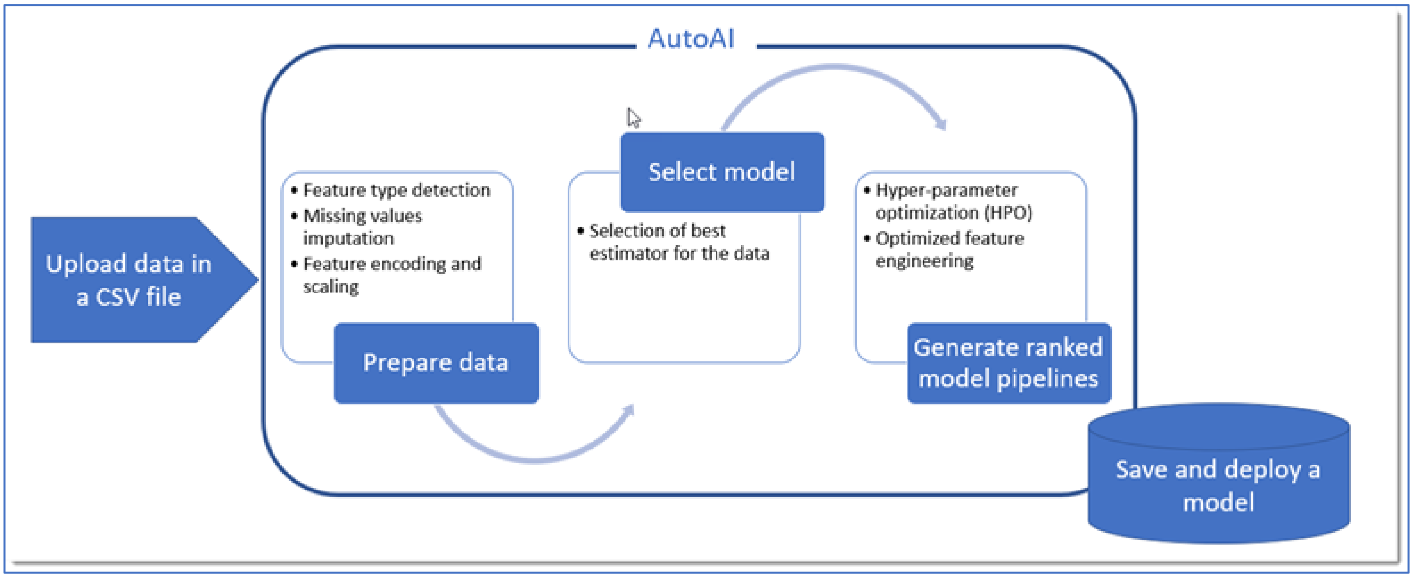
AutoAI Lab

This lab will demonstrate the new and exciting AutoAI capability to build and deploy an optimized model based on the Titanic data set.

AutoAI in Watson Studio automatically analyzes your data and generates candidate model pipelines customized for your predictive modeling problem. AutoAI algorithms analyze your dataset to discover data transformations, estimator algorithms, and parameter settings that work best for your problem setting. Results are displayed on a leaderboard, showing the automatically generated model pipelines ranked according to your problem optimization objective.

Using AutoAI, you can build and deploy a machine learning model with sophisticated training features and no coding. The tool does most of the work for you.


AutoAI
Upload data in a CSV file
• Feature type detection
• Missing values imputation
• Feature encoding and scaling
Prepare data
Select model
• Selection of best estimator for the data
• Hyper-parameter optimization (HPO)
• Optimized feature engineering
Generate ranked model pipelines
Save and deploy a model

The AutoAI process follows this sequence to build candidate pipelines:

- Data pre-processing
- Automated model selection
- Automated feature engineering
- Hyperparameter optimization

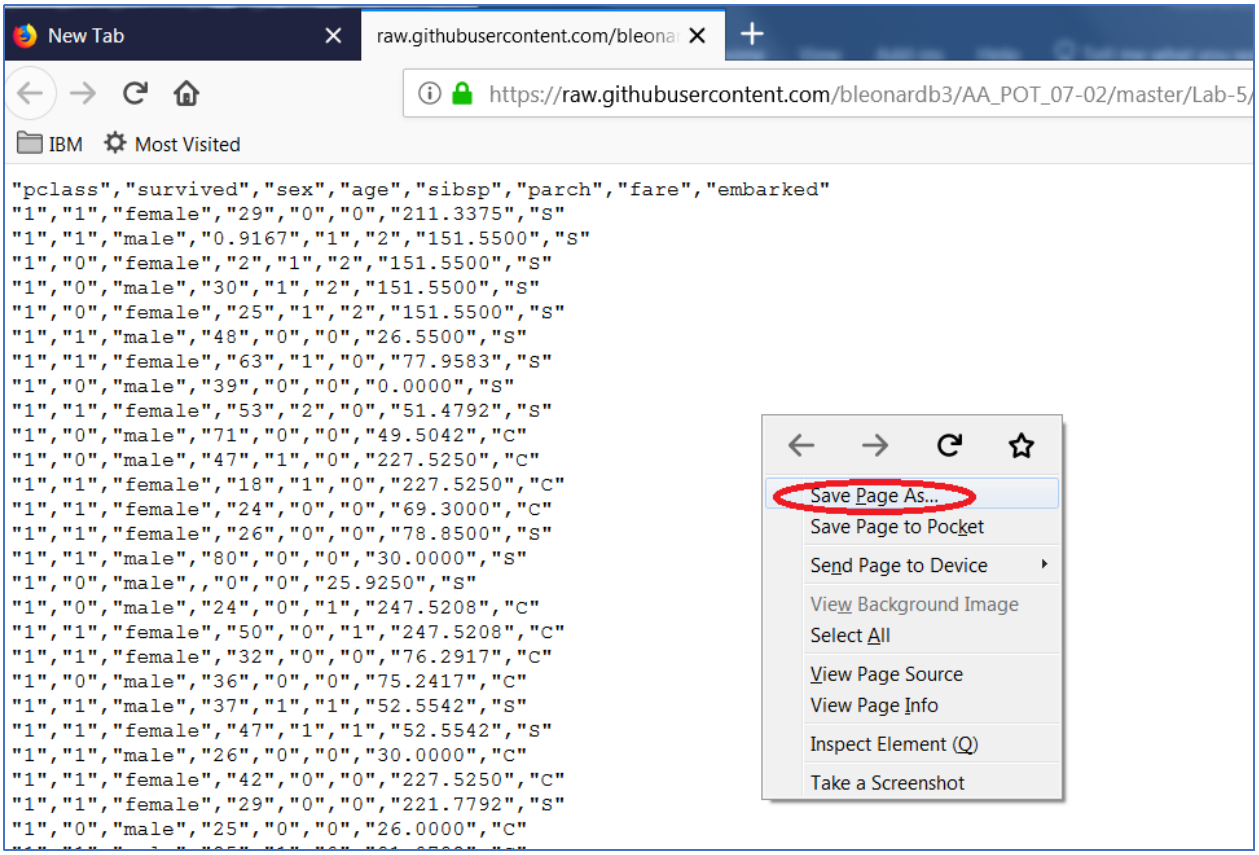
We will perform the following steps in this lab:

1. Download a Titanic cleansed data set
2. Add an Auto AI Experiment
3. Save and Deploy the selected model.

Step 1: Download the titanic_cleansed.csv data set

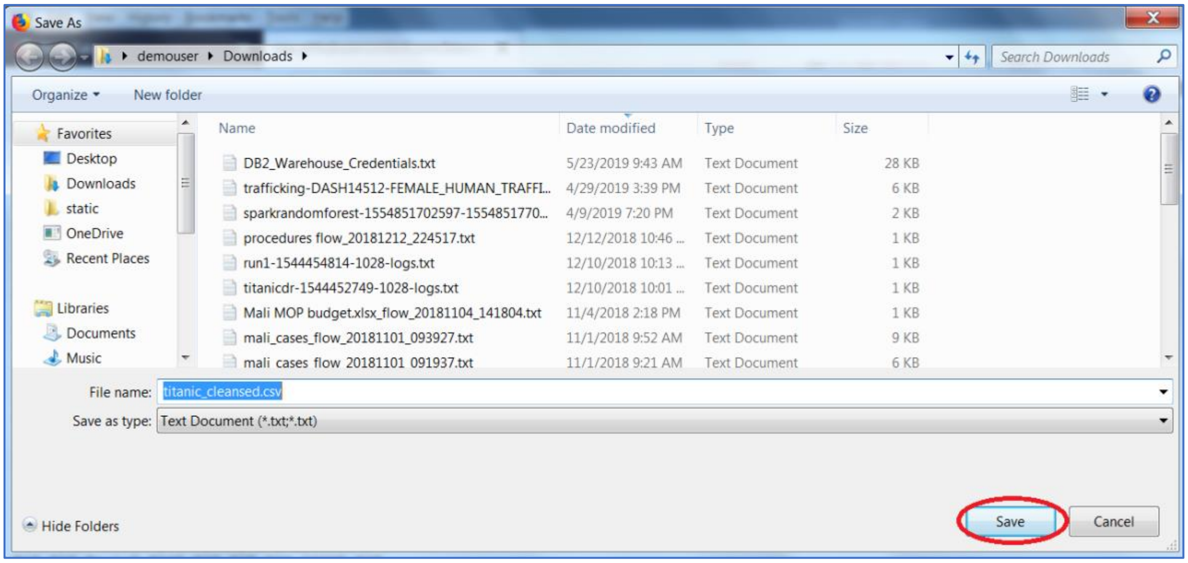
1. Download the **titanic_cleansed.csv** data file from the following location by clicking on the link here. Note this is a different file than used in the previous labs.

2. Right-click on the window, and click **Save Page As…**



```
"pclass","survived","sex","age","sibsp","parch","fare","embarked"
"1","1","female","29","0","0","211.3375","S"
"1","1","male","0.9167","1","2","151.5500","S"
"1","0","female","2","1","2","151.5500","S"
"1","0","male","30","1","2","151.5500","S"
"1","0","female","25","1","2","151.5500","S"
"1","1","male","48","0","0","26.5500","S"
"1","1","female","63","1","0","77.9583","S"
"1","0","male","39","0","0","0.0000","S"
"1","1","female","53","2","0","51.4792","S"
"1","0","male","71","0","0","49.5042","C"
"1","0","male","47","1","0","227.5250","C"
"1","1","female","18","1","0","227.5250","C"
"1","1","female","24","0","0","69.3000","C"
"1","1","female","26","0","0","78.8500","S"
"1","1","male","80","0","0","30.0000","S"
"1","0","male",,"0","0","25.9250","S"
"1","0","male","24","0","1","247.5208","C"
"1","1","female","50","0","1","247.5208","C"
"1","1","female","32","0","0","76.2917","C"
"1","0","male","36","0","0","75.2417","C"
"1","1","male","37","1","1","52.5542","S"
"1","1","female","47","1","1","52.5542","S"
"1","1","male","26","0","0","30.0000","C"
"1","1","female","42","0","0","227.5250","C"
"1","1","female","29","0","0","221.7792","S"
"1","0","male","25","0","0","26.0000","C"
```

3. Click on **Save**. Note, if the file is named titanic_cleansed.csv.txt, change it to be titanic_cleansed.csv.



Save As

demouser ▸ Downloads ▸

Search Downloads

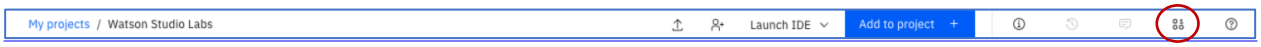
Organize ▾ New folder

Name	Date modified	Type	Size
DB2_Warehouse_Credentials.txt	5/23/2019 9:43 AM	Text Document	28 KB
trafficking-DASH14512-FEMALE_HUMAN_TRAFFI...	4/29/2019 3:39 PM	Text Document	6 KB
sparkrandomforest-1554851702597-1554851770...	4/9/2019 7:20 PM	Text Document	2 KB
procedures flow_20181212_224517.txt	12/12/2018 10:46 ...	Text Document	1 KB
run1-1544454814-1028-logs.txt	12/10/2018 10:13 ...	Text Document	1 KB
titanicdr-1544452749-1028-logs.txt	12/10/2018 10:01 ...	Text Document	1 KB
Mali MOP budget.xlsx_flow_20181104_141804.txt	11/4/2018 2:18 PM	Text Document	1 KB
mali_cases_flow_20181101_093927.txt	11/1/2018 9:52 AM	Text Document	9 KB
mali_cases_flow_20181101_091937.txt	11/1/2018 9:21 AM	Text Document	6 KB

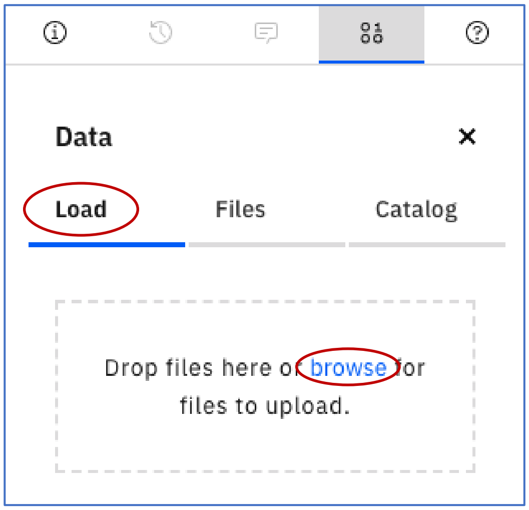
File name: titanic_cleansed.csv

Save as type: Text Document (*.txt;*.txt)

4. Go back to your Watson Studio Labs project. Click on the ⊞ icon.

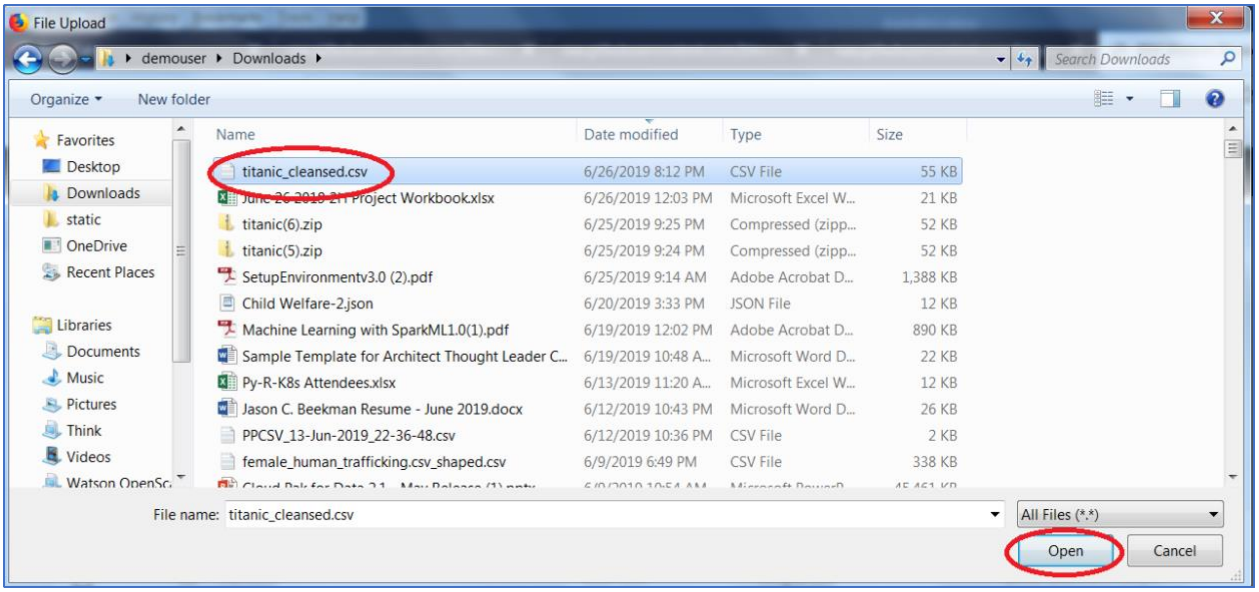


5. Click on the **Load** tab and then click on **browse**. If you don't see the **Load** tab, click on the ⊞ icon again.



Drop files here or browse for files to upload.

6. Go to the folder where the titanic_cleansed.csv file is stored. Select the titanic_cleansed.csv file and then click **Open**.

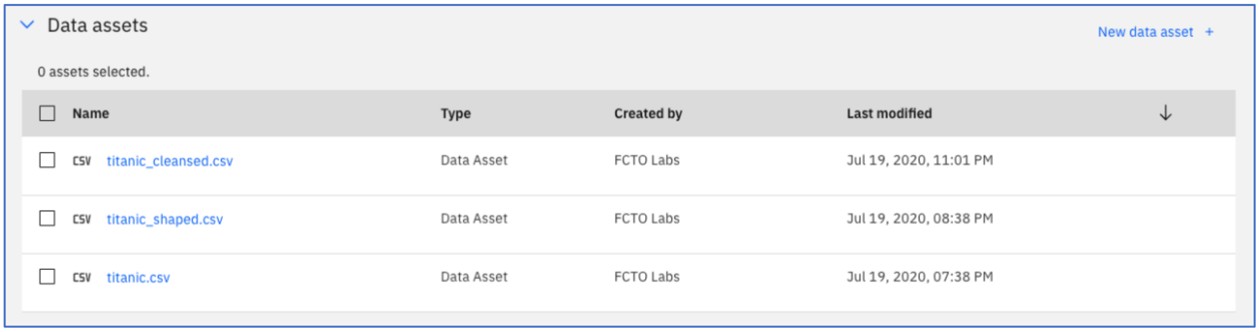


Name	Date modified	Type	Size
titanic_cleansed.csv	6/26/2019 8:12 PM	CSV File	55 KB
June 26 2019 2T Project Workbook.xlsx	6/26/2019 12:03 PM	Microsoft Excel W...	21 KB
titanic(6).zip	6/25/2019 9:25 PM	Compressed (zipp...	52 KB
titanic(5).zip	6/25/2019 9:24 PM	Compressed (zipp...	52 KB
SetupEnvironmentv3.0 (2).pdf	6/25/2019 9:14 AM	Adobe Acrobat D...	1,388 KB
Child Welfare-2.json	6/20/2019 3:33 PM	JSON File	12 KB
Machine Learning with SparkML1.0(1).pdf	6/19/2019 12:02 PM	Adobe Acrobat D...	890 KB
Sample Template for Architect Thought Leader C...	6/19/2019 10:48 A...	Microsoft Word D...	22 KB
Py-R-K8s Attendees.xlsx	6/13/2019 11:20 A...	Microsoft Excel W...	12 KB
Jason C. Beekman Resume - June 2019.docx	6/12/2019 10:43 PM	Microsoft Word D...	26 KB
PPCSV_13-Jun-2019_22-36-48.csv	6/12/2019 10:36 PM	CSV File	2 KB
female_human_trafficking.csv_shaped.csv	6/9/2019 6:49 PM	CSV File	338 KB
Cloud Pak for Data 2.1 - May Release (1).pptx	6/9/2019 10:54 AM	Microsoft PowerP...	45,461 KB

File name: titanic_cleansed.csv

All Files (*.*)

7. The titanic_cleansed.csv file is now added as a Data Asset.

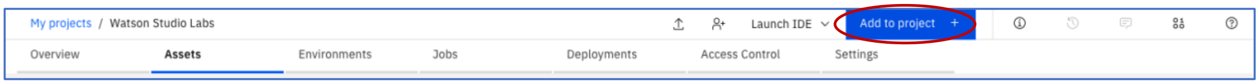


0 assets selected.

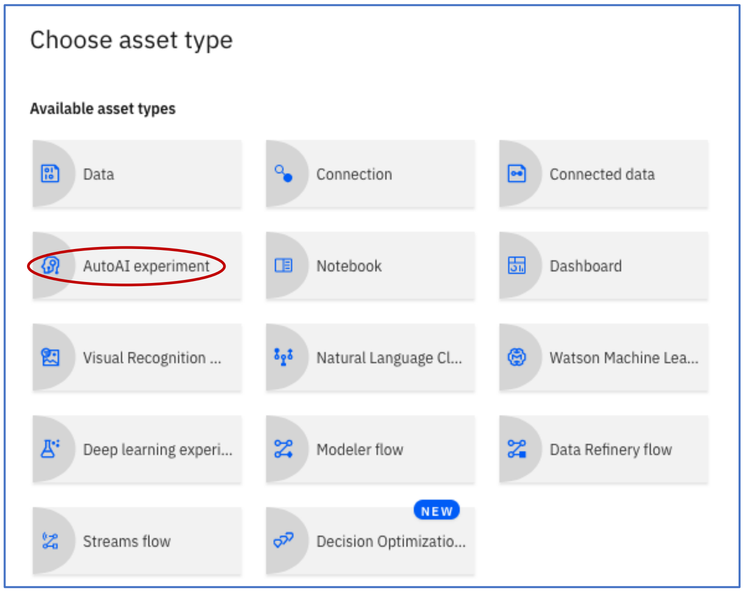
□ Name	Type	Created by	Last modified ↓
□ CSV titanic_cleansed.csv	Data Asset	FCTO Labs	Jul 19, 2020, 11:01 PM
□ CSV titanic_shaped.csv	Data Asset	FCTO Labs	Jul 19, 2020, 08:38 PM
□ CSV titanic.csv	Data Asset	FCTO Labs	Jul 19, 2020, 07:38 PM

Step 2: Add an AutoAI Experiment

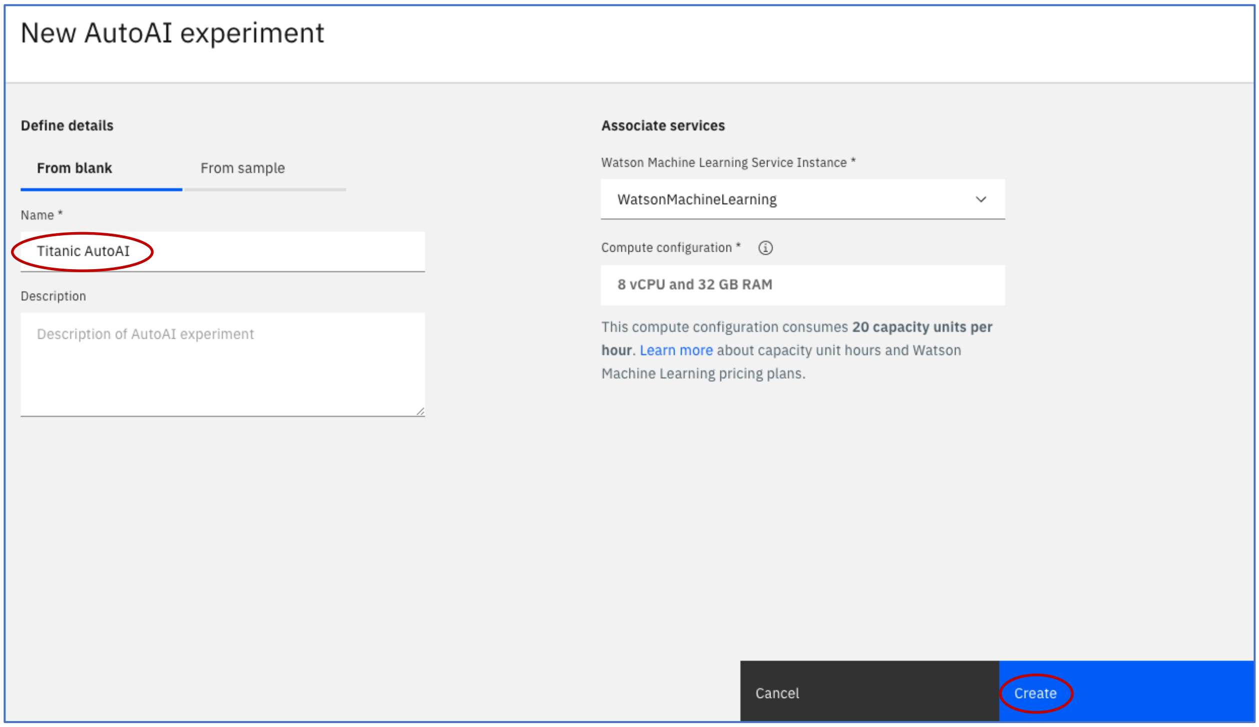
1. Click on **Add to project**.



2. Click on **AutoAI experiment**



3. Enter an **Asset name**, leave the defaults for the Watson Machine Learning and Compute configuration and click on **Create**.



New AutoAI experiment

Define details

From blank | From sample

Name *

Titanic AutoAI

Description

Description of AutoAI experiment

Associate services

Watson Machine Learning Service Instance *

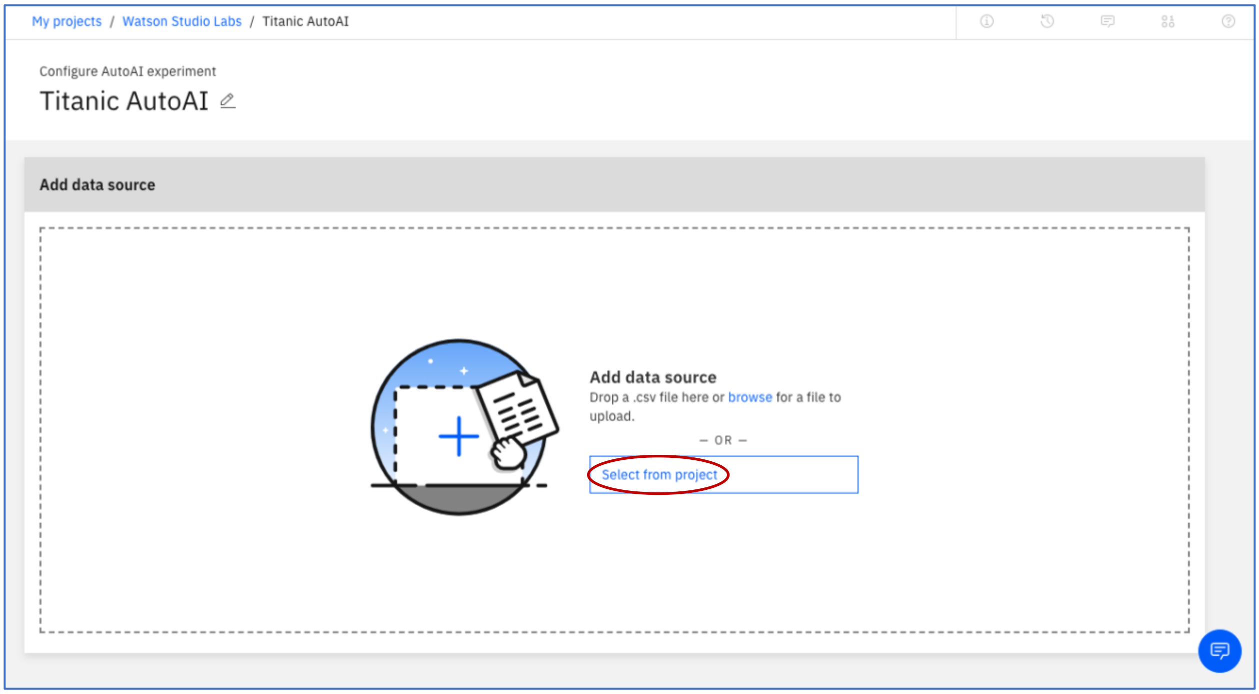
WatsonMachineLearning

Compute configuration *

8 vCPU and 32 GB RAM

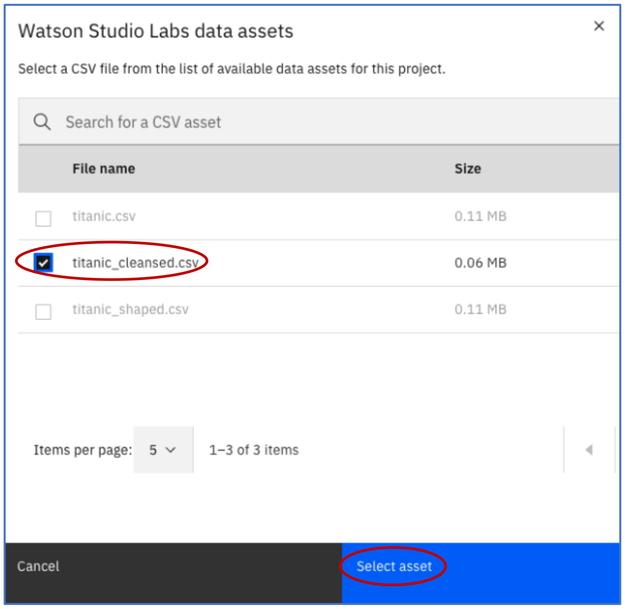
This compute configuration consumes **20 capacity units per hour**. Learn more about capacity unit hours and Watson Machine Learning pricing plans.

4. Click on **Select from project**.


My projects / Watson Studio Labs / Titanic AutoAI
Configure AutoAI experiment
Titanic AutoAI
Add data source
Add data source
— OR —
Select from project

Drop a .csv file here or browse for a file to upload.

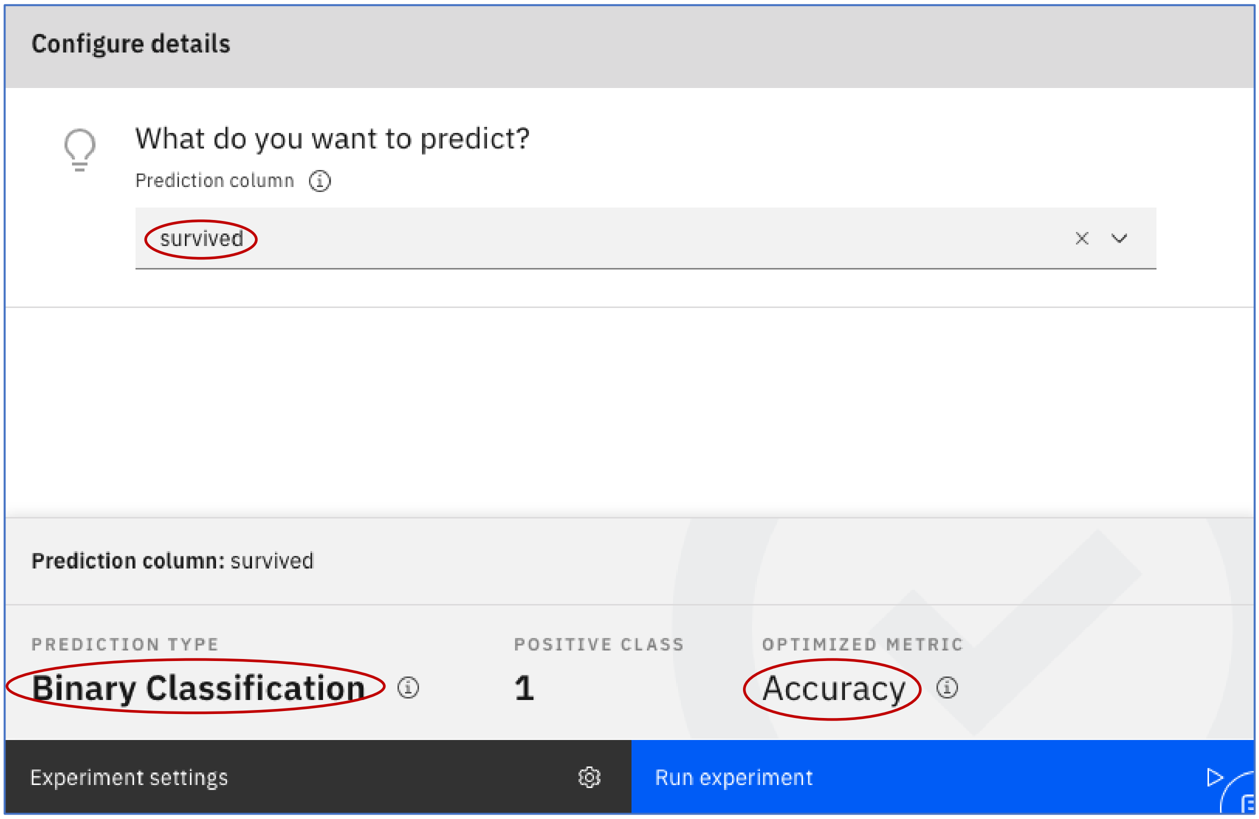
5. Click on **titanic_cleansed.csv** and then click on **Select asset**.



Select a CSV file from the list of available data assets for this project.

	File name	Size
□	titanic.csv	0.11 MB
☑	titanic_cleansed.csv	0.06 MB
□	titanic_shaped.csv	0.11 MB

6. Click on **survived** as **the column to predict**. Note, based on this selection, the **Prediction Type** is **Binary Classification**, and the **Optimized Metric** is **Accuracy**. Further note, the **Positive Class** is correctly defaulted as “1” – survived.



Configure details

What do you want to predict?

Prediction column: survived

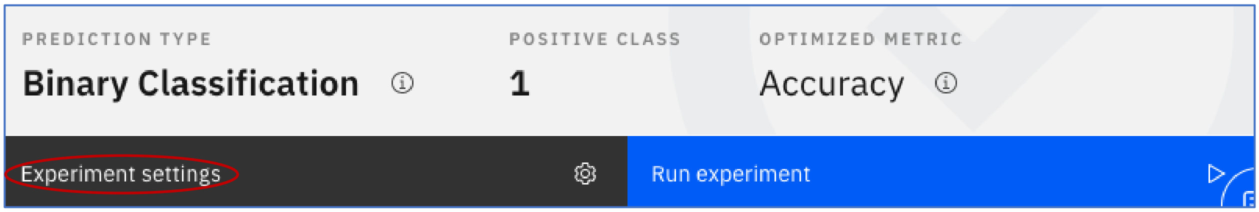
Prediction column: survived

PREDICTION TYPE: Binary Classification

POSITIVE CLASS: 1

OPTIMIZED METRIC: Accuracy

7. Click on **Experiment settings** to change the default optimized metric.

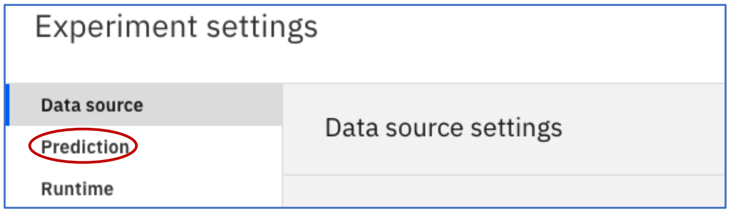


PREDICTION TYPE: Binary Classification

POSITIVE CLASS: 1

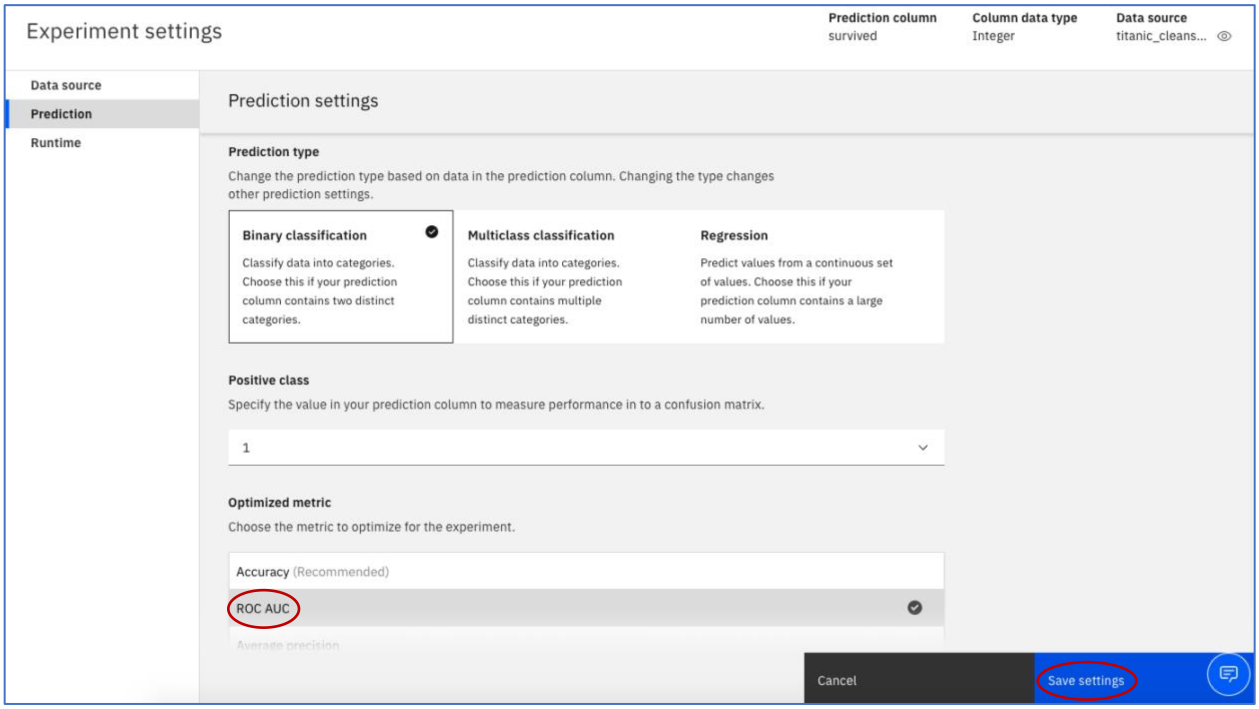
OPTIMIZED METRIC: Accuracy

8. Click on **Prediction**.



Experiment settings

9. Click on **ROC AUC** (Receiver Operating Characteristic Area Under the Curve) and then click on **Save Settings**



Experiment settings

Prediction column: survived

Column data type: Integer

Data source: titanic_cleans...

Prediction settings

Prediction type

Change the prediction type based on data in the prediction column. Changing the type changes other prediction settings.

● **Binary classification**
Classify data into categories. Choose this if your prediction column contains two distinct categories.

Multiclass classification
Classify data into categories. Choose this if your prediction column contains multiple distinct categories.

Regression
Predict values from a continuous set of values. Choose this if your prediction column contains a large number of values.

Positive class

Specify the value in your prediction column to measure performance in to a confusion matrix.

1

Optimized metric

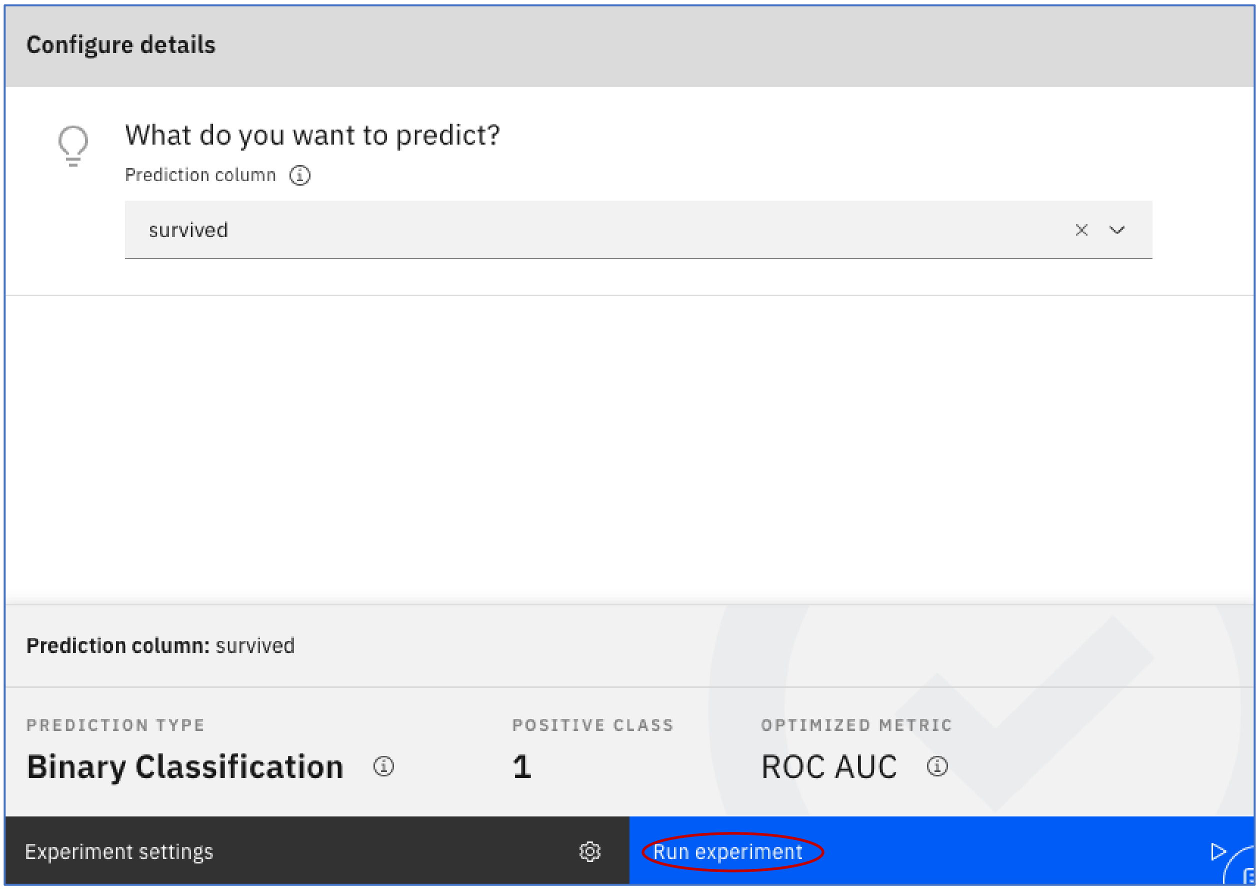
Choose the metric to optimize for the experiment.

Accuracy (Recommended)

ROC AUC ●

Average precision

10. Click on **Run experiment**.



Configure details

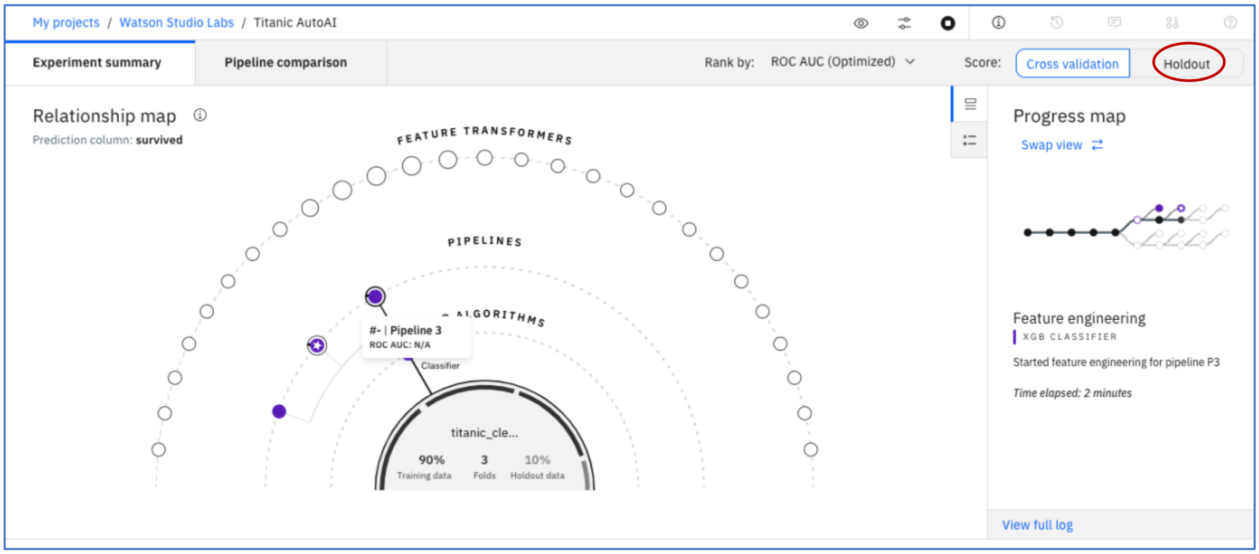
What do you want to predict?

Prediction column: survived

Prediction column: survived

PREDICTION TYPE	POSITIVE CLASS	OPTIMIZED METRIC
Binary Classification	1	ROC AUC

11. It will take several minutes for the eight alternative pipelines to be analyzed. The first pipeline picks the best algorithm. The second pipeline optimizes the hyper-parameters for the selected algorithm. The third pipeline does a feature transformation to try to improve the performance of the algorithm. The fourth pipeline repeats the hyper-parameter tuning with the new set of features. The next 4 pipelines do the same thing for the second best algorithm. Note, you can move ahead after 3 of the 8 pipelines have been completed. We are going to proceed as if Pipeline 2 is the best ranked pipeline. Yours could be different. Click on **Holdout**.

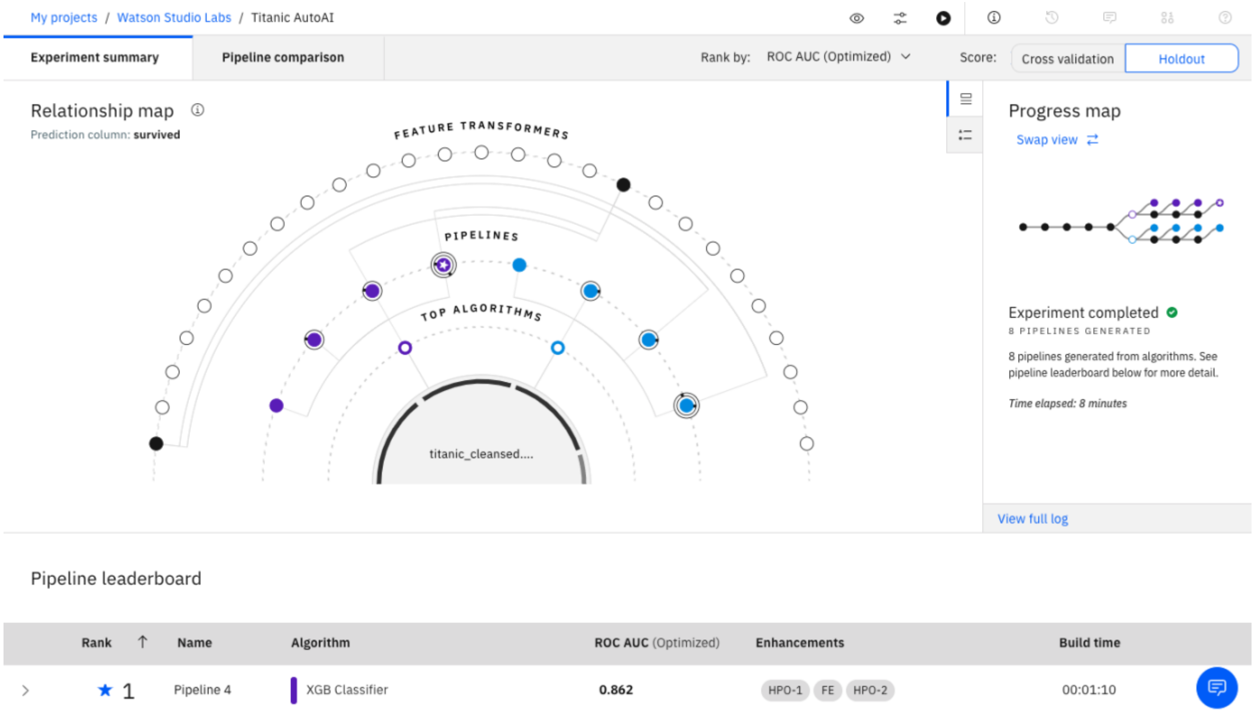


Feature engineering
XGB CLASSIFIER

Started feature engineering for pipeline P3

Time elapsed: 2 minutes

12. The best pipeline based on the performance metric (ROC AUC) is pipeline 4 for this run.



Experiment completed ✓
8 PIPELINES GENERATED

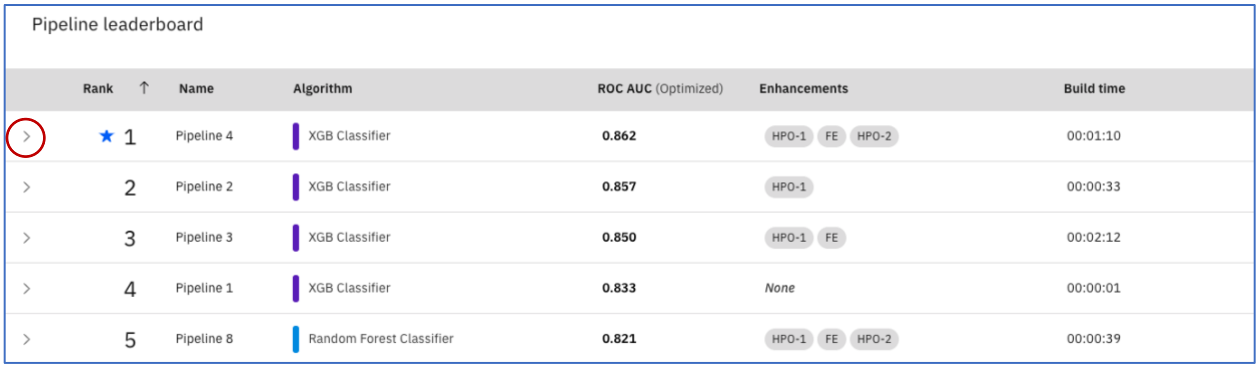
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 8 minutes

Pipeline leaderboard

Rank ↑	Name	Algorithm	ROC AUC (Optimized)	Enhancements	Build time
★ 1	Pipeline 4	XGB Classifier	0.862	HPO-1 FE HPO-2	00:01:10

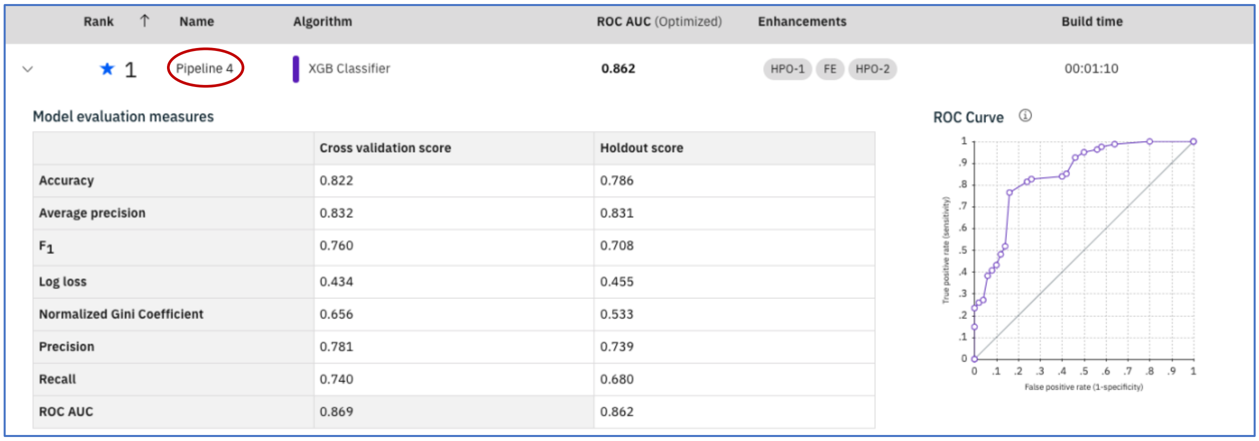
13. Scroll down to view the **Pipeline leaderboard**. The pipeline summary is then displayed. Click on the right arrow › next to the top ranked pipeline.



Pipeline leaderboard

	Rank ↑	Name	Algorithm	ROC AUC (Optimized)	Enhancements	Build time
›	★ 1	Pipeline 4	XGB Classifier	0.862	HPO-1 FE HPO-2	00:01:10
›	2	Pipeline 2	XGB Classifier	0.857	HPO-1	00:00:33
›	3	Pipeline 3	XGB Classifier	0.850	HPO-1 FE	00:02:12
›	4	Pipeline 1	XGB Classifier	0.833	*None*	00:00:01
›	5	Pipeline 8	Random Forest Classifier	0.821	HPO-1 FE HPO-2	00:00:39

14. Scores are displayed for different metrics for both the training sample and the holdout sample. Click on the **Pipeline** link for the top ranked pipeline.



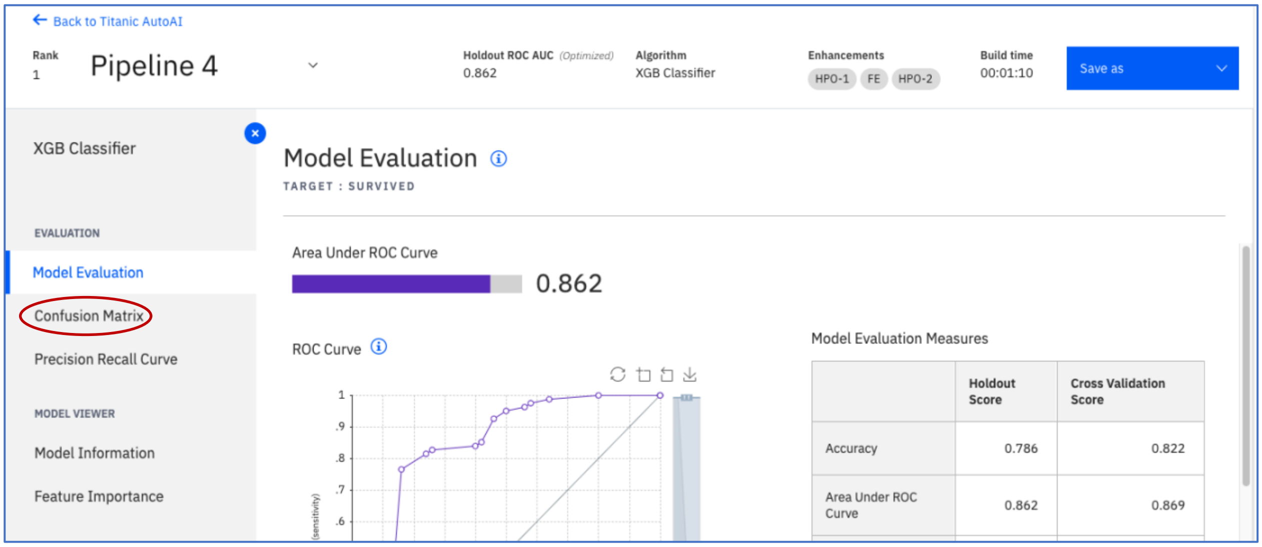
	Rank ↑	Name	Algorithm	ROC AUC (Optimized)	Enhancements	Build time
⌄	★ 1	Pipeline 4	XGB Classifier	0.862	HPO-1 FE HPO-2	00:01:10

Model evaluation measures

	Cross validation score	Holdout score
Accuracy	0.822	0.786
Average precision	0.832	0.831
F_1	0.760	0.708
Log loss	0.434	0.455
Normalized Gini Coefficient	0.656	0.533
Precision	0.781	0.739
Recall	0.740	0.680
ROC AUC	0.869	0.862

ROC Curve ⓘ
True positive rate (sensitivity)
False positive rate (1-specificity)
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1

15. The model evaluation metrics for the training and holdout sample are repeated. On the left are options for additional information. Click on the **Confusion Matrix**.



← Back to Titanic AutoAI

Rank		Holdout ROC AUC (Optimized)	Algorithm	Enhancements	Build time
1	Pipeline 4	0.862	XGB Classifier	HPO-1 FE HPO-2	00:01:10

Model Evaluation ⓘ

TARGET : SURVIVED

Area Under ROC Curve

0.862

Model Evaluation Measures

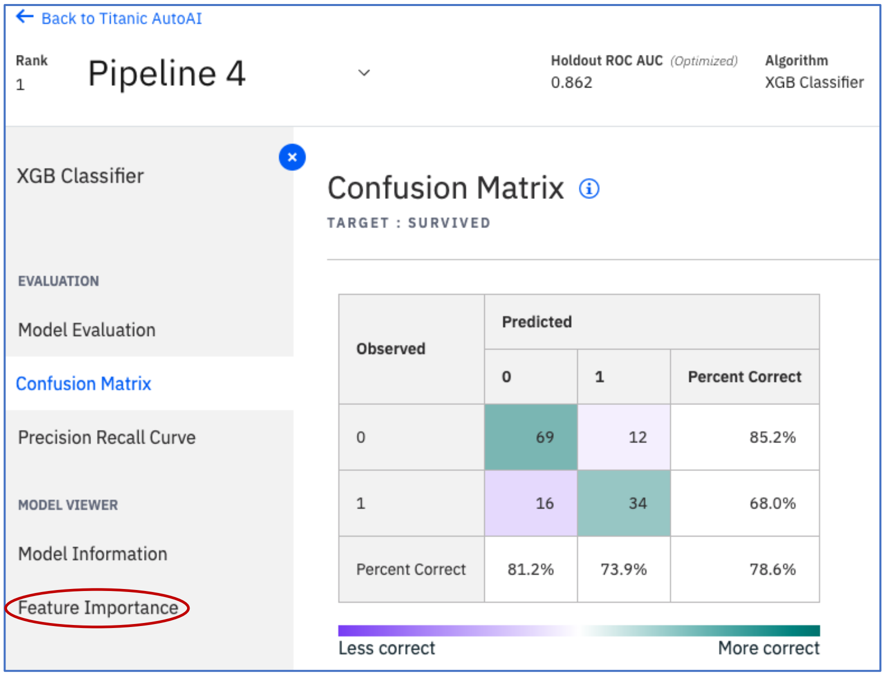
	Holdout Score	Cross Validation Score
Accuracy	0.786	0.822
Area Under ROC Curve	0.862	0.869

16. The Confusion Matrix is displayed for the holdout sample. The different metrics are computed based on the numbers in the Confusion Matrix. For example, Precision is defined by the percentage of predicted positives that are actually positive (i.e. the percentage of predicted survivors that survived). Recall is defined as the percentage of observed positives that the model predicts are positive (i.e. the percentage of survivors that the model predicted would survive). Note the higher the Precision the lower the Recall.

    Precision = True Positive/ (True Positive + False Positive) – shown inside blue rectangle on diagram below.

    Recall = True Positive/(True Positive + False Negative) - shown inside green rectangle on diagram below

    After viewing the Confusion Matrix, click on the **Feature Importance** option.



Rank
1

Pipeline 4

Holdout ROC AUC (Optimized)
0.862

Algorithm
XGB Classifier

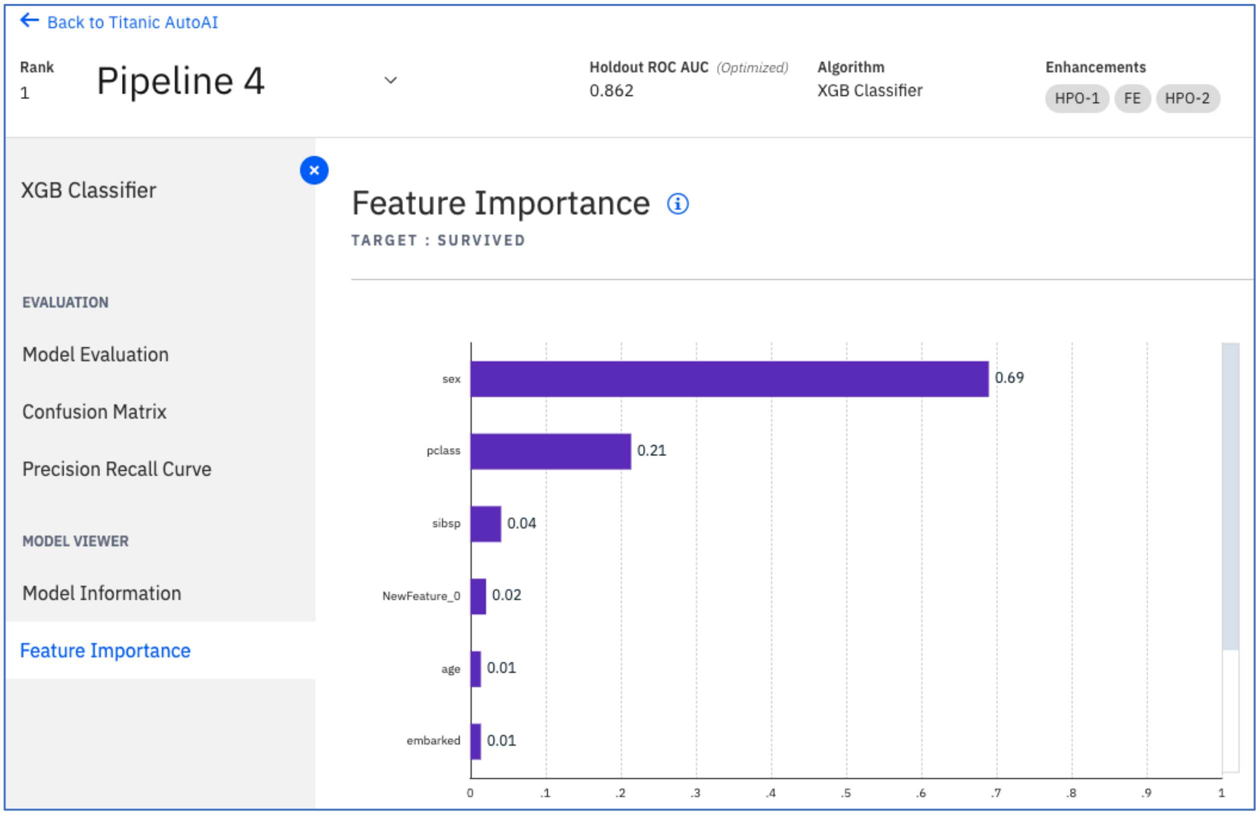
Confusion Matrix

TARGET : SURVIVED

Observed	Predicted		
	0	1	Percent Correct
0	69	12	85.2%
1	16	34	68.0%
Percent Correct	81.2%	73.9%	78.6%

Less correct More correct

17. According to the Feature Importance, the sex variable is considered the most important feature followed by the passenger class.



Rank 1 Pipeline 4

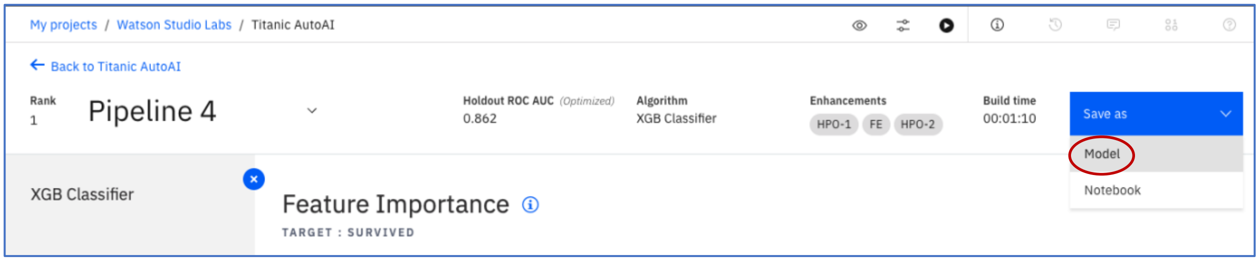
Holdout ROC AUC *(Optimized)* 0.862

Algorithm XGB Classifier

Enhancements HPO-1 FE HPO-2

Step 3 – Save and Deploy the Selected Model

1. Click on **Save as**. Click on **Model**. Note you also have the option to save a Notebook. The notebook contains the code used to generate the pipeline. In this way a data scientist could use this as a starting point to tune the model even further.



Rank 1 Pipeline 4

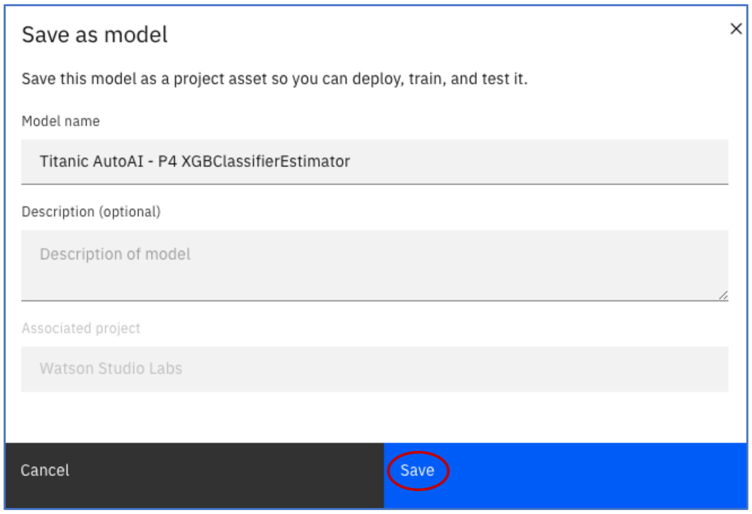
Holdout ROC AUC *(Optimized)* 0.862

Algorithm XGB Classifier

Enhancements HPO-1 FE HPO-2

Build time 00:01:10

2. Optionally change the default name and click on **Save**.



Save as model

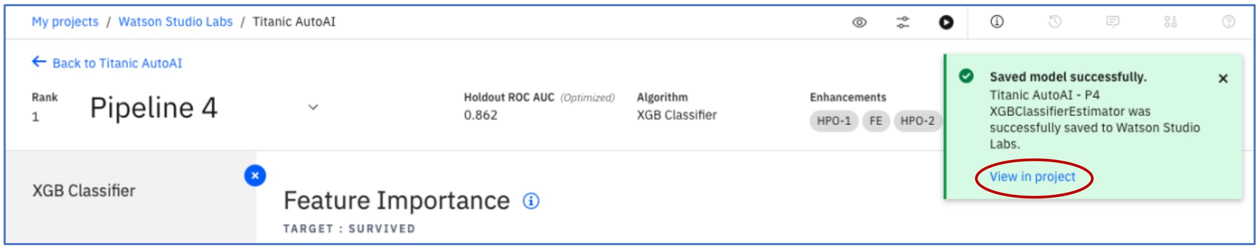
Save this model as a project asset so you can deploy, train, and test it.

Model name: Titanic AutoAI - P4 XGBClassifierEstimator

Description (optional): Description of model

Associated project: Watson Studio Labs

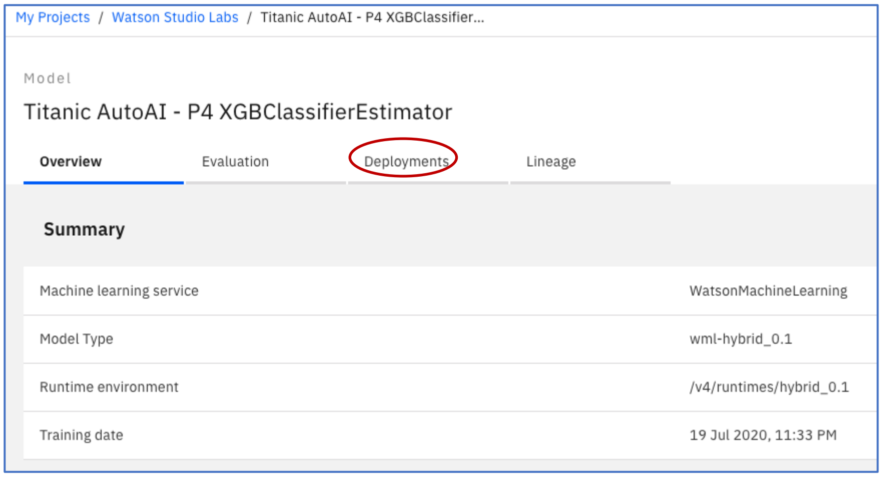
3. Click on **View in Project**.



Rank		Holdout ROC AUC (Optimized)	Algorithm	Enhancements
1	Pipeline 4	0.862	XGB Classifier	HPO-1 FE HPO-2

Saved model successfully. Titanic AutoAI - P4 XGBClassifierEstimator was successfully saved to Watson Studio Labs.

4. Click on **Deployments**



My Projects / Watson Studio Labs / Titanic AutoAI - P4 XGBClassifier...
Overview
Evaluation
Deployments
Lineage

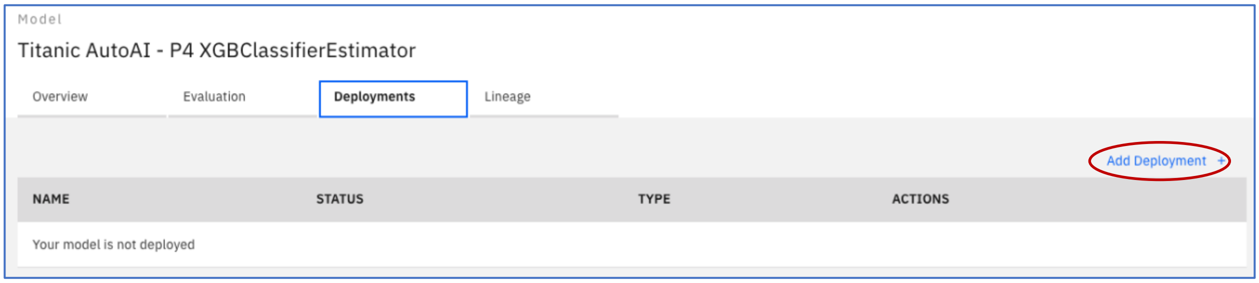
Model

Titanic AutoAI - P4 XGBClassifierEstimator

Summary

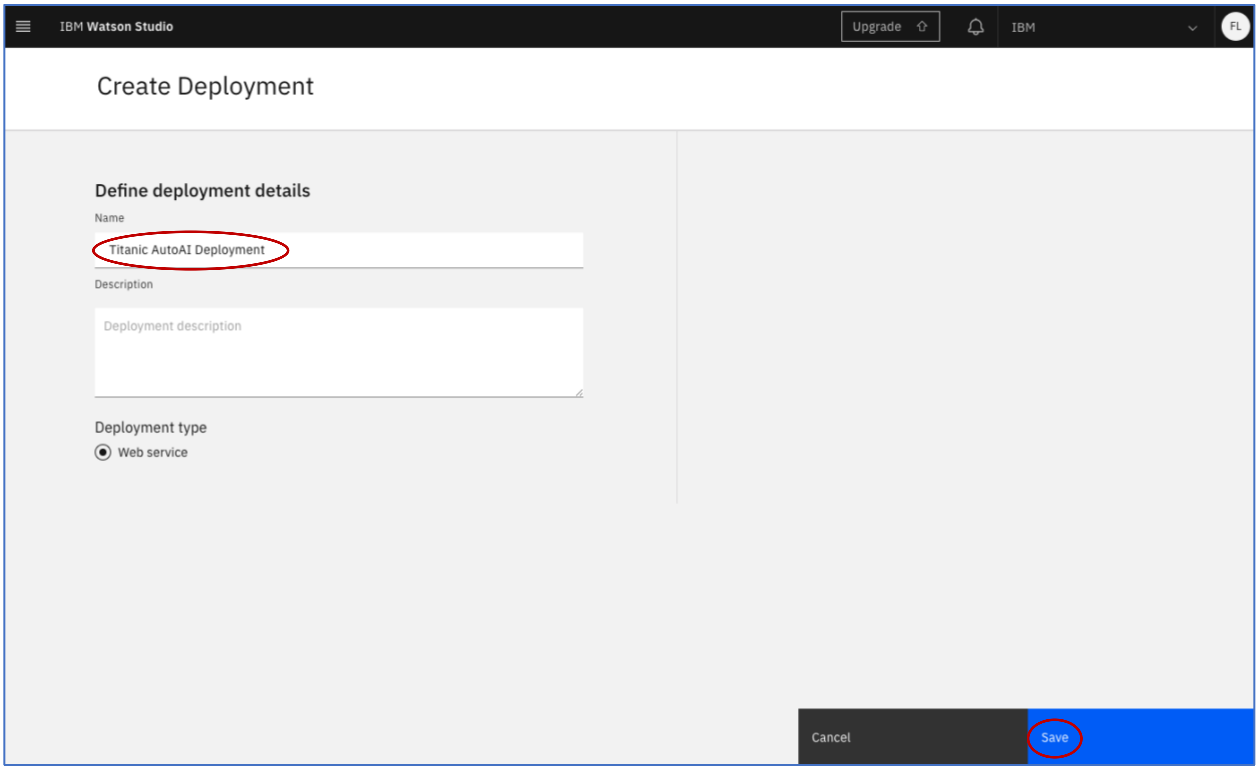
Machine learning service	WatsonMachineLearning
Model Type	wml-hybrid_0.1
Runtime environment	/v4/runtimes/hybrid_0.1
Training date	19 Jul 2020, 11:33 PM

5. Click on **Add Deployment**.



NAME	STATUS	TYPE	ACTIONS
Your model is not deployed			

6. Enter a **Name** and click **Save**. Note, if status stays on **initializing** for a while, click on the browser refresh.



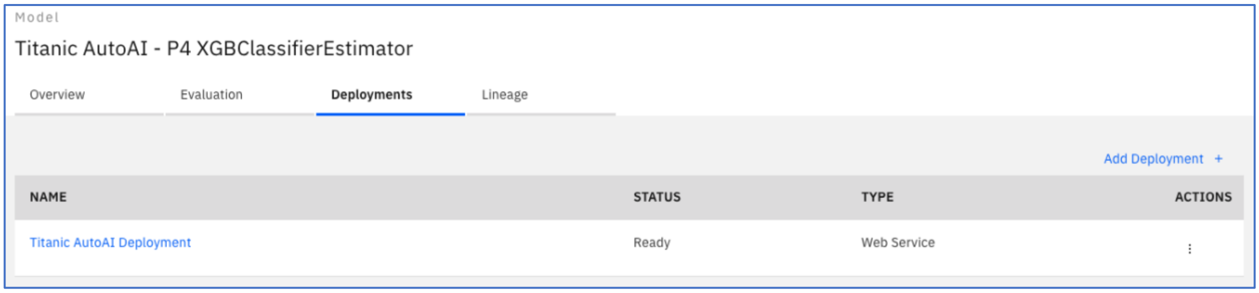
Name: Titanic AutoAI Deployment

Description: Deployment description

Deployment type

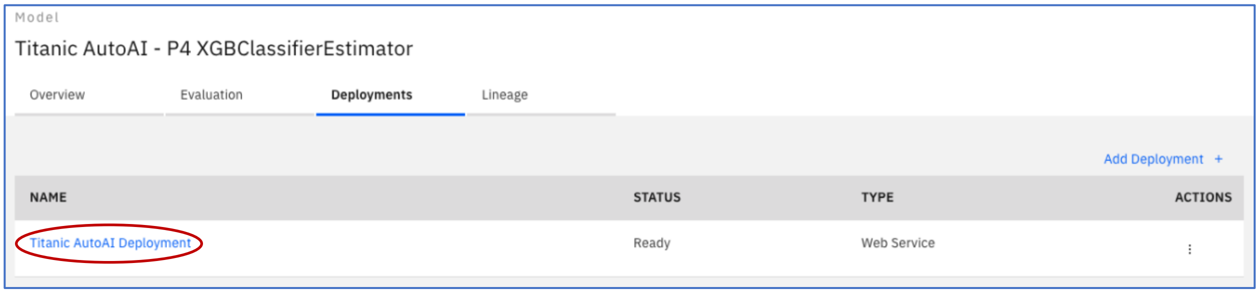
● Web service

7. The model is successfully deployed on the IBM Cloud.



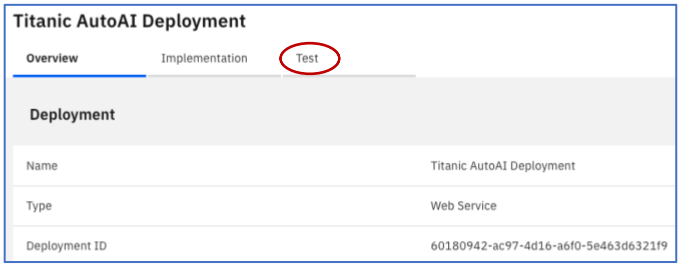
NAME	STATUS	TYPE	ACTIONS
Titanic AutoAI Deployment	Ready	Web Service	⋮

8. Click on **Titanic AutoAI Deployed**.



NAME	STATUS	TYPE	ACTIONS
Titanic AutoAI Deployment	Ready	Web Service	⋮

9. Click on **Test**.

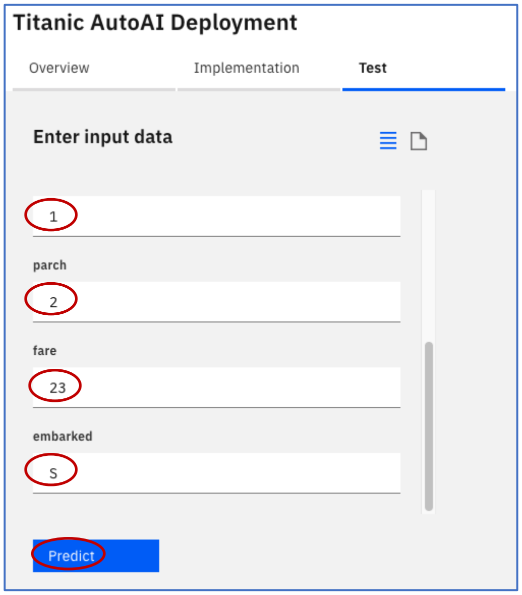


Deployment

Name	Titanic AutoAI Deployment
Type	Web Service
Deployment ID	60180942-ac97-4d16-a6f0-5e463d6321f9

10. Enter values for a passenger. For example,
    pclass – 2
    sex – male
    age – 5
    sibsp – 1
    parch – 2
    fare – 23
    embarked – S

    and click **Predict.**



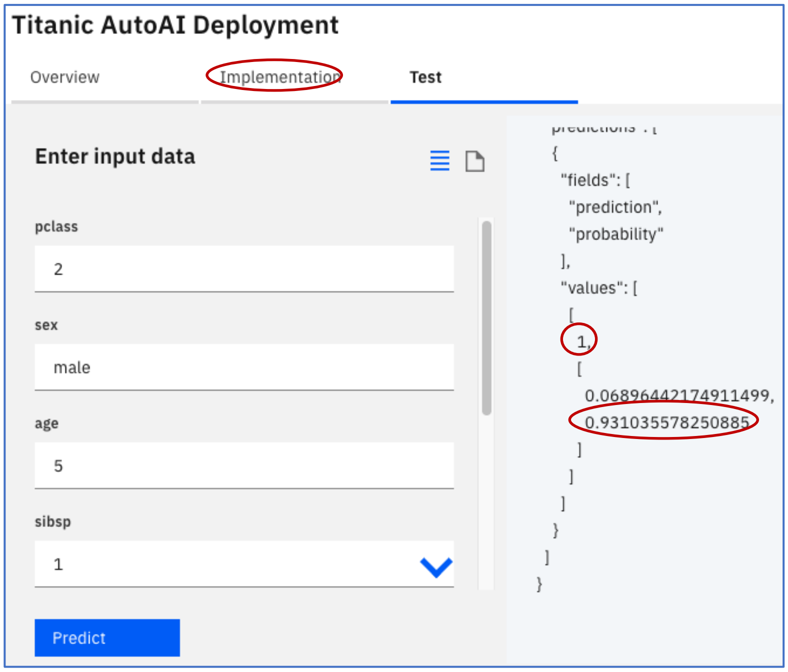
1

parch
2

fare
23

embarked
S

11. The model predicts this passenger would survive, with a very high confidence. Click on **Implementation**.



Titanic AutoAI Deployment

Enter input data

pclass: 2

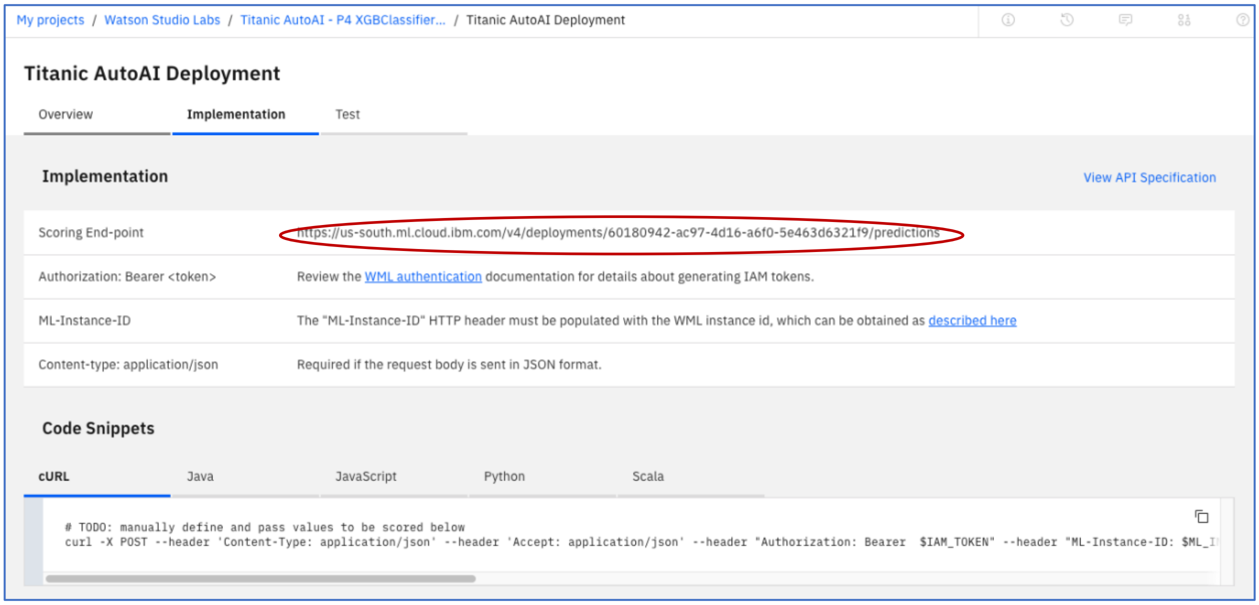
sex: male

age: 5

sibsp: 1

```
"predictions": [
{
  "fields": [
    "prediction",
    "probability"
  ],
  "values": [
    [
      1,
      [
        0.06896442174911499,
        0.931035578250885
      ]
    ]
  ]
}
]
}
```

12. The Implementation panel provides information for the application developers to invoke the deployed model. It includes sample code in various programming languages and the scoring endpoint to be used when invoking the web service.



My projects / Watson Studio Labs / Titanic AutoAI - P4 XGBClassifier... / Titanic AutoAI Deployment

Titanic AutoAI Deployment

Overview
Implementation
Test

Implementation

Scoring End-point	https://us-south.ml.cloud.ibm.com/v4/deployments/60180942-ac97-4d16-a6f0-5e463d6321f9/predictions
Authorization: Bearer <token>	Review the WML authentication documentation for details about generating IAM tokens.
ML-Instance-ID	The "ML-Instance-ID" HTTP header must be populated with the WML instance id, which can be obtained as described here
Content-type: application/json	Required if the request body is sent in JSON format.

Code Snippets

```
# TODO: manually define and pass values to be scored below
curl -X POST --header 'Content-Type: application/json' --header 'Accept: application/json' --header "Authorization: Bearer  $IAM_TOKEN" --header "ML-Instance-ID: $ML_I
```