

# AutoAI walkthrough using Historical Wildfires combined with Historical Weather data

Useful Links :

The Contest landing page <http://ibm.biz/cfcsc-wildfires>

The Contest GitHub <https://github.com/Call-for-Code/Spot-Challenge-Wildfires>

NOTE: Only the NSW – New South Wales Region from the Historical Wildfires data and the Temperature from the Historical Weather data have been combined for this walkthrough

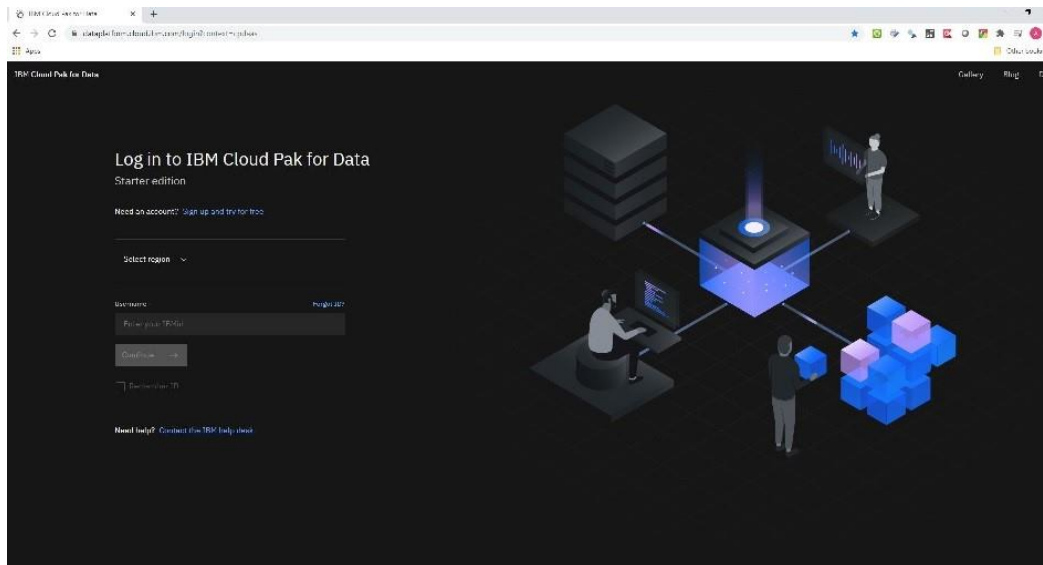
Please see the following [datasets](#), [accompanying readme](#) and [slides](#), which are available via GitHub <https://github.com/Call-for-Code/Spot-Challenge-Wildfires> together with a [starter notebook](#).

The short dataset was created by reading the Historical Wildfires data into a Pandas dataframe, only the NSW region was selected from the dataframe, the Date column was converted to a datetime format. Next, the Historical Weather data was read into another Pandas dataframe, the NSW region and the Temperature parameter were selected from the dataframe, the Date column in this dataframe was also converted to datetime format.

These two selected dataframes were then merged using the NSW Historical Wildfires dataframe as the anchor, since the wildfires measurements were not available for all corresponding days in the weather data.

NOTE: This raw merged data was used in this AutoAI experiment walkthrough without any further processing or wrangling.

Log in to: <https://dataplatfom.cloud.ibm.com/login?context=cpdaas>



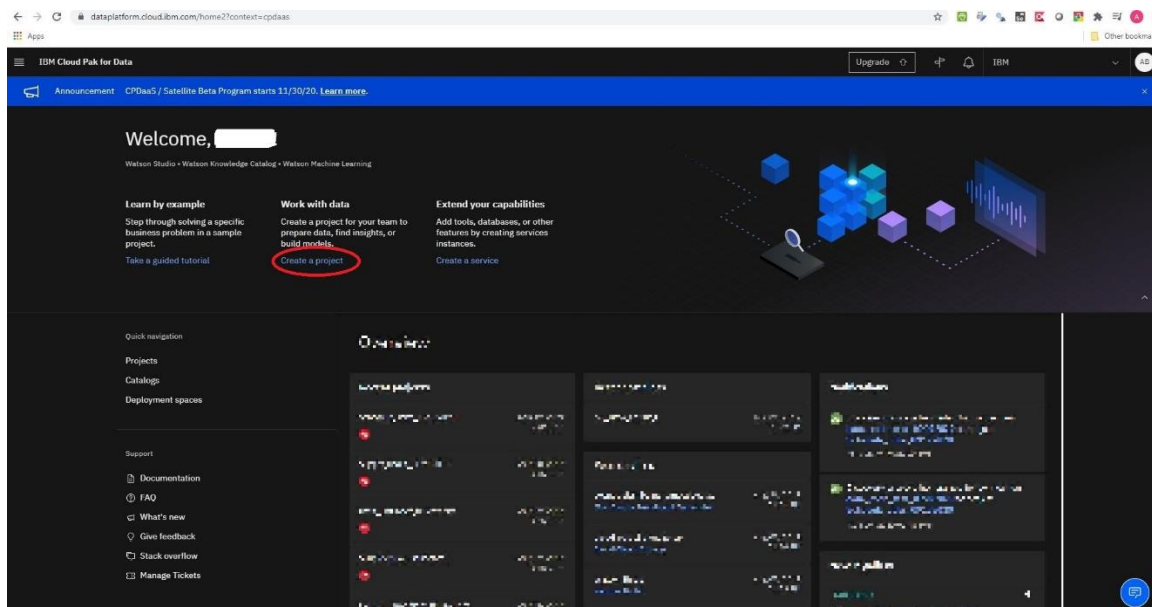
Select Region, closest to your location (Options: Dallas, London, Frankfurt, Tokyo)

If previously registered, provide username, then hit “Continue” and follow the instructions.

Else please use the link [Sign up and try for free](#) as shown in the screen above and follow the steps to register.

If you have used the IBM Cloud Pack for Data previously, you will be presented with a welcome screen as shown below, listing past projects, Notifications etc..

Click the “Create a project” link, as shown below, to create a new project.



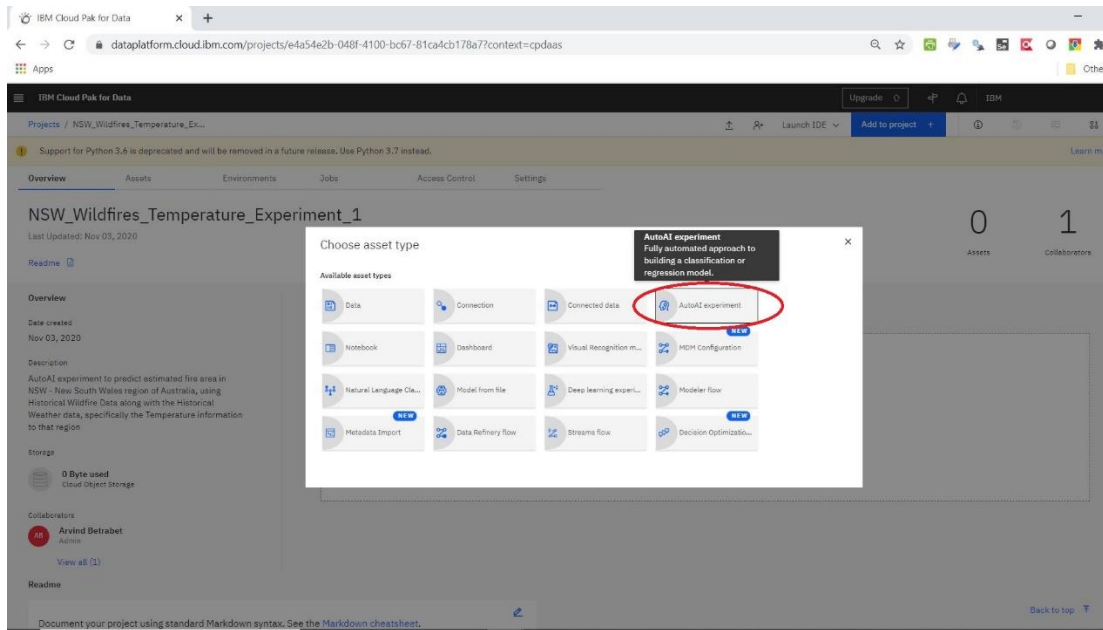
Fill in the appropriate information for the project (Name and Description), as shown below and click “Create”

The screenshot shows the 'New project' page in the IBM Cloud Pak for Data interface. The page is divided into two main sections: 'Define project details' and 'Storage'. In the 'Define project details' section, the 'Name' field contains 'NSW\_Wildfires\_Temperature\_Experiment\_1' and the 'Description' field contains 'AutoAI experiment to predict estimated fire area in NSW - New South Wales region of Australia, using Historical Wildfire Data along with the Historical Weather data, specifically the Temperature information to that region'. Below these fields, the 'Choose project options' section has a checked box for 'Restrict who can be a collaborator'. At the bottom right, there are 'Cancel' and 'Create' buttons, with the 'Create' button highlighted by a red circle.

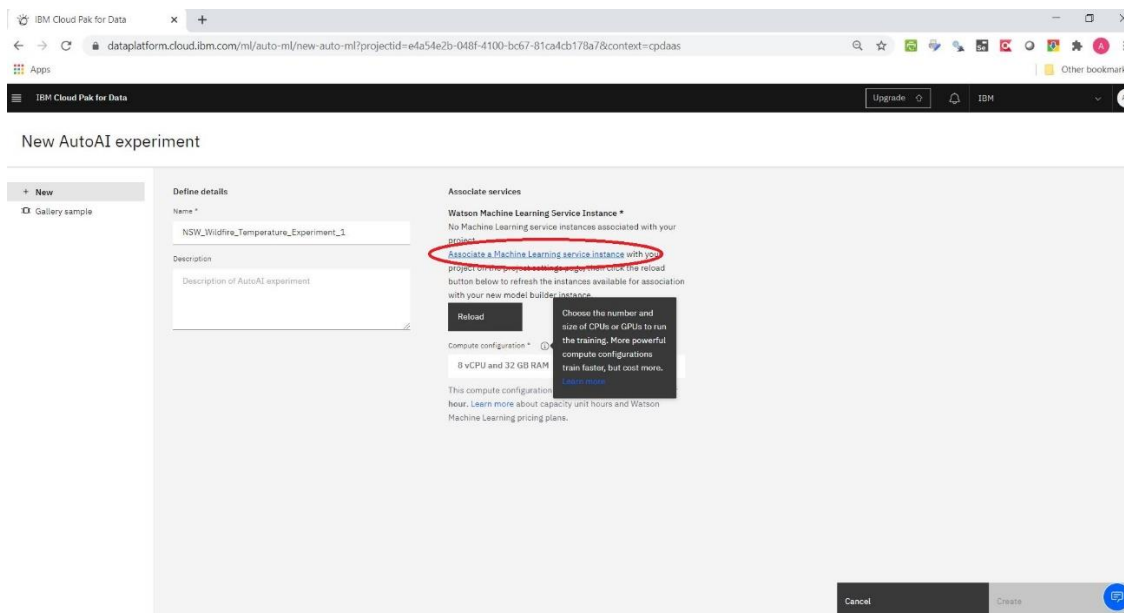
Takes a few seconds to create an environment.

The screenshot shows the project overview page for 'NSW\_Wildfires\_Temperature\_Experiment\_1'. The page has a top navigation bar with 'Add to project' and 'Launch IDE' buttons, both highlighted by red circles. Below the navigation bar, there is a warning message about Python 3.6 deprecation. The main content area is divided into two sections: 'Overview' and 'Recent activity'. The 'Overview' section shows the project's details, including the name, description, storage (0 Byte used), and collaborators (Arvind Betrabet). The 'Recent activity' section shows a message: 'Alerts related to this project appear here when the project is active.' At the bottom right, there is a 'Back to top' button.

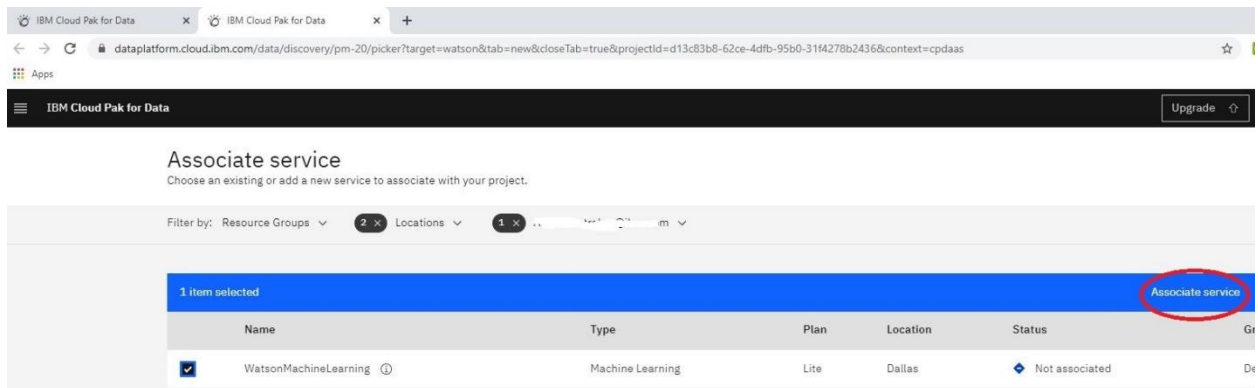
Click on Add to project on the top right, then select “AutoAI experiment” as shown:



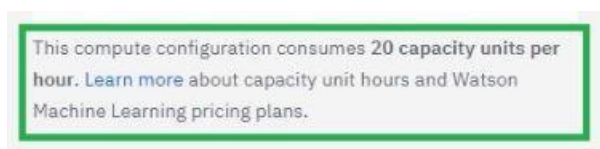
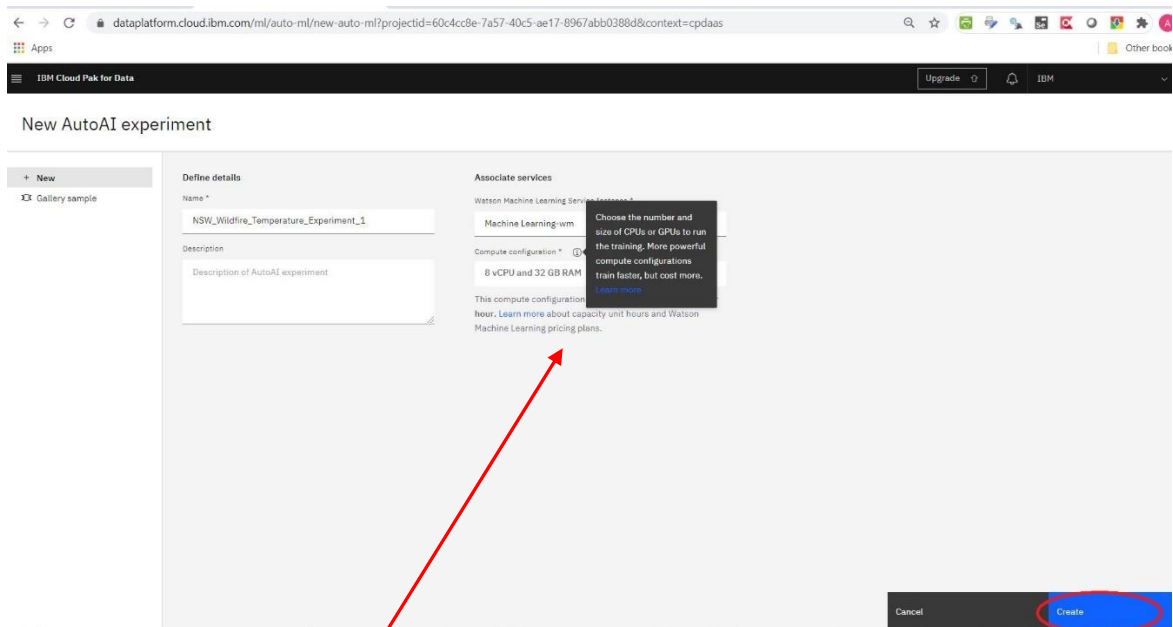
In the next screen, provide a name for the experiment, as well as read the information icon on the screen, as shown



Click on “Associate a Machine Learning service instance” link to select the resource required for your experiment (a new tab is opened). Select the “WatsonMachineLearning” to associate a service, click on the “Associate Service” Button.

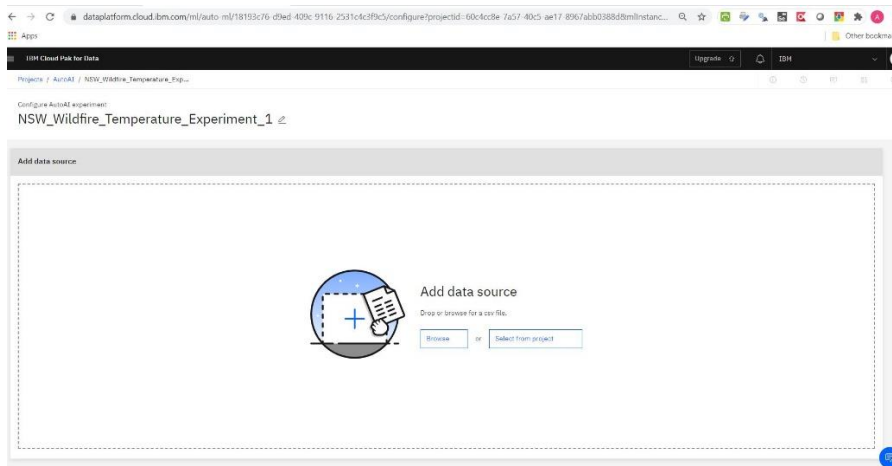


That window will close and take you back to the previous window, Click on “Reload” button to complete the association. Next click the “Create” button.

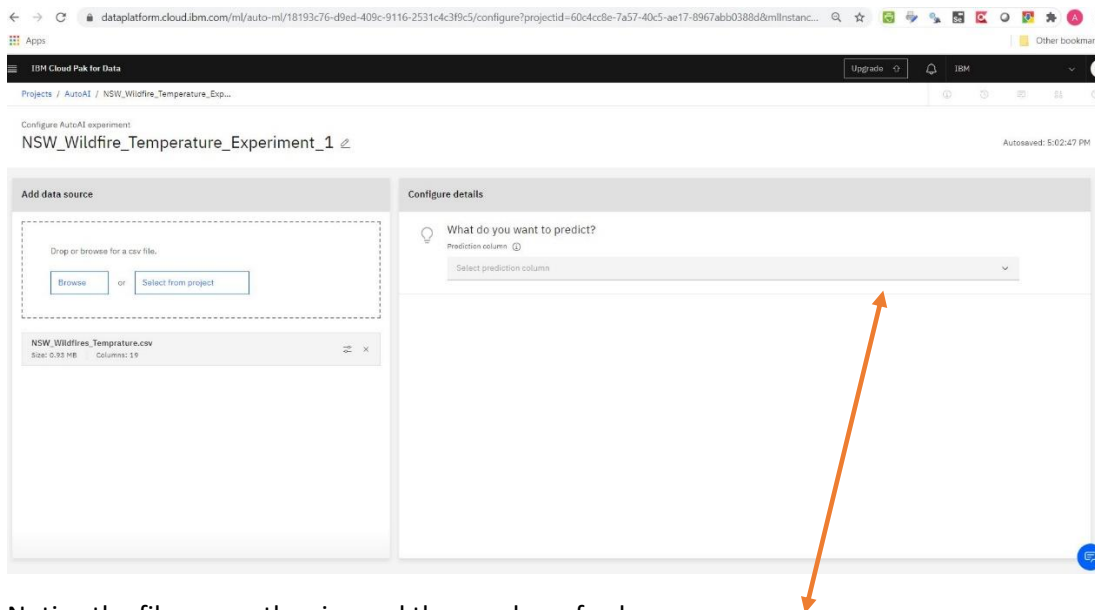


Please NOTE: the green box above informs how many CUH – Capacity Units per Hour will be consumed by this service/instance that has 8-Virtual CPUs and 32GB of RAM

Now add the data source

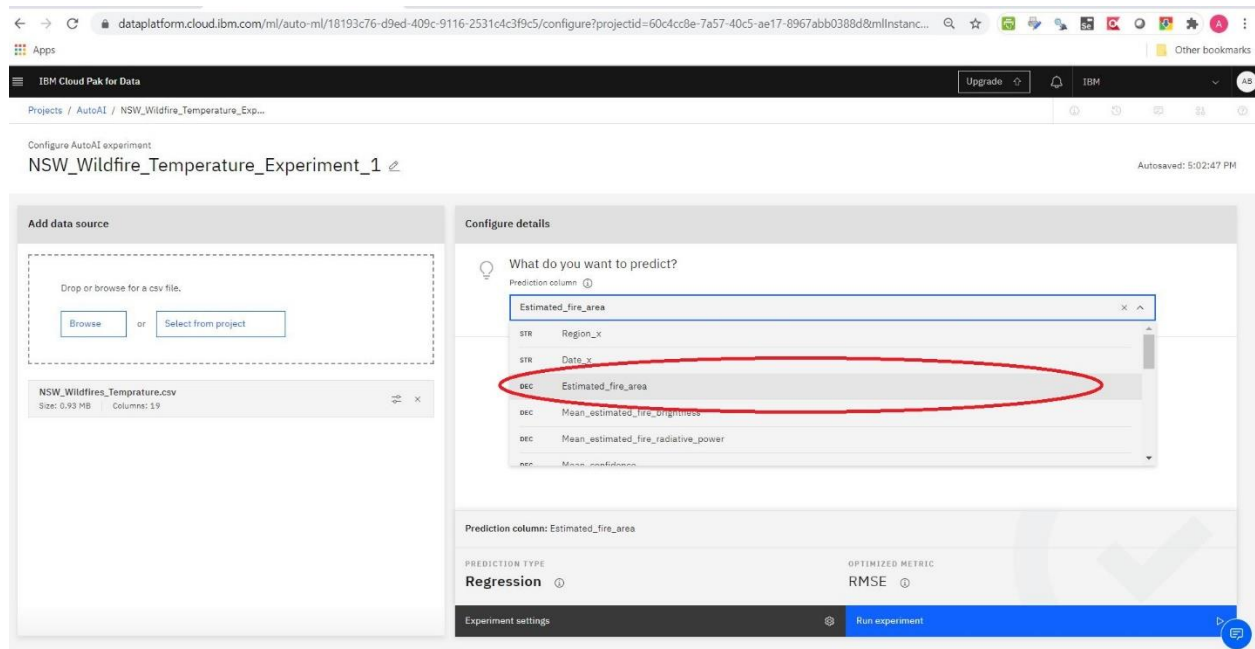


Browse to where the data is located, here the file is named NSW\_Wildfires\_Temprature.csv, once that is located and brought into the experiment it will show up as follows:



Notice the file name, the size and the number of columns.

Next, the drop down on the right allows you to select the column to predict. Notice that AutoAI engine has already determined the data types for each of the columns. Select "Estimated\_fire\_area" as shown.



Notice the AutoAI engine has already determined the prediction type as well as suggested the optimization metric to be RMSE.

Click on the “Run experiment” button at the bottom right to invoke the running of the Experiment.

The “Experiment setting” button will let you select the setting associated with the experiment.

# Experiment Settings

## Data Source:

Set the data splitting, column selection to include in the experiment

Experiment settings

Prediction column: Estimated\_fire\_area (DEC) Data source: NSW\_Wildfires\_Temperature.csv

Data source

Prediction

Runtime

Data source settings

**Subsample**  
For a large data set, use a subset of data to train the experiment. This speeds up results but may affect accuracy.  
☒ Subsample rows

**Training data split**  
You can optionally adjust the percentage of your data source to use for creating, optimizing, and validating pipelines. Only recommended for large data sets to avoid decreasing the quality of the pipelines.  
85% 95%  
Training data split: 90% - 3 folds Holdout data split: 10%

**Select columns to include** 14 / 19  
Select columns with data that support the prediction column.

Column name	Type
<input checked="" type="checkbox"/> Region_x	String
<input type="checkbox"/> Date_x	String

Cancel Save settings

Uncheck box to exclude column data from the experiment

Experiment settings

Prediction column: Estimated\_fire\_area (DEC) Data source: NSW\_Wildfires\_Temperature.csv

Data source

Prediction

Runtime

Data source settings

Column name	Type
<input checked="" type="checkbox"/> Region_x	String
<input type="checkbox"/> Date_x	String
<input checked="" type="checkbox"/> Estimated_fire_area	Decimal
<input checked="" type="checkbox"/> Mean_estimated_fire_brightness	Decimal
<input checked="" type="checkbox"/> Mean_estimated_fire_radiative_power	Decimal
<input checked="" type="checkbox"/> Mean_confidence	Decimal
<input checked="" type="checkbox"/> Std_confidence	Decimal
<input checked="" type="checkbox"/> Var_confidence	Decimal
<input type="checkbox"/> Count	Integer
<input checked="" type="checkbox"/> Replaced	String

Items per page: 10 1-10 of 19 items 1 1 of 2 pages

Cancel Save settings



Prediction column	Data source
Estimated_fire_area (b0c)	NSW_Wildfires_Temperature.csv

Cancel Save settings

## Prediction:

Experiment settings

Prediction column: Estimated\_fire\_area (DEC) | Data source: NSW\_Wildfires\_Temperature.csv

**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** (Selected)  
Predict values from a continuous set of values. Choose this if your prediction column contains a large number of values.

**Optimized metric**  
Choose the metric to optimize for the experiment.

- Root mean squared error (RMSE) (Recommended)** (Selected)
- Mean squared error (MSE)
- Mean absolute error (MAE)
- Median absolute error (MedAE)
- Root mean squared log error (RMSLE)
- Mean squared log error (MSLE)
- Explained variance

Cancel | Save settings

AutoAI engine uses Estimated\_fire\_area as the prediction column, as a Regression experiment with the choices of Optimized metrics, shown. RMSE is selected in this case.

Experiment settings

Prediction column: Estimated\_fire\_area (DEC) | Data source: NSW\_Wildfires\_Temperature.csv

**Optimized metric**  
Choose the metric to optimize for the experiment.

- Root mean squared error (RMSE) (Recommended)** (Selected)
- Mean squared error (MSE)
- Mean absolute error (MAE)
- Median absolute error (MedAE)
- Root mean squared log error (RMSLE)
- Mean squared log error (MSLE)
- Explained variance

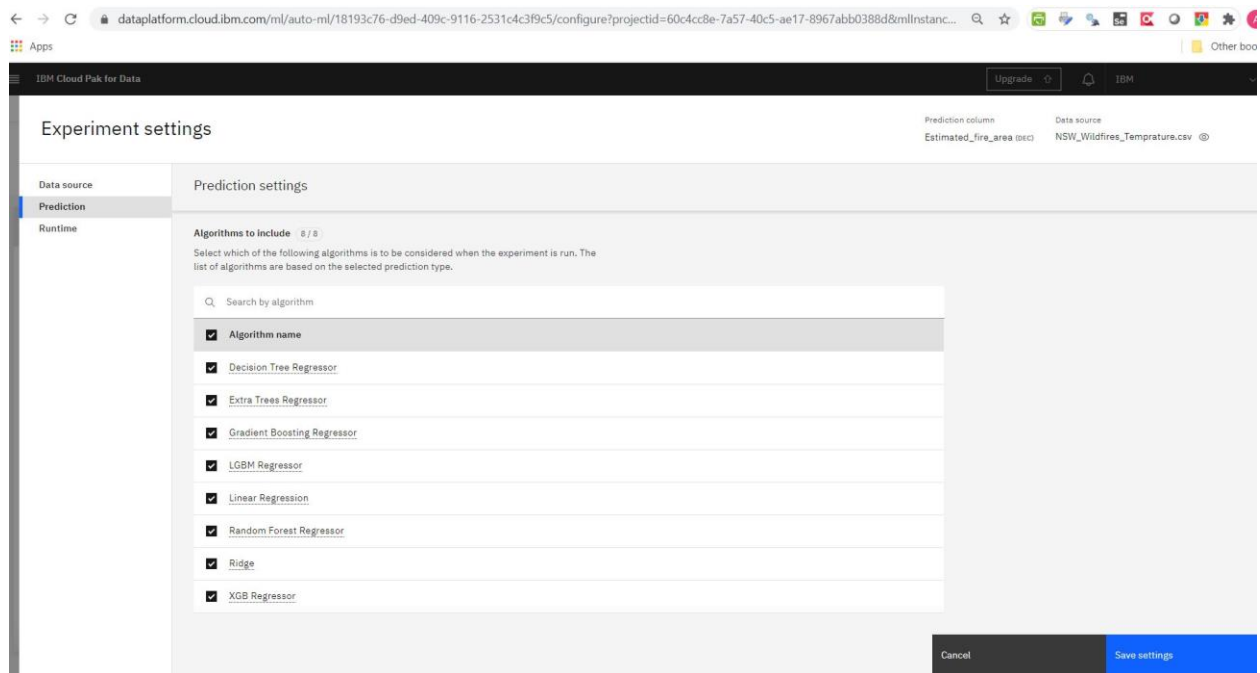
**Algorithms to include** 8 / 8  
Select which of the following algorithms is to be considered when the experiment is run. The list of algorithms are based on the selected prediction type.

Search by algorithm

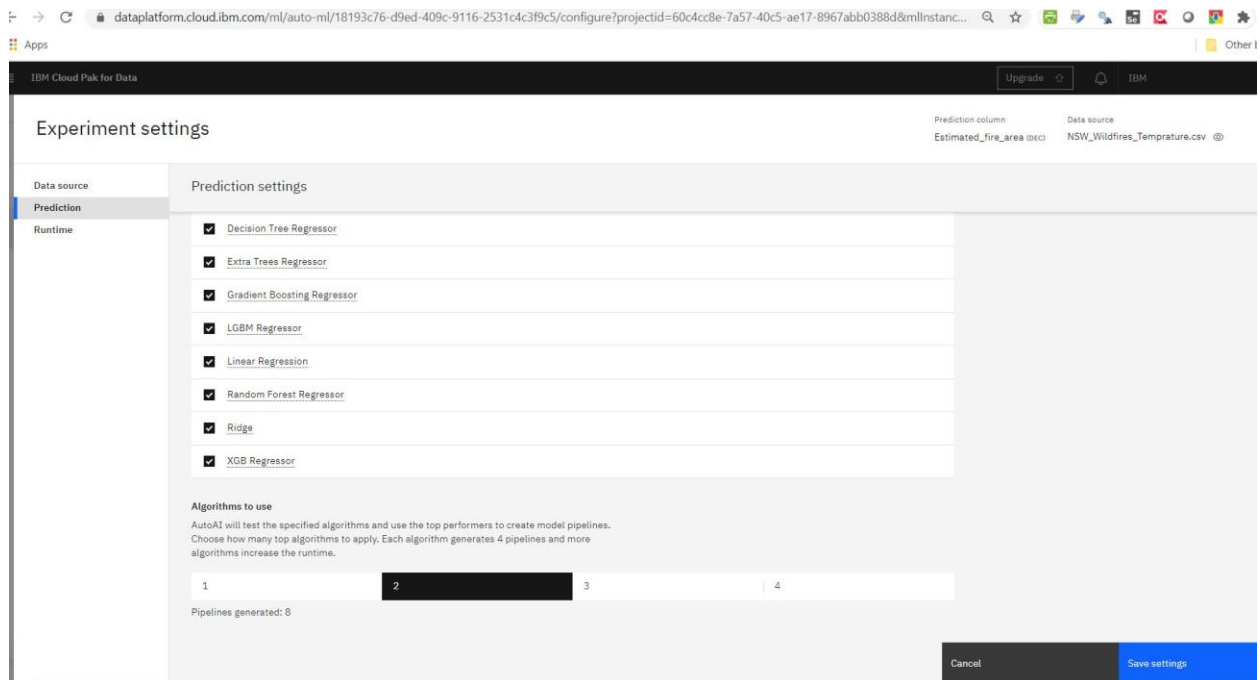
- ☒ Algorithm name

Cancel | Save settings

Algorithms that AutoAI will use is listed for selection next:



Select the number of top algorithms that AutoAI will test and use as shown: More algorithms increase the runtime as well as CUH.



## Experiment settings

Prediction column	Data source
Estimated_fire_area (DEC)	NSW_Wildfires_Temperature.csv

Data source

Prediction

Runtime

Prediction settings

☒ [Decision Tree Regressor](#)

☒ [Extra Trees Regressor](#)

☒ [Gradient Boosting Regressor](#)

☒ [LGBM Regressor](#)

☒ [Linear Regression](#)

☒ [Random Forest Regressor](#)

☒ [Ridge](#)

☒ [XGB Regressor](#)

Algorithms to use

AutoAI will test the specified algorithms and use the top performers to create model pipelines. Choose how many top algorithms to apply. Each algorithm generates 4 pipelines and more algorithms increase the runtime.

1234

Pipelines generated: 12

Cancel

Save settings

## Runtime Settings:

The screenshot shows the 'Experiment settings' page in the IBM Cloud Pak for Data interface. The 'Runtime' tab is selected in the left sidebar. The main content area is titled 'Runtime settings' and contains several sections:

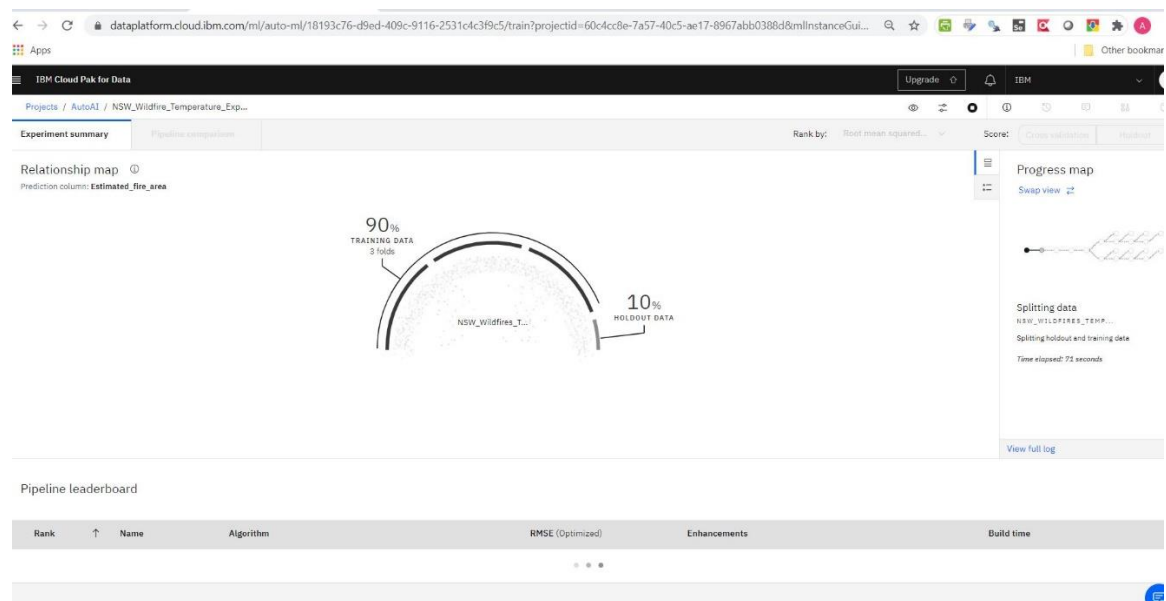
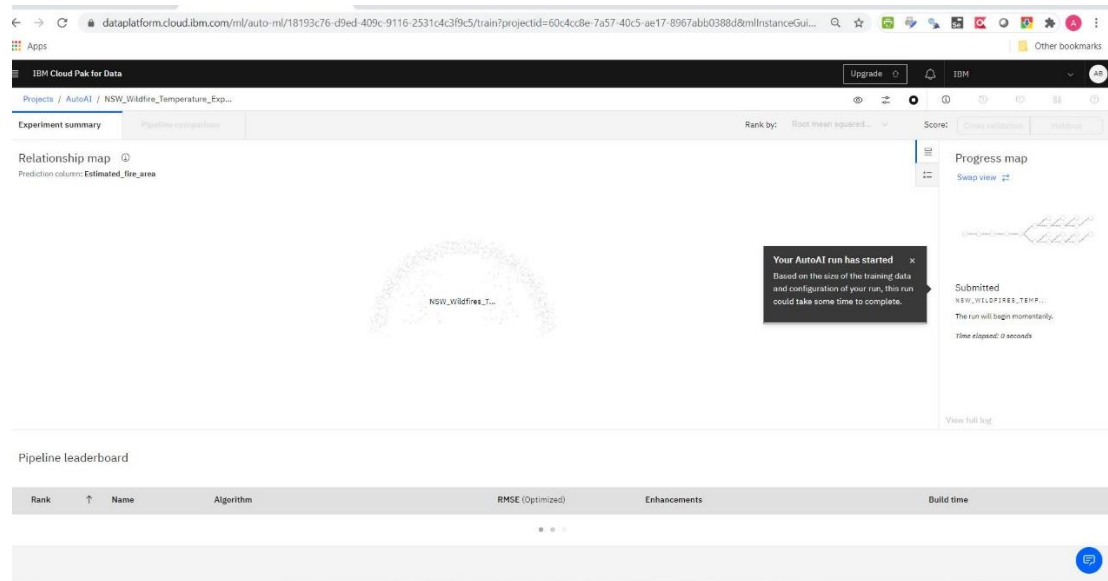
- Experiment details:** Review additional details about your experiment. These cannot be changed.
  - Initial model tuning iterations: 25
  - Feature engineering iterations: 60
  - Final model tuning iterations: 50
- Watson Machine Learning service:** Machine Learning-wm
- Compute configuration:** Customize the resources allocated for the AutoAI experiment. Learn more about capacity unit hours and Watson Machine Learning pricing plans.
- Model creation:** The selected runtime has 8 CPU and 32 GB RAM and consumes 20 capacity units per hour.

At the bottom right, there are two buttons: 'Cancel' and 'Save settings'. The 'Save settings' button is highlighted with a red circle.

Save settings, before progressing to the next step.

# Run Experiment

The next few screen captures detail the progress of the experiment.



← → ↻ dataplatform.cloud.ibm.com/ml/auto-ml/18193c76-d9ed-409c-9116-2531c4c3f9c5/train?projectId=60c4cc8e-7a57-40c5-ae17-8967abb0388d&mlInstanceGui...


Apps Other bookmarks

IBM Cloud Pak for Data Upgrade IBM


Projects / AutoAI / NSW\_Wildfire\_Temperature\_Exp...

Experiment summary Pipeline comparison Rank by: Root mean squared... Score: Cross validation Holdout

Relationship map ⓘ  
Prediction column: **Estimated\_fire\_area**



Progress map  
Swap view ⓘ



Model selection  
NSW\_WILDFIRE\_TEMPERATURE...  
Selecting algorithms for pipeline generation using 10% of training data. Discarding underperforming algorithms and keeping the top 2 algorithms.  
Time elapsed: 93 seconds  
View full log

Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
------	---	------	-----------	------------------	--------------	------------

← → ↻ dataplatform.cloud.ibm.com/ml/auto-ml/18193c76-d9ed-409c-9116-2531c4c3f9c5/train?projectId=60c4cc8e-7a57-40c5-ae17-8967abb0388d&mlInstanceGui...

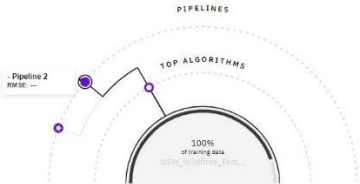
Apps Other bookmarks

IBM Cloud Pak for Data Upgrade IBM

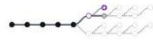
Projects / AutoAI / NSW\_Wildfire\_Temperature\_Exp...

Experiment summary Pipeline comparison Rank by: Root mean squared... Score: Cross validation Holdout

Relationship map ⓘ  
Prediction column: **Estimated\_fire\_area**



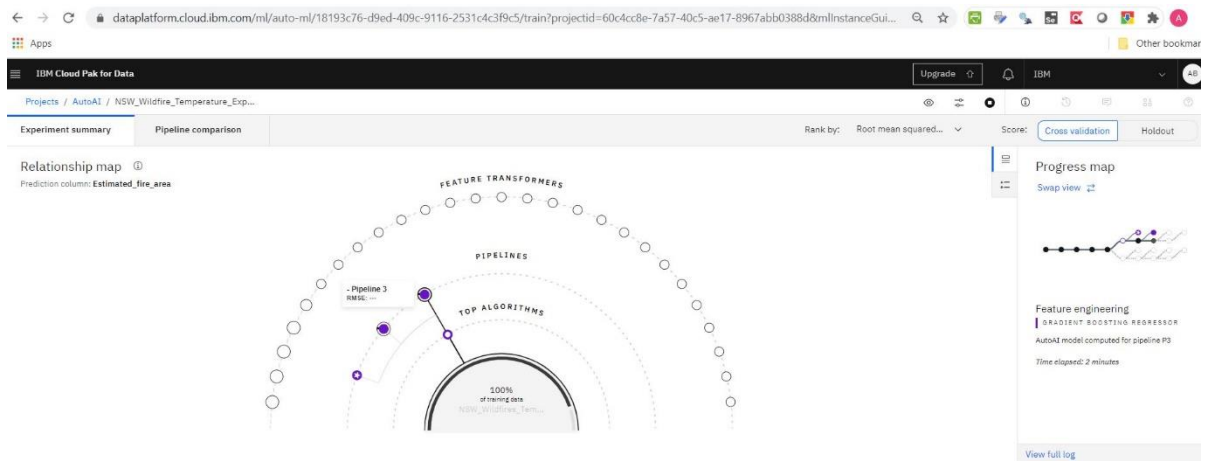
Progress map  
Swap view ⓘ



Hyperparameter optimization  
GRADIENT BOOSTING REGRESSOR  
AutoAI model computed for pipeline P2  
Time elapsed: 114 seconds  
View full log

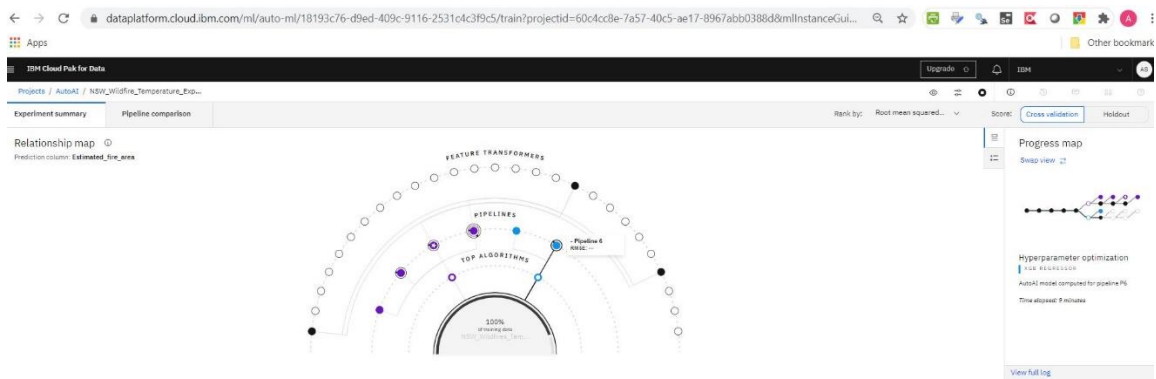
Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
★ 1		Pipeline 1	Gradient Boosting Regressor	344652.767	None	00:00:05



#### Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
★ 1		Pipeline 1	Gradient Boosting Regressor	344652.767	None	00:00:05
2		Pipeline 2	Gradient Boosting Regressor	344652.767	HPO-1	00:00:28



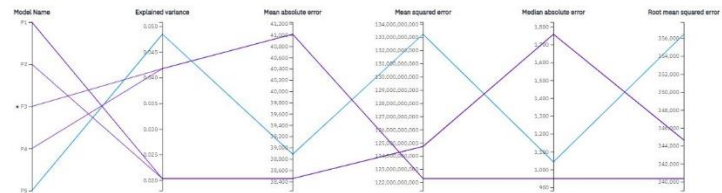
#### Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
★ 1		Pipeline 3	Gradient Boosting Regressor	340359.907	HPO-1 FE	00:04:22
2		Pipeline 4	Gradient Boosting Regressor	340359.907	HPO-1 FE HPO-2	00:01:08
3		Pipeline 1	Gradient Boosting Regressor	344652.767	None	00:00:05
4		Pipeline 2	Gradient Boosting Regressor	344652.767	HPO-1	00:00:28
5		Pipeline 5	XGB Regressor	356436.558	None	00:00:01



Metric chart

Prediction column: Estimated\_fire\_area



Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Explained variance	Mean absolute error (MAE)	Mean squared error (MSE)	Median absolute error (Me...)	R <sup>2</sup>
1	★	Pipeline 3	Gradient Boosting Regressor	340359.907	0.042	41952.752	122290835356.673	949.340	0.041
2		Pipeline 4	Gradient Boosting Regressor	340359.907	0.042	41952.752	122290835356.673	949.340	0.041
3		Pipeline 1	Gradient Boosting Regressor	344652.767	0.020	38453.033	124725419168.493	1758.102	0.020
4		Pipeline 2	Gradient Boosting Regressor	344652.767	0.020	38453.033	124725419168.493	1758.102	0.020
5		Pipeline 5	XGB Regressor	356438.518	-0.048	38885.389	133227394389.333	1042.307	-0.049

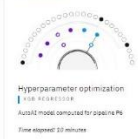
Progress map

Prediction column: Estimated\_fire\_area



Relationship map

Swap view

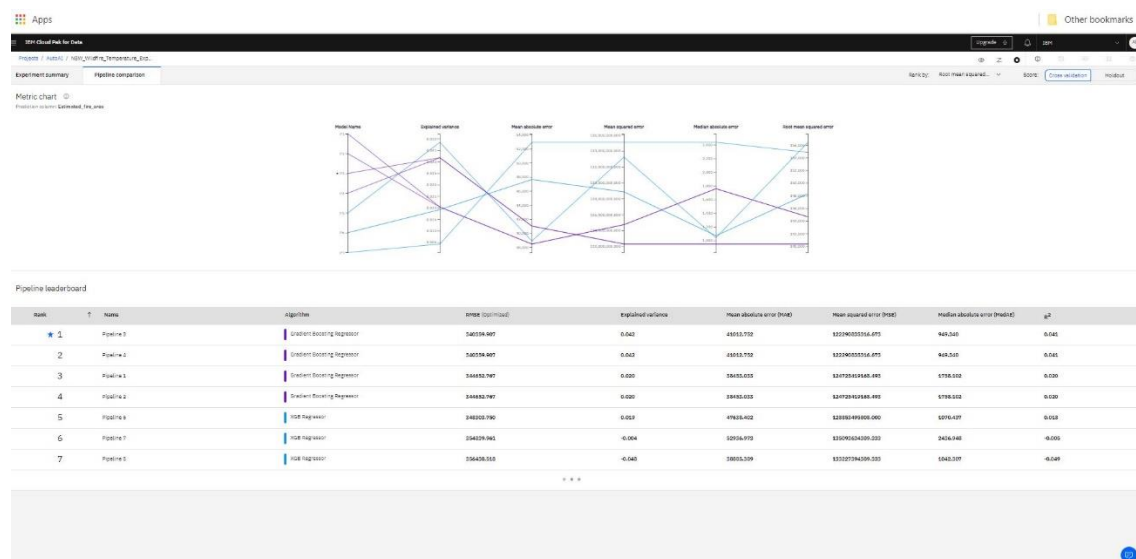
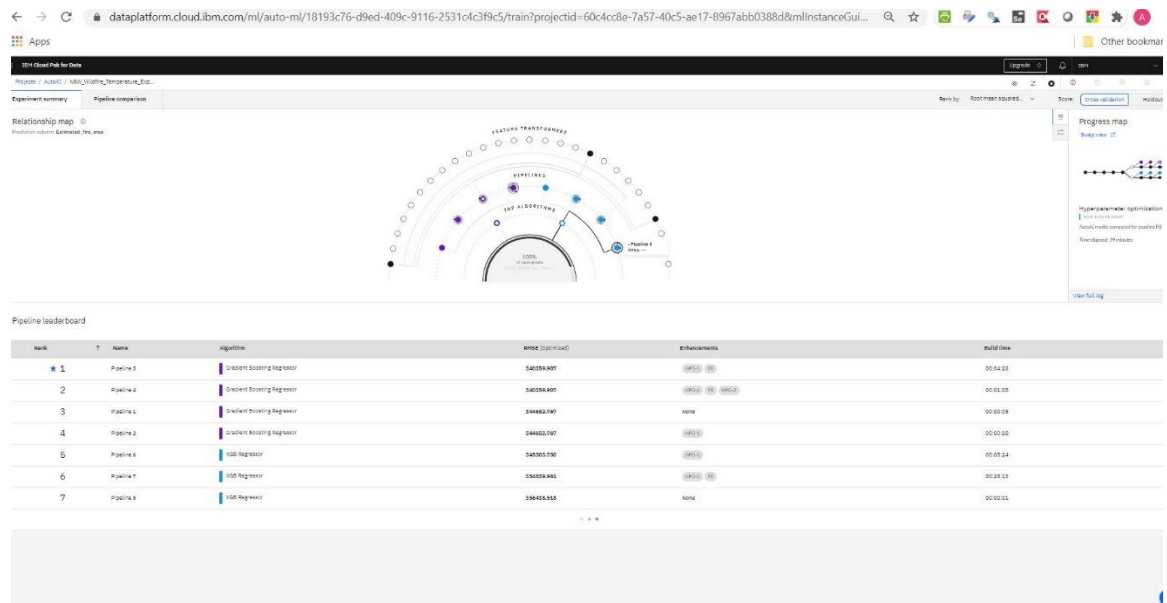


Hyperparameter optimization  
AutoML mode computed for pipeline P6  
Time elapsed: 10 minutes

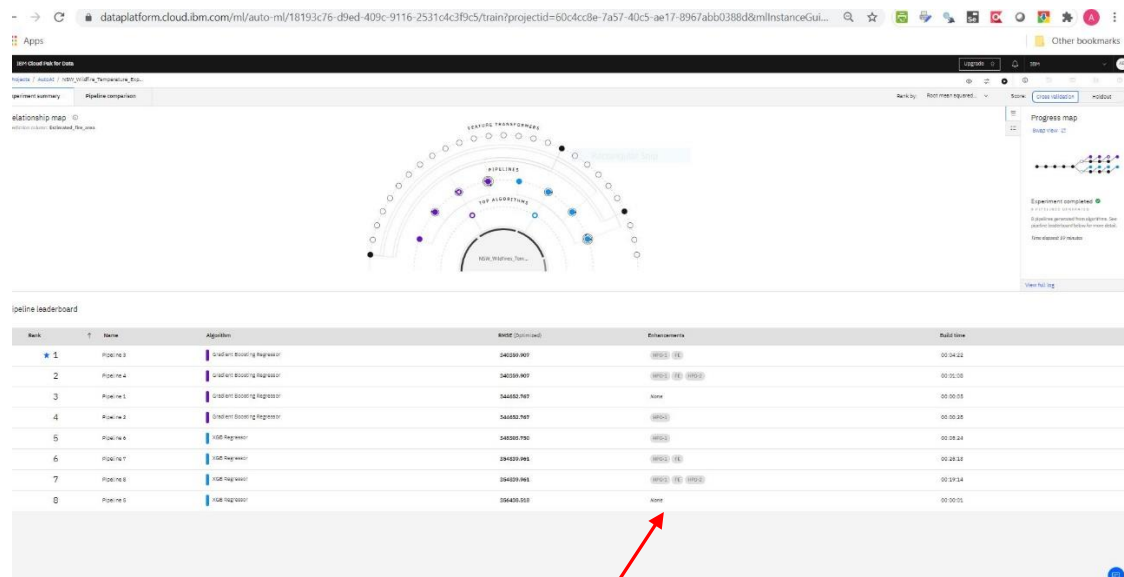
View full log

Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
1	★	Pipeline 3	Gradient Boosting Regressor	340359.907	HPO-1, FE	00:04:22
2		Pipeline 4	Gradient Boosting Regressor	340359.907	HPO-1, FE, HPO-3	00:02:08
3		Pipeline 1	Gradient Boosting Regressor	344652.767	None	00:00:06
4		Pipeline 2	Gradient Boosting Regressor	344652.767	HPO-1	00:00:26
5		Pipeline 5	XGB Regressor	356438.518	None	00:00:01



Experiment is Completed, as shown in the next screen.

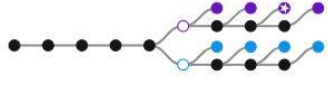


Showing the top pipelines, the algorithms and the value of RMSE. The enhancements performed are also listed. HPO-Hyper Parameter Optimization and FE – Feature Engineering

Score: Cross validation Holdout

Progress map

[Swap view](#)



Experiment completed ✓

8 PIPELINES GENERATED

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

*Time elapsed: 59 minutes*

[View full log](#)

AutoAI experiment full log


NSW\_Wildfire\_Temperature\_Experiment\_1

Date	Start time	End time	Time elapsed
11/3/2020	5:15:24 PM	6:14:24 PM	59 minutes

- 5:16:15 PM Starting the AutoAI experiment
- 5:16:31 PM AutoAI model computed for pipeline P1
- 5:16:32 PM Selecting algorithms for pipeline generation using 10% of training data. Discarding underperforming algorithms and keeping the top 2 algorithms.
- 5:16:32 PM AutoAI model computed for pipeline P1
- 5:16:38 PM Selecting algorithms for pipeline generation using 10% of training data. Discarding underperforming algorithms and keeping the top 2 algorithms.
- 5:16:38 PM AutoAI model computed for pipeline P1

Progress map

[Swap view](#)



Experiment completed ✓

8 PIPELINES GENERATED

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

*Time elapsed: 59 minutes*

[View full log](#)

FEATURE TRANSFORMERS

AutoAI experiment full log

NSW\_Wildfire\_Temperature\_Experiment\_1

Date	Start time	End time	Time elapsed
11/3/2020	5:15:24 PM	6:14:24 PM	59 minutes

6:13:45 PM

Completed data training

6:13:47 PM

AutoAI model computed for pipeline P8

6:14:04 PM

Completed composing pipelines

6:14:08 PM

AutoAI model computed for pipeline P8

6:14:24 PM

AutoAI experiment completed.

6:14:26 PM

Pipeline creation completed

6:14:26 PM

AutoAI experiment completed.

Swap view

Experiment completed

8 PIPELINES GENERATED

8 pipelines generated from algorithms. 3 pipeline leaderboard below for more details.

Time elapsed: 59 minutes

View full log

m

RMSE (Optimized)

Enhancements

Build time

## Saving the work

Save the top pipeline as required.

← → ↺ dataplatform.cloud.ibm.com/ml/auto-ml/18193c/6-d9ed-409c-9116-2531c43f9c3/train/projectid=60c4cc8e-/a5/-40c3-ae1/-896/abb0388d&mlInstanceGu... Apps Other bookmarks

IBM Cloud Pak for Data Upgrade IBM

Projects / AutoAI / NSW\_Wildfire\_Temperature\_Exp...

Experiment summary Pipeline comparison Rank by: Root mean squared... Score: Cross validation Holdout

Time elapsed: 59 minutes

View full log

Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
★ 1		Pipeline 3	Gradient Boosting Regressor	340359.907	HPO-1 FE	00:04:22 <a href="#">Save as</a>
2		Pipeline 4	Gradient Boosting Regressor	340359.907	HPO-1 FE HPO-2	00:01:08
3		Pipeline 1	Gradient Boosting Regressor	344652.767	None	00:00:05
4		Pipeline 2	Gradient Boosting Regressor	344652.767	HPO-1	00:00:28
5		Pipeline 6	XGB Regressor	348303.750	HPO-1	00:05:24
6		Pipeline 7	XGB Regressor	354839.961	HPO-1 FE	00:25:13

Save as

Select asset type

**Model**

Create a Watson Machine Learning model asset that you can test with new data, deploy to generate predictions, and trace lineage activity.

**Notebook beta**

Create a notebook if you want to view the code that created this model pipeline or interact with the model programmatically.

**Define details**

Notebook type

☐ WML notebook ⓘ

☒ AutoAI\_lib notebook ⓘ

Name

NSW\_Wildfire\_Temperature\_Experiment\_1 - P3 notebook

Description (optional)

First run of the NSW Wildfires , Estimated fire area versus Temperature. Experiment 1

Tags

Add tags to make assets easier to find.

Tags

Cancel Create

Apps

Other bookmarks

IBM Cloud Pak for Data

Upgrade

IBM

AB

Projects / AutoAI / NSW\_Wildfire\_Temperature\_Exp...

Experiment summary

Pipeline comparison

Rank by: Root mean squared...

NSW\_Wildfires\_Tem...

Notebook saved successfully.  
NSW\_Wildfire\_Temperature\_Exp...  
nt\_1 - P3 notebook was successfully  
saved to AutoAI.

View in project

View full log

Pipeline leaderboard

Rank	↑	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
★ 1		Pipeline 3	Gradient Boosting Regressor	340359.907	HPO-1 FE	00:04:22
2		Pipeline 4	Gradient Boosting Regressor	340359.907	HPO-1 FE HPO-2	00:01:08
3		Pipeline 1	Gradient Boosting Regressor	344652.767	None	00:00:05

Apps

Other bookmark

IBM Cloud Pak for Data

Upgrade

IBM

AB

Projects / AutoAI

Support for Python 3.6 is deprecated and will be removed in a future release. Use Python 3.7 instead.

Overview

Assets

Environments

Jobs

Access Control

Settings

What assets are you looking for?

Data assets

0 assets selected

<input type="checkbox"/>	Name	Type	Created by	Last modified
<input checked="" type="checkbox"/>	NSW_Wildfires_Temperature_Exp...	Data Asset	J	Nov 03, 2020, 09:03 PM
<input type="checkbox"/>	spambase_reduced_dim	Data Asset	E	Oct 22, 2020, 12:16 PM

AutoAI experiments

New AutoAI experiment

Name	Status	Model type	Last modified
NSW_Wildfires_Temperature_Exp..._3	Completed	Regression	Nov 03, 2020, 08:54 PM
AutoAI	Completed	Binary Classification	Oct 23, 2020, 08:58 AM
spambase-new	Completed	Binary Classification	Oct 22, 2020, 12:56 PM

Notebooks

New notebook

Name	Shared	Scheduled	Status	Language	Last editor	Last modified
spambase-new - P1 notebook				Python 3.7		Oct 22, 2020
spambase-new - P3 notebook				Python 3.7		Oct 22, 2020
spambase-new - P7 notebook				Python 3.7		Oct 22, 2020
NSW_Wildfires_Temperature_Exp..._3 - P3 notebook				Python 3.7		Nov 03, 2020

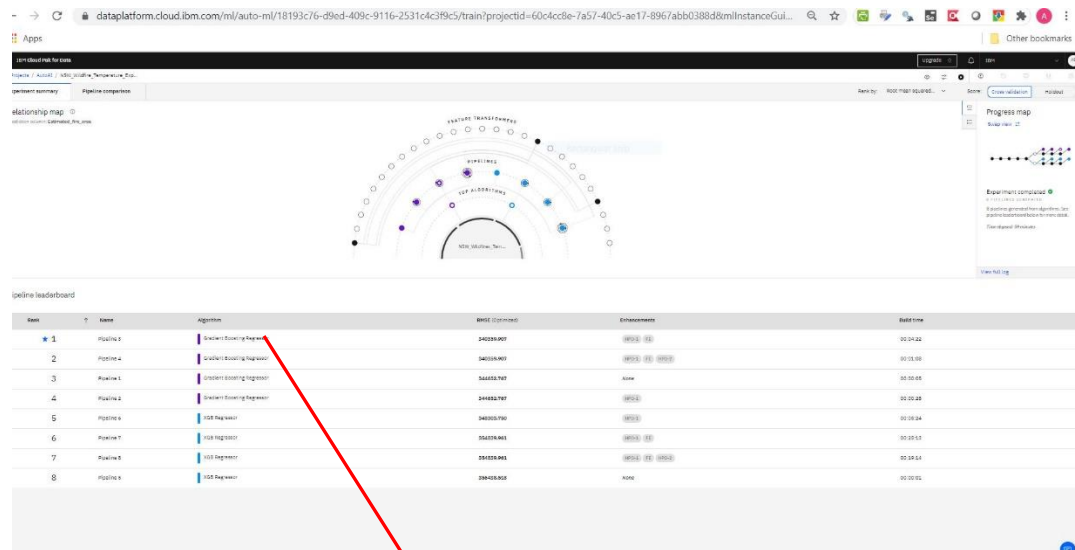
Deep learning experiments

New deep learning experiment

Name	Status	Last modified
You don't have any deep learning experiments yet.		

## Details of the Algorithm

Click the link in the link in the Algorithm column of the top pipeline to drill down into the working of the pipeline as shown.



Rank	Pipeline 1	Holdout RMSE (Optimized)	Algorithm	Enhancements	Build time	Save as
3		136156.641	Gradient Boosting Regressor	None	00:00:05	

Model Evaluation Measures		
TARGET : ESTIMATED_FIRE_AREA		
	Holdout Score	Cross Validation Score
Root Mean Squared Error (RMSE)	136,156.641	344,652.767
R <sup>2</sup>	-0.794	0.020
Explained Variance	-0.785	0.020
Mean Squared Error (MSE)	18,538,630,937.799	124,725,419,168.493
Mean Absolute Error (MAE)	26,718.360	38,453.033
Median Absolute Error (MedAE)	1,543.148	1,758.102



Rank  
3

Pipeline 1

Holdout RMSE (Optimized)  
136156.641

Algorithm  
Gradient Boosting Regressor

Enhancements  
None

Build time  
00:00:05

Save as

GradientBoostingRegressor

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Importance

Model Information

TARGET : ESTIMATED\_FIRE\_AREA

Label (Target)	Estimated_fire_area
Number of Features	9
Created At	11/3/2020, 5:17:11 PM

Rank  
3

Pipeline 1

Holdout RMSE (Optimized)  
136156.641

Algorithm  
Gradient Boosting Regressor

Enhancements  
None

Build time  
00:00:05

Save as

GradientBoostingRegressor

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Importance

Feature Importance

TARGET : ESTIMATED\_FIRE\_AREA

Feature	Importance
Mean_confidence	0.73
2nd moment	0.07
Mean_estimated_f_	0.06
Mean_estimated_f_	0.05
Var_confidence	0.04
min0	0.02

Rank  
3

Pipeline 1

Holdout RMSE (Optimized)  
136156.641

Algorithm  
Gradient Boosting Regressor

Enhancements  
None

Build time  
00:00:05

Save as

GradientBoostingRegressor

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Importance

Feature Importance

TARGET : ESTIMATED\_FIRE\_AREA

Feature	Importance
Mean_estimated_f_	0.05
Var_confidence	0.04
min0	0.02
Std_confidence	0.01
mean0	0.01
min0	0.00

Rank  
8

Pipeline 5

Holdout RMSE (Optimized)  
151113.907

Algorithm  
XGB Regressor

Enhancements  
None

Build time  
00:00:01

Save as

XGBRegressor

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Importance

Model Evaluation Measures

TARGET : ESTIMATED\_FIRE\_AREA

	Holdout Score	Cross Validation Score
Root Mean Squared Error (RMSE)	151,113.907	356,438.518
R <sup>2</sup>	-1.210	-0.049
Explained Variance	-1.198	-0.048
Mean Squared Error (MSE)	22,835,412,992.000	133,227,394,389.333
Mean Absolute Error (MAE)	26,708.814	38,885.389
Median Absolute Error (MedAE)	1,051.590	1,042.367

Rank  
8

Pipeline 5

Holdout RMSE (Optimized)  
151113.907

Algorithm  
XGB Regressor

Enhancements  
None

Build time  
00:00:01

Save as

XGBRegressor

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Importance

Model Information

TARGET : ESTIMATED\_FIRE\_AREA

Label (Target)	Estimated_fire_area
Number of Features	9
Created At	11/3/2020, 5:23:32 PM

Rank  
8

Pipeline 5

Holdout RMSE (Optimized)  
151113.907

Algorithm  
XGB Regressor

Enhancements  
None

Build time  
00:00:01

Save as

XGBRegressor

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Importance

Feature Importance

TARGET : ESTIMATED\_FIRE\_AREA

Feature	Mean_confidence Importance
Mean_confidence	0.55
Std_moment	0.24
Mean_estimated_f	0.09
Std0	0.08
Mean_estimated_f	0.07
Std_confidence	0.06

Rank  
5

Pipeline 5

Holdout RMSE (Optimized)  
151113.997

Algorithm  
XGB Regressor

Enhancements  
None

Build time  
00:00:01

Save as

XGBRegressor

EVALUATION

Model Evaluation Measures

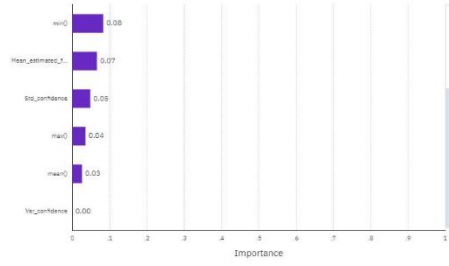
MODEL VIEWER

Model Information

Feature Importance

## Feature Importance ⓘ

TARGET : ESTIMATED\_FIRE\_AREA



IBM Cloud Pak for Data

Upgrade ⓘ

🔔

IBM

Rank  
1

Pipeline 3

Holdout RMSE (Optimized)  
23666.301

Algorithm  
Extra Trees Regressor

Enhancements  
HPD-1 FE

Build time  
00:01:26

Save as

ExtraTreesRegressor

Model Evaluation Measures ⓘ

TARGET : SALEPRICE

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Transformations

Feature Importance

	Holdout Score	Cross Validation Score
Root Mean Squared Error (RMSE)	23,666.301	27,792.471
R <sup>2</sup>	0.899	0.879
Explained Variance	0.899	0.879
Mean Squared Error (MSE)	560,093,826.510	781,905,658.128
Mean Squared Log Error (MSLE)	0.016	0.022
Mean Absolute Error (MAE)	15,419.735	16,806.971
Median Absolute Error (MedAE)	10,250.000	10,470.833
Root Mean Squared Log Error (RMSLE)	0.125	0.149

IBM Cloud Pak for Data

Upgrade ⓘ

🔔

IBM

Rank  
1

Pipeline 3

Holdout RMSE (Optimized)  
23666.301

Algorithm  
Extra Trees Regressor

Enhancements  
HPD-1 FE

Build time  
00:01:26

Save as

ExtraTreesRegressor

Model Information ⓘ

TARGET : SALEPRICE

EVALUATION

Model Evaluation Measures

MODEL VIEWER

Model Information

Feature Transformations

Label (Target)	SalePrice
Number of Features	116
Created At	10/28/2020, 3:24:04 PM

## Feature Engineering

IBM Cloud Pak for Data

Rank 1 Pipeline 3 Holdout RMSE (Optimized) 23666.301 Algorithm Extra Trees Regressor Enhancements HPO-1 FE Build time 00:01:26 Save as

ExtraTreesRegressor

Feature Transformations ⓘ

TARGET : SALEPRICE

New Feature	Original Feature	Transformation
NewFeature_10	Year.Remod.Add,GrLiv.Area	$\text{product}(\text{square}(\text{Year.Remod.Add}), \text{GrLiv.Area})$
NewFeature_22	Year.Remod.Add,GrLiv.Area	$\text{sqrt}(\text{product}(\text{Year.Remod.Add}, \text{square}(\text{GrLiv.Area})))$
NewFeature_15	Year.Remod.Add,Garage.Area	$\text{product}(\text{square}(\text{Garage.Area}), \text{square}(\text{Year.Remod.Add}))$
NewFeature_30	Year.Built,Garage.Yr.Blt	$\text{sqrt}(\text{product}(\text{square}(\text{Year.Built}), \text{Garage.Yr.Blt}))$
NewFeature_0	Year.Built,GrLiv.Area	$\text{product}(\text{Year.Built}, \text{GrLiv.Area})$
NewFeature_24	X1st.Fir.SF,Garage.Area	$\text{sqrt}(\text{product}(\text{X1st.Fir.SF}, \text{Garage.Area}))$
NewFeature_19	GrLiv.Area	$\text{sqrt}(\text{GrLiv.Area})$
NewFeature_21	Year.Built,GrLiv.Area	$\text{sqrt}(\text{product}(\text{Year.Built}, \text{GrLiv.Area}))$

## Feature Importance

IBM Cloud Pak for Data

Rank 1 Pipeline 3 Holdout RMSE (Optimized) 23666.301 Algorithm Extra Trees Regressor Enhancements HPO-1 FE Build time 00:01:26 Save as

ExtraTreesRegressor

Feature Importance ⓘ

TARGET : SALEPRICE

Feature	Importance
Overall.Quiet	0.29
NewFeature_10	0.11
NewFeature_22	0.09
Garage.Cars	0.04
NewFeature_19	0.04
NewFeature_30	0.04

