Lab - 10

BDA : Big Data Analysis

AIM : Mining Complex Types of Data

Name : Ambalia Harshit

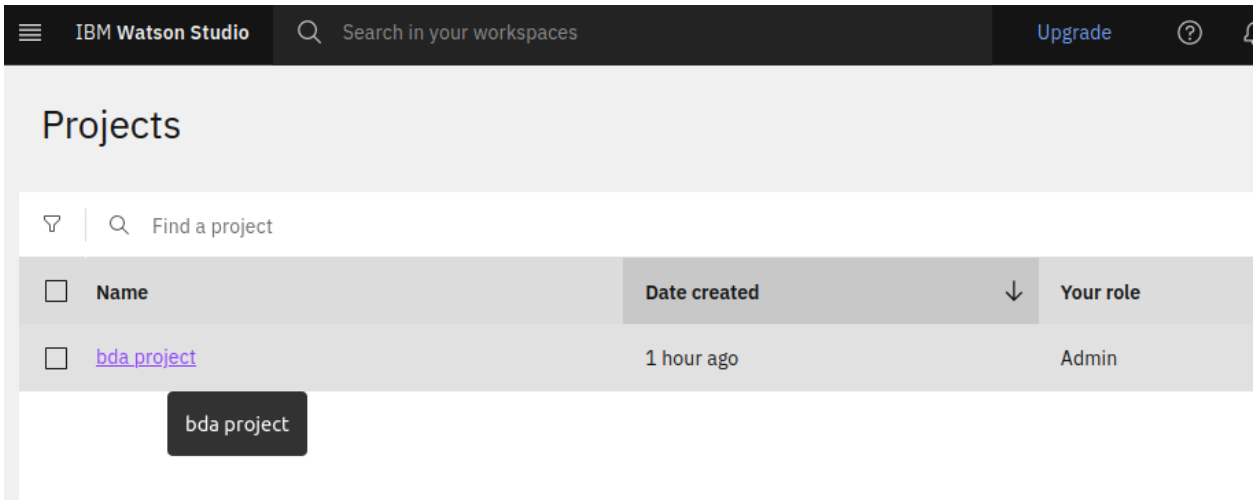
Department : Computer Engineering (M.Tech sem II)

Roll no : MT001

Date : 19 March 2024

Title : Using IBM WatSon to train data using ML

- Firstly creating New project on IBM Watson

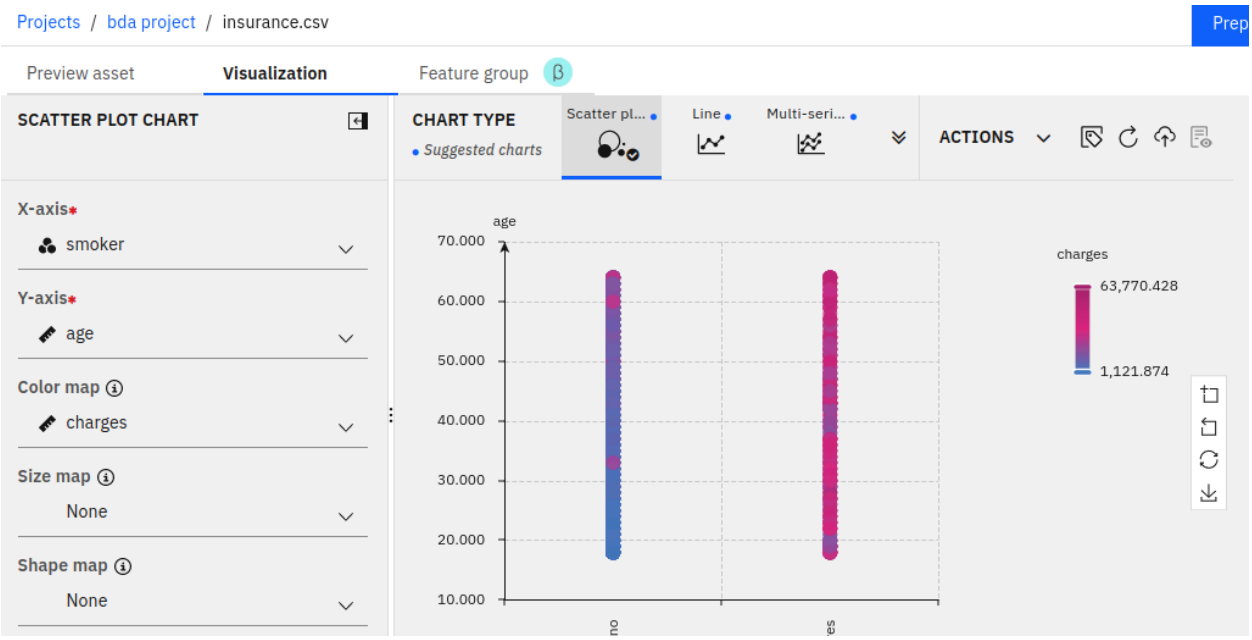


- This is our data which will be trained.

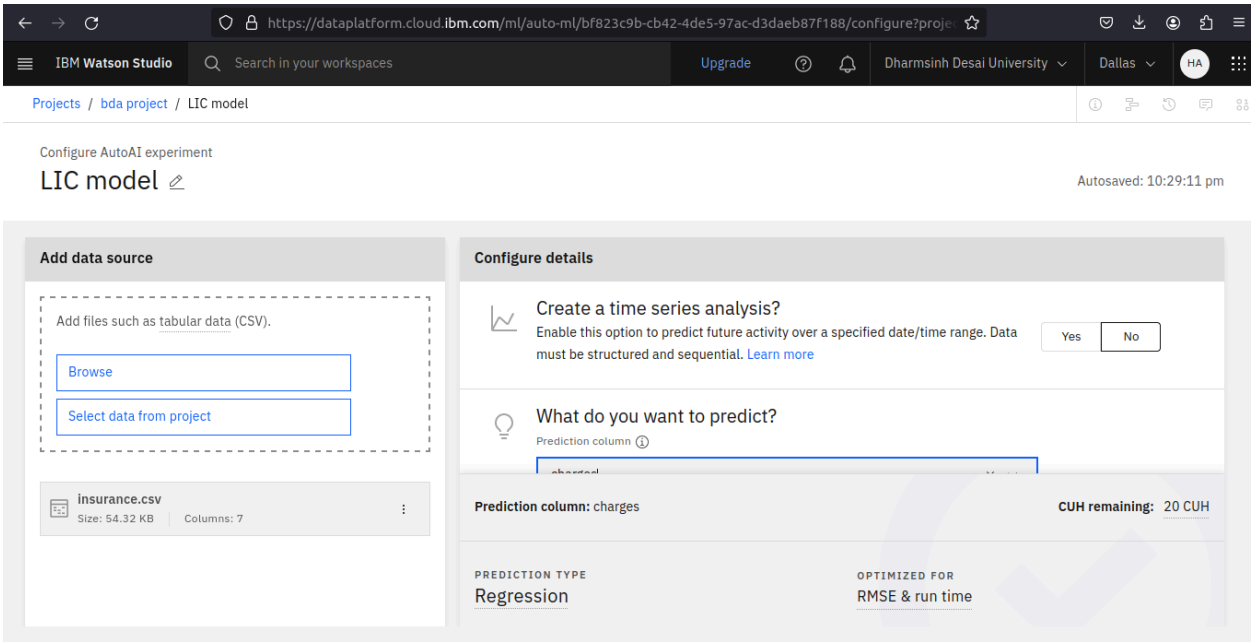
A1 age

	A	B	C	D	E	F	G	H
1	age	sex	bmi	children	smoker	region	charges	
2	19	female	27.9	0	yes	southwest	16884.924	
3	18	male	33.77	1	no	southeast	1725.5523	
4	28	male	33	3	no	southeast	4449.462	
5	33	male	22.705	0	no	northwest	21984.47061	
6	32	male	28.88	0	no	northwest	3866.8552	
7	31	female	25.74	0	no	southeast	3756.6216	
8	46	female	33.44	1	no	southeast	8240.5896	
9	37	female	27.74	3	no	northwest	7281.5056	
10	37	male	29.83	2	no	northeast	6406.4107	
11	60	female	25.84	0	no	northwest	28923.13692	
12	25	male	26.22	0	no	northeast	2721.3208	
13	62	female	26.29	0	yes	southeast	27808.7251	
14	23	male	34.4	0	no	southwest	1826.843	
15	56	female	39.82	0	no	southeast	11090.7178	
16	27	male	42.13	0	yes	southeast	39611.7577	
17	19	male	24.6	1	no	southwest	1837.237	
18	52	female	30.78	1	no	northeast	10797.3362	
19	23	male	23.845	0	no	northeast	2395.17155	
20	56	male	40.3	0	no	southwest	10602.385	
21	30	male	35.3	0	yes	southwest	36837.467	
22	60	female	36.005	0	no	northeast	13228.84695	
23	30	female	32.4	1	no	southwest	4149.736	
24	18	male	34.1	0	no	southeast	1137.011	
25	34	female	31.92	1	yes	northeast	37701.8768	

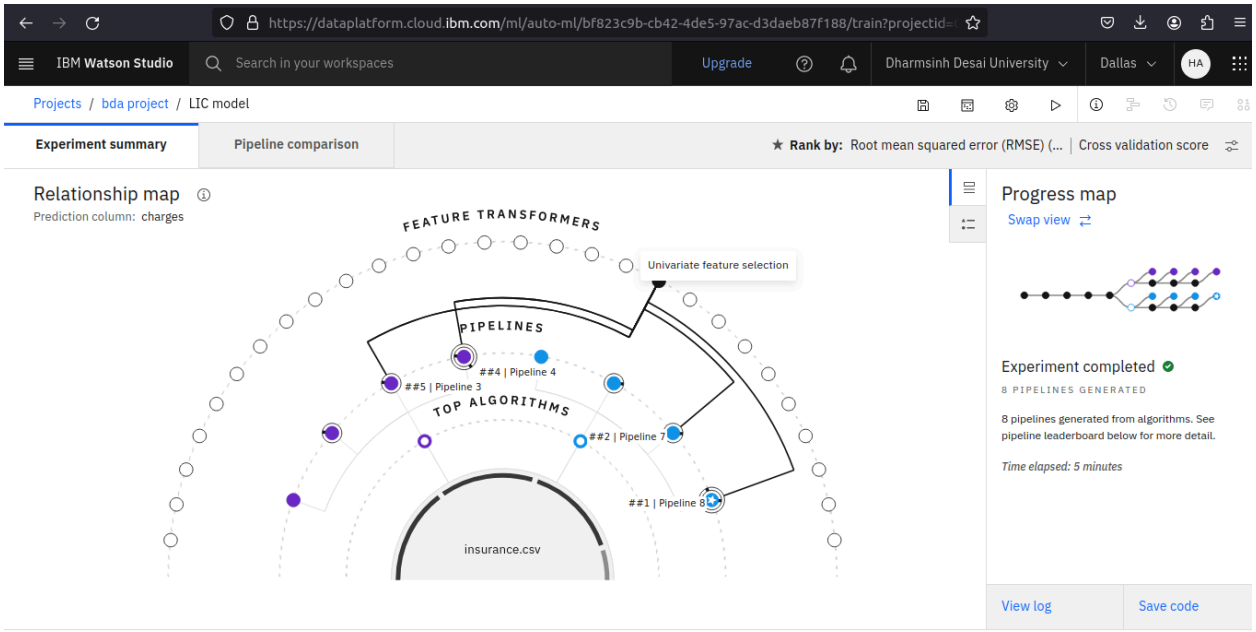
- We can even use this platform to visualize the data we have used to train this model.



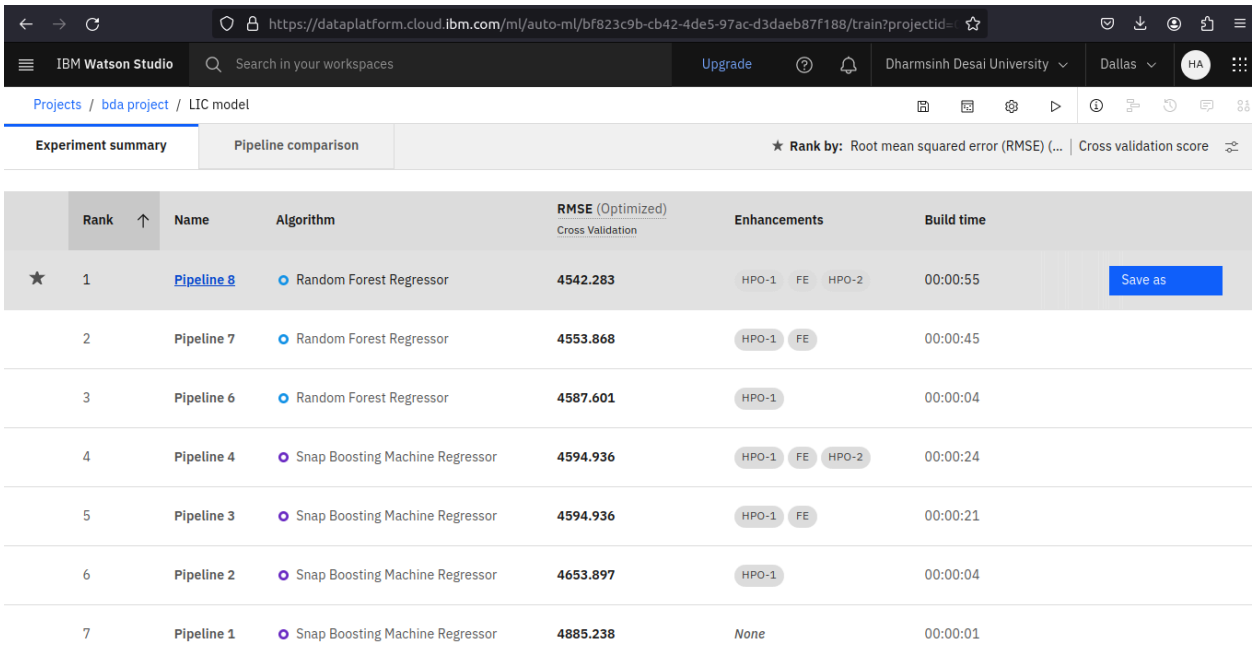
- After Visualizing data, now we are training the data after specifying the output/prediction collumn.



- This platform will use all types of ML methods to train the model by specifying all the kinds of hyper parameters.

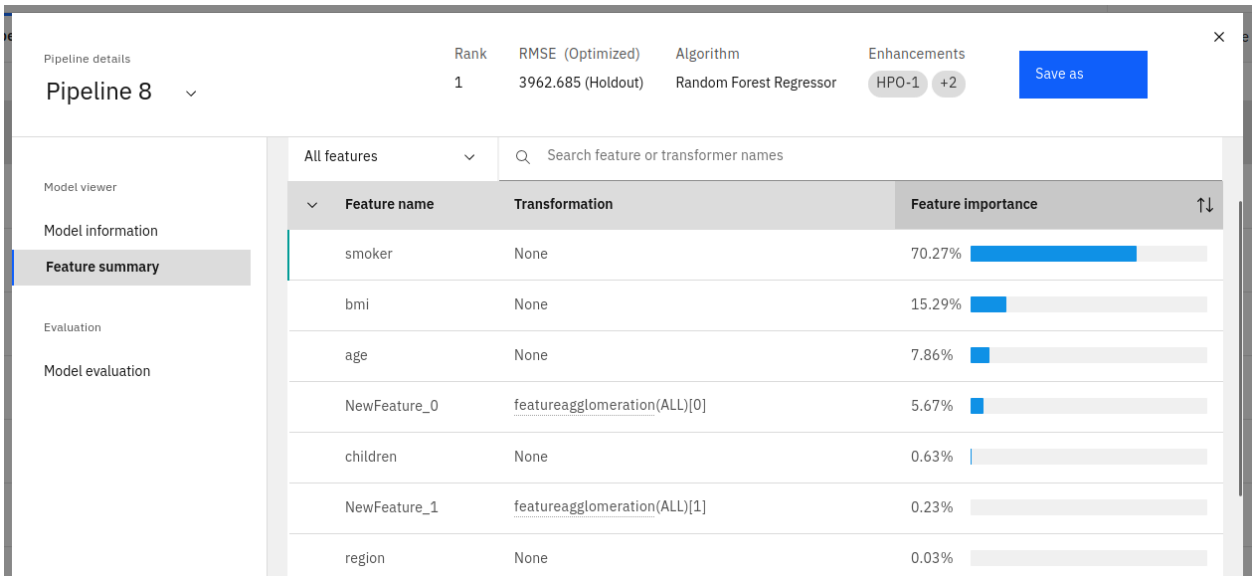


- As you can see, it have trained the various types of models by different pipelines. And it is ranked by the accuracy score of all ML models.

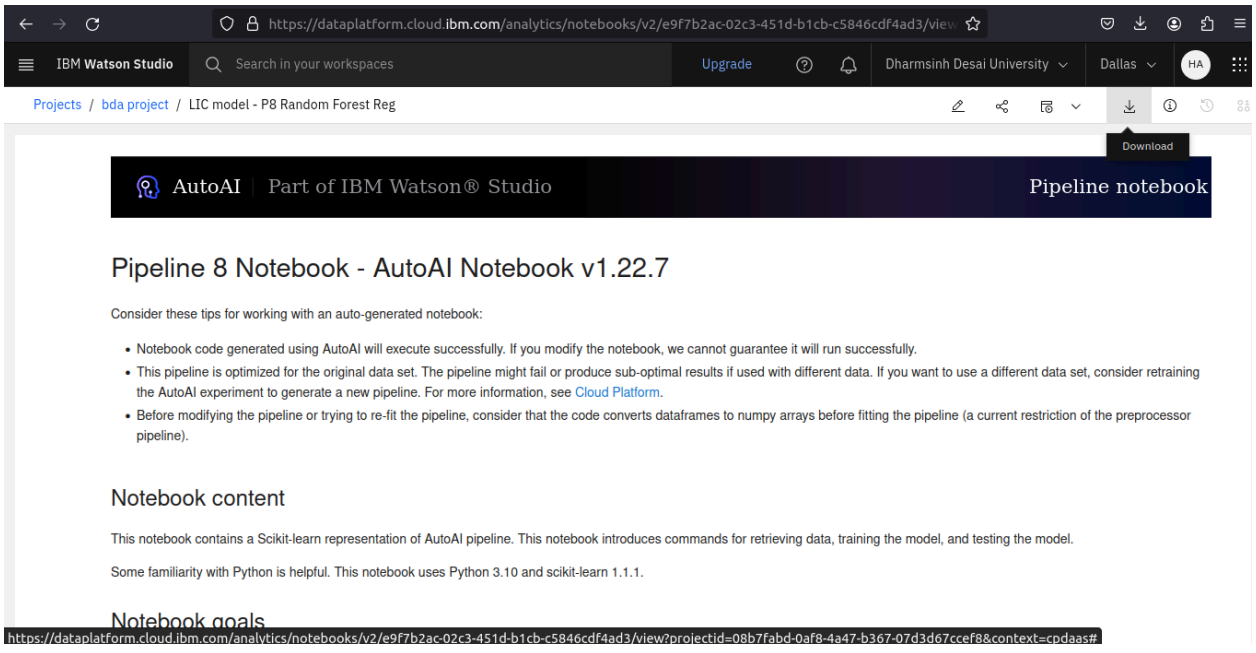


	Rank	Name	Algorithm	RMSE (Optimized) Cross Validation	Enhancements	Build time	
★	1	Pipeline 8	Random Forest Regressor	4542.283	HPO-1 FE HPO-2	00:00:55	Save as
	2	Pipeline 7	Random Forest Regressor	4553.868	HPO-1 FE	00:00:45	
	3	Pipeline 6	Random Forest Regressor	4587.601	HPO-1	00:00:04	
	4	Pipeline 4	Snap Boosting Machine Regressor	4594.936	HPO-1 FE HPO-2	00:00:24	
	5	Pipeline 3	Snap Boosting Machine Regressor	4594.936	HPO-1 FE	00:00:21	
	6	Pipeline 2	Snap Boosting Machine Regressor	4653.897	HPO-1	00:00:04	
	7	Pipeline 1	Snap Boosting Machine Regressor	4885.238	None	00:00:01	

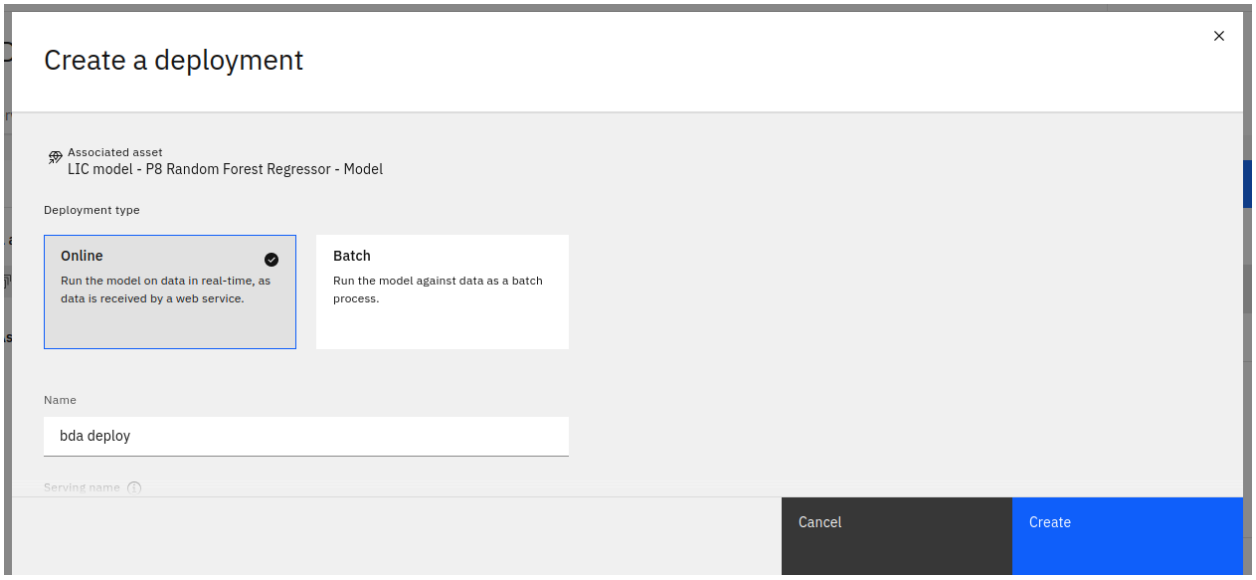
- As we can see Pipeline 8 has the highest accuracy, we can see specifications for that model.



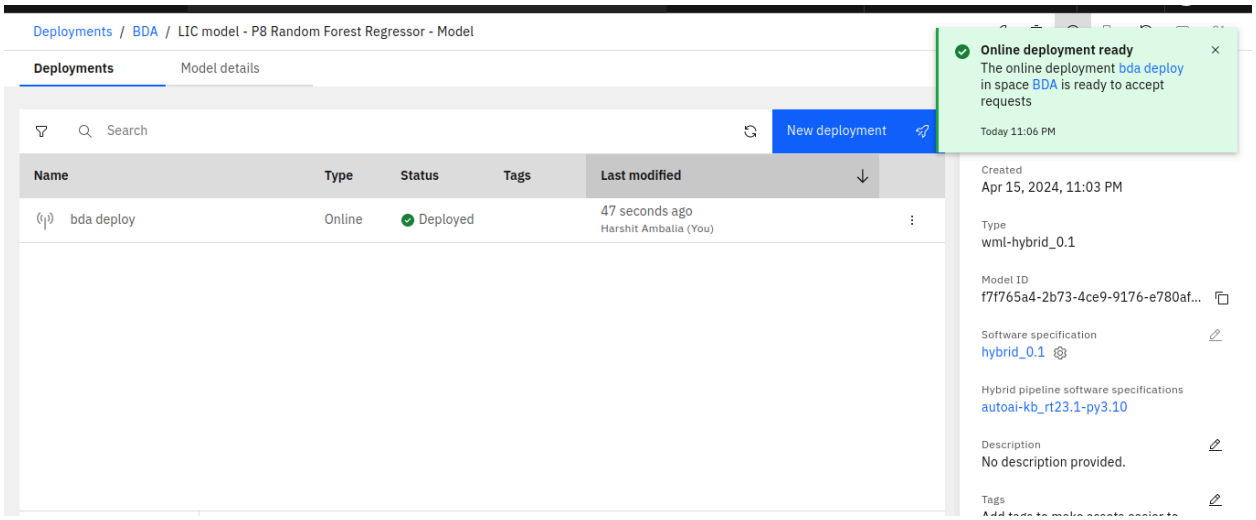
- We can also download the ipynb code/ source code file which is used to train that model.



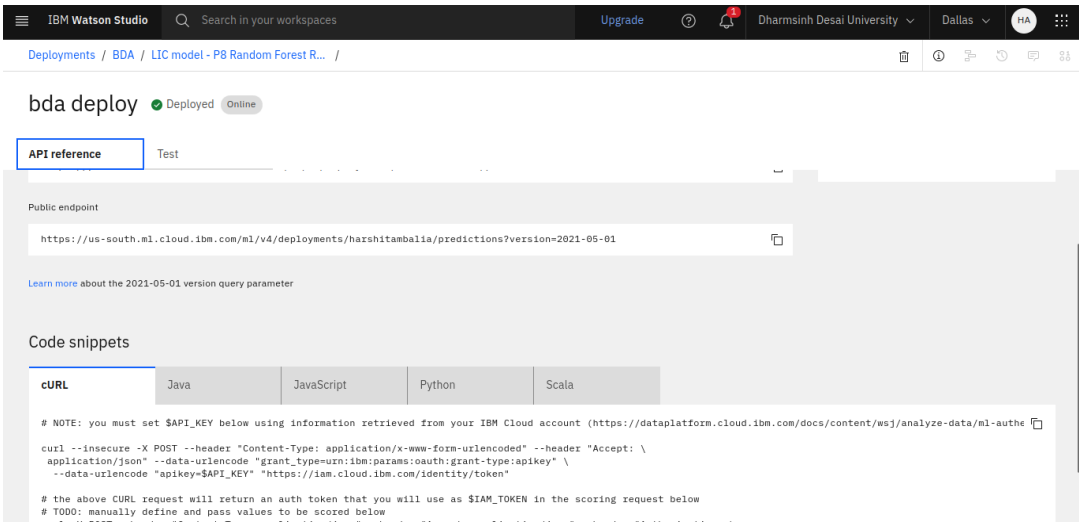
- Now as the model has been created, we can deploy this model online as so that public can use this model.
    - So now we are creating the new deployment for this model.



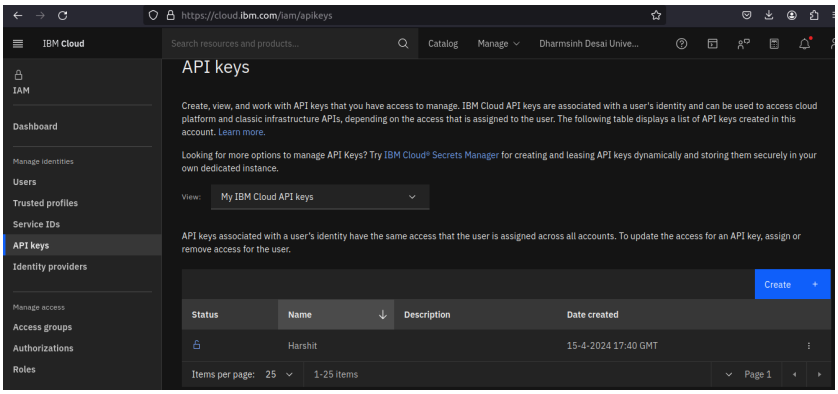
- So now we have created deployment named ‘bda deploy’.



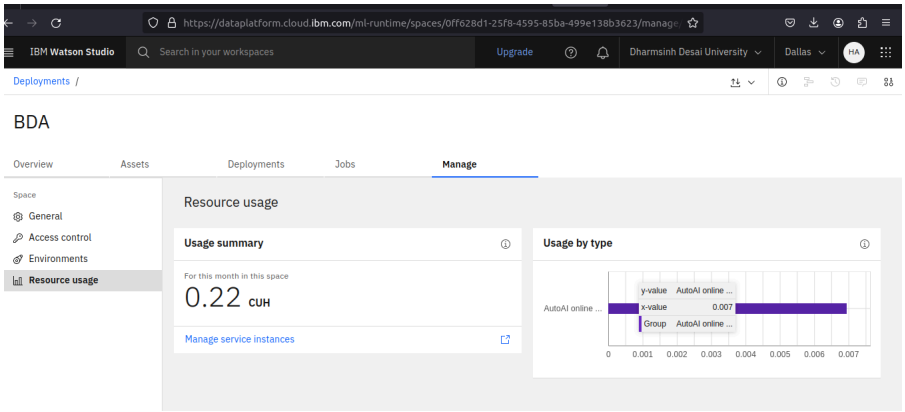
- We can use various language end points to access this model.



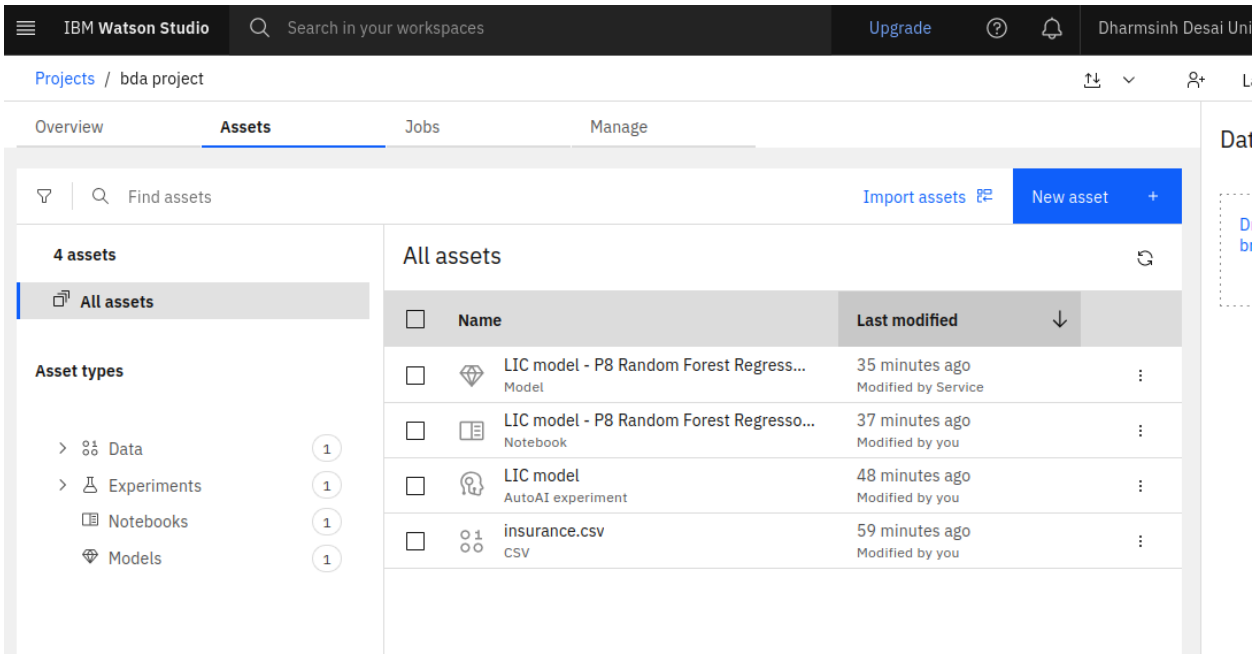
- To access this end points, we have to get the API keys, so now we are creating the API key to access our model through public network.



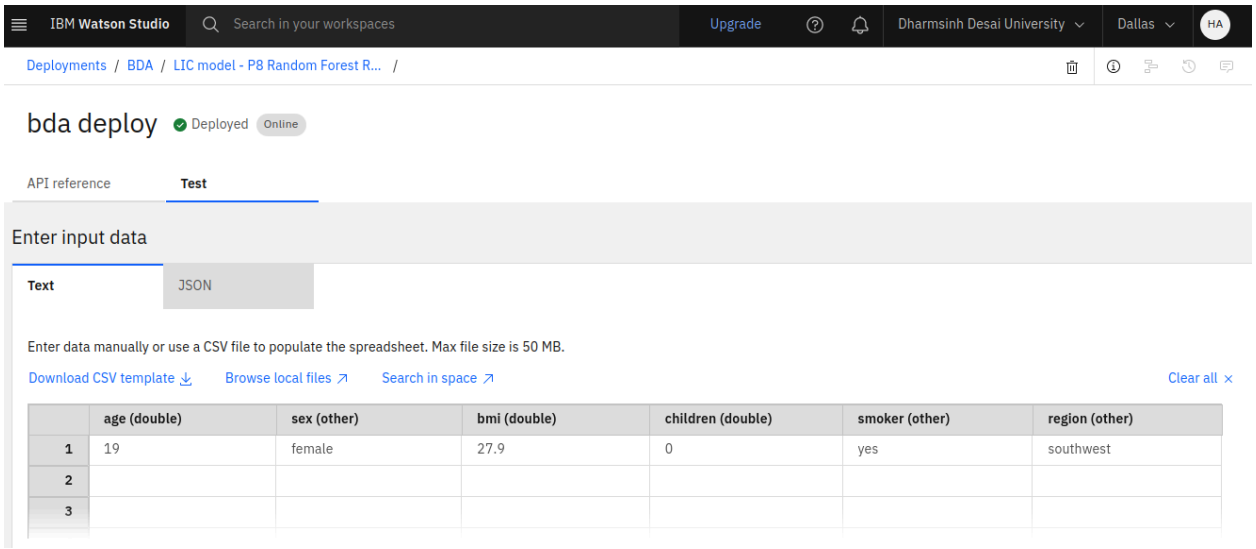
- We can also use ‘Resource usage’ direcly by accessing the feature provided.



- If we go on assets of the project we have created, we can see all the assets that has been used in this project.



- We can use the inbuild testing feature provided by itself. As we can see, it is asking for the data on which our model will try to generate the output.



- After proving the input, we can predict the output generated by this model for given input.