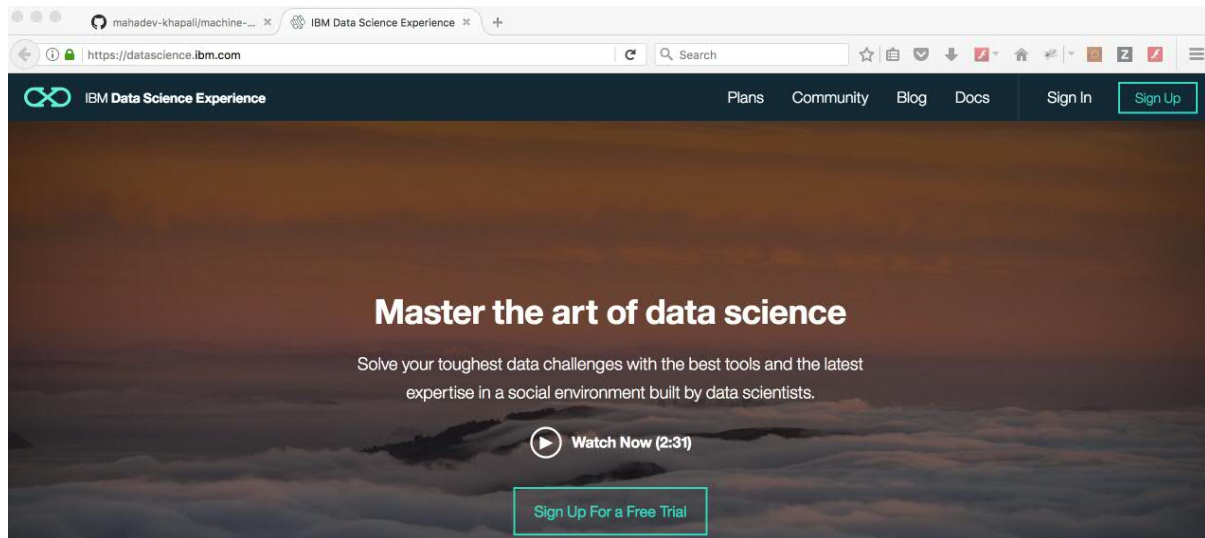
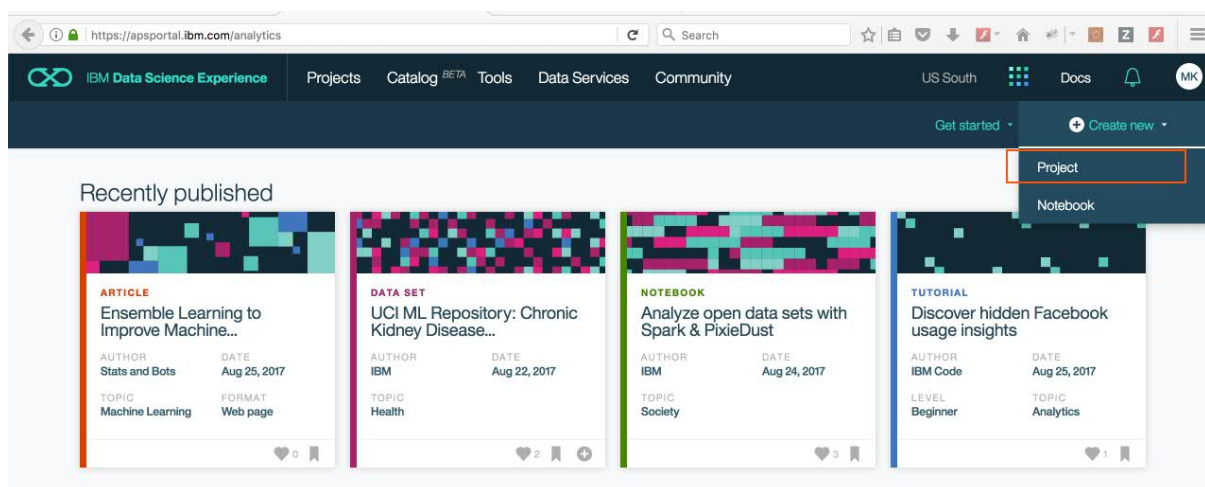


## ML Model Building hand-on lab steps:

Log-in to <https://datascience.ibm.com/> and sign-in if you have already registered else sign-up.



➤ After launching the the data-science portal lick on “Create new” → Project.



➤ In the create project page, you need to provide:

- Project name
- Spark-Service instance
- Object Storage instance

If you have already have the drop-down menu shows the list to choose from, else you are asked to provision a new Spark-Service and Object Storage instance as shown below:

**Name**  
ISTC-Lab1

**Description**  
Project for ISTC Lab

**Spark Service**  
 ⚠️ If you associate the same Spark service with multiple projects, the Spark history server will display job history information for all the projects.  
 No Spark instances found.  
 Create a new IBM Analytics for Apache Spark instance, then return to this page to create a new project.  
 Reload

**Target Object Storage Instance**  
 No object storage instances found.  
 Create a new instance, then return to this page.

Cancel Create

- Now Project is created, browse through different Tabs like “Analytics Assets”, “Data Assests” , “Deployments” , “Collaborators” and “Settings”.

**My Projects > TestLab1**

Overview Analytics Assets Data Assets Bookmarks Deployments Collaborators Settings

**Notebooks** view all (0) + add notebooks

NAME	SHARED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
you currently have no notebooks						

**Data Assets** view all (0) + add data assets

NAME	TYPE	SERVICE	LAST MODIFIED	ACTIONS
you currently have no data assets				

**Bookmarks** view all (0) + explore community

you currently have 0 bookmarks

**Files** Connections

Drop file here or browse your files to add a new file

Find in Storage

No files found.

- Now click on “Analytics Assests” and click on “add models” to create a ML model

Overview **Analytics Assets** Data Assets Bookmarks Deployments Collaborators Settings

Find in My Notebooks

### Notebooks

+ add notebooks

NAME	SHARED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
you currently have no notebooks						

### Models

+ add models

NAME	STATUS	RUNTIME	LAST MODIFIED	ACTIONS
ISTC-Flow-Model-DibeticPrediction	trained	spark-2.0	28 Aug 2017	...

- Provide the “Model name”, “Machine Learning” and “Spark Service” instance name. You need to provision a “Machine Learning” service instance (if not provisioned already).
- You can select data preparation as “Automatic” or “Manual” according to your requirement.

Create new model BETA

**Name**  
ISTCLabTest

**Description**  
Model description

**Machine Learning Service**  
ML-ISTCLab2017

**Spark Service**  
Spark-ISTCLab2017

**Automatic**  
Prepare my data and create a model automatically

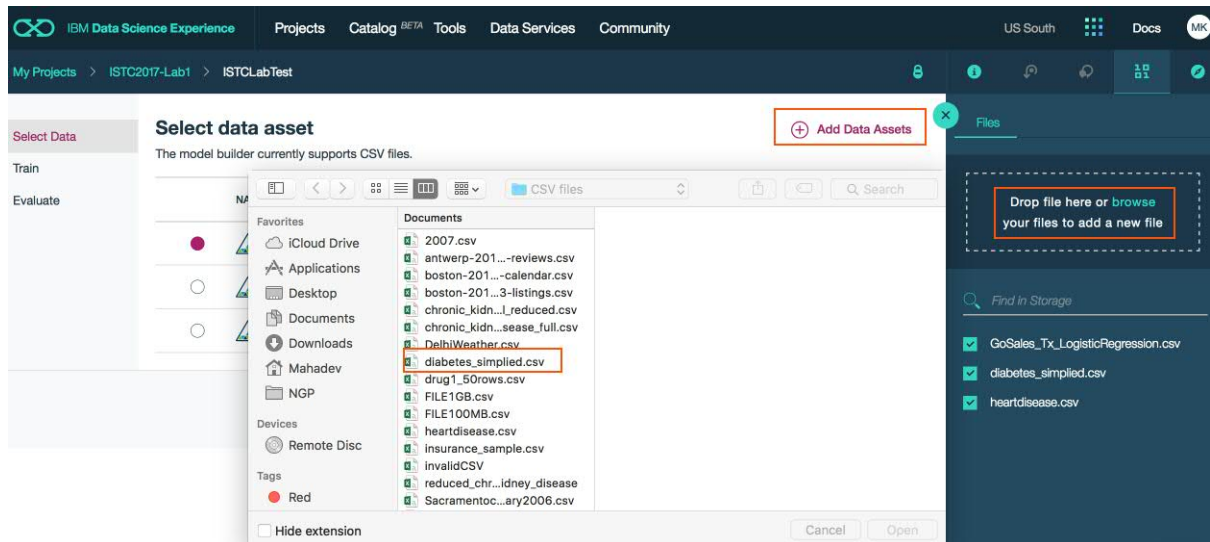
**Manual**  
Let me prepare my data and select which models to train

Need something more flexible? Create a notebook.

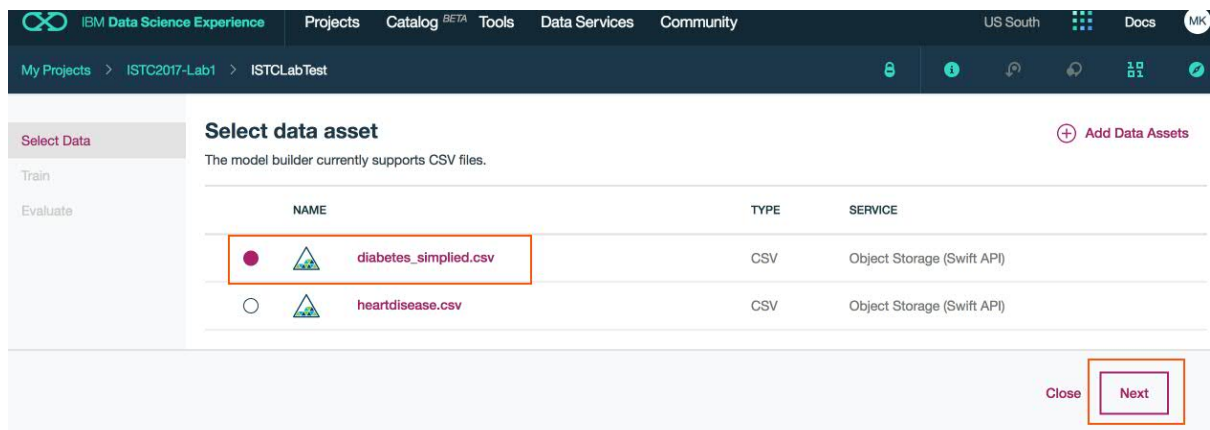
Cancel Create

- Once the model is created you will land to “Select data asset” page. You can add your dataset to the “Data Assets” by clicking on “Add Data Assets”.

- Download the “diabetes\_simplified.csv” dataset from the github link → <https://github.com/mahadev-khapali/machine-learning>



- Once the data set is added to the “Data Assets” it lists in the “Select data asset” page. Now user can select the required data set and click next.



- Once the data is loaded, in the next page user can select the “label column” and “feature column” and select the which spark ML algorithm and technique to be used in case of Manual data preparation.
- In case of “Automatic Data Preparation” technique is auto selected based on the “label column” , “feature column” and “dataset” etc.

- Adjust the amount of input data that you want to use for “Train” “Test” and “Hold-out” and click on the Next button.

IBM Data Science Experience

Projects Catalog <sup>BETA</sup> Tools Data Services Community

US South Docs

My Projects > ISTC2017-Lab1 > ISTCLabTest

Select Data

Train

Evaluate

### Select a technique

Column value to predict (Label Col)

Outcome (Integer)

Feature columns

Glucose (Integer), BloodPressure (Integer), Insulin (Integer), BMI (Decimal), Age (Integer)

Suggested technique.

**Binary Classification**

Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.

**Multiclass Classification**

Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.

**Regression**

Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split

Train: 60 Test: 20 Holdout: 20

Close Previous Next

- In the next page Model will be trained and evaluated. User can save the trained model by clicking on “Save”.

My Projects > ISTC2017-Lab1 > ISTCLabTest

Select Data

Train

Evaluate

### Select model

	ESTIMATOR TYPE	STATUS	PERFORMANCE	AREA UNDER ROC CURVE	AREA UNDER PR CURVE	LAST EVALUATION	ACTIONS
	LogisticRegression	Training...	...	...	...	...	...

Close Previous Save

- User can deploy the model by clicking on “Add Deployment” and Select type as “Online”

My Projects > ISTC2017-Lab1 > ISTCLab1

Details Predictions

**ISTCLab1**

Machine learning service	ML-ISTCLab2017
Label column	Outcome
Model builder details	<a href="#">View</a>
Training data schema	<a href="#">View</a>
Input data schema	<a href="#">View</a>
Runtime environment	spark-2.0
Training date	28 Aug 2017, 3:05 PM

**Deployments** [+ Add Deployment](#)

NAME	DEPLOYMENT TYPE	ACTIONS
DibeticPrediction	Online	...

- Before using the online deployed model, user can test the prediction with test datasets from the “predictions” section in below page.

My Projects > ISTC2017-Lab1 > ISTCLab1

Details Predictions

The predictor simulates scoring your prediction data with an undeployed model.

Spark Service  
Spark-ISTCLab2017

Prediction input data

Glucose  
148

BloodPressure  
72

Insulin  
0

BMI  
33.6

[Predict](#)


Predicted value for Outcome 1.00

Outcome	Predicted Value (%)
1	65.28%
0	34.72%

- Open the online deployments that just deployed to view the online scoring endpoint for integrating in user application or stand-alone online scoring using CURL.

Details Test API

Deployment Details

Name	DabeticPrediction
Type	online
Scoring End Point	<p>https://ibm-watson-ml.mybluemix.net/v3/wml_instances/cb5bd88b-ea9c-4854-b515-55339ef7be7f/published_models/b31e58c4-55ad-4490-8a2d-8a33ece3eefc/deployments/9355a5b1-f1fe-4996-9b8f-8eb1d0771da0/online</p> <p>Submit a POST request to retrieve scoring output. <a href="#">View API specification</a> </p>
Status	ACTIVE
Associated Model	ISTCLab1