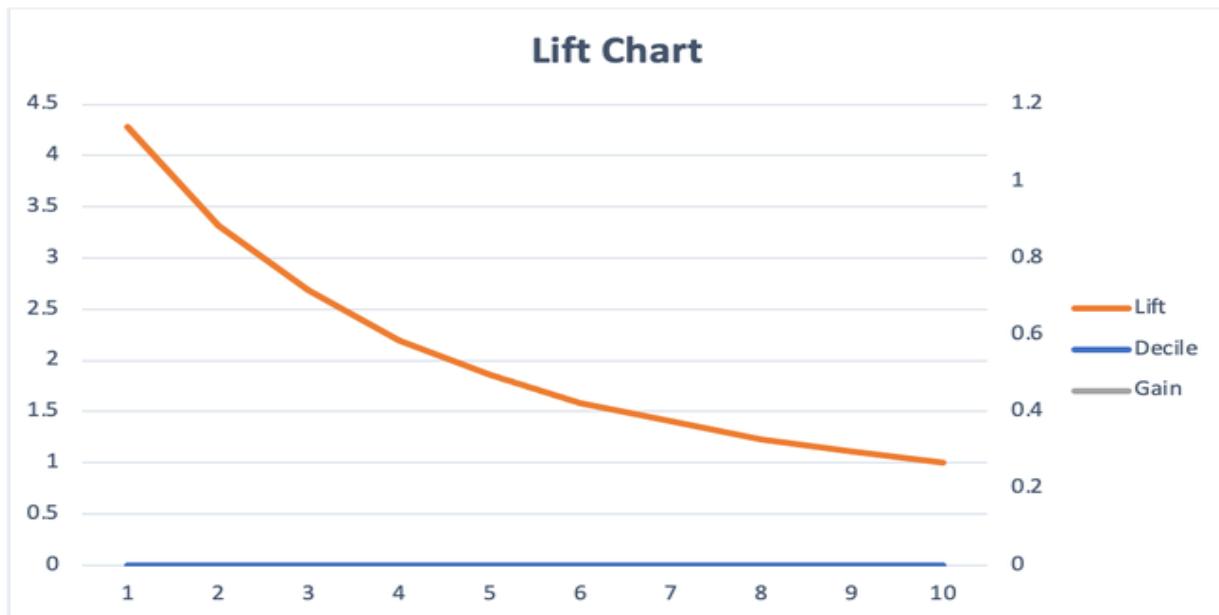
Ratio between the cumulative number of positive response up to a decile to the total number of positive responses in the data.



Lift Curve:

Decile	Cumulative Positive Response Without Model [A]	Positive Response using Model [x]	Cumulative Positive Response using Model [B]	Gain	Lift [B/A]
1	52.1	223	223	0.4280	4.2802
2	104.2	122	345	0.6622	3.3109
3	156.3	74	419	0.8042	2.6807
4	208.4	38	457	0.8772	2.1929
5	260.5	27	484	0.9290	1.8580
6	312.6	11	495	0.9501	1.5835
7	364.7	18	513	0.9846	1.4066
8	416.8	3	516	0.9904	1.2380
9	468.9	4	520	0.9981	1.1090
10	521	1	521	1.000	1.000

Ratio of the number of positive responses up to decile i using the model to the expected number of positives up to that decile i based on a random model



- Cumulative gains and lift charts are visual aids for measuring model performance.
- Both charts consist of Lift Curve (In Lift Chart) / Gain Chart (In Gain Chart) and Baseline (Blue Line for Lift, Orange Line for Gain).
- The Greater the area between the Lift / Gain and Baseline, the **Better** the model.