MACHINE LEARNING MODEL DEPLOYMENT WITH IBM CLOUD WATSON STUDIO [PHASE-3]

INTRODUCTION:

IBM Cloud Watson Studio is a comprehensive platform for data science and machine learning that provides various tools and services to streamline the model development process. In this guide, we will create a machine learning model using Watson Studio to address a predictive use case.

1. Define the Predictive Use Case:

The first step is to define the problem you want to solve using machine learning. This could be anything from predicting customer churn to image classification. For this example, let's assume we want to build a model to predict customer churn for a telecommunications company.

2. Select a Relevant Dataset:

Next, you need to find and select a dataset that is relevant to your use case. You can use publicly available datasets or your own data if applicable. In this case, you might use a dataset containing customer information and historical churn data.

3. Import the Dataset:

In Watson Studio, you can import the dataset into your project. Watson Studio provides a user-friendly interface to upload and manage data. The dataset should be in a compatible format (e.g., CSV, Excel).

4. Preprocess the Data:

Data preprocessing is a crucial step. You may need to handle missing values, clean the data, and transform it to make it suitable for modelling. Watson Studio provides tools for data cleaning and transformation.

5. Select Features:

In this step, you choose which features (attributes) from the dataset will be used as input for your machine learning model. Feature selection can have a significant impact on model performance.

6. Train the Machine Learning Model:

Watson Studio offers a variety of machine learning algorithms and libraries. You can build and train your model using these tools. For customer churn prediction, you might use classification algorithms like logistic regression, decision trees, or a neural network.

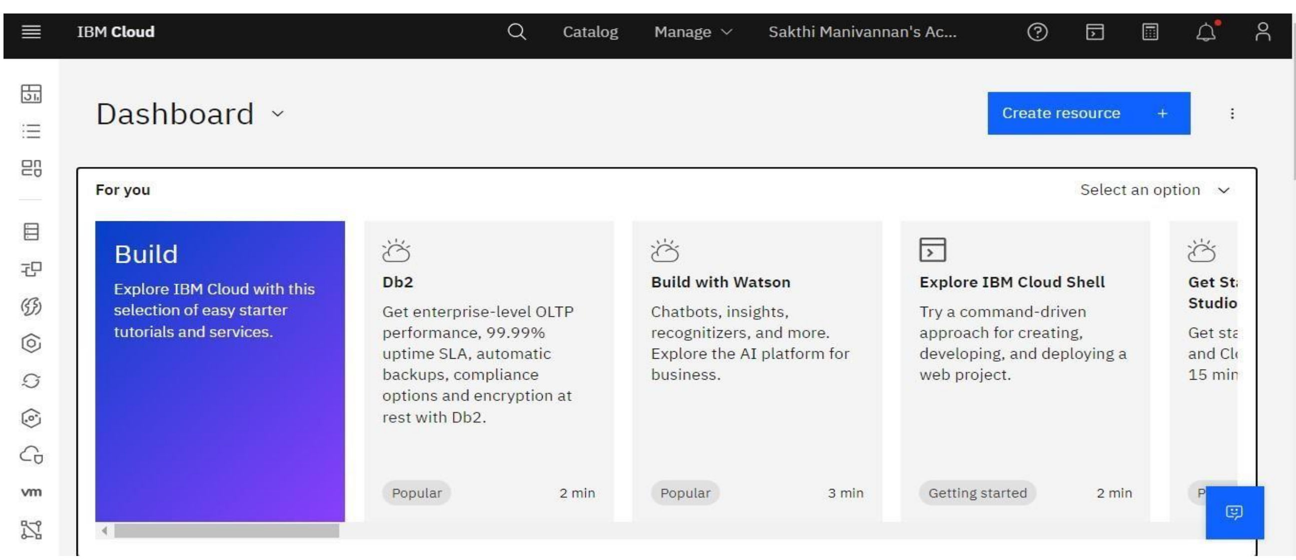
7. Evaluate and Tune the Model:

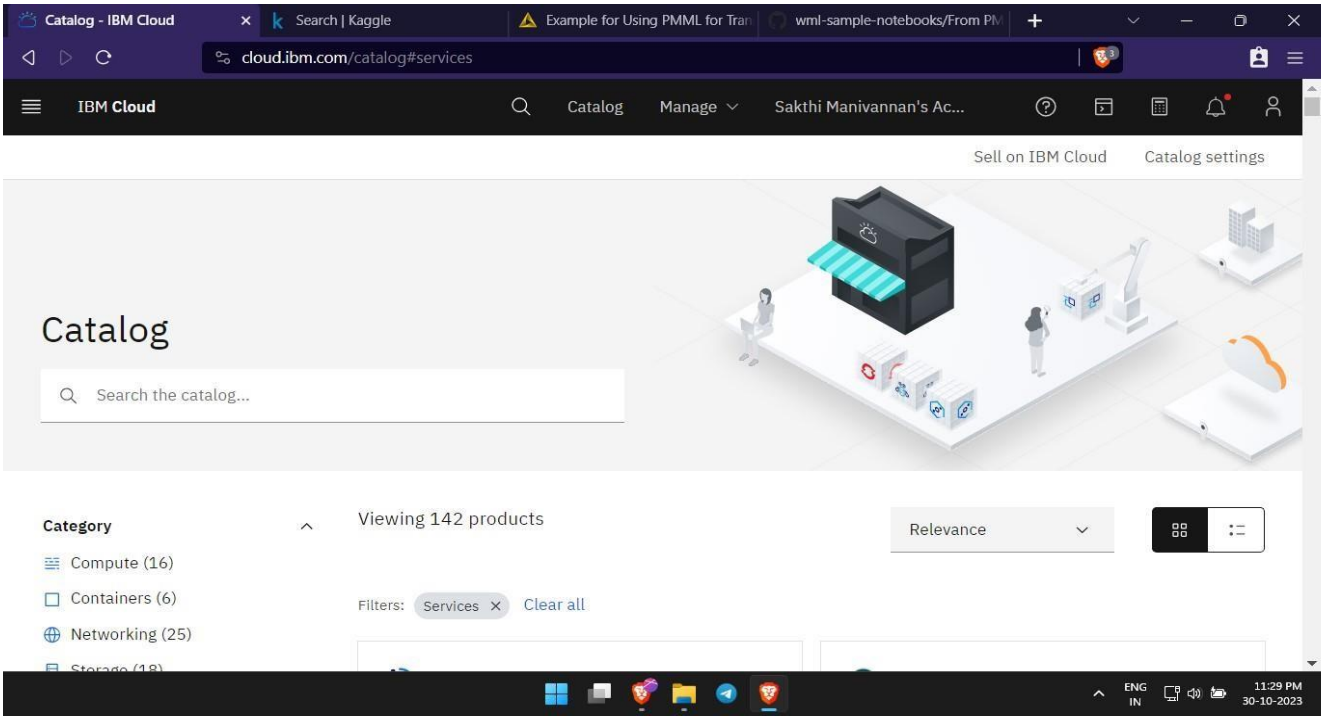
After training, it's essential to evaluate the model's performance using metrics such as accuracy, precision, recall, and F1-score. You can iterate on your model by tuning hyperparameters and selecting the best-performing one.

8. Deploy the Model:

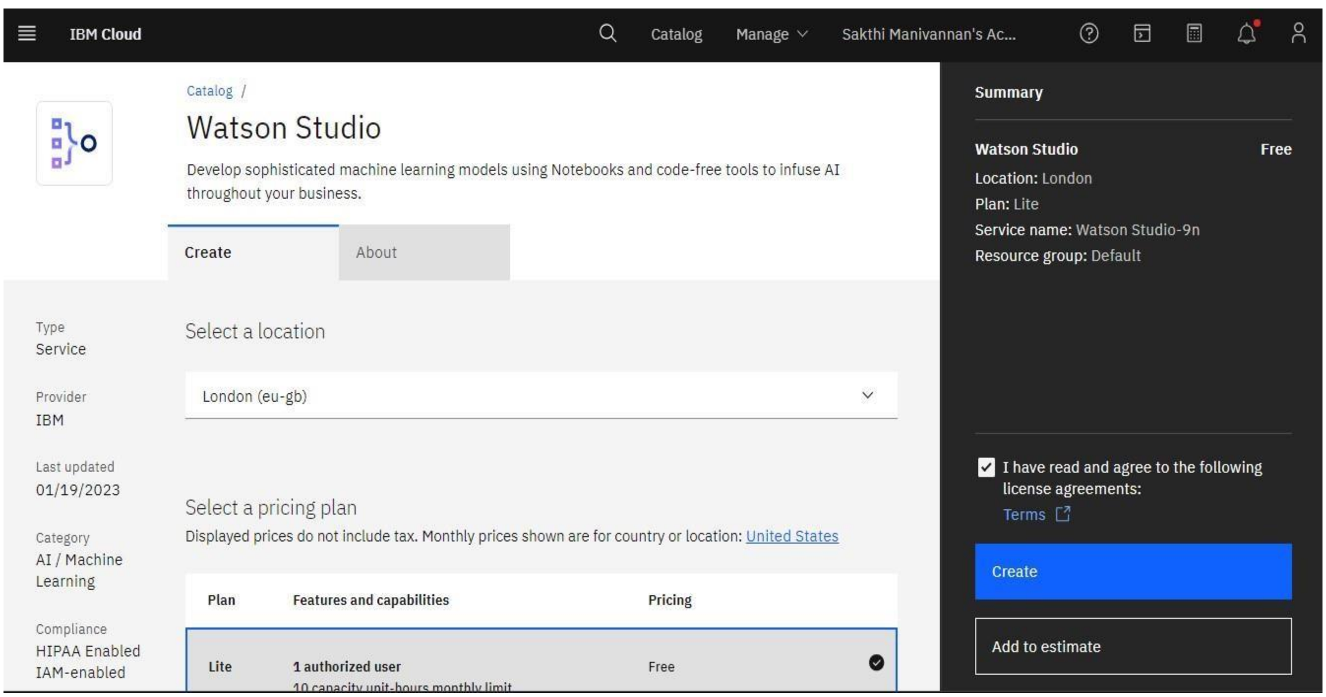
Once you have a satisfactory model, you can deploy it to make predictions on new data. Watson Studio provides deployment options, including APIs for real-time predictions.

Step1:

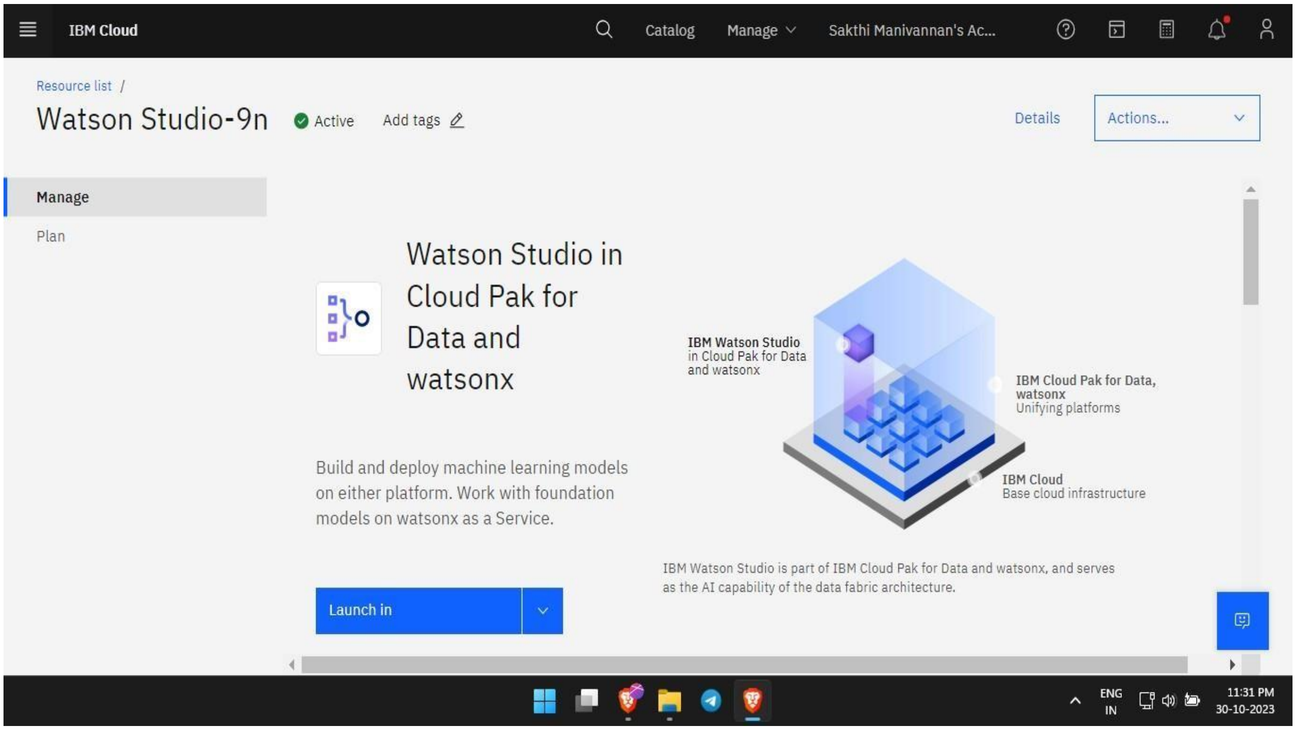


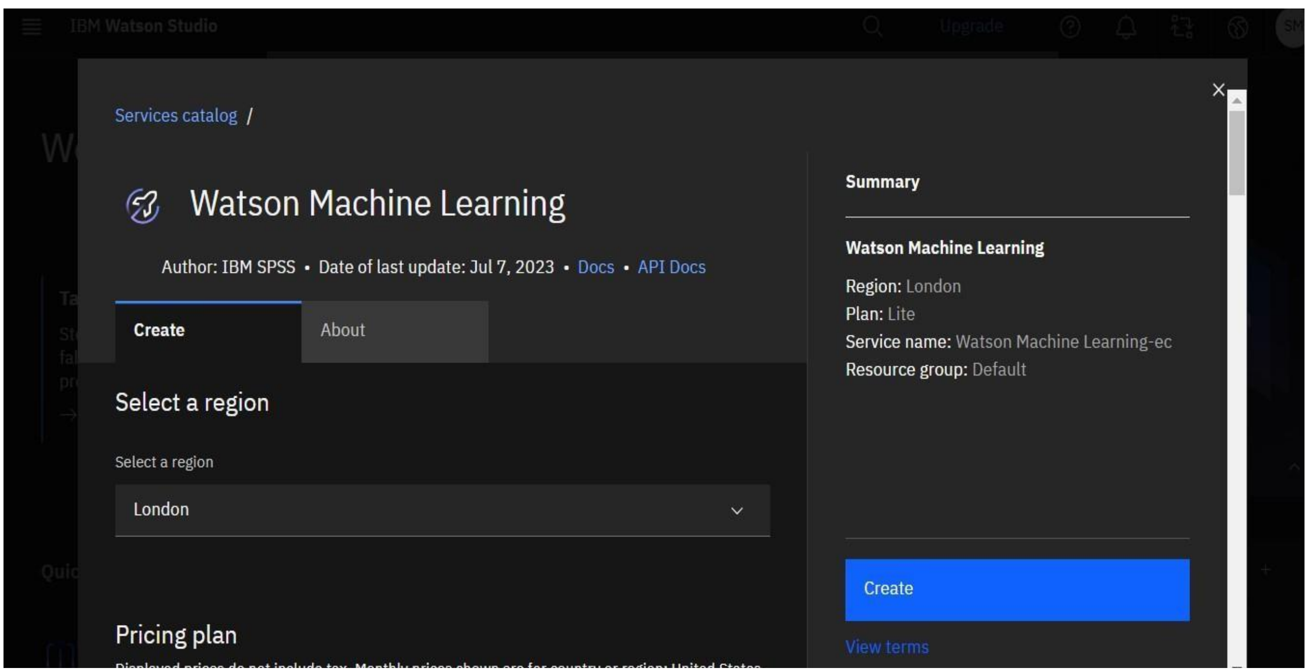

Catalog - IBM Cloud
Search | Kaggle
Example for Using PMML for Tran
wml-sample-notebooks/From PM
cloud.ibm.com/catalog#services
IBM Cloud
Catalog
Manage
Sakthi Manivannan's Ac...
Sell on IBM Cloud
Catalog settings
Catalog
Search the catalog...
Category
Compute (16)
Containers (6)
Networking (25)
Viewing 142 products
Relevance
Filters:
Services
Clear all
ENG
IN
11:29 PM
30-10-2023

Step2:

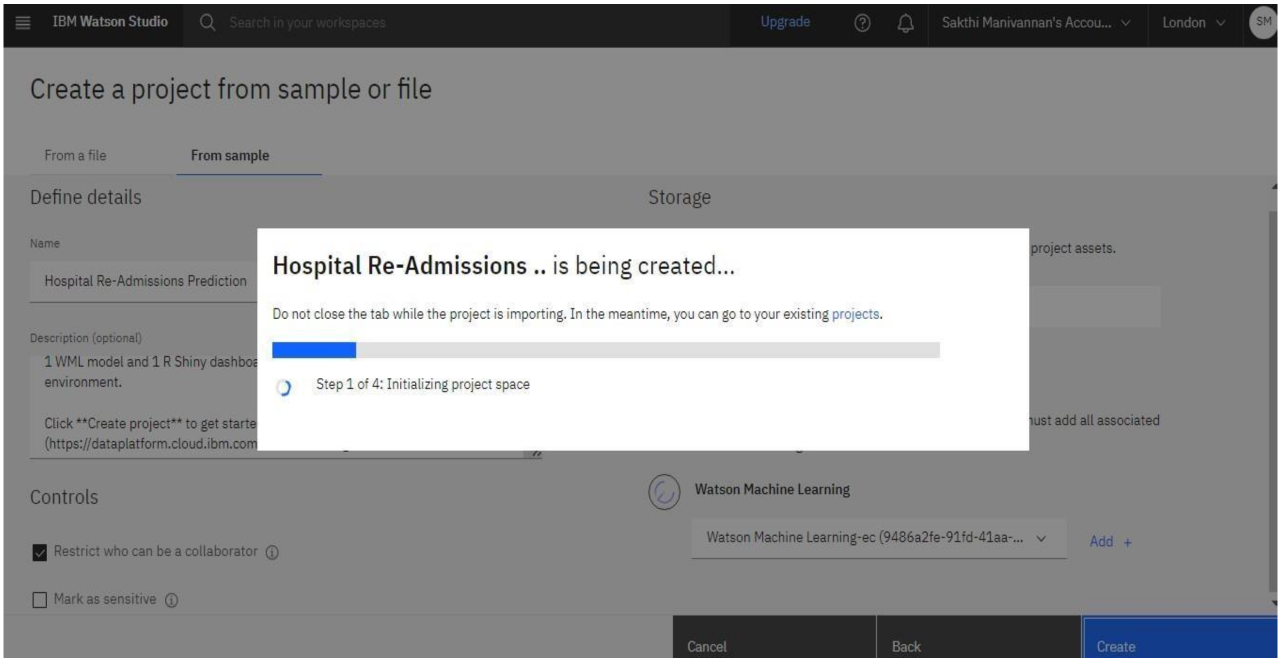

IBM Cloud
Catalog
Manage
Sakthi Manivannan's Ac...
Catalog /
Watson Studio
Develop sophisticated machine learning models using Notebooks and code-free tools to infuse AI throughout your business.
Create
About
Type
Service
Provider
IBM
Last updated
01/19/2023
Category
AI / Machine Learning
Compliance
HIPAA Enabled
IAM-enabled
Select a location
London (eu-gb)
Select a pricing plan
Displayed prices do not include tax. Monthly prices shown are for country or location: United States
Plan
Features and capabilities
Pricing
Lite
1 authorized user
Free
Summary
Watson Studio
Free
Location: London
Plan: Lite
Service name: Watson Studio-9n
Resource group: Default
I have read and agree to the following license agreements:
Terms
Create
Add to estimate

Step3:

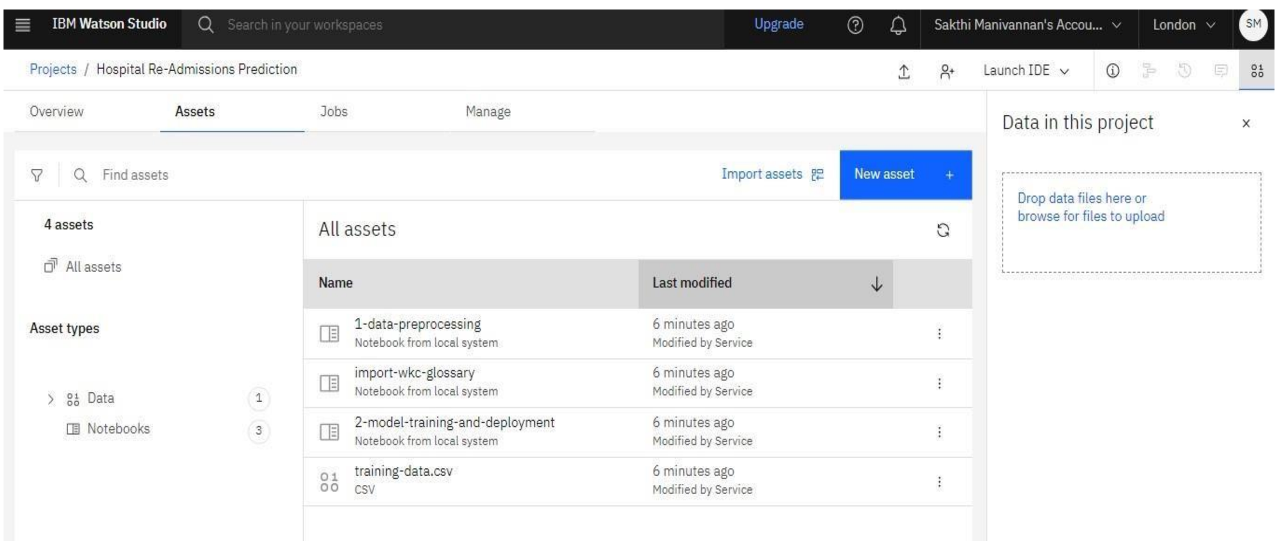

IBM Cloud
Catalog
Manage
Sakthi Manivannan's Ac...
Resource list /
Watson Studio-9n
Active
Add tags
Details
Actions...
Manage
Plan
Watson Studio in Cloud Pak for Data and watsonx
IBM Watson Studio in Cloud Pak for Data and watsonx
IBM Cloud Pak for Data, watsonx
Unifying platforms
IBM Cloud
Base cloud infrastructure
Build and deploy machine learning models on either platform. Work with foundation models on watsonx as a Service.
IBM Watson Studio is part of IBM Cloud Pak for Data and watsonx, and serves as the AI capability of the data fabric architecture.
Launch in
ENG
IN
11:31 PM
30-10-2023


Services catalog /
Watson Machine Learning
Author: IBM SPSS • Date of last update: Jul 7, 2023 • Docs • API Docs
Create
About
Select a region
Select a region
London
Pricing plan
Summary
Watson Machine Learning
Region: London
Plan: Lite
Service name: Watson Machine Learning-ec
Resource group: Default
Create
View terms

Step4:


IBM Watson Studio
Search in your workspaces
Upgrade
Sakthi Manivannan's Accou...
London
Create a project from sample or file
From a file
From sample
Define details
Storage
Name
Hospital Re-Admissions Prediction
Description (optional)
Hospital Re-Admissions .. is being created...
Do not close the tab while the project is importing. In the meantime, you can go to your existing projects.
Step 1 of 4: Initializing project space
Controls
Restrict who can be a collaborator
Mark as sensitive
Watson Machine Learning
Add +
Cancel
Back
Create

Step5:


IBM Watson Studio
Search in your workspaces
Upgrade
Sakthi Manivannan's Accou...
London
Projects / Hospital Re-Admissions Prediction
Launch IDE
Overview
Assets
Jobs
Manage
Data in this project
Find assets
Import assets
New asset +
Drop data files here or browse for files to upload
4 assets
All assets
Asset types
Data 1
Notebooks 3
Name
Last modified
1-data-preprocessing
Notebook from local system
6 minutes ago
Modified by Service
import-wkc-glossary
Notebook from local system
6 minutes ago
Modified by Service
2-model-training-and-deployment
Notebook from local system
6 minutes ago
Modified by Service
training-data.csv
CSV
6 minutes ago
Modified by Service

IBM Watson Studio   Search in your workspaces   Upgrade   Sakthi Manivannan's Accou...   London   SM

Projects / Hospital Re-Admissions Prediction / 1-data-preprocessing

Data Preparation

We will go through the end-to-end process of importing raw data, preparing the data using the patient details available, and save out the training data for further analysis in the next notebook `2-model-training-and-deployment`.

In the cell below we import the python libraries that we will use throughout the notebook.

In [4]:
```
import pandas as pd
pd.set_option('display.max_columns', None)
import numpy as np


import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
```

Steps to prepare the data

1. Import the data sets required for the analysis and merge the datasets.
2. Map the column headers to values.
3. Handle null values.
4. Filter the data based on matching conditions
5. Map Diagnoses types using icd-9 codes.
6. Handle target column

Step6:

Projects / Hospital Re-Admissions Prediction / 1-data-preprocessing

1. Import the datasets

The raw dataset `diabetic_data.csv` contains all patients personal, hospital admission/discharge and medication details in it. The data was downloaded from UCI repo.

In [5]:
```
hospital_details = project.get_file("diabetic_data.csv")
hospital_details.seek(0)
df_hospital = pd.read_csv(hospital_details)

df_hospital.head()
```

Out[5]:

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	payer_code	medical_specialty	num_lab_procedures	nui
0	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1	1	?	Pediatrics-Endocrinology	41	
1	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7	3	?	?	59	
2	64410	86047875	AfricanAmerican	Female	[20-30)	?	1	1	7	2	?	?	11	
3	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7	2	?	?	44	
4	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7	1	?	?	51	

Step7:

5. Identify Diagnosis types

1. Remove the diagnosis types starts with `V` and `E`. (`V` and `E` codes represent external causes of injury and supplemental classification.)
2. Convert Diagnosis columns to float type
3. Map diagnosis codes to diagnosis string using a dictionary of icd-9 codes and the diagnosis types.

In [12]:

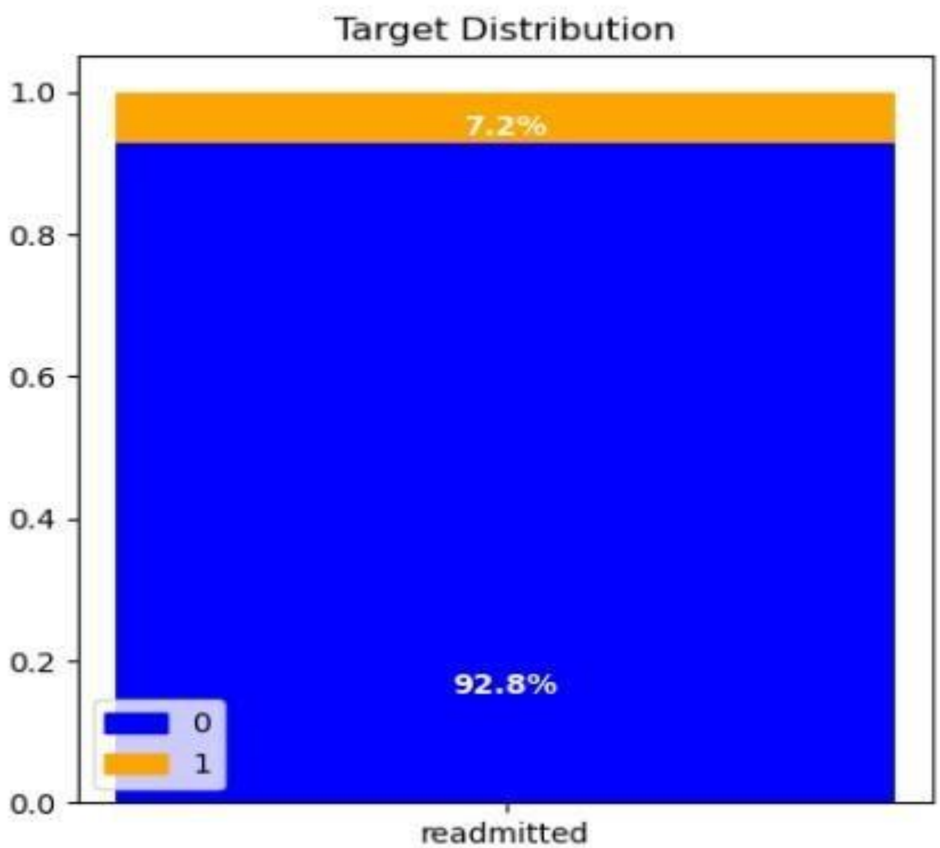
```
df_diag=df_diag[~df_diag.diag_1.str.startswith("V")]
df_diag=df_diag[~(df_diag.diag_2.str.startswith("V")|df_diag.diag_2.str.startswith("E"))]
df_diag=df_diag[~(df_diag.diag_3.str.startswith("V")|df_diag.diag_3.str.startswith("E"))]
df_diag.diag_1=df_diag.diag_1.astype('float64')
df_diag.diag_2=df_diag.diag_2.astype('float64')
df_diag.diag_3=df_diag.diag_3.astype('float64')
```

In [13]:

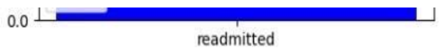
```
icd9coderange_dict = [(1, 140), (140, 240), (240, 280), (280, 290), (290, 320), (320, 390),
              (390, 460), (460, 520), (520, 580), (580, 630), (630, 680), (680, 710),
              (710, 740), (740, 760), (760, 780), (780, 800), (800, 1000), (1000, 2000)]


diagnosis_dict = {0: 'Infectious And Parasitic', 1: 'Neoplasms', 2: 'Endocrine, Nutritional And Metabolic', 3: 'Blood Related/Diabetes',
            4: 'Mental', 5: 'Nervous System And Sense Organs', 6: 'Circulatory System', 7: 'Respiratory System',
            8: 'Digestive System', 9: 'Genitourinary System', 10: 'Pregnancy, Childbirth, And The Puerperium', 11: 'Skin And Subcutaneous Tissue',
            12: 'Musculoskeletal System And Connective Tissue', 13: 'Congenital', 14: 'Prenatal', 15: 'Ill-Defined Conditions',
            16: 'Injury And Poisoning', 17: 'Misc'}


for num, cat_range in enumerate(icd9coderange_dict):
```



Step8:




```
In [16]: df_diag.columns=[x.upper() for x in df_diag.columns.tolist()]
         project.save_data('training-data.csv', df_diag.to_csv(index=False), overwrite=True)

Out[16]: {'file_name': 'training-data.csv',
          'message': 'File saved to project storage.',
          'bucket_name': 'hospitalreadmissionspredictionrel-donotdelete-pr-w4kl9yicbyk3wg',
          'asset_id': 'ff0d1894-a19c-41b1-a357-addc5ed9d830'}
```

The data is prepared and stored in the file `training-data.csv`, which we can view in the project's data assets. This data set is used in `2-model-training-and-deployments` to predict and deploy hospital readmissions using sci-kit and wml libraries. Alternatively, if you want to experiment with the auto AI capability, this data can be used in an auto-ai pipeline to predict hospital readmissions.

CONCLUSION:

we've outlined the process of building a machine learning model using IBM Cloud Watson Studio. Starting from defining the predictive use case, importing, and preprocessing the data, selecting features, training the model, and deploying it, Watson Studio offers a comprehensive set of tools to streamline the entire process.

PRESENTED BY

1. C .Vinothini
2. M. Sakthi
3. M. Swetha
4. B. Jayasudha