

AWS Tutorial: Build a Machine Learning Model

Step 1: Prepare Your Data

Upload Data in bucket

<input type="checkbox"/>	Name	Last modified	Size	Storage class
<input type="checkbox"/>	banking-batch.csv	Nov 12, 2018 2:31:48 PM GMT-0500	469.2 KB	Standard
<input type="checkbox"/>	banking.csv	Nov 12, 2018 2:31:27 PM GMT-0500	4.7 MB	Standard

Viewing 1

Step 2: Create a Training Datasource

Create the training datasource

Review and make any changes, and then click Finish.

Input data

Edit

Datasource ID ds-zdaGt8l9xuA
Datasource name Banking Data 1
Creation time Nov 12, 2018 2:37:45 PM
Status Completed

ML model settings

Edit

ML model Name ML model: Banking Data 1
ML model parameters Default (includes the ML model evaluation) - See Advanced settings below.
Evaluation name Evaluation: ML model: Banking Data 1
Evaluation data Amazon ML will split your training datasource into 70% for training and will reserve the remaining 30% for evaluation.

Recipe

Recipe

Recipes help Amazon Machine Learning find patterns in your data. If you did not provide a recipe, Amazon ML will generate one for you.
[Learn more.](#)

Advanced settings

Maximum ML model Size 100MB
Maximum number of data p... 10
Shuffle type for training data Auto
Regularization type L2
Regularization amount 1e-6 - Mild

Tags

Amazon ML copies a maximum of 10 tags from parent objects. Edit the list to keep the tags you need.

No tags

Step 3: Create an ML Model

Q ML model name or ID

Items per page: 10 << < 1 - 1 of 1 ML models > >>

	Name	ID	Status	Real time predictions	Creation time	Completion time	Datasource ID
<input type="checkbox"/>	ML model: Banking Data 1	ml-V7abUpT21XD	Completed	Not enabled	Nov 12, 2018 2:39:27 PM	4 mins.	ds-1yHOxc2pKr2

ML model name

ML model: Banking Data 1

ML model ID

ml-V7abUpT21XD

ML model type

Binary classification

Creation time

Nov 12, 2018 2:39:27 PM

Status

Completed

Datasource ID

ds-1yHOxc2pKr2

Log

Download log

Input schema

View input schema

Target attribute

Target

Target type

BINARY

Number of attributes

21

Evaluations created

1

Latest evaluation result

Not available

Batch predictions created

0

Tags

Add or edit tags

No tags

<< < 1 - 1 of 1 ML models > >>

Step 4: Review the ML Model's Predictive Performance and Set a Score Threshold

To review the summary

ML model performance

This chart shows the distributions of your predicted answers for the actual "1" and "0" records in your evaluation data. Any overlap of the actual "1" & "0" is where your ML model guesses wrong. [Learn more.](#)

Adjust the slider to indicate how much error you can tolerate from your ML model based on your needs. Moving the score threshold to the right decreases the number of false positives and increases the number of false negatives.

Explain this chart

Trade-off based on score threshold

0.5

Reset score threshold (0.5)

- 91% are correct
 - 581 true positive
 - 10,677 true negative
- 9% are errors
 - 315 false positive
 - 782 false negative

7% of the records are predicted as "1"

93% of the records are predicted as "0"

Save score threshold at 0.50

Advanced metrics

False positive rate

0.0287

0

1

Precision

0.6484

0

1

Recall

0.4263

0

1

Accuracy



0.9112

0

1

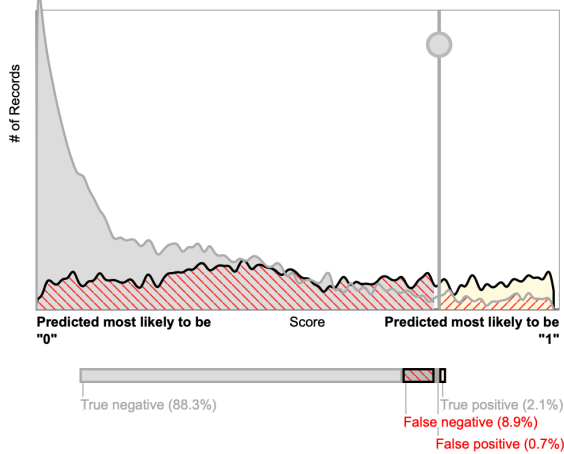
To set a score threshold for your ML model

ML model performance

This chart shows the distributions of your predicted answers for the actual "1" and "0" records in your evaluation data. Any overlap of the actual "1"  & "0"  is where your ML model guesses wrong. [Learn more.](#)

Adjust the slider to indicate how much error you can tolerate from your ML model based on your needs. Moving the score threshold to the right decreases the number of false positives and increases the number of false negatives.

Explain this chart



Score threshold saved successfully.

Trade-off based on score threshold [Reset score threshold \(0.77\)](#)

- **90% are correct**
263 true positive
10,906 true negative
- **10% are errors**
86 false positive
1,100 false negative

- 3% of the records are predicted as "1"
- 97% of the records are predicted as "0"

Save score threshold at 0.77

Advanced metrics

False positive rate 0.0078	0	<input type="range"/>	1
Precision 0.7536	0	<input type="range"/>	1
Recall 0.193	0	<input type="range"/>	1
Accuracy 0.904	0	<input type="range"/>	1

Step 5: Use the ML Model to Generate Predictions

Try real-time predictions

Try real-time predictions

You submitted 20 out of 20 data values for this prediction.

Try generating real-time predictions for free using the web browser on this page. To request a real-time prediction, complete the following form or provide a single data record in CSV format. To provide a data record, choose the **Paste a record** button.

Paste a record

Attribute name				Items per page: 10	<< < 1 - 10 of 21 > >>
	Name	Type	Value		
1	Var01	Numeric	32.0		
2	Var02	Categorical	services		
3	Var03	Categorical	divorced		
4	Var04	Categorical	basic.9y		
5	Var05	Categorical	no		
6	Var06	Categorical	unknown		
7	Var07	Categorical	yes		
8	Var08	Categorical	cellular		
9	Var09	Categorical	dec		
10	Var10	Categorical	mon		

Prediction results

Target name

ML model type BINARY

Predicted label 0

```
{
  "Prediction": {
    "details": {
      "Algorithm": "SGD",
      "PredictiveModelType": "BINARY"
    },
    "predictedLabel": "0",
    "predictedScores": {
      "0": 0.042566876858472824
    }
  }
}
```

Next steps