MODY
UNIVERSITY
OF SCIENCE AND TECHNOLOGY
LAKSHMANGARH ★ RAJASTHAN
A Leading Women's University

DOCUMENTATION

On

Fake News Classification

Submitted by

Yashika Singal (180150)

Aanchal Saxena (180006)

Uchita Gupta (180143)

CS 18.338

B.Tech (CSE+BDA)

Under the Guidance of

IBM

Ms. Khushbu Datwani

School of Science and Technology

Mody University of Science and Technology, Lakshmangarh

ACKNOWLEDGEMENT

We respect and thank Ms. Khushbu Datwani for providing us an opportunity to do the project work and giving us all their guidance and support.

We are thankful to and fortunate enough to receive constant support and guidance from all Teaching Staff of CSE which helped us in successfully completing our project work. Also, we would like to extend our sincere thanks to all staff in labs for their timely support.

TABLE OF CONTENT

Sr. no.	**Topics**	**Page no.**
	Acknowledgement	1
	List of Figures/Tables/Abbreviations	2
1.	Introduction	3
2.	Related Work	4-6
3.	Hardware and Software Details	7
4.	Implementation Work details	8
5.	Dashboard	9
6.	Individual Contribution	28
7.	Conclusion	29

INTRODUCTION

FAKE NEWS CLASSIFICATION (Description of Project)

This is a ML problem where the task is to classify Fake News in an article. Jupyter notebook consists of various stages needed for identifying fake news such as data preprocessing, model experimentation, and evaluation of results.

The datasets consist of news.csv file. It consists of a label column for the Fake News Classification part. First the data is pre-processed, then visualized graphically and then finally evaluated by calculating accuracy of different models.

The Project includes evaluating results of various models and getting the result from the best model by checking the accuracy.

1. Data Set: - Fake News Classification

There are countless sources of fake news nowadays, mostly coming from programmed bots, that can’t get tired and continue to spread false information 24/7. This project is basically regarding analysis of the fake news that is nowadays spread very oftenly by some people and people trust them easily. A fake are those news stories that are false: the story itself is fabricated, with no verifiable facts, sources, or quotes.

The features of our project are given below –

1. Framing the data sector wise for better understanding.

2. Data munging for cleansing or removing irrelevant data.

3. Visualization by different techniques

4. Graphical visualization

5. Advance Plotting Techniques.

6. Training Model.

7. Deployment of Model.

8. Job Creation.

1.1. Purpose

This project is to analyse a data set related to fake news. It includes information about different types of fake news. It gives us the perception of all kinds of news around the world.

1.2. Project Scope

1. It could be used to explain several ways to detect fake news using collected data from different articles.

2. It is valid for descriptive analysis and sneak peak of predictive analysis.

GOALS: -

1. How does the spread rate of fake news change with time?

2. Do people play a major role in spreading fake news?

3. What is the percentage of news that are fake amongst the whole section of

news?

1.3. References

https://drive.google.com/drive/folders/1DunoRz2T6NjZ8at2ID5IS5am_q9Gomko?usp=sharing

1.4 About Data Set

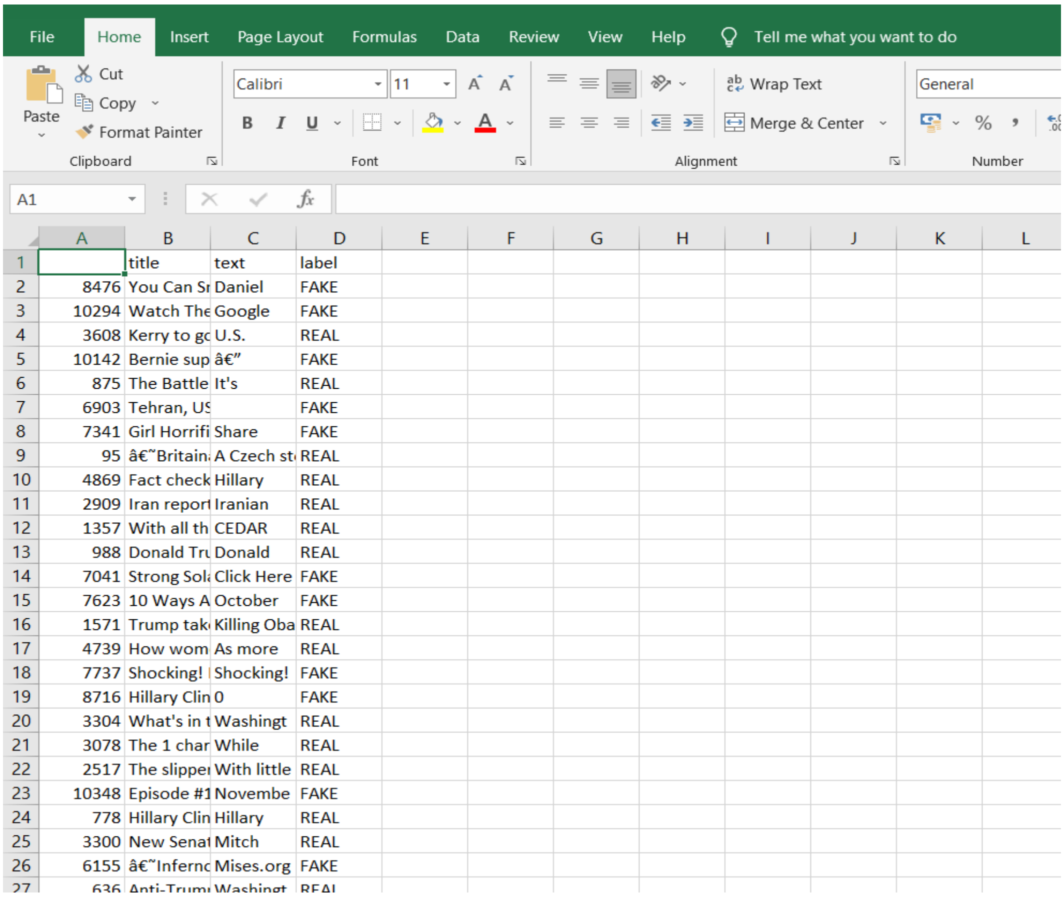
This data set contains information related to fake and real news. It contains the news number, title of the news, text of the news and label for news. This POC contains python code to manipulate and analyse the data. The data set contains approx. 7800 entries.

1.5 The details of the data set are given below:

Format:

The data set contains the following information about different news taken from google. They contain news numbers, title of the news, text of the news and label for news in the data set.

1.6 Screenshots



	A	B	C	D
1		title	text	label
2	8476	You Can Sı	Daniel	FAKE
3	10294	Watch The	Google	FAKE
4	3608	Kerry to gc	U.S.	REAL
5	10142	Bernie sup	â€”	FAKE
6	875	The Battle	It's	REAL
7	6903	Tehran, US		FAKE
8	7341	Girl Horrifi	Share	FAKE
9	95	â€˜Britain	A Czech st	REAL
10	4869	Fact check	Hillary	REAL
11	2909	Iran report	Iranian	REAL
12	1357	With all th	CEDAR	REAL
13	988	Donald Tru	Donald	REAL
14	7041	Strong Sol	Click Here	FAKE
15	7623	10 Ways A	October	FAKE
16	1571	Trump tak	Killing Oba	REAL
17	4739	How wom	As more	REAL
18	7737	Shocking!	Shocking!	FAKE
19	8716	Hillary Clin	0	FAKE
20	3304	What's in t	Washingt	REAL
21	3078	The 1 char	While	REAL
22	2517	The slipper	With little	REAL
23	10348	Episode #1	Novembe	FAKE
24	778	Hillary Clin	Hillary	REAL
25	3300	New Senat	Mitch	REAL
26	6155	â€˜Infernc	Mises.org	FAKE
27	636	Anti-Trum	Washingt	REAL

1.7 Attribute Information

ATTRIBUTE NAME	DESCRIPTION
News Number	Specifies the number of particular news
Title	Headline of the news
Text	What does that news say about
Label	States that whether the news is fake or real

2. TOOLS AND WORKING ENVIRONMENT:

Project will be made using IBM Watson Studio working with Python code.

Working environment will be the Anaconda Distribution (Jupyter Notebook). It includes data manipulation and visualization libraries such as

1. Numpy

2. Pandas

3.Sklearn

4.Matplotlib

5. Seaborn

2.1 Assumptions and Dependencies

The dependencies that could be involved throughout the project is IBM Watson Studio and AutoAI tools and ML algorithms insights.

Statically view, assumption plotting and training of models for prediction from the dataset.

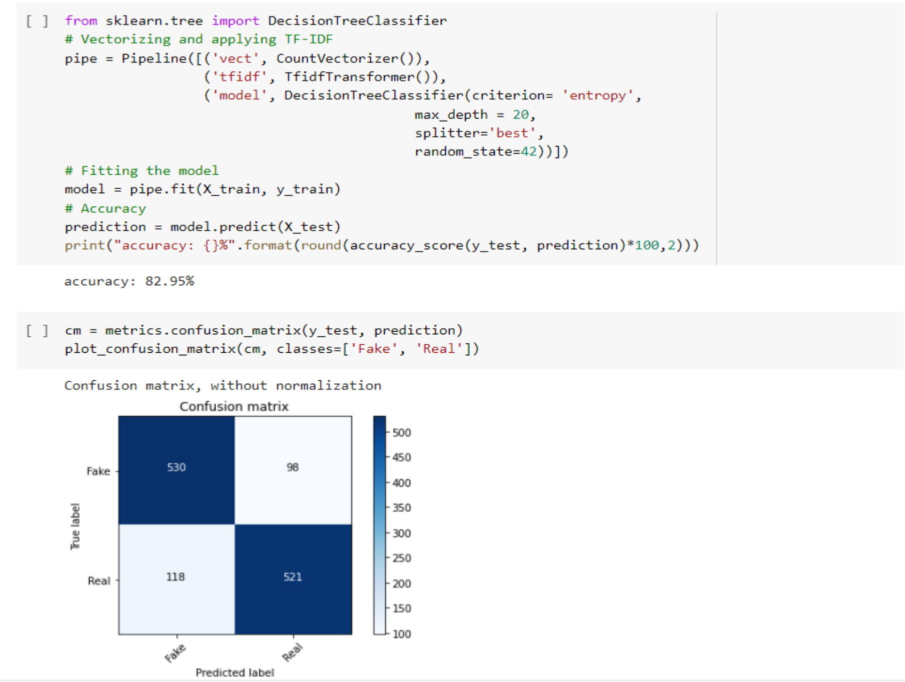
3. PYTHON CODE USING ML AND DEEP LEARNING:

The Source Code is available here in this file:

https://drive.google.com/file/d/1ngbZ3tEpPJs-p9pxeNKhB040_Rx6Pxl1/view?usp=sharing

Output of the models trained are as follows:

1. Using DecisionTreeClassifier



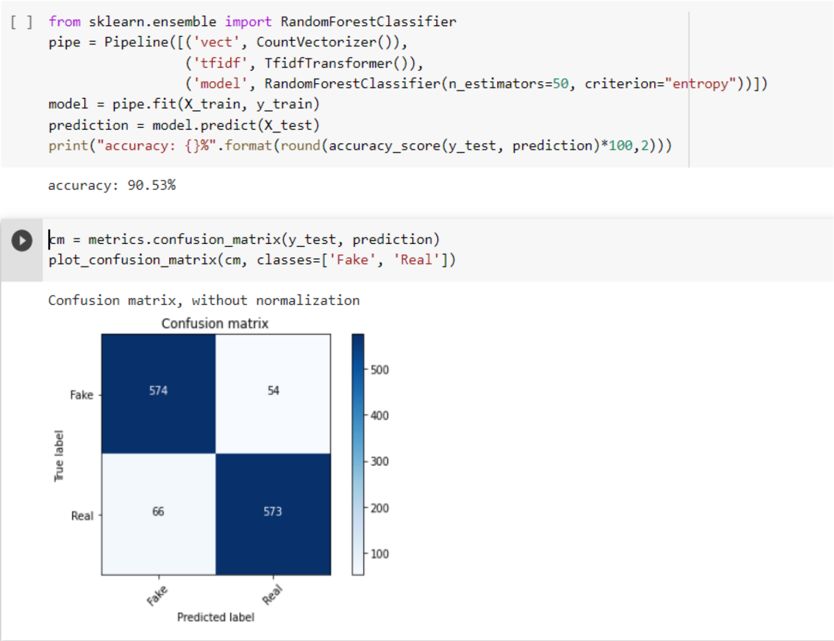
```
from sklearn.tree import DecisionTreeClassifier
# Vectorizing and applying TF-IDF
pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', DecisionTreeClassifier(criterion= 'entropy',
                                                  max_depth = 20,
                                                  splitter='best',
                                                  random_state=42))])


# Fitting the model
model = pipe.fit(X_train, y_train)
# Accuracy
prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

```
accuracy: 82.95%
```

```
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

```
Confusion matrix, without normalization
```

2. Using RandomForestClassifier



```
from sklearn.ensemble import RandomForestClassifier
pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', RandomForestClassifier(n_estimators=50, criterion="entropy"))])
model = pipe.fit(X_train, y_train)
prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

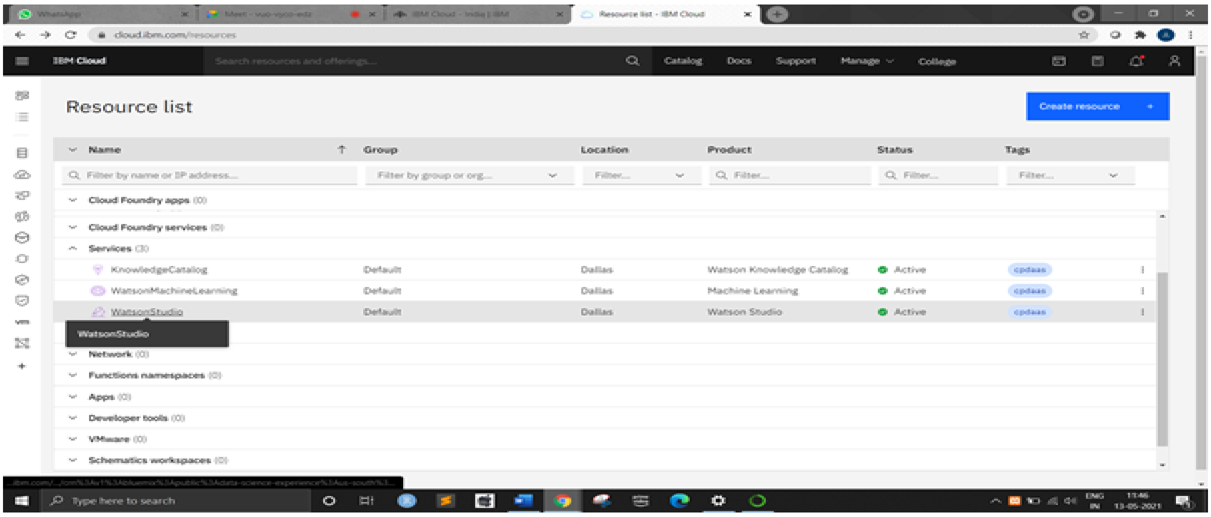
```
accuracy: 90.53%
```

```
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

```
Confusion matrix, without normalization
```

4. WORKING WITH IBM WATSON STUDIO USING AUTOAI:

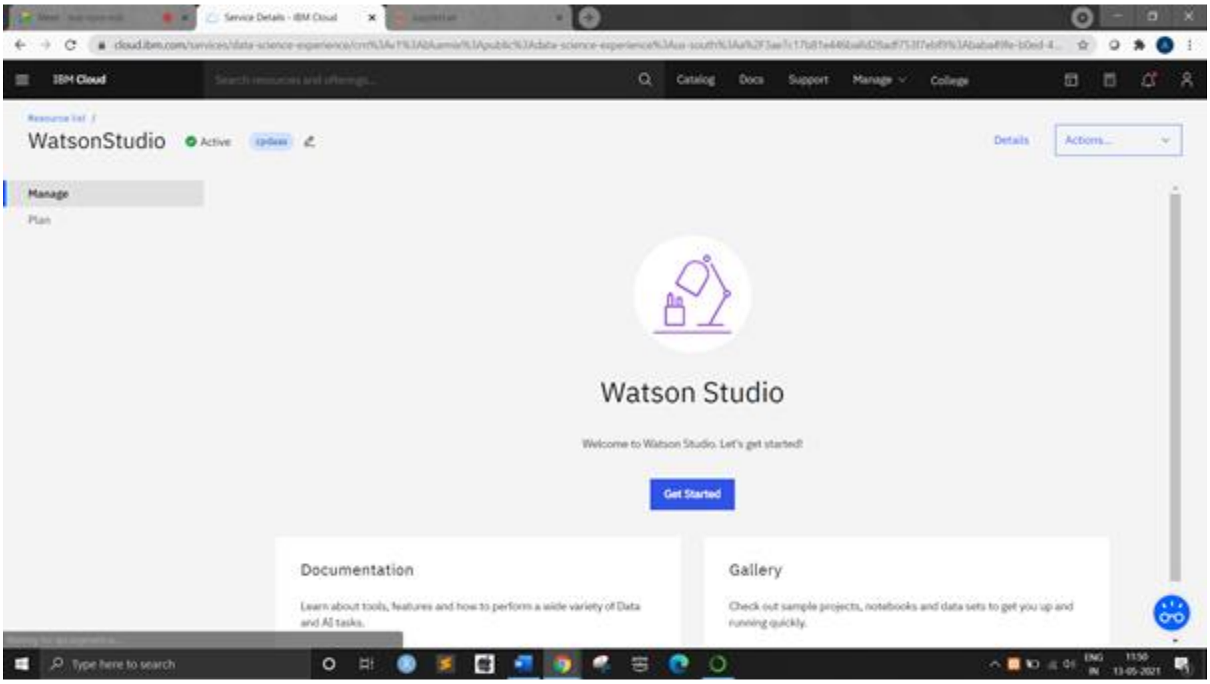
1. First, login in the IBM Cloud and then in the resource list in the services section, click on WatsonStudio Service, If not created then first create it.



Resource list

Name	Group	Location	Product	Status	Tags
Cloud Foundry apps (0)					
Cloud Foundry services (0)					
Services (3)					
KnowledgeCatalog	Default	Dallas	Watson Knowledge Catalog	Active	cpdaas
WatsonMachineLearning	Default	Dallas	Machine Learning	Active	cpdaas
WatsonStudio	Default	Dallas	Watson Studio	Active	cpdaas
Network (0)					
Functions namespaces (0)					
Apps (0)					
Developer tools (0)					
VMware (0)					
Schematics workspaces (0)					

2. Start the WatsonStudio by clicking at the center on Get Started.



WatsonStudio Active cpdaas

Watson Studio

Welcome to Watson Studio. Let's get started!

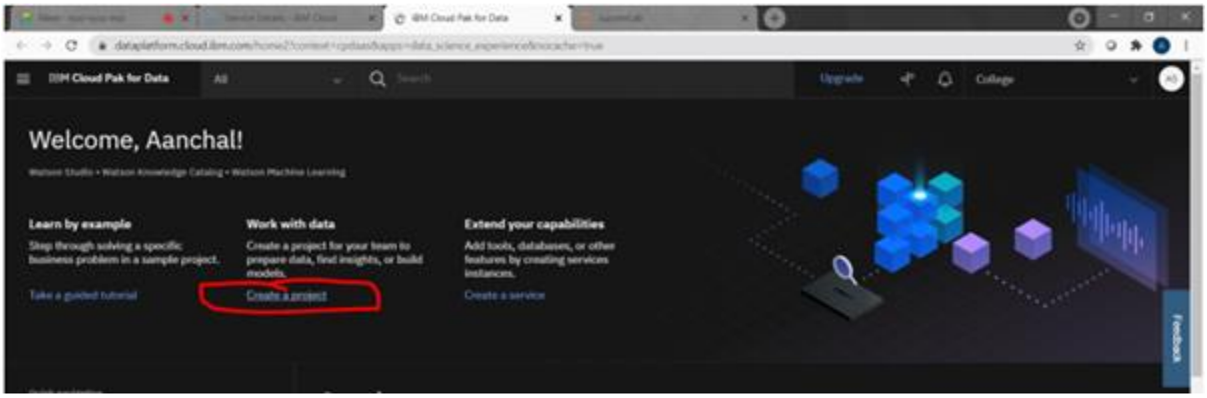
Documentation

Learn about tools, features and how to perform a wide variety of Data and AI tasks.

Gallery

Check out sample projects, notebooks and data sets to get you up and running quickly.

3. Click on Create a Project to create the project you want to work on.



IBM Cloud Pak for Data

Welcome, Aanchal!

Watson Studio • Watson Knowledge Catalog • Watson Machine Learning

Learn by example

Step through solving a specific business problem in a sample project.

Take a guided tutorial

Work with data

Create a project for your team to prepare data, find insights, or build models.

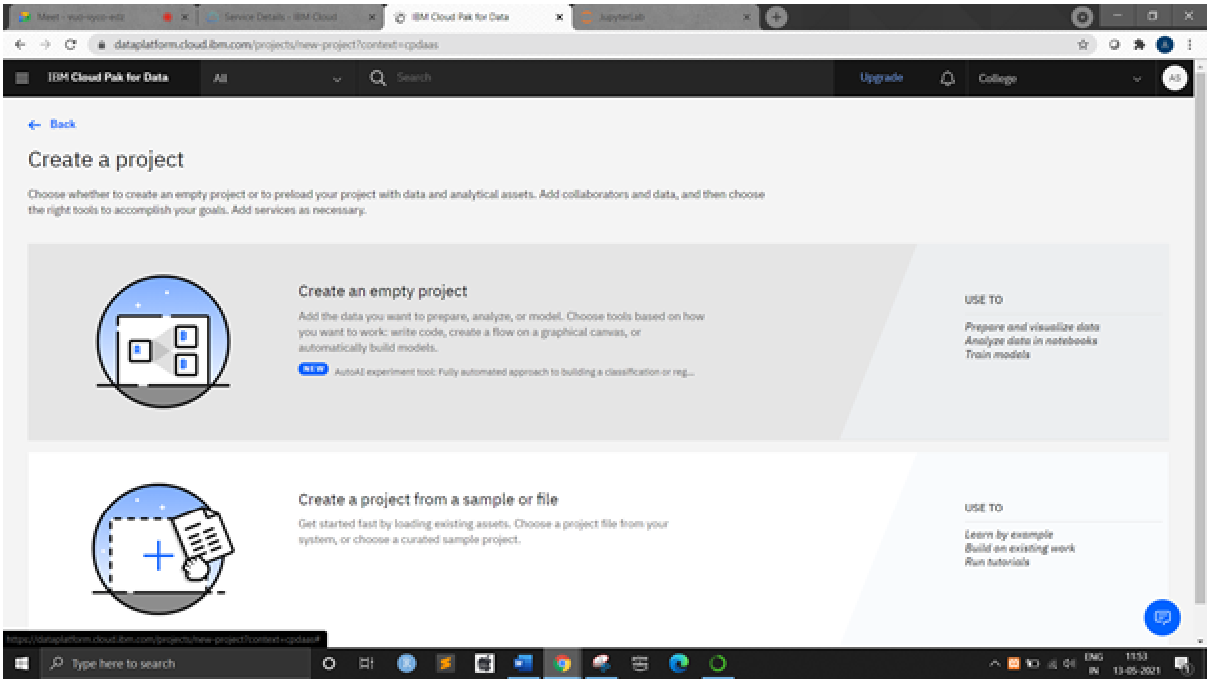
Create a project

Extend your capabilities

Add tools, databases, or other features by creating services instances.

Create a service

4. Click on Create an empty project.



dataplatform.cloud.ibm.com/projects/new-project?context=cpdaas

IBM Cloud Pak for Data

Back

Create a project

Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.

Create an empty project

Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

NEW AutoAI experiment tool: Fully automated approach to building a classification or reg...

USE TO

Prepare and visualize data
Analyze data in notebooks
Train models

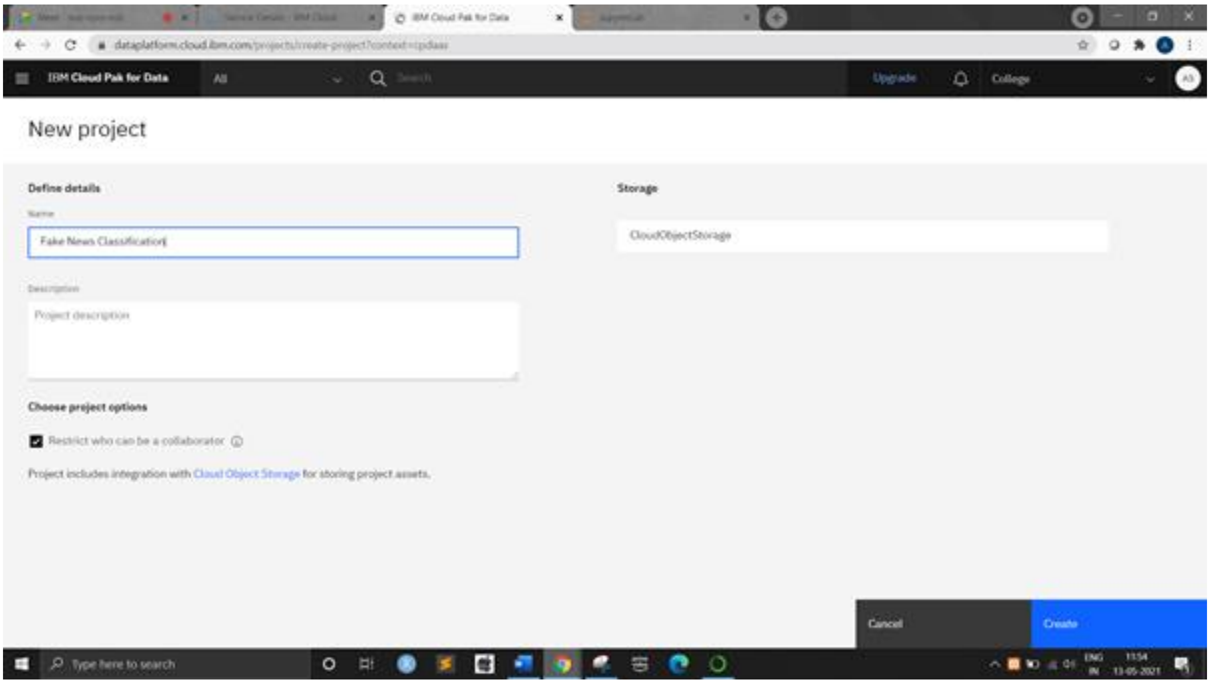
Create a project from a sample or file

Get started fast by loading existing assets. Choose a project file from your system, or choose a curated sample project.

USE TO

Learn by example
Build on existing work
Run tutorials

5. Give a Project Name and click on Create in the bottom-right corner.



New project

Define details

Name

Fake News Classification

Description

Project description

Storage

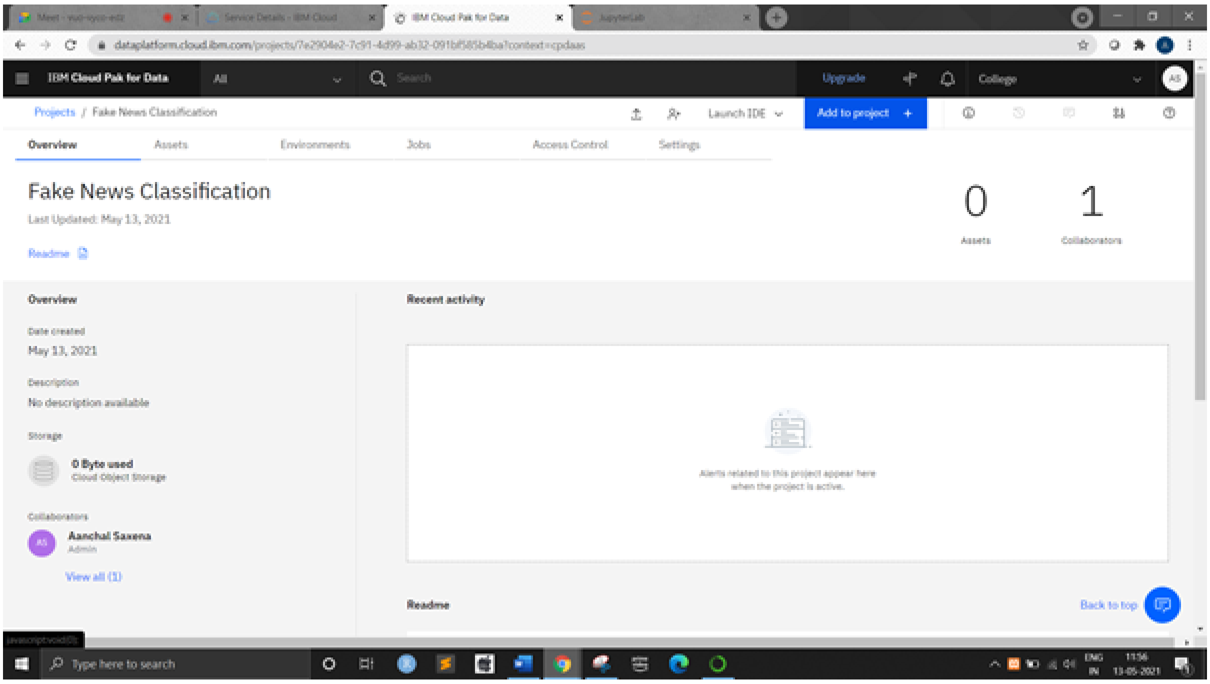
CloudObjectStorage

Choose project options

☑ Restrict who can be a collaborator

Project includes integration with Cloud Object Storage for storing project assets.

6. Click on Add to project + button at the top-right corner.



Fake News Classification

Last Updated: May 13, 2021

0 Assets

1 Collaborators

Overview

Date created

May 13, 2021

Description

No description available

Storage

0 Byte used

Cloud Object Storage

Collaborators

Aanchal Saxena

Admin

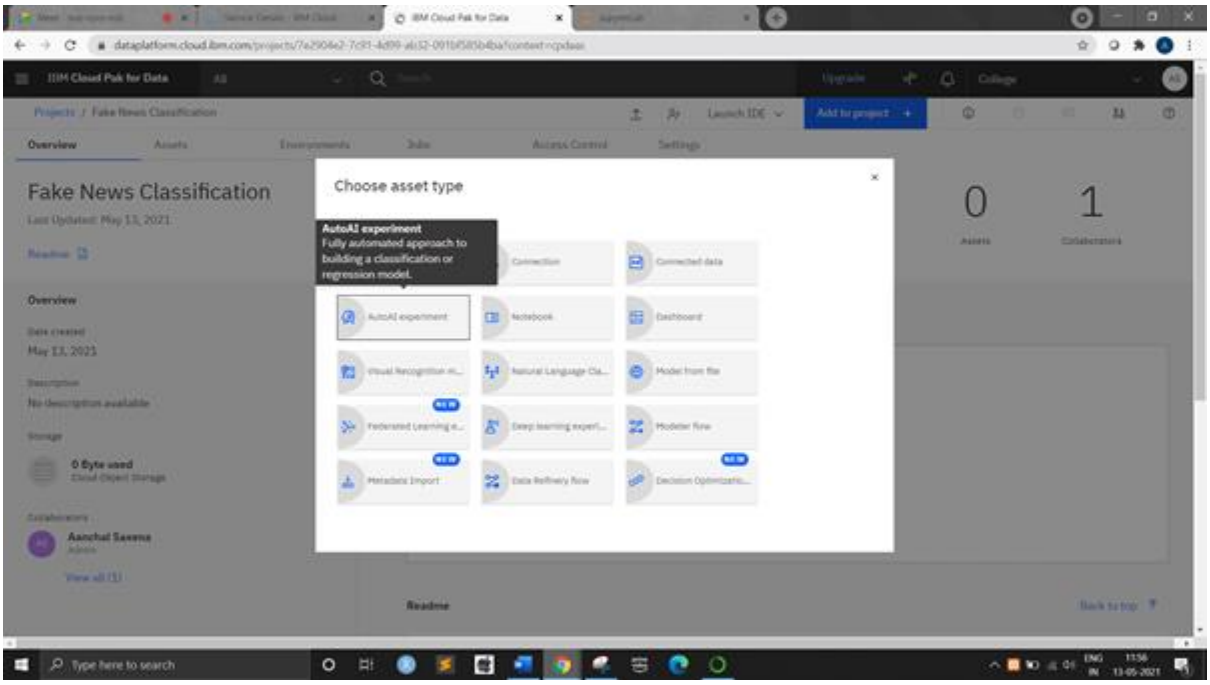
View all (1)

Recent activity

Alerts related to this project appear here when the project is active.

Readme

7. Click on AutoAI experiment.



Fake News Classification

Last Updated: May 13, 2021

Overview

Date created

May 13, 2021

Description

No description available

Storage

0 Byte used

Cloud Object Storage

Collaborators

Aanchal Saxena

Admin

View all (1)

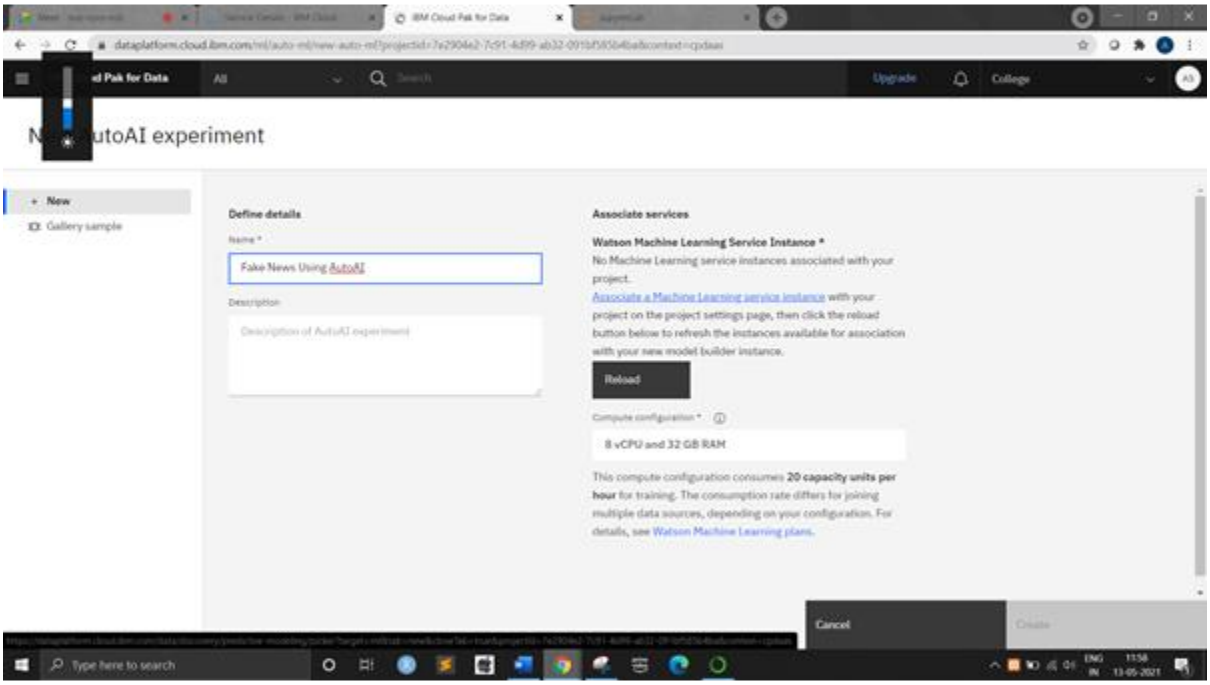
0 Assets

1 Collaborators

Choose asset type

AutoAI experiment
Fully automated approach to building a classification or regression model.

8. Click on Associate a ML instance breadcrumb link.



New AutoAI experiment

Define details

Name *

Fake News Using AutoAI

Description

Description of AutoAI experiment

Associate services

Watson Machine Learning Service Instance *

No Machine Learning service instances associated with your project.

Associate a Machine Learning service instance with your project on the project settings page, then click the reload button below to refresh the instances available for association with your new model builder instance.

Reload

Compute configuration *

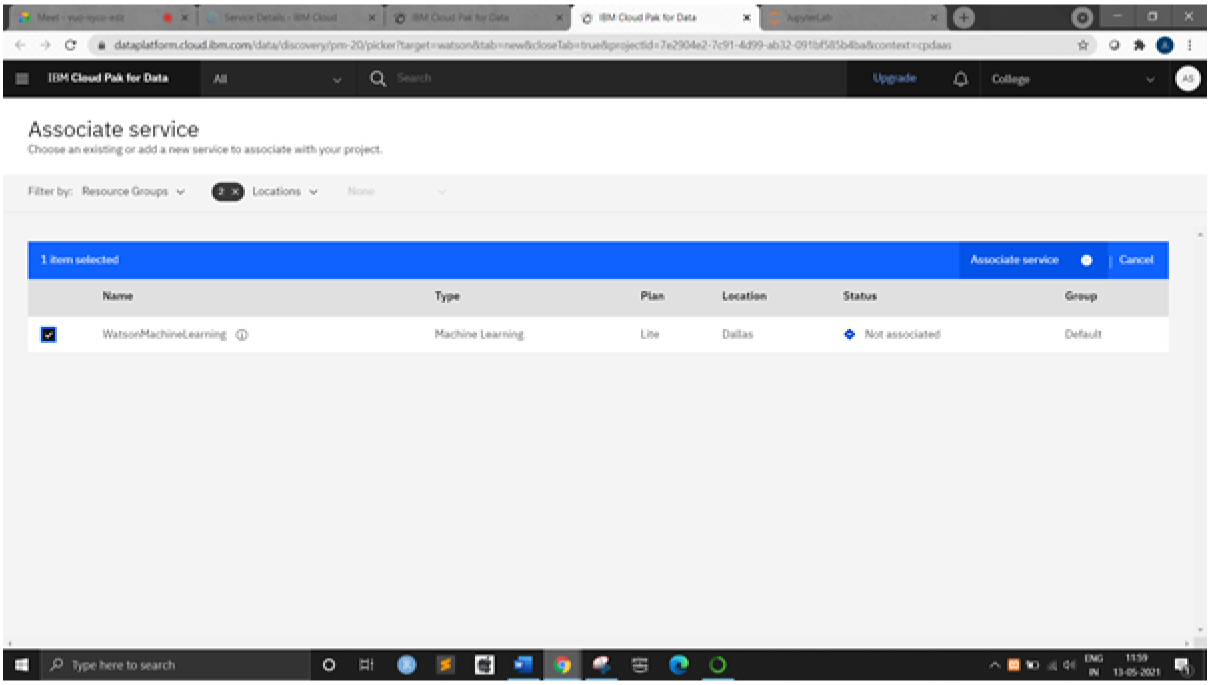
8 vCPU and 32 GB RAM

This compute configuration consumes **20 capacity units per hour** for training. The consumption rate differs for joining multiple data sources, depending on your configuration. For details, see Watson Machine Learning plans.

Cancel

Create

9. Select WatsonMachineLearning Service and click on Associate Service button present at right-corner of the table appearing.



dataplatform.cloud.ibm.com/data/discovery/pm-20/picker?target=watson&tab=new&closeTab=true&projectid=7e2904e2-7c91-4d99-ab32-091bf585b4ba&context=cpdaas

Associate service

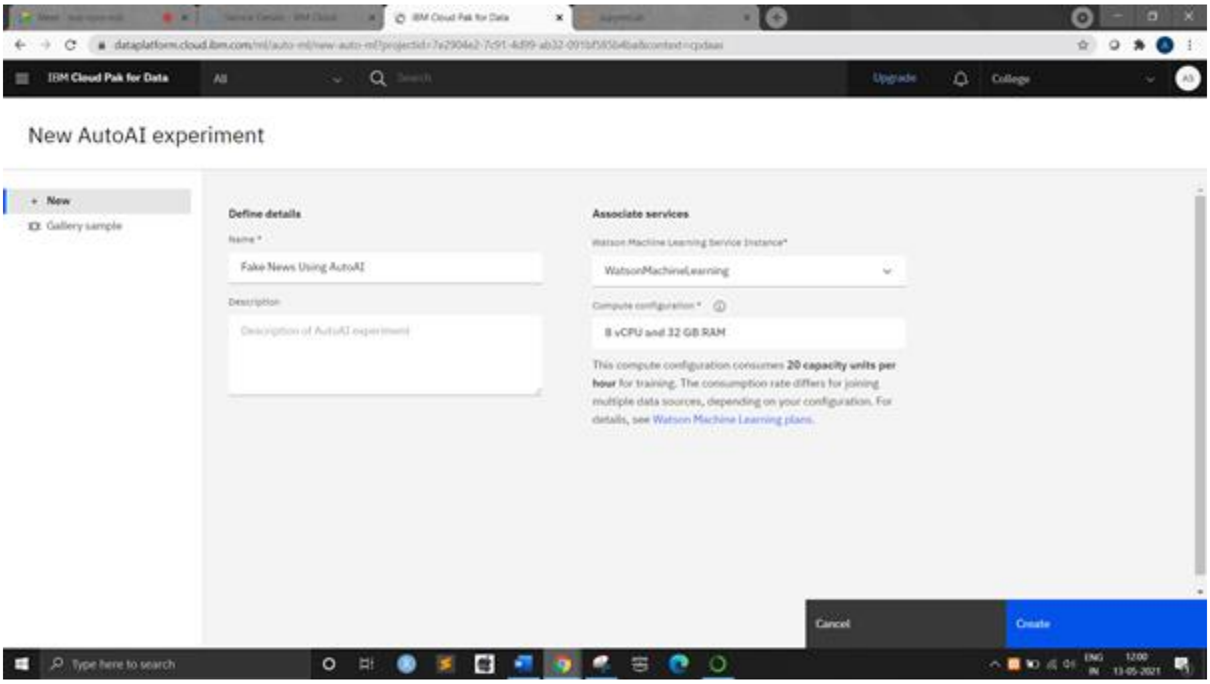
Choose an existing or add a new service to associate with your project.

Filter by: Resource Groups · Locations

1 item selected — Associate service | Cancel

	Name	Type	Plan	Location	Status	Group
☑	WatsonMachineLearning	Machine Learning	Lite	Dallas	Not associated	Default

10. You'll be directed to this page, click on Reload and Associate services will have an option of the given ML service, just click on Create.



dataplatform.cloud.ibm.com/ml/auto-ml/new-auto-ml?projectid=7e2904e2-7c91-4d99-ab32-091bf585b4ba&context=cpdaas

New AutoAI experiment

Define details

Name *: Fake News Using AutoAI

Description: Description of AutoAI experiment

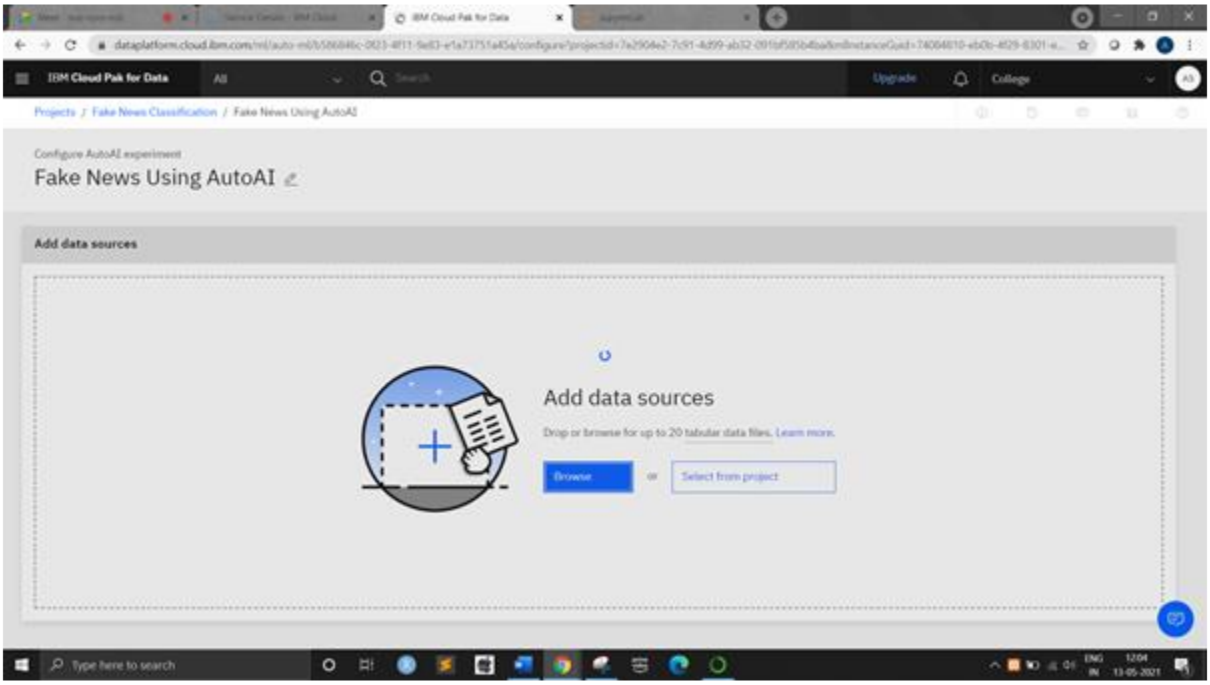
Associate services

Watson Machine Learning Service Instance*: WatsonMachineLearning

Compute configuration *: 8 vCPU and 32 GB RAM

This compute configuration consumes **20 capacity units per hour** for training. The consumption rate differs for joining multiple data sources, depending on your configuration. For details, see Watson Machine Learning plans.

11. Click on Browse and add your dataset there.


Projects / Fake News Classification / Fake News Using AutoAI

Configure AutoAI experiment

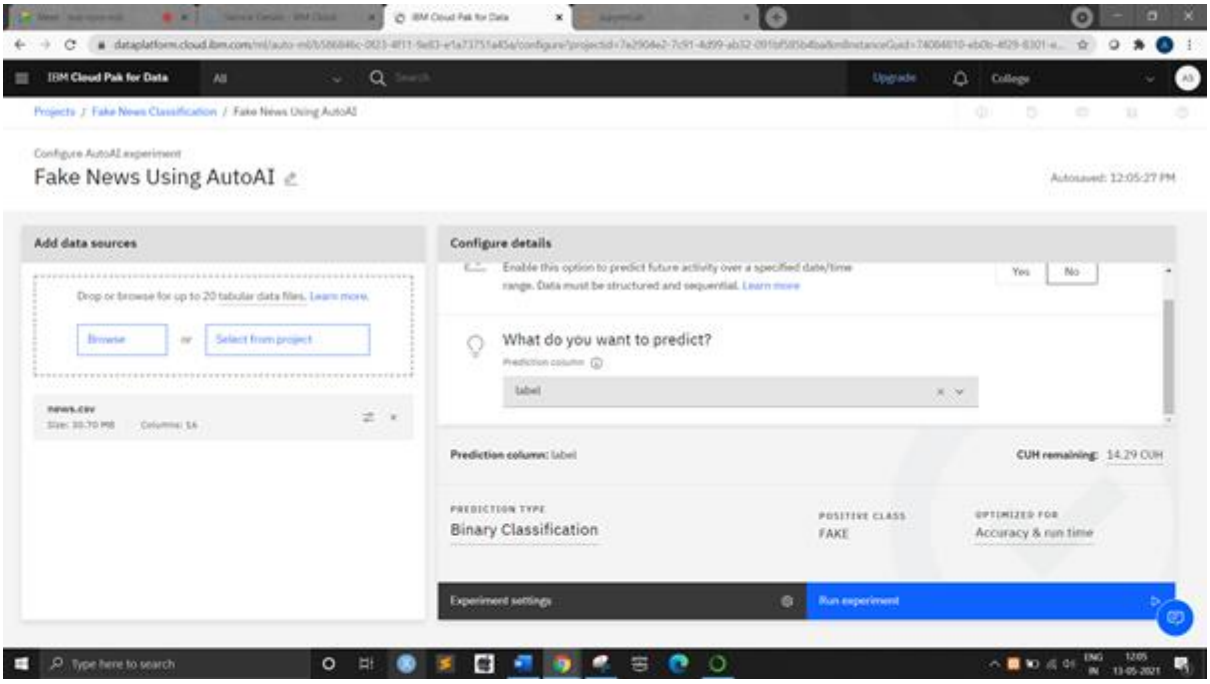
Fake News Using AutoAI

Add data sources

Add data sources

Drop or browse for up to 20 tabular data files. Learn more.

12. Fill in the details required as per your convenience. Here I filled label in What do you want to predict? as we have label to be predicted. Change the Experiment settings if you want and After that click on Run experiment.


Projects / Fake News Classification / Fake News Using AutoAI

Configure AutoAI experiment

Fake News Using AutoAI

Autosaved: 12:05:27 PM

Add data sources

Drop or browse for up to 20 tabular data files. Learn more.

news.csv

Size: 30.70 MB Columns: 16

Configure details

Enable this option to predict future activity over a specified date/time range. Data must be structured and sequential. Learn more

What do you want to predict?

Prediction column

label

Prediction column: label

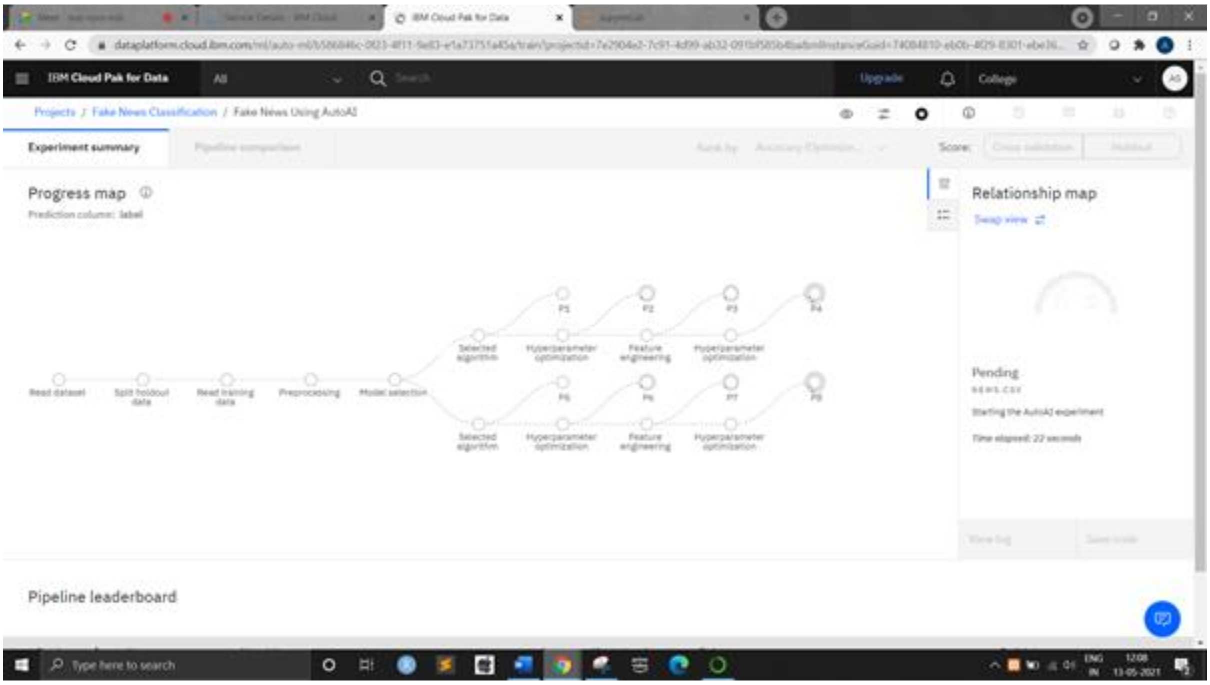
CUH remaining: 14.29 CUH

PREDICTION TYPE: Binary Classification

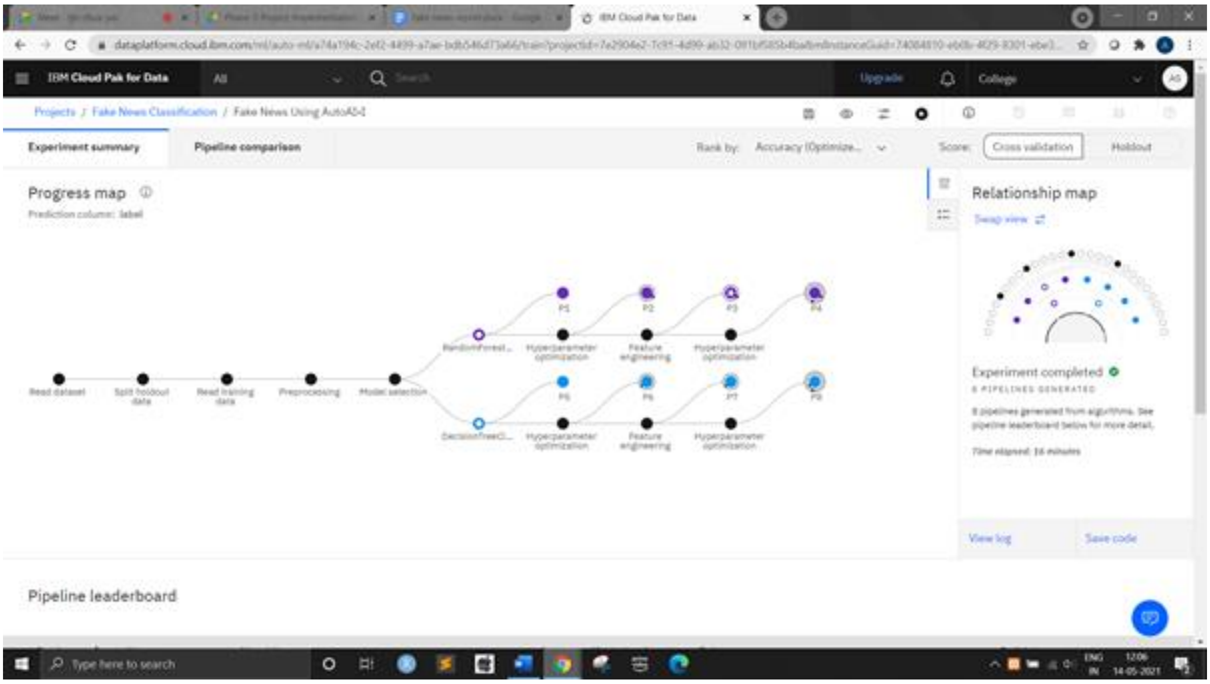
POSITIVE CLASS: FAKE

OPTIMIZED FOR: Accuracy & run time

13. This will start running the experiment. It will take 10-20 minutes to complete successfully.

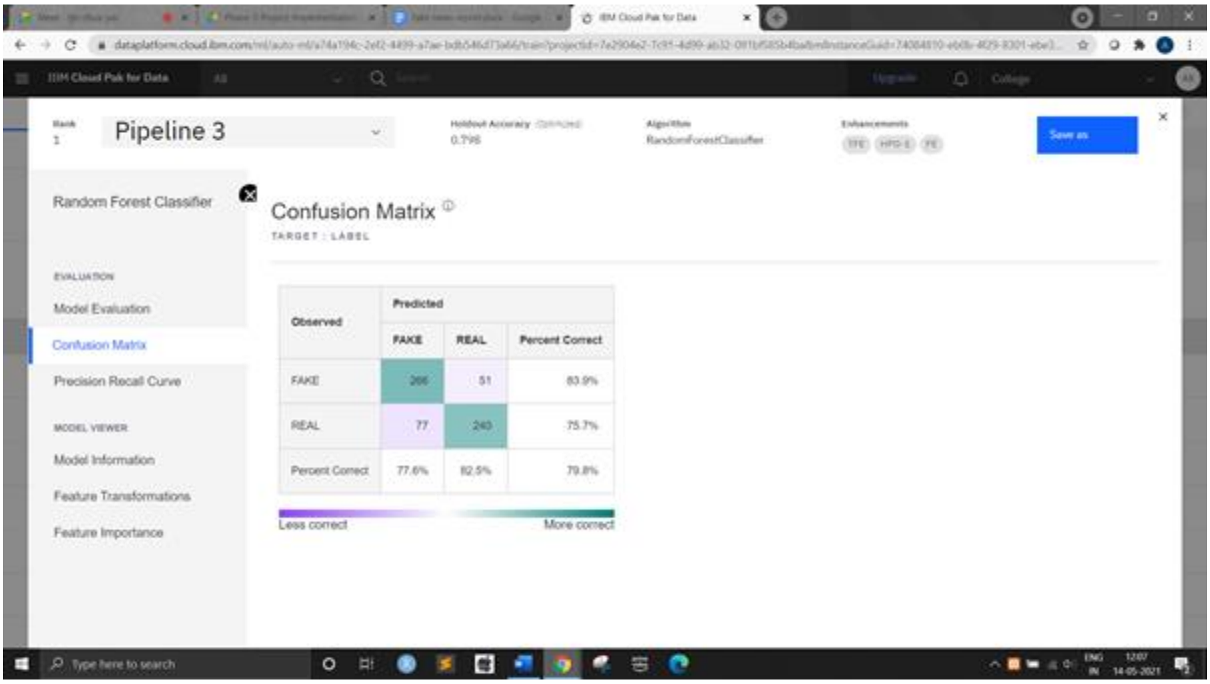


14. The Experiment is now completed and the best pipeline result is displayed in the Pipeline Leaderboard.



15. Clicking on Pipeline 3 will display this. Just simply see all the results and then click on Save as button present in the top-right corner of the page.

16. This is the Confusion Matrix. Click on Save as button present in the top-right corner of the page.



Rank 1 Pipeline 3

Holdout Accuracy (Optimized) 0.798

Algorithm RandomForestClassifier

Enhancements TFE HPO-1 FE

Random Forest Classifier

EVALUATION

Model Evaluation

Confusion Matrix

Precision Recall Curve

MODEL VIEWER

Model Information

Feature Transformations

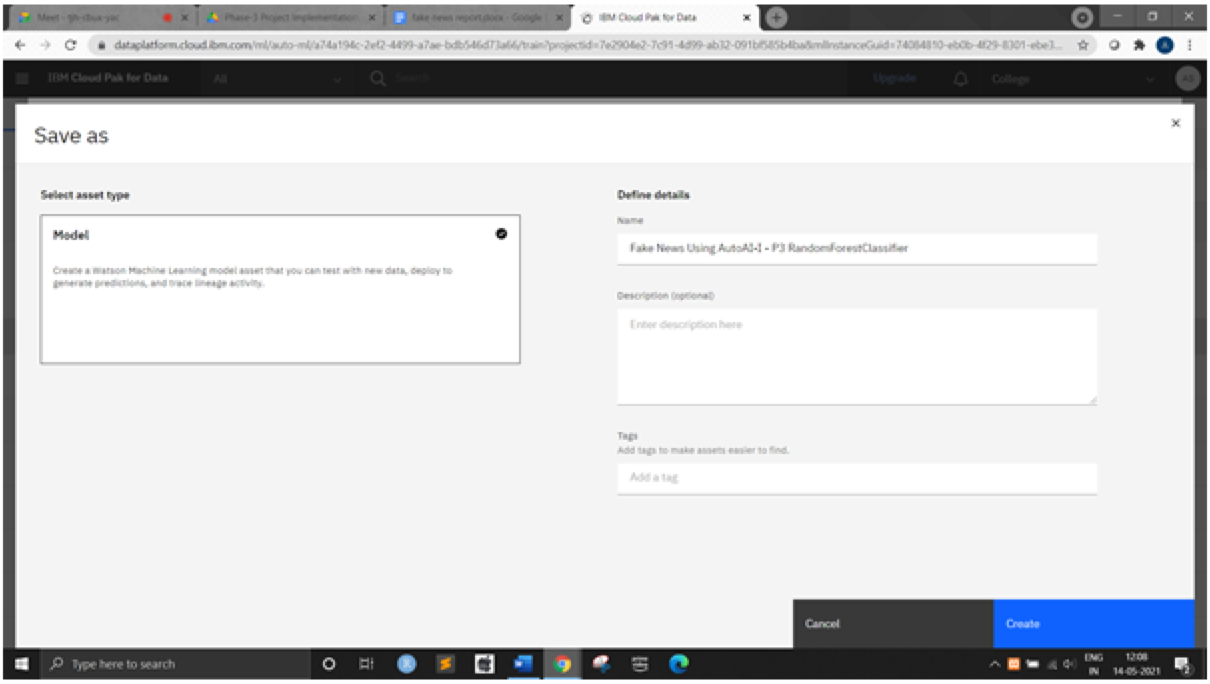
Feature Importance

Confusion Matrix

TARGET : LABEL

Observed	Predicted		
	FAKE	REAL	Percent Correct
FAKE	266	51	83.9%
REAL	77	240	75.7%
Percent Correct	77.6%	82.5%	79.8%

17. Enter all the Define details here and click on Create.



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.

Define details

Name

Fake News Using AutoAI-1 - P3 RandomForestClassifier

Description (optional)

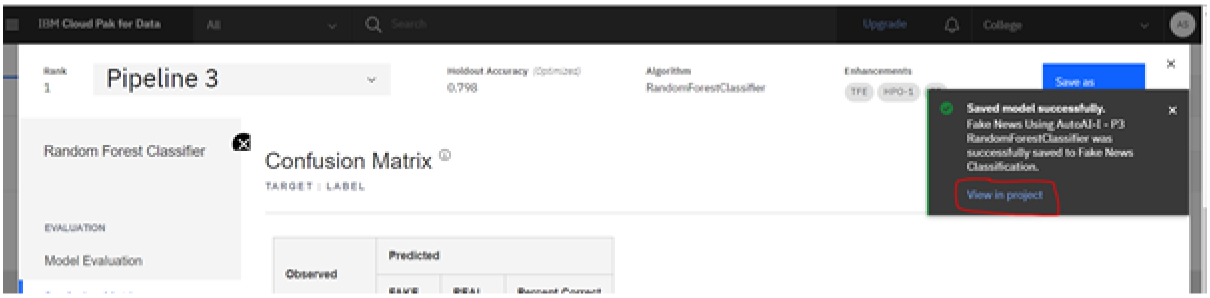
Enter description here

Tags

Add tags to make assets easier to find.

Add a tag

18. Click on View in project.



Rank
1

Pipeline 3

Holdout Accuracy (Optimized)
0.798

Algorithm
RandomForestClassifier

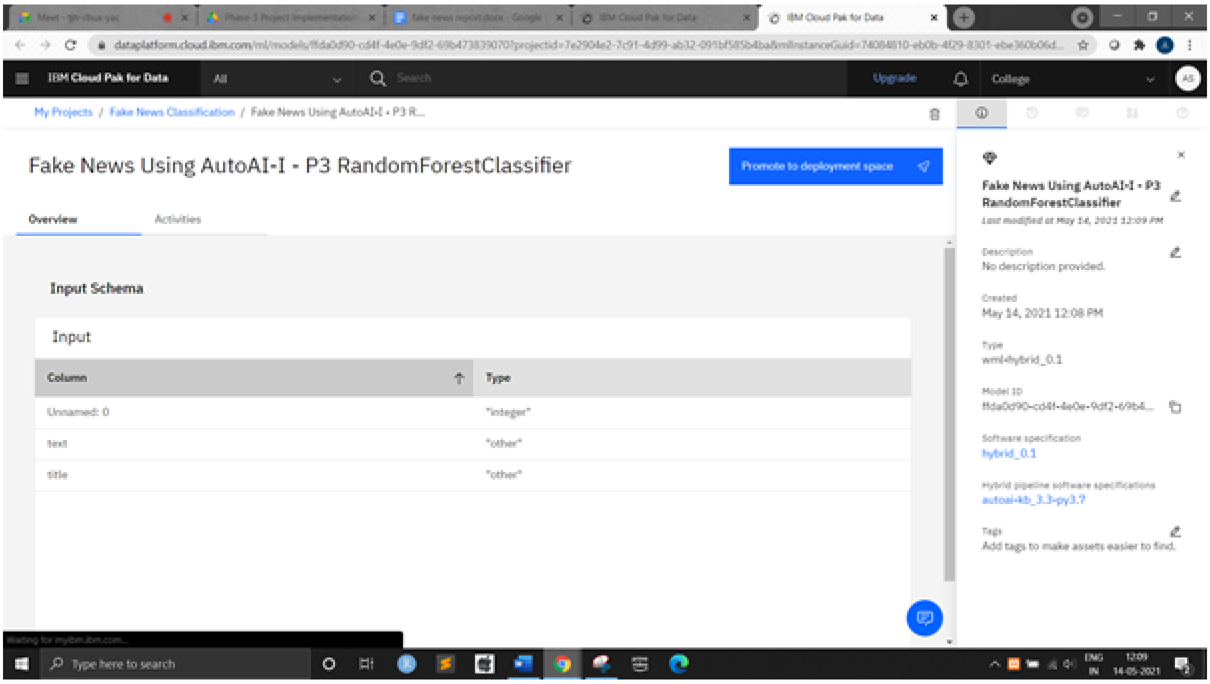
Enhancements
TFE HPO-1

Saved model successfully.
Fake News Using AutoAI-I - P3 RandomForestClassifier was successfully saved to Fake News Classification.
View in project

Confusion Matrix
TARGET : LABEL

Observed	Predicted		
	FAKE	REAL	Percent Correct

19. Click on Promote to deployment space.



dataplatform.cloud.ibm.com/ml/models/ffda0d90-cd4f-4e0e-9df2-69b473839070?projectid=7e2904e2-7c91-4d99-ab32-091bf585b4ba&instanceGuid=74084810-eb0b-4f29-8301-ebe360b06d...

My Projects / Fake News Classification / Fake News Using AutoAI-I - P3 R...

Fake News Using AutoAI-I - P3 RandomForestClassifier

Input Schema

Input

Column	Type
Unnamed: 0	"integer"
text	"other"
title	"other"

Fake News Using AutoAI-I - P3 RandomForestClassifier
Last modified at May 14, 2021 12:09 PM

Description
No description provided.

Created
May 14, 2021 12:08 PM

Type
wml-hybrid_0.1

Model ID
ffda0d90-cd4f-4e0e-9df2-69b4...

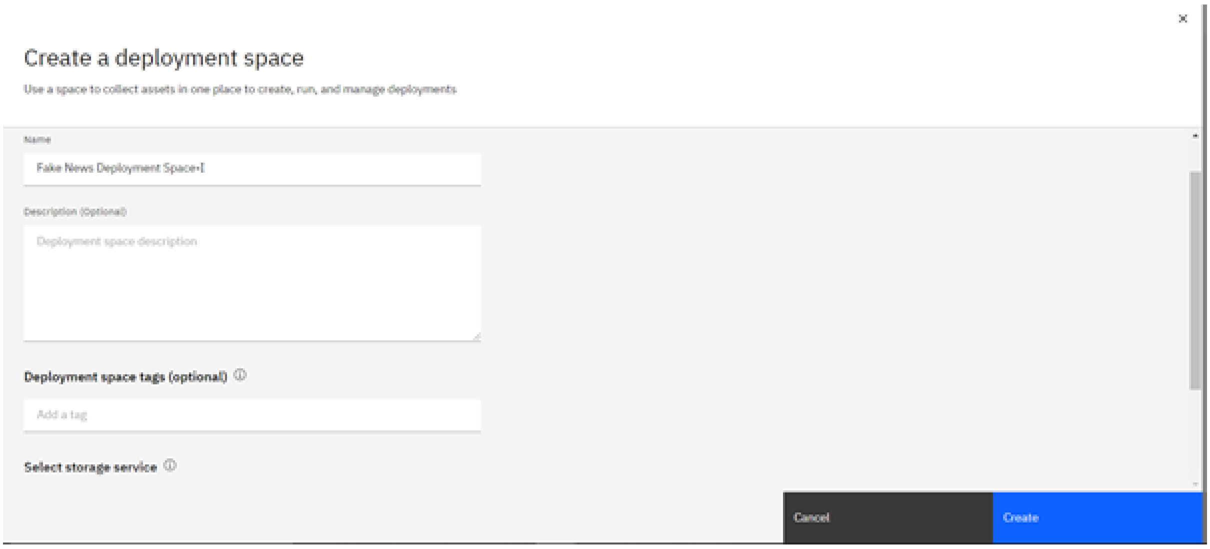
Software specification
hybrid_0.1

Hybrid pipeline software specifications
autoai-kb_3.3-py3.7

Tags
Add tags to make assets easier to find.

Waiting for myibm.ibm.com...

20. Create a deployment space and select storage service of CloudObjectStorage and WatsonMachineLearning space.



Create a deployment space

Use a space to collect assets in one place to create, run, and manage deployments

Name

Fake News Deployment Space-I

Description (Optional)

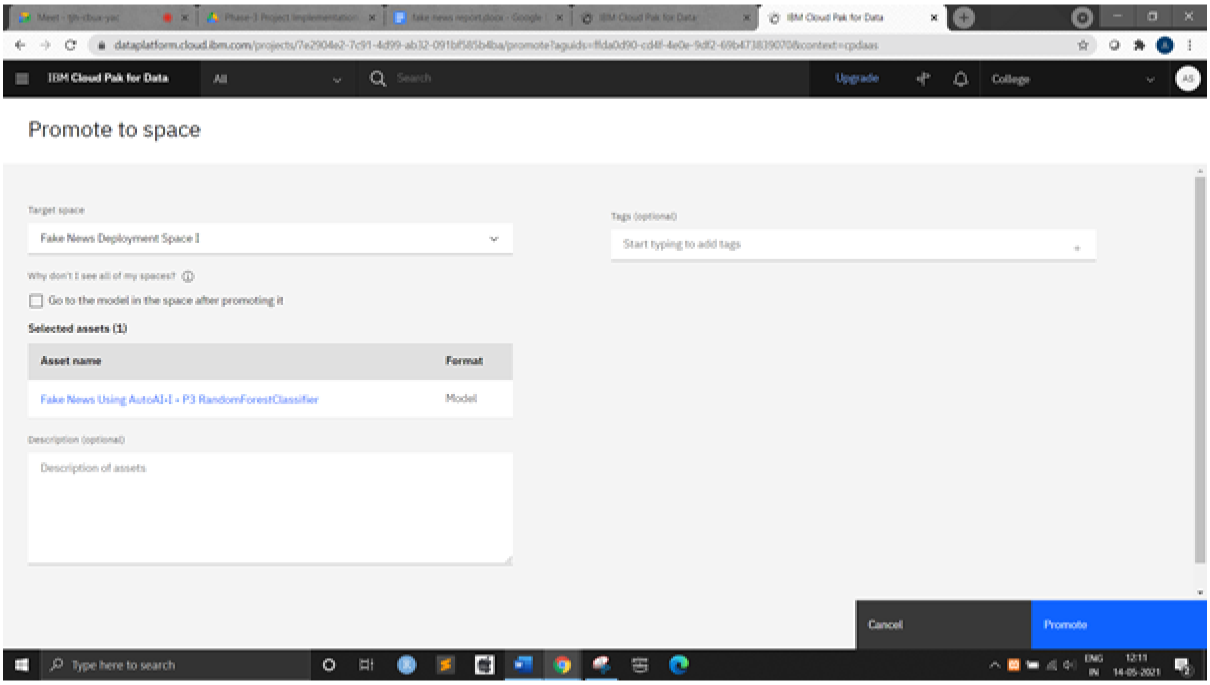
Deployment space description

Deployment space tags (optional)

Add a tag

Select storage service

21. Now click on Promote button present at bottom-right corner.



dataplatform.cloud.ibm.com/projects/7e2904e2-7c91-4d99-ab32-091bf585b4ba/promote?aguids=ffda0d90-cd4f-4e0e-9df2-69b4738390708context=cpdaas

Promote to space

Target space

Fake News Deployment Space I

Why don't I see all of my spaces?

□ Go to the model in the space after promoting it

Selected assets (1)

Asset name	Format
Fake News Using AutoAI-I - P3 RandomForestClassifier	Model

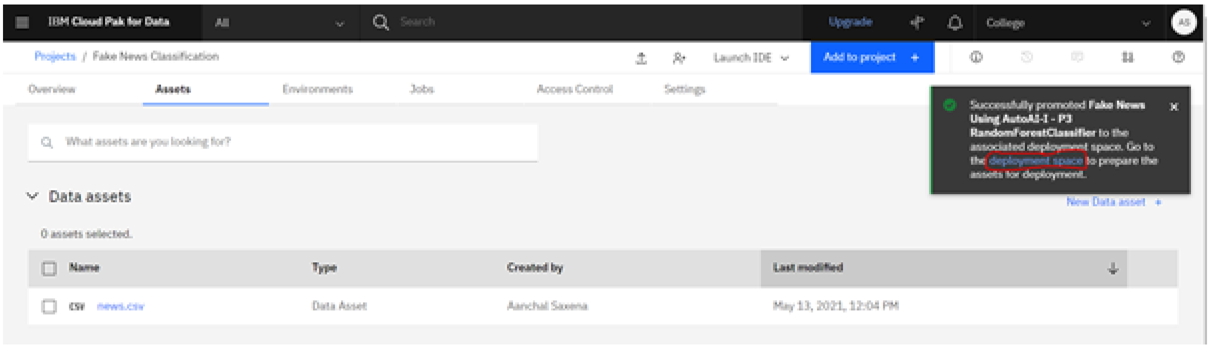
Description (optional)

Description of assets

Tags (optional)

Start typing to add tags

22. Once done, a pop-up will be displayed and simply click on deployment space.



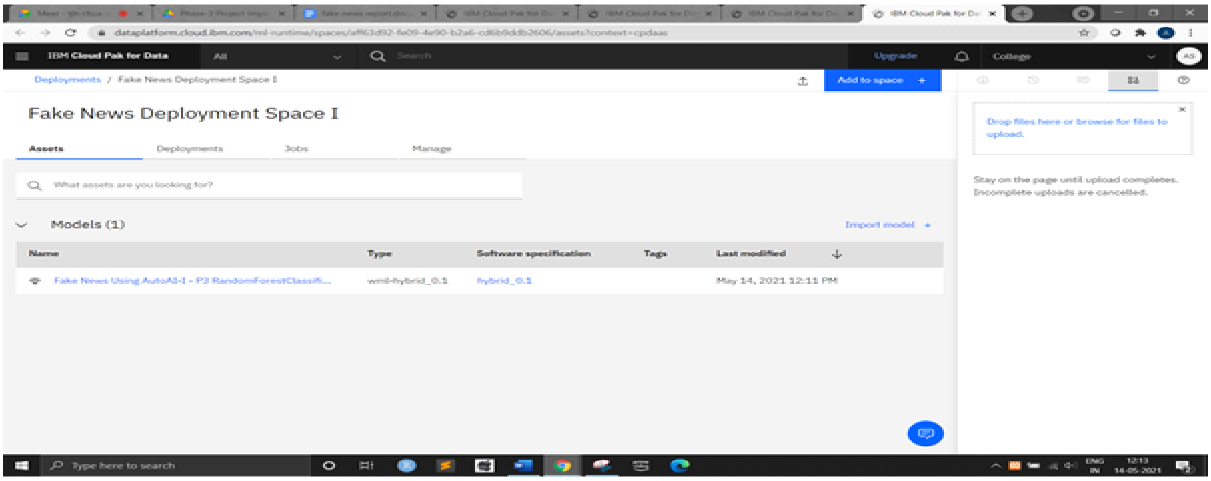
Successfully promoted **Fake News Using AutoAI-I - P3 RandomForestClassifier** to the associated deployment space. Go to the deployment space to prepare the assets for deployment.

Data assets

0 assets selected.

Name	Type	Created by	Last modified
CSV news.csv	Data Asset	Aanchal Saxena	May 13, 2021, 12:04 PM

23. Click on the Model.



Fake News Deployment Space I

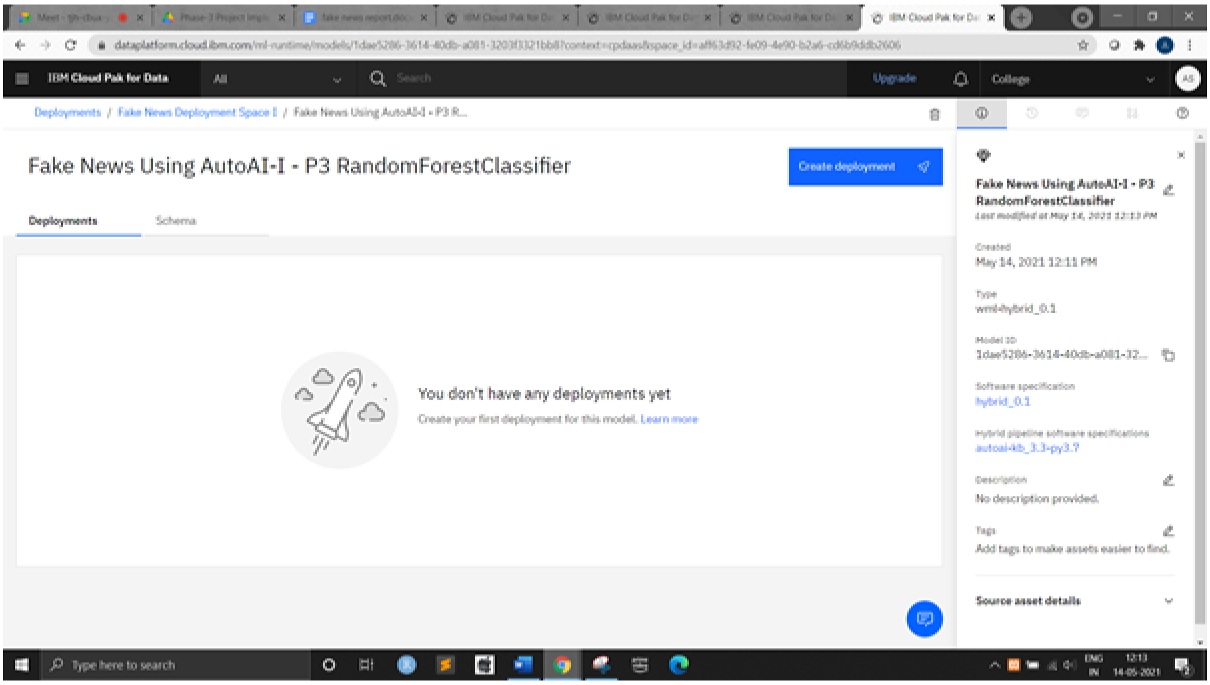
Models (1)

Name	Type	Software specification	Tags	Last modified
Fake News Using AutoAI-I - P3 RandomForestClassifi...	wml-hybrid_0.1	hybrid_0.1		May 14, 2021 12:11 PM

Drop files here or browse for files to upload.

Stay on the page until upload completes. Incomplete uploads are cancelled.

24. Click on Create deployment.


IBM Cloud Pak for Data
Deployments / Fake News Deployment Space I / Fake News Using AutoAI-I - P3 R...
Create deployment
Deployments
Schema
Type here to search

Fake News Using AutoAI-I - P3 RandomForestClassifier

You don't have any deployments yet

Create your first deployment for this model. Learn more

Fake News Using AutoAI-I - P3 RandomForestClassifier
Last modified at May 14, 2021 12:13 PM

Created
May 14, 2021 12:11 PM

Type
wml-hybrid_0.1

Model ID
1dae5286-3614-40db-a081-32...

Software specification
hybrid_0.1

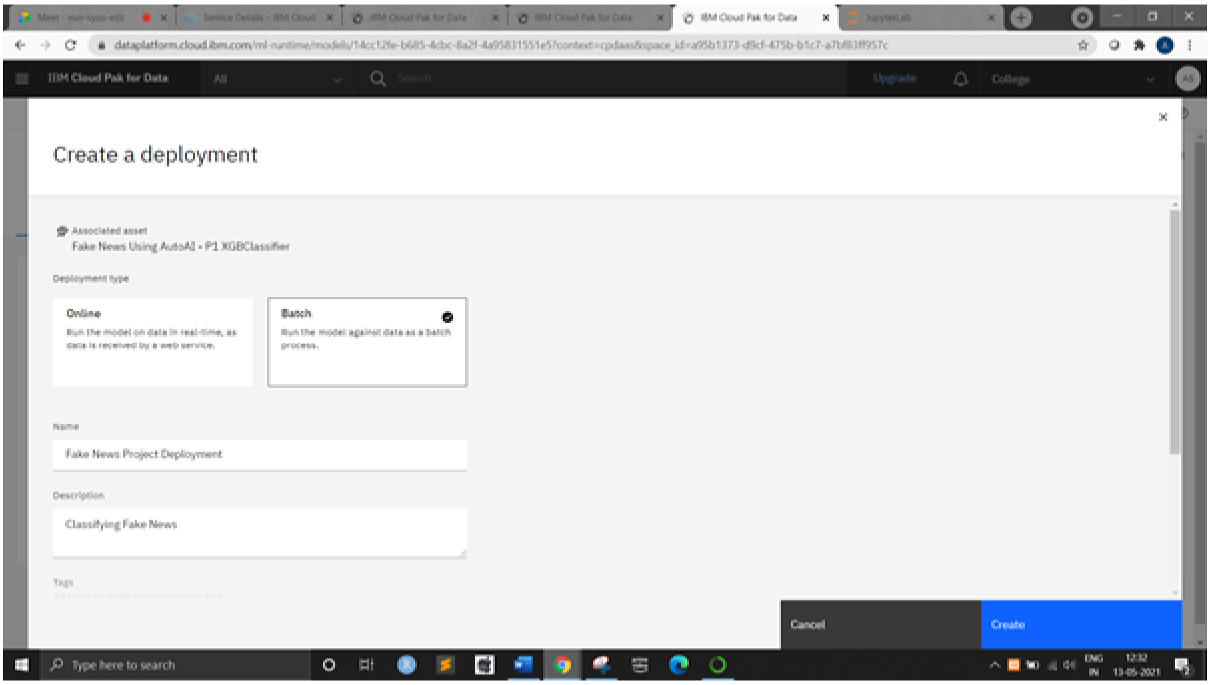
Hybrid pipeline software specifications
autoai-kb_3.3-py3.7

Description
No description provided.

Tags
Add tags to make assets easier to find.

Source asset details

25. Fill in the Details given and click on Create.



Create a deployment

Associated asset
Fake News Using AutoAI - P1 XGBClassifier

Deployment type

○ Online — Run the model on data in real-time, as data is received by a web service.

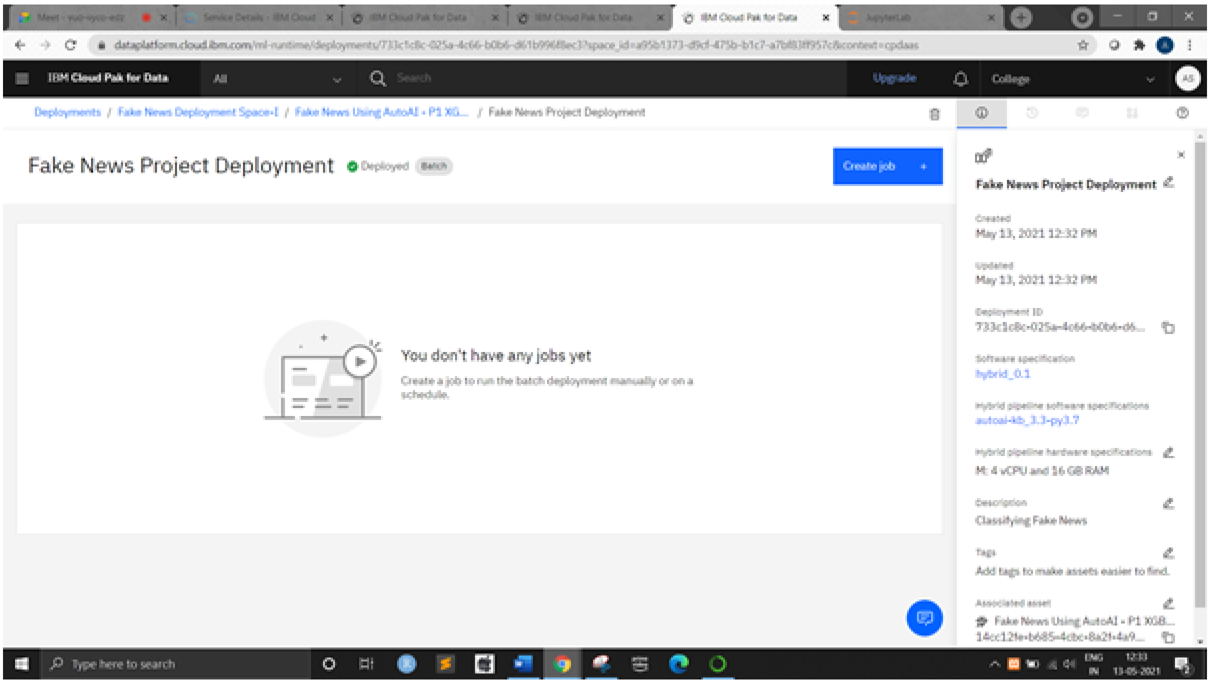
● Batch — Run the model against data as a batch process.

Name: Fake News Project Deployment

Description: Classifying Fake News

Tags

26. The model is deployed. Now we need to create a job by clicking on Create Job button present in the top-right corner of the page.



Deployments / Fake News Deployment Space-I / Fake News Using AutoAI - P1 XG... / Fake News Project Deployment

Fake News Project Deployment ● Deployed Batch

You don't have any jobs yet
Create a job to run the batch deployment manually or on a schedule.

Fake News Project Deployment

Created: May 13, 2021 12:32 PM

Updated: May 13, 2021 12:32 PM

Deployment ID: 733c1c8c-025a-4c66-b0b6-d6...

Software specification: hybrid_0.1

Hybrid pipeline software specifications: autoai-kb_3.3-py3.7

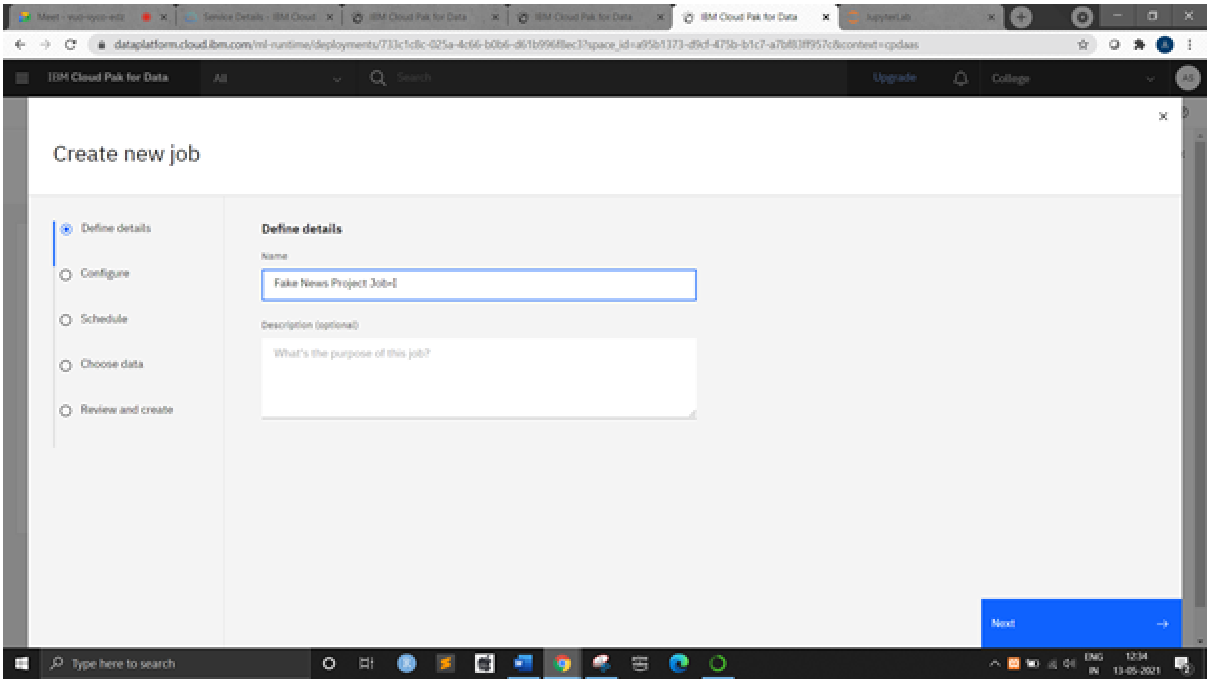
Hybrid pipeline hardware specifications: M: 4 vCPU and 16 GB RAM

Description: Classifying Fake News

Tags: Add tags to make assets easier to find.

Associated asset: Fake News Using AutoAI - P1 XGB... 14cc12fe-b685-4cbc-8a2f-4a9...

27. Fill all the details in the below pages displayed.

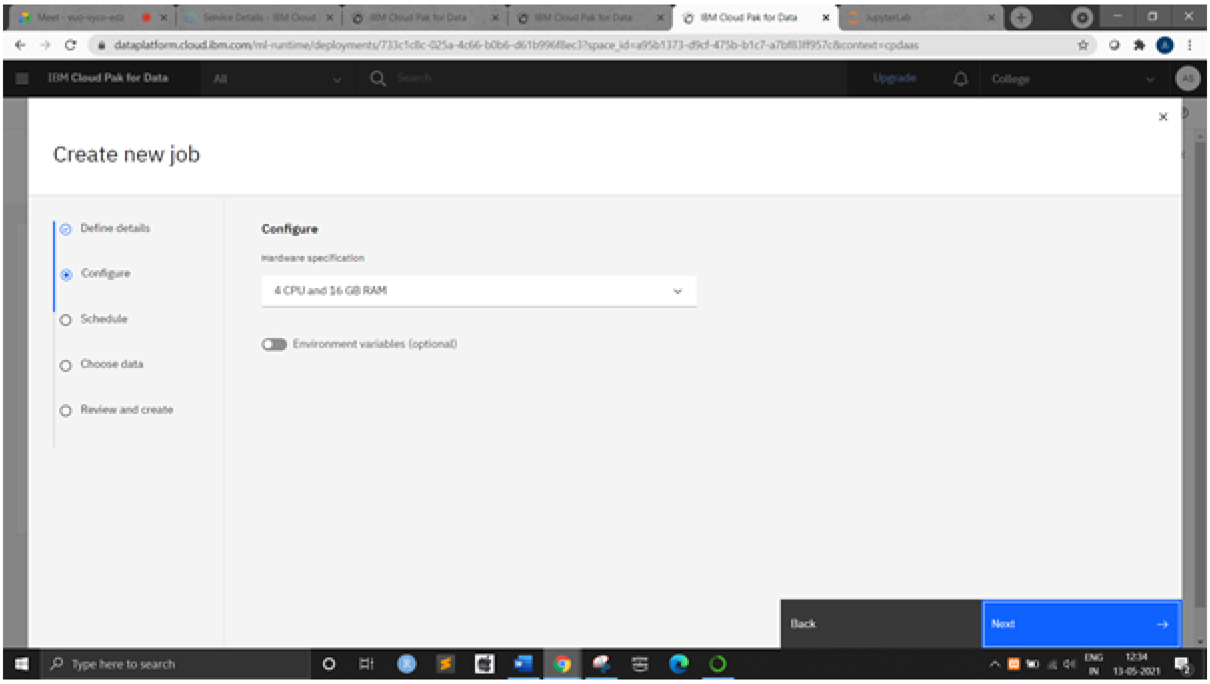


Create new job

Define details

Name: Fake News Project Job-I

Description (optional): What's the purpose of this job?

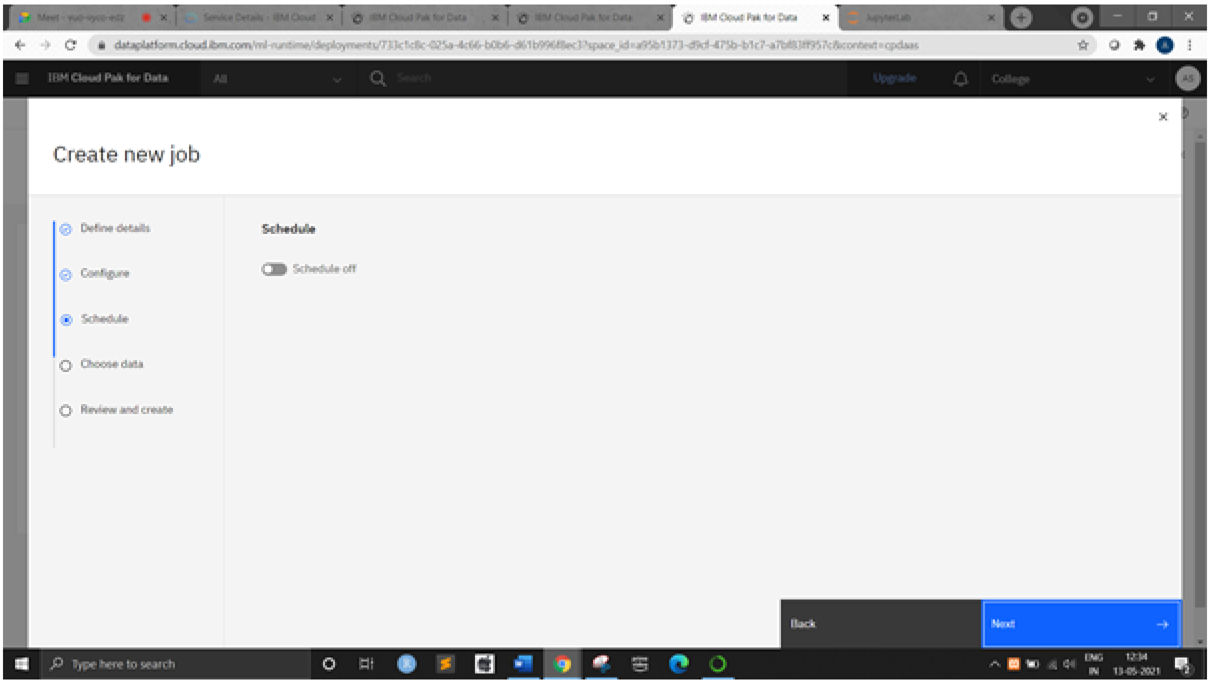


Create new job

Configure

Hardware specification: 4 CPU and 16 GB RAM

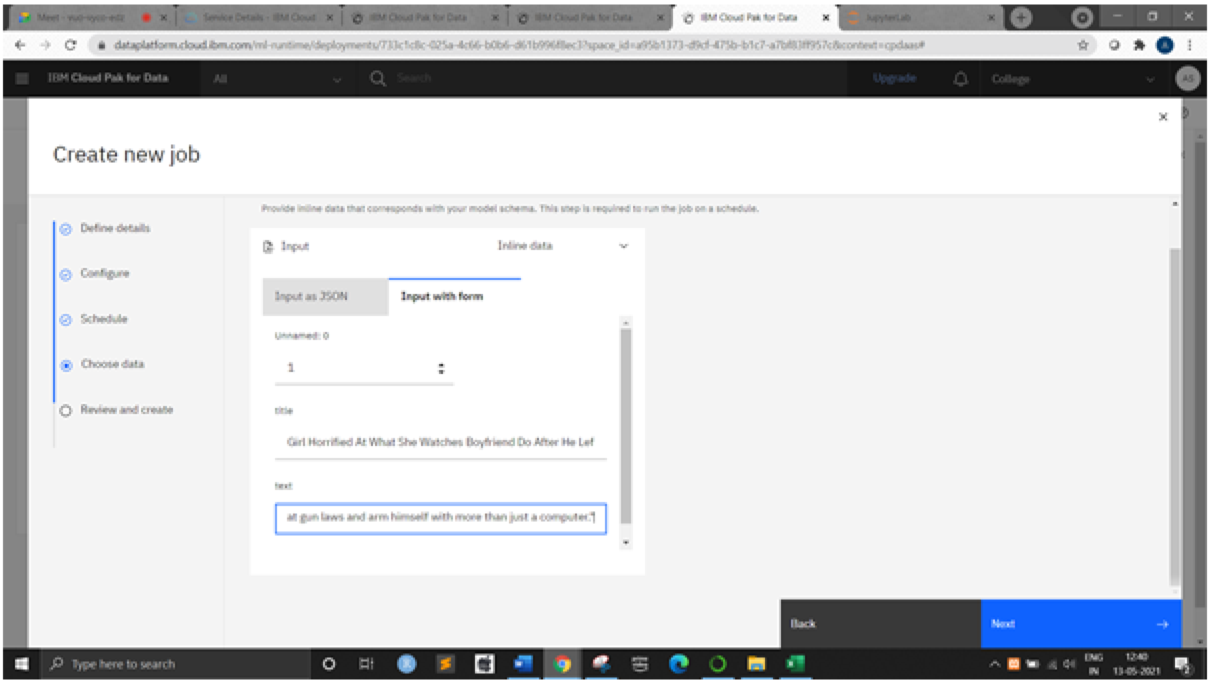
Environment variables (optional)



Create new job

Schedule

Schedule off



Create new job

Provide inline data that corresponds with your model schema. This step is required to run the job on a schedule.

Input: Inline data

Unnamed: 0

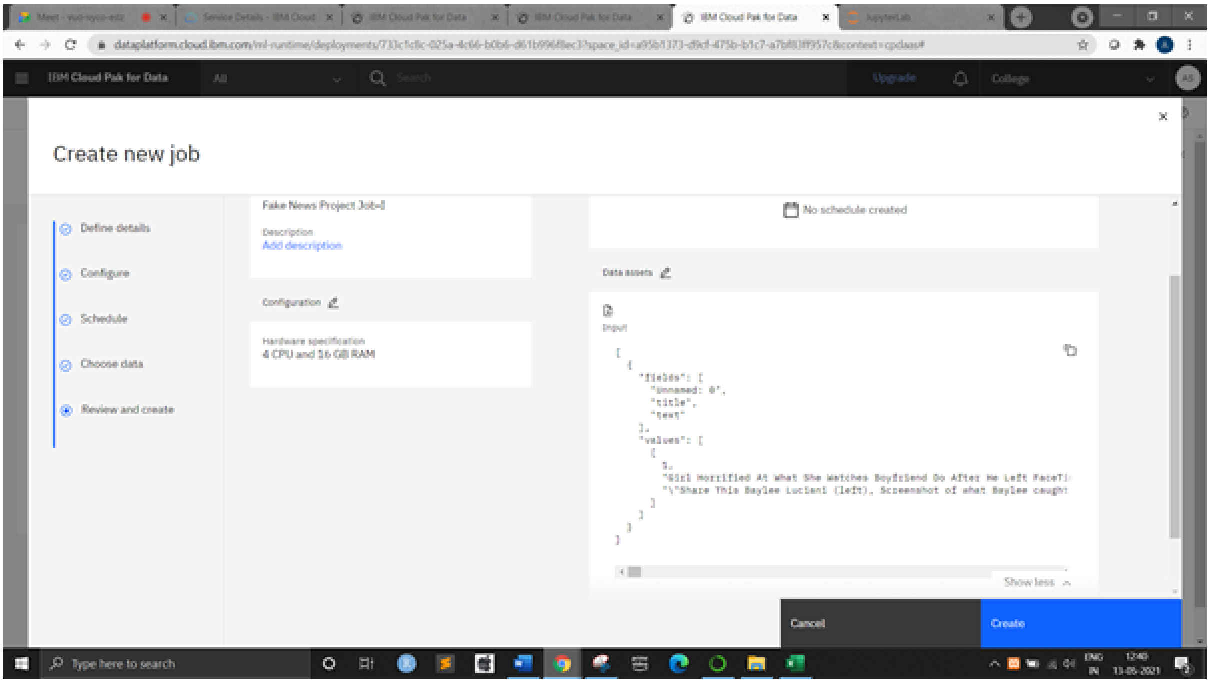
1

title

Girl Horrified At What She Watches Boyfriend Do After He Lef

text

at gun laws and arm himself with more than just a computer."



Create new job

- Define details
- Configure
- Schedule
- Choose data
- Review and create

Fake News Project Job-I

Description
Add description

Configuration

Hardware specification
4 CPU and 16 GB RAM

No schedule created

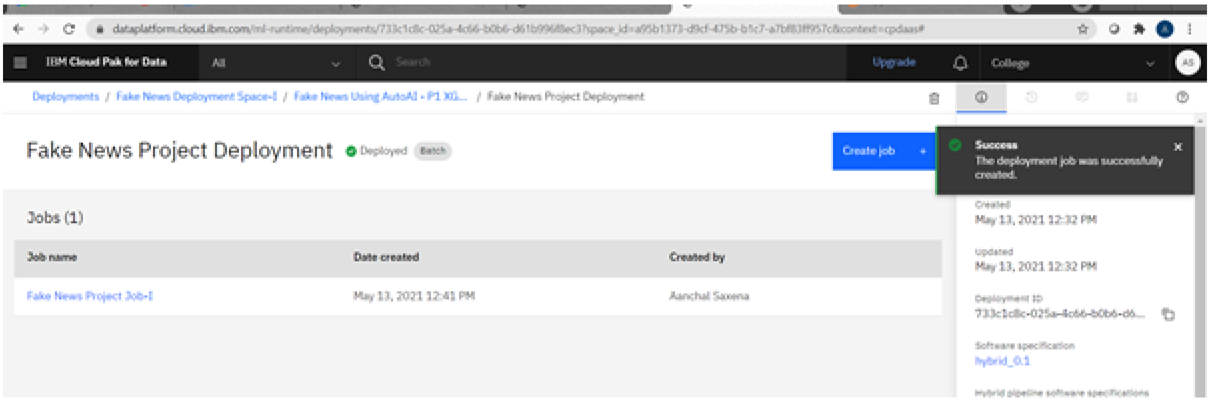
Data assets

Input

```
[
  {
    "fields": [
      "Unnamed: 0",
      "title",
      "text"
    ],
    "values": [
      [
        1,
        "Girl Horrified At What She Watches Boyfriend Do After He Left FaceTi
        "\"Share This Baylee Luciani (left), Screenshot of what Baylee caught
      ]
    ]
  }
]
```

28. Simply click on Create.

29. Once done, the pop-up will be displayed. Just click on the job created.



Deployments / Fake News Deployment Space-I / Fake News Using AutoAI - P1 XG... / Fake News Project Deployment

Fake News Project Deployment ● Deployed Batch

Success
The deployment job was successfully created.

Jobs (1)

Job name	Date created	Created by
Fake News Project Job-I	May 13, 2021 12:41 PM	Aanchal Saxena

Created
May 13, 2021 12:32 PM

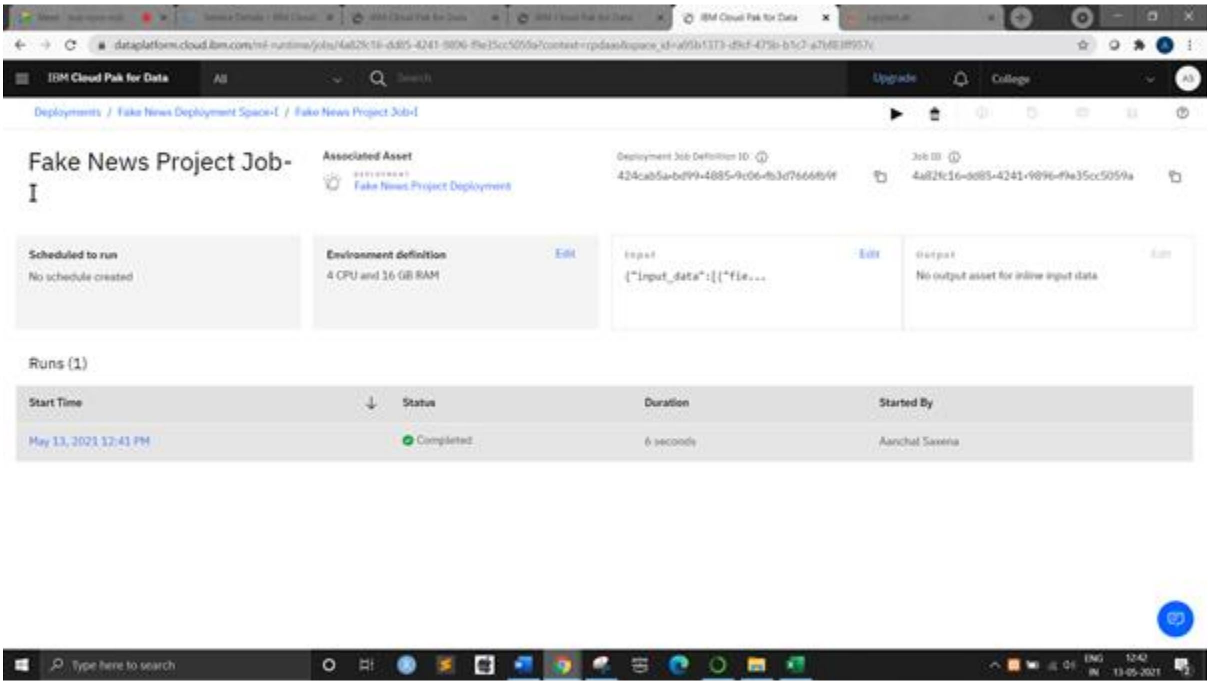
Updated
May 13, 2021 12:32 PM

Deployment ID
733c1c8c-025a-4c66-b0b6-d6...

Software specification
hybrid_0.1

Hybrid pipeline software specifications

30. The job is successfully created and run. Click on the link present in the Runs table.



Deployments / Fake News Deployment Space-I / Fake News Project Job-I

Fake News Project Job-I

Associated Asset: Fake News Project Deployment

Deployment Job Definition ID: 424cab5a-bd99-4885-9c06-fb3d7666fb9f

Job ID: 4a82fc16-dd85-4241-9896-f9e35cc5059a

Scheduled to run: No schedule created

Environment definition: 4 CPU and 16 GB RAM

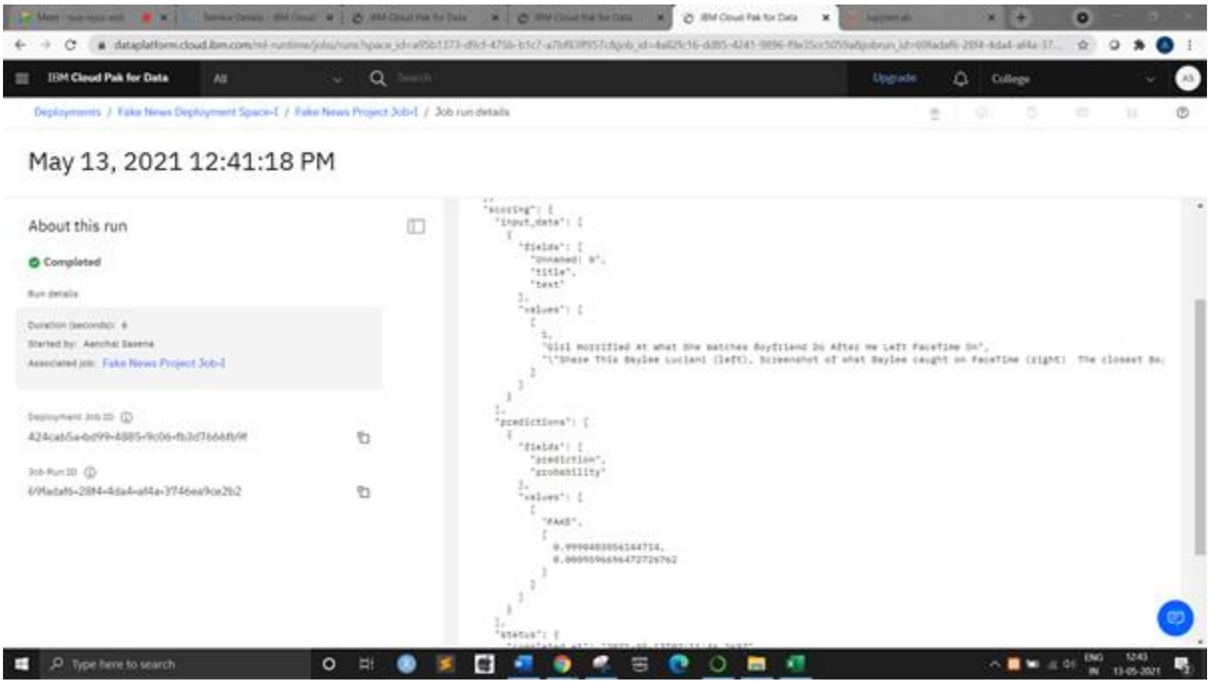
Input: {"input_data":[{"fie...

Output: No output asset for inline input data

Runs (1)

Start Time	Status	Duration	Started By
May 13, 2021 12:41 PM	Completed	6 seconds	Aanchal Saxena

31. This is the Job run detail and here we input the values which is displayed in values Json form and prediction is given just below it which predicted that given input news is 99.99 % Fake and 0.001 % Real. Hence, our model is successfully implemented.



Deployments / Fake News Deployment Space-I / Fake News Project Job-I / Job run details

May 13, 2021 12:41:18 PM

About this run

Completed

Run details

Duration (seconds): 6

Started by: Aanchal Saxena

Associated job: Fake News Project Job-I

Deployment Job ID: 424cab5a-bd99-4885-9c06-fb3d7666fb9f

Job Run ID: 69fadaf6-28f4-4da4-af4a-3746ea9ce2b2

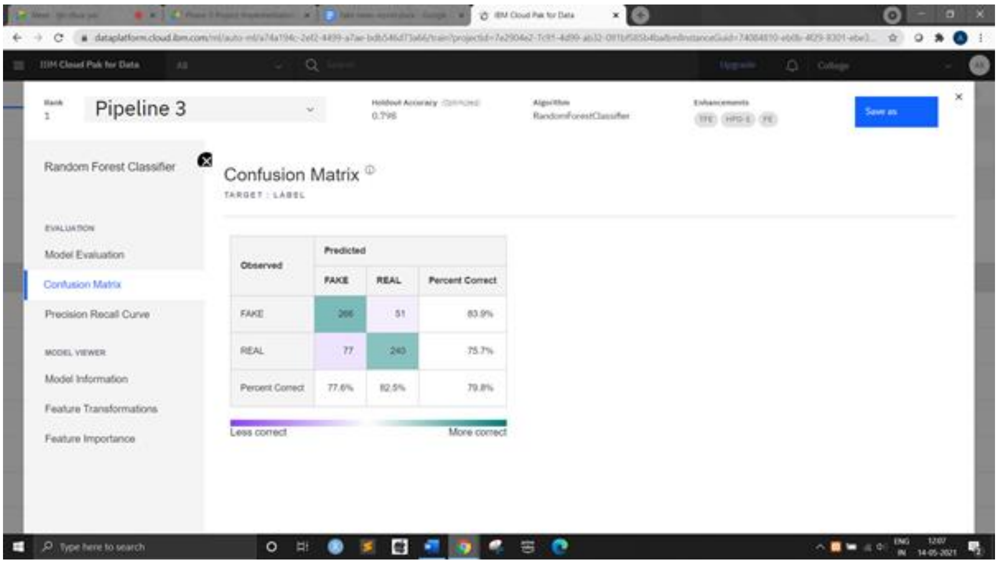
```
"scoring": {
  "input_data": [
    {
      "fields": [
        "Unnamed: 0",
        "title",
        "text"
      ],
      "values": [
        [
          1,
          "Girl Horrified At What She Watches Boyfriend Do After He Left FaceTime On",
          "\"Share This Baylee Luciani (left), Screenshot of what Baylee caught on FaceTime (right)  The closest Ba
        ]
      ]
    }
  ],
  "predictions": [
    {
      "fields": [
        "prediction",
        "probability"
      ],
      "values": [
        [
          "FAKE",
          [
            0.9999403856144714,
            0.000059669647272676 2
          ]
        ]
      ]
    }
  ],
  "status": {
```

32. This is the log that was displayed there.

```
{
  "deployment": {
    "id": "733c1c8c-025a-4c66-b0b6-d61b996f8ec3"
  },
  "platform_job": {
    "job_id": "4a82fc16-dd85-4241-9896-f9e35cc5059a",
    "run_id": "69fadaf6-28f4-4da4-af4a-3746ea9ce2b2"
  },
  "scoring": {
    "input_data": [
      {
        "fields": [
          "Unnamed: 0",
          "title",
          "text"
        ],
        "values": [
          [
            1,
            "Girl Horrified At What She Watches Boyfriend Do After He Left FaceTime On",
            "\"Share This Baylee Luciani (left), Screenshot of what Baylee caught on FaceTime (right)  The closest Baylee Luciani could get to her boyfriend, whoâ€™s attending college...\""
          ]
        ]
      }
    ],
    "predictions": [
      {
        "fields": [
          "prediction",
          "probability"
        ],
        "values": [
          [
            "FAKE",
            [
              0.9990403056144714,
              0.0009596696472726762
            ]
          ]
        ]
      }
    ],
    "status": {
      "completed_at": "2021-05-13T07:11:46.269Z",
      "running_at": "2021-05-13T07:11:41.328Z",
      "state": "completed"
    }
  }
}
```

5. COMPARING COMPARISON MATRIX OF CODE AND IBM WATSON:

Here, as we can see here is the confusion matrix generated using **IBM Watson Studio AutoAI**



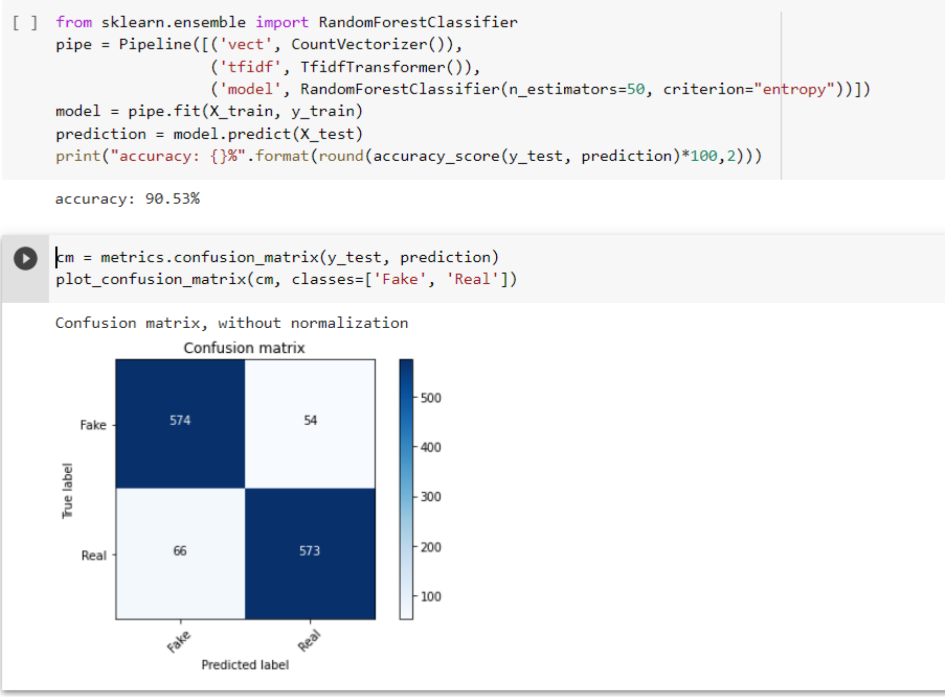
Confusion Matrix

TARGET : LABEL

Observed	Predicted		
	FAKE	REAL	Percent Correct
FAKE	266	51	83.9%
REAL	77	240	75.7%
Percent Correct	77.6%	82.5%	79.8%

Less correct More correct

V/S the one generated using **Python Coding**



```
from sklearn.ensemble import RandomForestClassifier
pipe = Pipeline([('vect', CountVectorizer()),
                 ('tfidf', TfidfTransformer()),
                 ('model', RandomForestClassifier(n_estimators=50, criterion="entropy"))])
model = pipe.fit(X_train, y_train)
prediction = model.predict(X_test)
print("accuracy: {}%".format(round(accuracy_score(y_test, prediction)*100,2)))
```

```
accuracy: 90.53%
```

```
cm = metrics.confusion_matrix(y_test, prediction)
plot_confusion_matrix(cm, classes=['Fake', 'Real'])
```

```
Confusion matrix, without normalization
```

6. INDIVIDUAL CONTRIBUTION:

Yashika Singal: Helped in searching the content for the project online and contributed in writing the code and performing experiment in IBM Watson Studio.

Aanchal Saxena: Worked with the model-training part while coding in Python and contributed in performing the practical implementation of the project in IBM Watson Studio. Guided the team to proceed with the work.

Uchita Gupta: Searched for the dataset online. Helped in writing the description of the project and contributed in the visualization part in coding and job creation part in IBM Watson Studio.

7. CONCLUSION:

We classified news as FAKE and REAL using 2 models: DecisionTreeClassifier and RandomForestClassifier. And we coded it in Python in Jupyter Lab and practically implemented it using IBM Watson Studio AutoAI feature.

After deploying the model in IBM Watson Studio, we created a job and predicted the news by inputting the values in the input form and we got the predictions as **FAKE 99.999%** which was an accurate result.

```
"FAKE",
 [
   0.9990403056144714,
   0.0009596696472726762
  ]
```