SPRINT – 2 PROJECT DOCUMENT

Date	13 November 2022
Team ID	PNT2022TMID15779
Project Name	Flight Delay Prediction Using Machine Learning

DEVELOPMENT PHASE:

SPRINT-2:

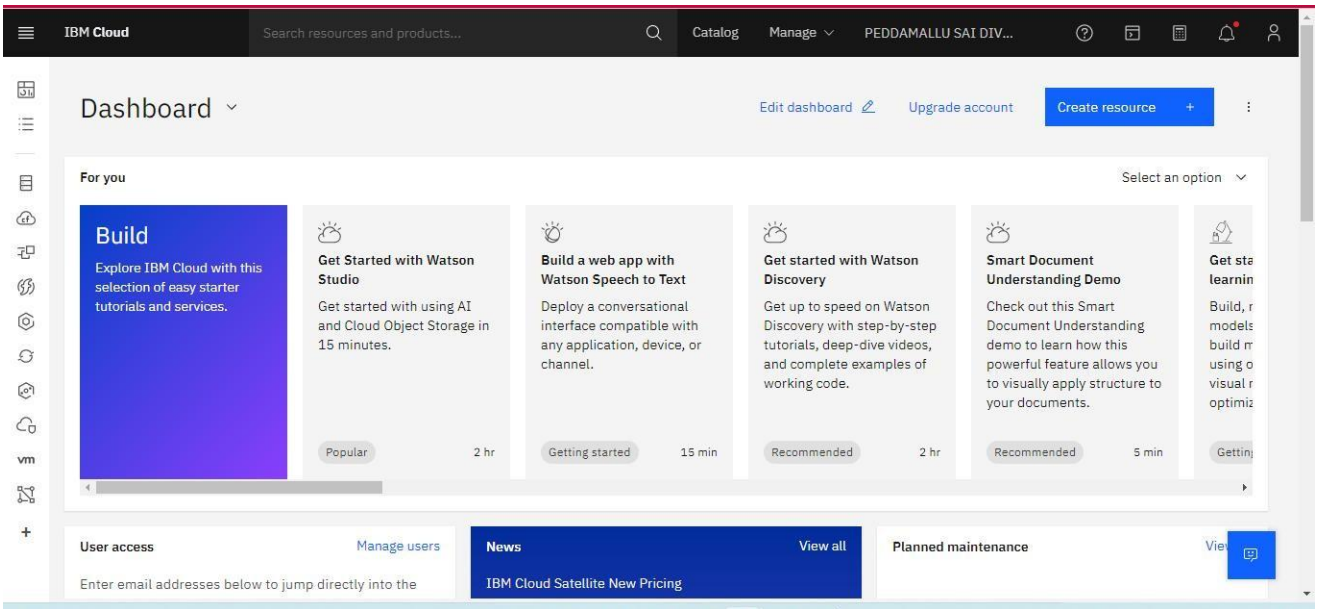
- Creating IBM cloud account & Required Resources
- Importing jupyter notebook file in ibm watson
- Deploy our model in IBM Watson
- Predict the result

Creating IBM cloud account & Required Resources:

Creating IBM cloud account:

Frist, need to create IBM Cloud account by using SI Mail Id and SI Password which is provided by IBM in profile.

Below dashboard of an account after created,



Creating IBM Cloud Required Resources:

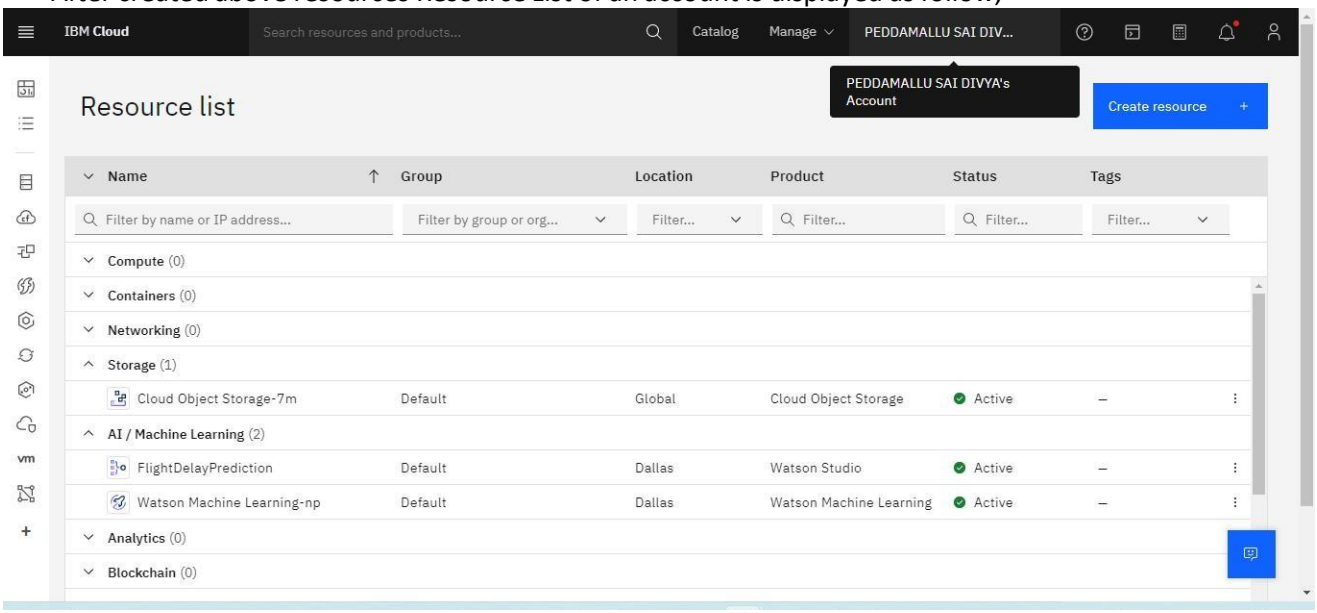
After creating IBM cloud account, to deploy ML model, need to create following resources such as,

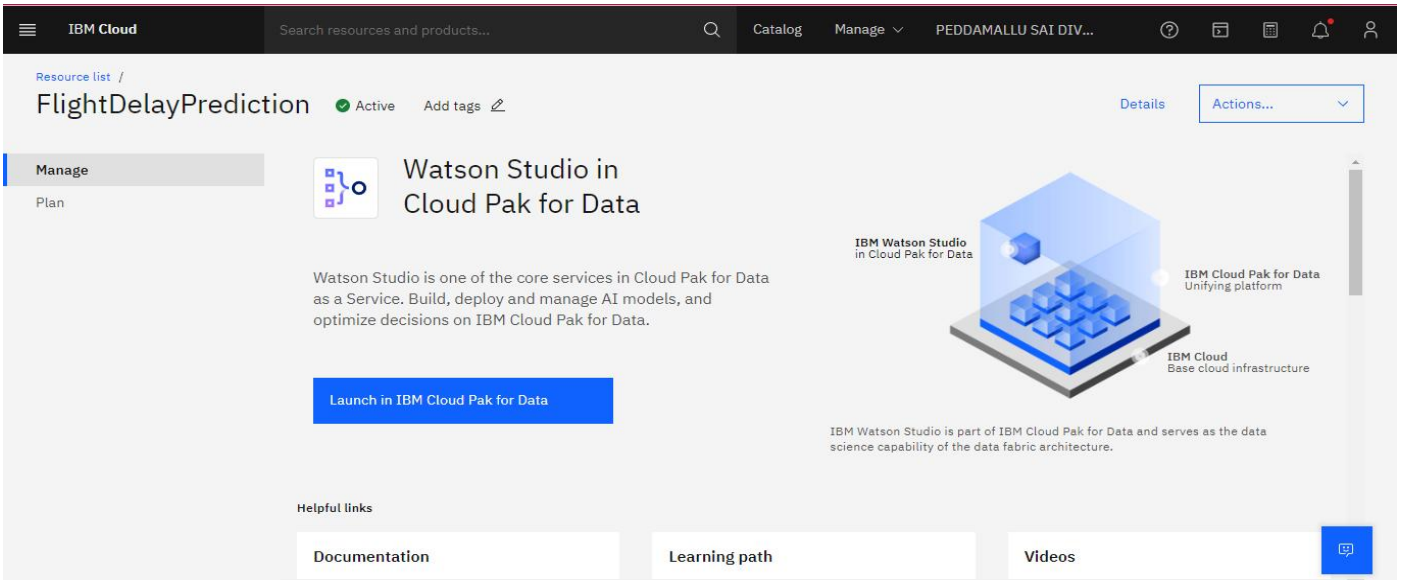
Cloud Object Storage

Watson Machine Learning

Watson Studio

After created above resources Resource List of an account is displayed as follow,

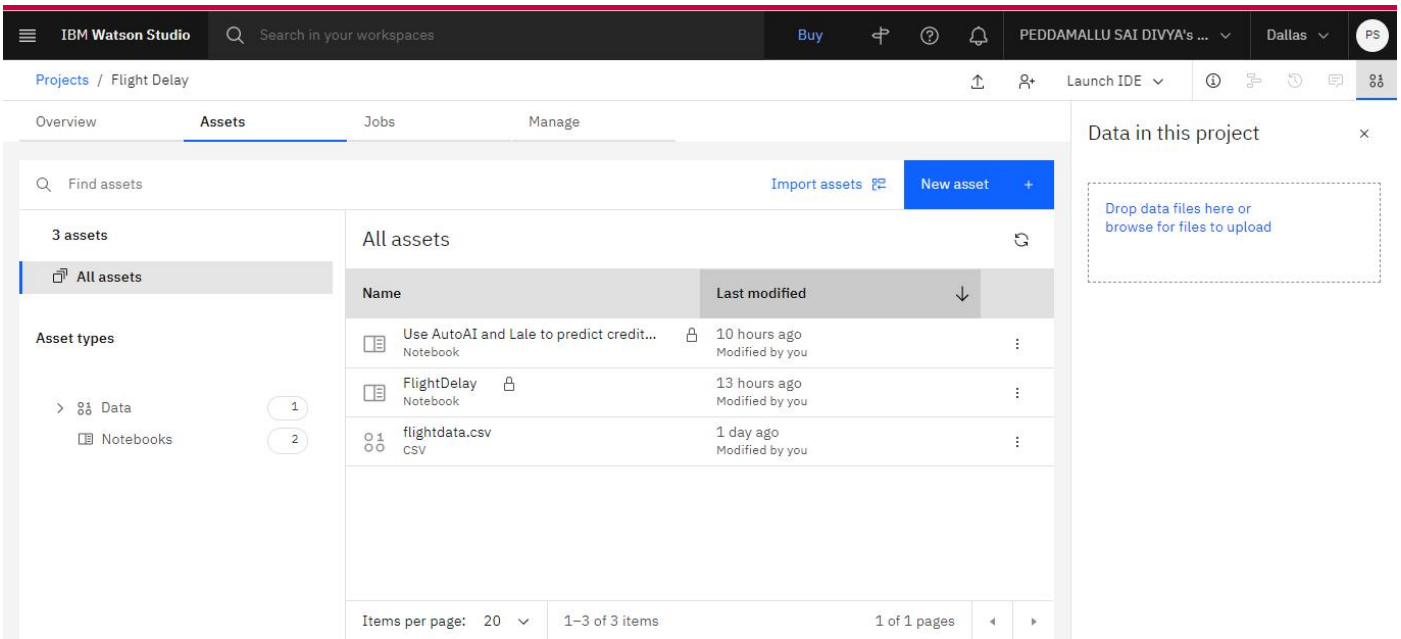




All the resource are in active state.

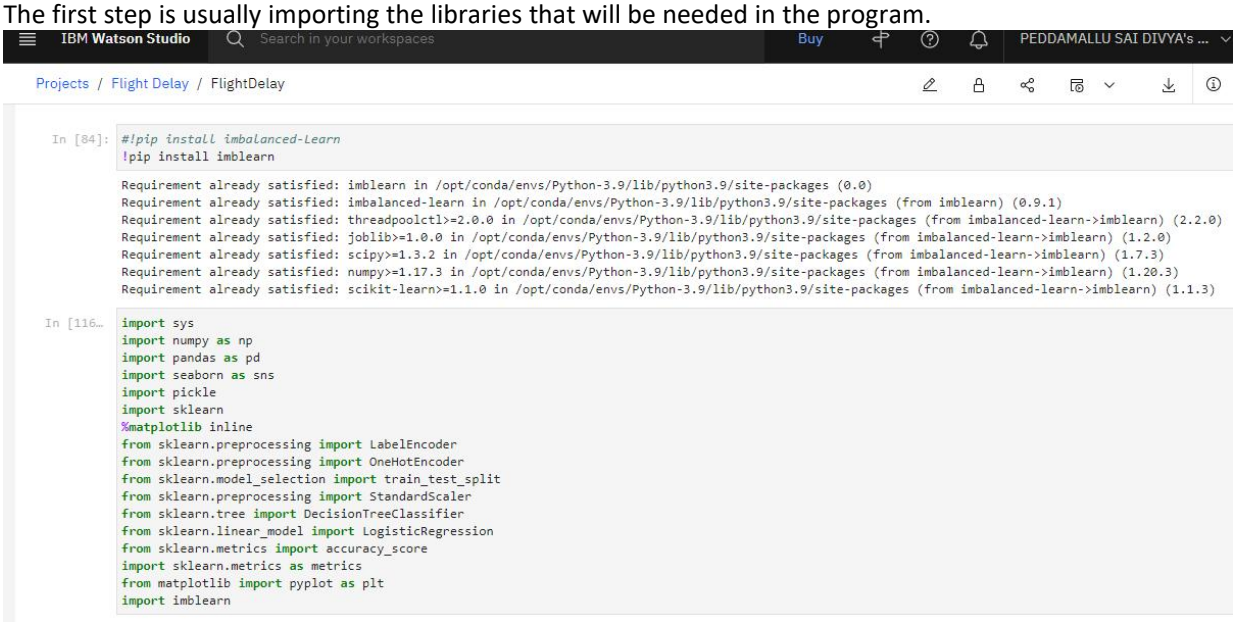
All the required cloud resources are created successfully.

Import .ipynb file of sprint-1 which ML models are build in Jupyter notebook.

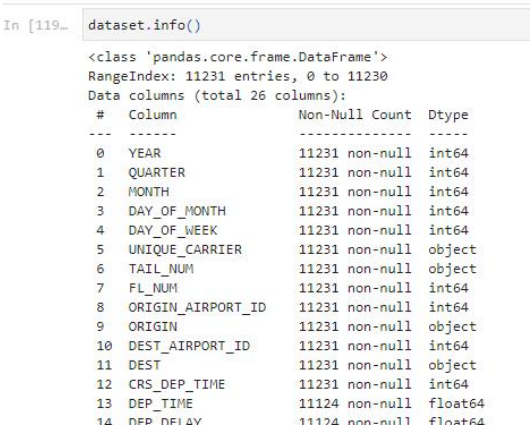


Import Required Libraries

The first step is usually importing the libraries that will be needed in the program.



Importing The Dataset and Analyze The Data set



Handling Missing Values

IBM Watson Studio | Search in your workspaces

Projects / Flight Delay / FlightDelay

In [121…
```
dataset.isnull().sum()
```

Out[121]:
```
YEAR                     0
QUARTER                  0
MONTH                    0
DAY_OF_MONTH             0
DAY_OF_WEEK              0
UNIQUE_CARRIER           0
TAIL_NUM                 0
FL_NUM                   0
ORIGIN_AIRPORT_ID        0
ORIGIN                   0
DEST_AIRPORT_ID          0
DEST                     0
CRS_DEP_TIME             0
DEP_TIME               107
DEP_DELAY              107
DEP_DEL15              107
CRS_ARR_TIME             0
ARR_TIME               115
ARR_DELAY              188
ARR_DEL15              188
CANCELLED                0
DIVERTED                 0
CRS_ELAPSED_TIME         0
```

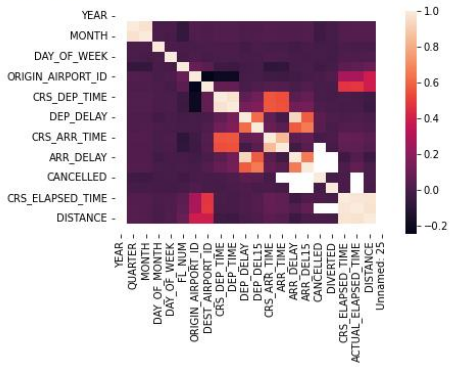
Data Visualization

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ... | Dallas | PS

Projects / Flight Delay / FlightDelay

In [124…
```
#sns.cataplot(x='ARR_DELAY',y='ARR_DEL15', data=flightdata)
sns.heatmap(dataset.corr())
```

Out[124]: <AxesSubplot:>



Dropping Un-Necessary Columns

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ... | Dallas | PS

Projects / Flight Delay / FlightDelay

In [125…
```
dataset = dataset.drop('Unnamed: 25', axis=1)
```

In [126…
```
dataset.isnull().sum()
```

Out[126]:
```
YEAR                     0
QUARTER                  0
MONTH                    0
DAY_OF_MONTH             0
DAY_OF_WEEK              0
UNIQUE_CARRIER           0
TAIL_NUM                 0
FL_NUM                   0
ORIGIN_AIRPORT_ID        0
ORIGIN                   0
DEST_AIRPORT_ID          0
DEST                     0
CRS_DEP_TIME             0
DEP_TIME               107
DEP_DELAY              107
DEP_DEL15              107
CRS_ARR_TIME             0
ARR_TIME               115
ARR_DELAY              188
ARR_DEL15              188
CANCELLED                0
DIVERTED                 0
CRS_ELAPSED_TIME         0
```

Label Encoding & One Hot Encoding

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ... | Dallas | PS

Projects / Flight Delay / FlightDelay

```
dtype: int64
```

In [128…
```
dataset = dataset.fillna({'ARR_DEL15':1})
dataset = dataset.fillna({'DEP_DEL15':0})
dataset.iloc[177:185]
```

Out[128]:

	FL_NUM	MONTH	DAY_OF_MONTH	DAY_OF_WEEK	ORIGIN	DEST	CRS_ARR_TIME	DEP_DEL15	ARR_DEL15
177	2834	1	9	6	MSP	SEA	852	0.0	1.0
178	2839	1	9	6	DTW	JFK	1724	0.0	0.0
179	86	1	10	7	MSP	DTW	1632	0.0	1.0
180	87	1	10	7	DTW	MSP	1649	1.0	0.0
181	423	1	10	7	JFK	ATL	1600	0.0	0.0
182	440	1	10	7	JFK	ATL	849	0.0	0.0
183	485	1	10	7	JFK	SEA	1945	1.0	0.0
184	557	1	10	7	MSP	DTW	912	0.0	1.0

In [129…
```
import math

for index, row in dataset.iterrows():
    dataset.loc[index,'CRS_ARR_TIME'] = math.floor(row['CRS_ARR_TIME']/100)
dataset.head()
```

Splitting The Dataset Into Dependent And Independent Variables

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ... | Da

Projects / Flight Delay / FlightDelay

In [174…
```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)
```

In [175…
```
!pip install imblearn
```

```
Requirement already satisfied: imblearn in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (0.0)
Requirement already satisfied: imbalanced-learn in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from imblearn) (0.9.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from imbalanced-learn->imblearn) (2.2.0)
Requirement already satisfied: scikit-learn>=1.1.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from imbalanced-learn->imblearn) (1.1.3)
Requirement already satisfied: scipy>=1.3.2 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from imbalanced-learn->imblearn) (1.7.3)
Requirement already satisfied: joblib>=1.0.0 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from imbalanced-learn->imblearn) (1.2.0)
Requirement already satisfied: numpy>=1.17.3 in /opt/conda/envs/Python-3.9/lib/python3.9/site-packages (from imbalanced-learn->imblearn) (1.20.3)
```

In [176…
```
from imblearn.over_sampling import SMOTE
smote = SMOTE()
```

In [177…
```
x_train_smote, y_train_smote = smote.fit_resample(x_train, y_train)
```

In [178…
```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(random_state = 0)
classifier.fit(x_train_smote,y_train_smote)
```

Out[178]:
```
▾ DecisionTreeClassifier
DecisionTreeClassifier(random_state=0)
```

Splitting The Dataset Into Dependent And Independent Variables

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ... | Dallas | PS

Projects / Flight Delay / FlightDelay

```
In [179…  decisiontree = classifier.predict(x_test)

In [180…  from sklearn.metrics import accuracy_score
          acc = accuracy_score(y_test,decisiontree)
          acc

Out[180]: 1.0

In [181…  from sklearn.metrics import confusion_matrix
          cm = confusion_matrix(y_test,decisiontree)

In [182…  cm

Out[182]: array([[1938,    0],
                 [   0,  309]])

In [183…  import pickle
          pickle.dump(classifier,open("flight.pkl","wb"))

In [157…  pwd

Out[157]: '/home/wsuser/work'
```

Train And Test The Model Using Decision Tree Classifier

IBM Watson Studio | Search in your workspaces | Buy

Projects / Flight Delay / FlightDelay

```
In [179…  decisiontree = classifier.predict(x_test)

In [180…  from sklearn.metrics import accuracy_score
          acc = accuracy_score(y_test,decisiontree)
          acc

Out[180]: 1.0

In [181…  from sklearn.metrics import confusion_matrix
          cm = confusion_matrix(y_test,decisiontree)

In [182…  cm
```


Model Evaluation


```
In [180…  from sklearn.metrics import accuracy_score
          acc = accuracy_score(y_test,decisiontree)
          acc

Out[180]: 1.0

In [181…  from sklearn.metrics import confusion_matrix
          cm = confusion_matrix(y_test,decisiontree)

In [182…  cm
```


Deploy our model in IBM Watson:

To deploy ML model in IBM cloud, need to create project in IBM Watson. After successful creation of project import .ipynb file of sprint-1 which ML models are build in Jupyter notebook.

Upload required datasets and import it.

Deploy model using following code,

```
!pip install -U ibm-watson-machine-learning
```

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ...

Projects / Flight Delay / FlightDelay

Authenticate and set Space

In [160…
```
wml_credentials = {
    "url":"https://us-south.ml.cloud.ibm.com",
    "apikey":"S2icgoOGNZUt5Wkoxztgx1yJ6JHZQkJVBm3Opy1NUCuu"
}
```

In [161…
```
wml_client=APIClient(wml_credentials)
wml_client.spaces.list()
```

```
Note: 'limit' is not provided. Only first 50 records will be displayed if the number of records exceed 50
------------------------------------  ------  ------------------------
ID                                    NAME    CREATED
b6783a9b-4513-4465-b052-a15d738cf69d  flight  2022-11-16T18:19:00.423Z
------------------------------------  ------  ------------------------
```

In [162…
```
space_id="b6783a9b-4513-4465-b052-a15d738cf69d"
```

In [163…
```
wml_client.set.default_space(space_id)
```

Out[163]: 'SUCCESS'

In [164…
```
wml_client.software_specifications.list()
```

```
----------------------------  ------------------------------------  ----
NAME                          ASSET_ID                              TYPE
default_py3.6                 0062b8c9-8b7d-44a0-a9b9-46c416adcbd9  base
kernel-spark3.2-scala2.12     020d69ce-7ac1-5e68-ac1a-31189867356a  base
```

```
from ibm_watson_machine_learning import APIClient
import  json
import numpy as np
wml_cred={
   "apikey":"okbr7ARnOQjyplTOyvNFC2QVkCF6q7afpci065Hucby8",
   "url":"https://us-south.ml.cloud.ibm.com"
}
wml_clients=APIClient(wml_cred)
wml_clients.spaces.list()
space_id="6d7c1218-3aca-4256-be3d-d610732530b1"
```

IBM Watson Studio | Search in your workspaces | Buy | PEDDAMALLU SAI DIVYA's ... | Dallas

Projects / Flight Delay / FlightDelay

```
    'name': 'flight',
    'owner': 'IBMid-6630042GJI',
    'resource_key': '1185cdbf-9975-49a4-ac41-7f372c902d5f',
    'space_id': 'b6783a9b-4513-4465-b052-a15d738cf69d'},
  'system': {'warnings': []}}
```

In [171…
```
model_uid = wml_client.repository.get_model_id(model_details)
model_uid
```

Out[171]: 'b124f04c-391d-4bea-8bb5-6c9d9e875765'

In [172…
```
deployment_props={
    wml_client.deployments.ConfigurationMetaNames.NAME:DEPLOYMENT_NAME,
    wml_client.deployments.ConfigurationMetaNames.ONLINE: {}

}
```

In [173…
```
deployment = wml_client.deployments.create(
artifact_uid=model_uid,
meta_props = deployment_props)
```

```
###################################################################################

Synchronous deployment creation for uid: 'b124f04c-391d-4bea-8bb5-6c9d9e875765' started

###################################################################################
```

```
wml_clients.set.default_space(space_id)
wml_clients.software_specifications.list(500)
MODEL_NAME="randomforest"
DEPLOYMENT_NAME="rf_deployment"
DEMO_MODEL=rf
soft_sepc_id=wml_clients.software_specifications.get_id_by_name("runtime-22.1-py3.9")
```

In [115]:

```
model_props={ wml_clients.repository.ModelMetaNames.NAME:MODEL_NAME,
   wml_clients.repository.ModelMetaNames.TYPE:"scikit-learn_1.0",
   wml_clients.repository.ModelMetaNames.SOFTWARE_SPEC_UID: soft_sepc_id

}
```

In [116]:

```
model_details=wml_clients.repository.store_model(model=DEMO_MODEL,meta_props=model_props,trai
ning_data=x_train,
                                 training_target=y_train.values.ravel())
```

In [117]:

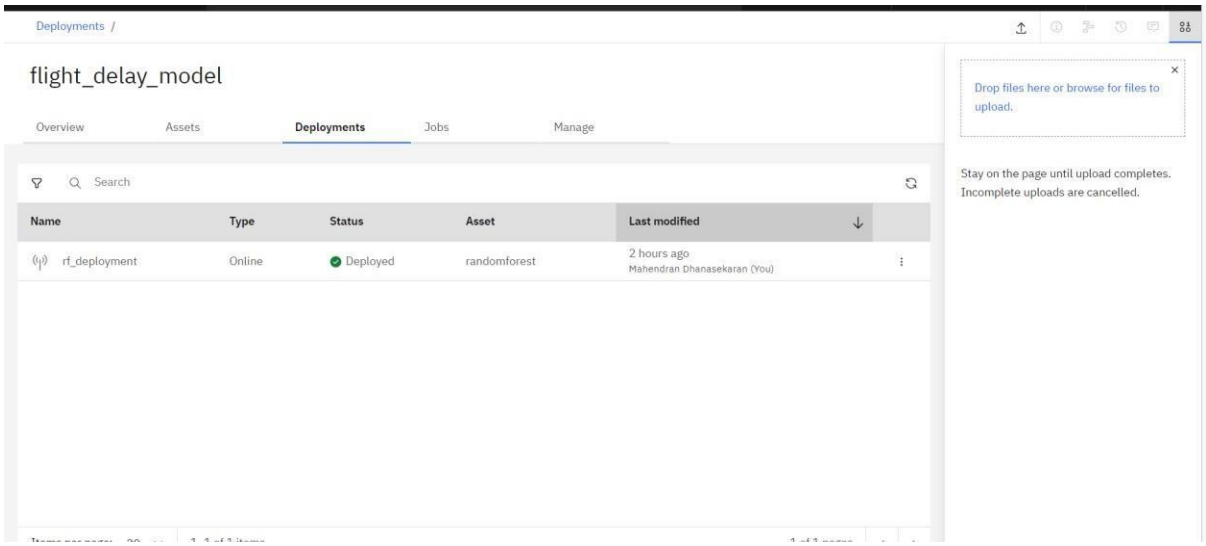
```
model_details
model_id=wml_clients.repository.get_model_id(model_details)
dep_props={
   wml_clients.deployments.ConfigurationMetaNames.NAME:DEPLOYMENT_NAME,
   wml_clients.deployments.ConfigurationMetaNames.ONLINE:{}
}
```

In [125]:

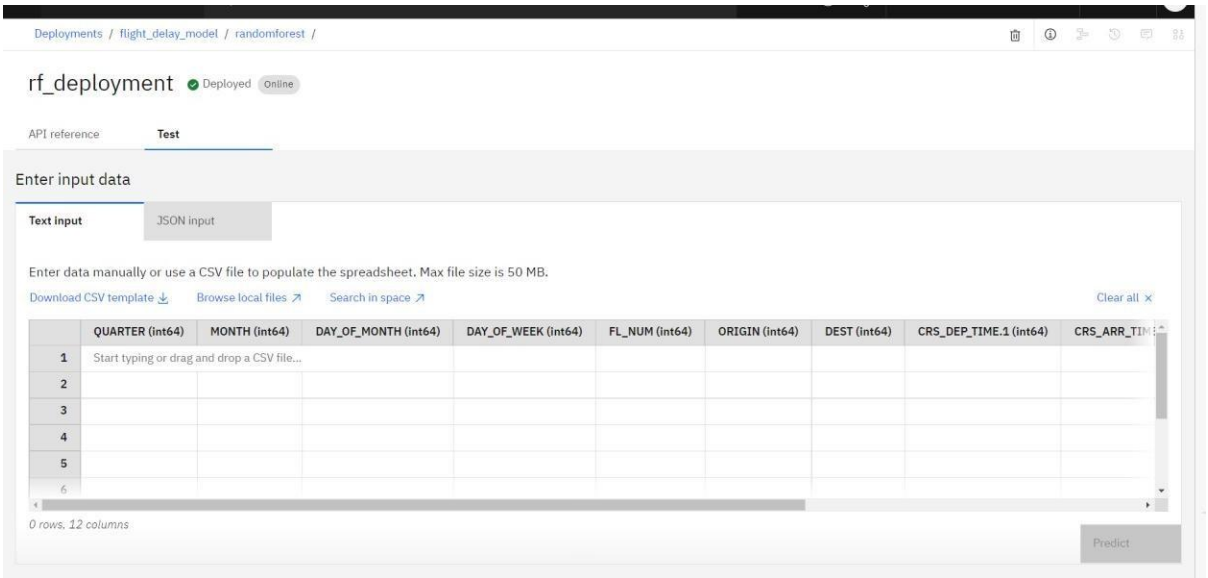
```
deployment=wml_clients.deployments.create(artifact_uid=model_id,meta_props=dep_props)
```

NOTE: APIKey must need to create to deploy and connect API

After successful of deployment, deployed is appeared in Deployment section as follow,



Testing of deployed model as follow, by giving values of all the features and it gives prediction.



Output is predicted by ML model successfully.