Machine Learning Model for SMS SPAM Detection Using AWS SageMaker Canvas

odels

Search models

+ New model

List

Filter by: Ready

Last viewed

✓ Ready

Banknote Authentication Project
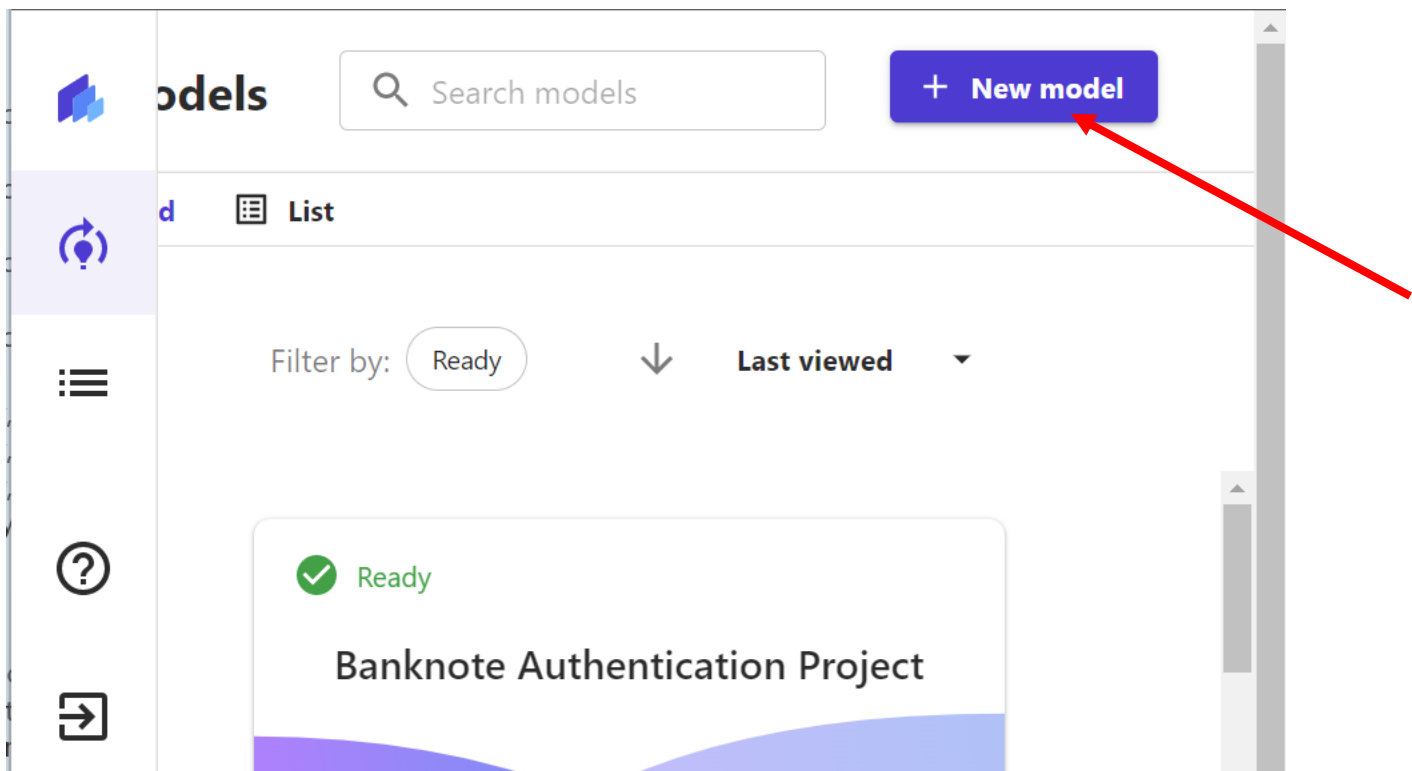
Models

Search models

+ New model

Grid   List

Create new model

×

Model name

SMS SPAM DETECTION

Use only letters, numbers, and underscores up to 32 characters.

Cancel   **Create**

SMS SPAM DETECTION

V1 • Draft ▾   ⊕ Add version   👥 Share   🕑   ⋮

Select   Build   Analyze   Predict

Select dataset

⇥ Join   + Import

You can import a tabular dataset or choose one that has already been imported. Your dataset must contain at least one input column and a target column.

🔍 Search datasets in Canvas

All   Joined

Name	Source	Columns	Rows	Cells

Import

Upload | S3 | Add Redshift or Snowfla[ke]

Upload files to import

Drag a CSV file here

or

⬆ Select files from your computer

Cancel   Import data

Import

×

Choose files to import

Amazon S3 / newmlbucket123 /
SMS SPAM DETECTION

🔍 Search Amazon S3

☐	Name
☐	📄 Spam+SMS+Detection+
☐	📄 Spam+SMS+Detection+

Import

×

Choose files to import

Amazon S3 / newmlbucket123 /
SMS SPAM DETECTION

🔍 Search Amazon S3

▬	Name
☑	📄 Spam+SMS+Detection+-+Train.csv
☑	📄 Spam+SMS+Detection+-+Test.csv

2 new datasets   Preview all   Cancel   **Import data**

SMS SPAM DETECTION

| V1 | ● Draft ▾ |

⊕ Ad
vers

| **Select** | **Build** | **Analyze** | Pi |

Select dataset

You can import a tabular dataset or choose one that has already been imported. Your dataset must contain at least one input column and a target column.

🔍 Search datasets in Canvas

| **All** | **Joined** |

Name

Name	**Source**
⦿ Spam+SMS+Detection+-+Train(1).csv	S3
○ Spam+SMS+Detection+-+Test(1).csv	S3
○ Bank+Note+Authentication+-+Test.csv	S3

SMS SPAM DETECTION

Select   **Build**   Analyze   Predict

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

◎ | Target column ▾ |

Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

To see a recommended model type, specify a value for the target column.

| Quick build | ▾ |

| Preview model |

SMS SPAM DETECTION

V1 ● Draft ▾   ⊕ Add version   👥 Share   🕐   ⋮

Select   **Build**

Spam+SMS+Detection+-+Trai
Full dataset: 1.2k rows ✏️

☑ **Column name** ↓

☑ SMS Text

☑ Check

Random sample ⓘ   ✕

Sample size: 1199 rows
                  ‾‾‾‾

                                    ┌─────────┐
                                    │1,199 rows│
                                    └─────────┘
●━━━━━━━━━●━━━━━━━━━●━━━━━━━━━●
500   1,199
                                  (recommended)

                              Cancel   Update

▦ Total columns: 2   ≡ Total rows: 1,199   ▦ Total cells: 2,398   ☑ Show dropped columns

⊞ Data visualizer ⌃

Unique ⓘ   **Mean / Mode**

,185   Sorry, I'll call later

               ham

SMS SPAM DETECTION

Select	**Build**	Analyze	Predict

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

⊙ **Target column**
Check   ▾

Value distribution

ham

spam

Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

💡 **2 category prediction**

Your model classifies Check into two categories.

Change type

Spam+SMS+Detection+-+Train(1).csv

Full dataset: 1.2k rows ✏️

Data visualizer

	Column name ↓		Data type	Missing ⓘ	Mismatched ⓘ	Unique ⓘ	Mean / Mode
☑	SMS Text		Text	0.00% (0)	0.00% (0)	1,185	Sorry, I'll call later
◎	Check	Target	Binary	0.00% (0)	0.00% (0)	2	ham

Spam+SMS+Detection+-+Train(1).csv

Full dataset: 1.2k rows ✏️

Data visualizer

☑ SMS Text	A ⋮	◎ Check	Abc ⋮	
1185 Values		2 Categories		
This is the 2nd time we have tried...		spam		
He's just gonna worry for nothing...		ham		
Did you get any gift? This year i d...		ham		
somewhere out there beneath th...		ham		
Well there's a pattern emerging o...		ham		

▦ Total columns: 2   ≡ Total rows: 1,199   ▦ Total cells: 2,398   ⓘ Previewing first 100 rows   ☑ Show dropped columns

SMS SPAM DETECTION

V1 ● Draft ▾ ⊕ Add version 👥 Share 🕐 ⋮

Select   **Build**   Analyze   Predict

ⓘ **Validate your data**
It might take several minutes, depending on the dataset size.

Validate data

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

◎ ┌─ Target column ──────┐
   │ Check   ▾ │
   └──────────────────────┘

Value distribution

ham

Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

💡 2 category prediction

Your model classifies Check into two categories.

Change type

Quick build ▾

Preview model

Dataset issues

✕

You have **1** issues with **Spam+SMS+Detection+-+Train(1).csv**. We recommend you resolve them to ensure the model is built successfully:

All feature columns are text columns	All of the feature columns in your data are text columns	^

All of the feature columns that you've selected in your data are text columns. Text columns are dropped for standard builds, and this issue will cause your build to fail. You can select columns of a different data type, or you can start a quick build instead.

Back to build

SMS SPAM DETECTION

V1 • Dra

Model type   ×

Recommended

◉ 2 category model   ∧

Classify rows into two categories.

Example business questions

- Will a delivery be on time, yes or no?
- Will a customer churn, yes or no?

Cancel   **Change type**

▥ Total columns: 2   ≡ Total rows: 1,199   ▦ Total cells: 2,398   ☑ Show

Quick build ▼ | **Preview model**

✧ **Preview model** ✕

Estimated accuracy ⓘ

95%

The model predicts the correct target (Check) 95% of the time.

Column impact ⓘ

🔍 Search columns...

Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

2 category prediction

Your model classifies Check into two categories.

Change type

Quick build ▼

Preview model

98 ⓘ Previewing first 100 rows ☑ Show dropped columns

SMS SPAM DETECTION

V1 ● Ready ▾   ⊕ Add version   ♙ Share   ↺   ⋮

Select   Build   **Analyze**   Predict

Model status

98.776%

The model predicts the correct Check 98.776% of the time. ⓘ

[**Predict**]

Overview   Scoring

▦ Spam+SMS+Detection+-+Train(1).csv   ▥ Total columns: 2   ☰ Total rows: 1,199   ▦ Total cells: 2,398   ◎ Check   ♡ 2 category prediction   [**Predict**]

Advanced metrics

×

Positive Class		F1 ⓘ	Accuracy ⓘ	Precision ⓘ	Recall ⓘ	AUC ⓘ
spam	ham	95.385%	98.776%	100%	91.176%	0.993

Predicted values

	spam	ham
spam	True positive (TP) **31** 12.7% of predicted results	False negative (FN) **3** 1.2% of predicted results
ham	False positive (FP) **0** 0.0% of predicted results	True negative (TN) **211** 86.1% of predicted results

Actual Values

Advanced metrics   ✕

Positive Class	F1 ⓘ	Accuracy ⓘ	Precision ⓘ	Recall ⓘ	AUC ⓘ
spam **ham**	99.294%	98.776%	98.598%	100%	0.993

Predicted values

	ham	spam
ham	True positive (TP) **211** 86.1% of predicted results	False negative (FN) **0** 0.0% of predicted results
spam	False positive (FP) **3** 1.2% of predicted results	True negative (TN) **31** 12.7% of predicted results

Actual Values

SMS SPAM DETECTION

Select	Build	Analyze	**Predict**

Predict target values

Batch prediction	Single prediction

Generates predictions for an entire dataset.

Select a dataset to generate predictions

Select dataset

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 data**

🔍 Search datasets in Canvas

Name	Source	Columns	Rows	Cells	Created	Status
⚪ Spam+SMS+Detection+-+Test(1).csv	S3	2	374		12/27/2022 11:54 AM	Ready
⚪ Spam+SMS+Detection+-+Train(1).csv	S3	2	1,199		12/27/2022 11:54 AM	Ready

Close **Generate predictions**

batchInfer-SMS SPAM DETECTION-Spam+SMS+Detection+-+Test(1).csv-1672151277 ✕

Prediction (Check)	Probability	SMS Text
ham	94.9%	She.s fine. I have had difficulties with her phone. It works w...
ham	97.6%	Ugh my leg hurts. Musta overdid it on mon.
spam	94.8%	Call Germany for only 1 pence per minute! Call from a fixe...
spam	96.9%	YOU VE WON! Your 4* Costa Del Sol Holiday or å£5000 aw...
ham	97.8%	WOT STUDENT DISCOUNT CAN U GET ON BOOKS?

⤓ **Download CSV**

SMS SPAM DETECTION

V1 • Ready ▾   ⊕ Add version   👥 Share

Select   Build   Analyze   **Predict**

Predict target values

Batch prediction	**Single prediction**

Column ↑	Feature importance	Value
SMS Text	�— 100%	Sorry, I'll call later

Check Prediction   ▥ Copy

ham

■ Average prediction

ham   97.568% ⓘ

████████████████████████

spam   2.432% ⓘ

■

24