#ibmdevconnect

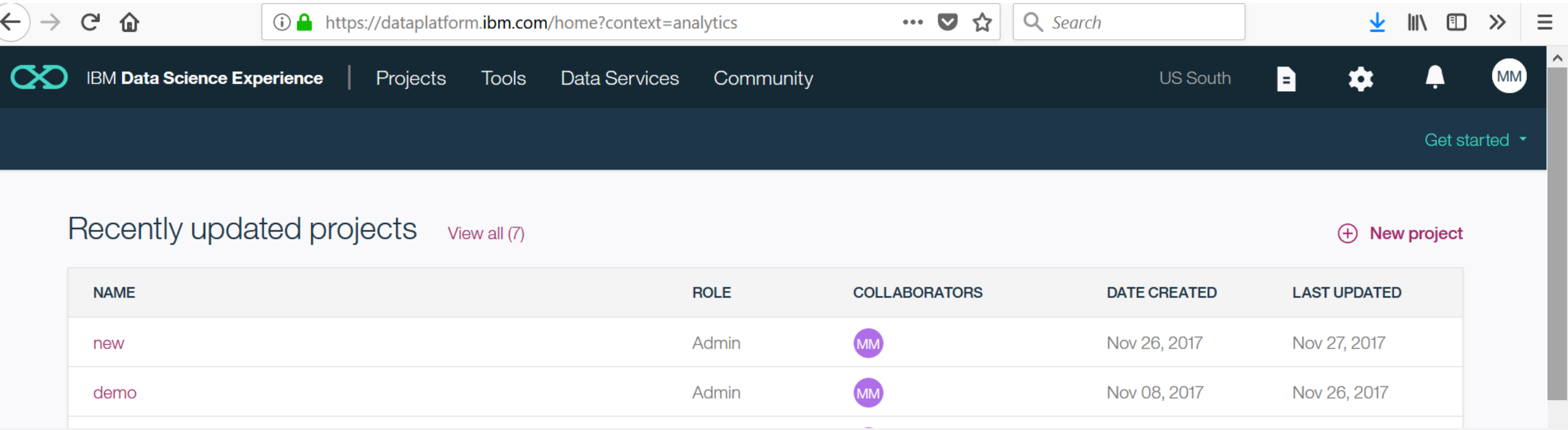
Using IBM DSX

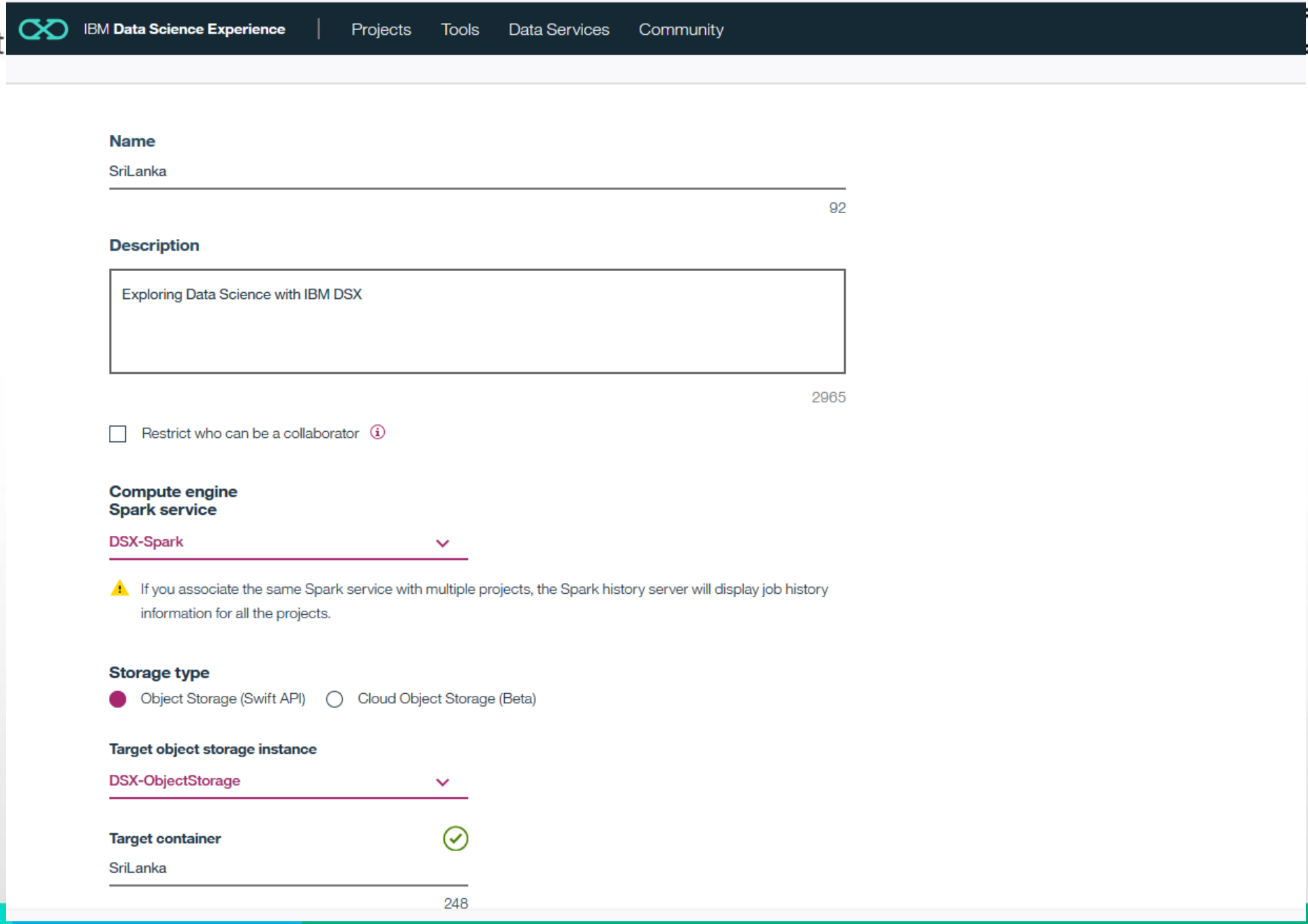
Mani Madhukar
@manimadhukar
https://www.linkedin.com/in/manimadhukar/

Takeaways

- Quick tour of IBM DSX
- Introduction to Jupyter Notebook
- Introduction to Classification And Regression
- Classification using IRIS Dataset
- Introduction to Model
- Predicting using financial data

Started with IBM Data Science Experience




IBM Data Science Experience
Projects
Tools
Data Services
Community
Name
SriLanka
92
Description
Exploring Data Science with IBM DSX
2965
Restrict who can be a collaborator
Compute engine
Spark service
DSX-Spark
If you associate the same Spark service with multiple projects, the Spark history server will display job history information for all the projects.
Storage type
Object Storage (Swift API)
Cloud Object Storage (Beta)
Target object storage instance
DSX-ObjectStorage
Target container
SriLanka
248

Overview | Assets | Bookmarks | Deployments | Collaborators | Settings

SriLanka

Last Updated: Nov 29 2017

0	0	1
Assets	Bookmarks	Collaborators

Date created
Nov 29 2017

Description
Exploring Data Science with IBM DSX

Storage
0% of 5 GB used

Collaborators View all (1)

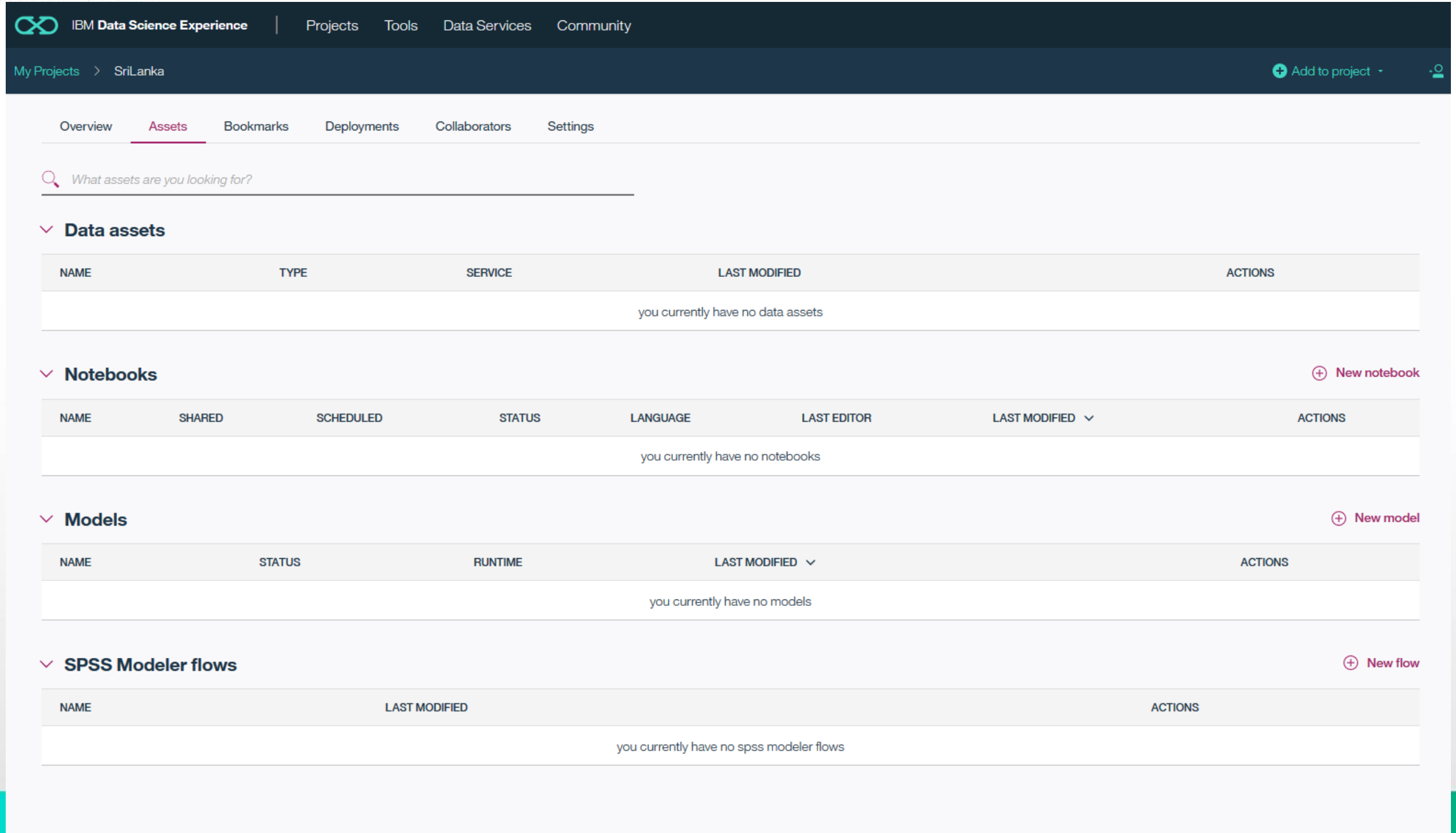
MM Mani Madhukar
Admin

Bookmarks View all (0)

You currently have 0 bookmarks

Recent activity

Alerts related to this project will show here when the project is active.


IBM Data Science Experience
Projects
Tools
Data Services
Community
My Projects
SriLanka
Add to project
Overview
Assets
Bookmarks
Deployments
Collaborators
Settings
What assets are you looking for?
Data assets
NAME
TYPE
SERVICE
LAST MODIFIED
ACTIONS
you currently have no data assets
Notebooks
New notebook
NAME
SHARED
SCHEDULED
STATUS
LANGUAGE
LAST EDITOR
LAST MODIFIED
ACTIONS
you currently have no notebooks
Models
New model
NAME
STATUS
RUNTIME
LAST MODIFIED
ACTIONS
you currently have no models
SPSS Modeler flows
New flow
NAME
LAST MODIFIED
ACTIONS
you currently have no spss modeler flows

IBM **Data Science Experience** | Projects Tools Data Services Community

New model *BETA*

Name

Model name

100

Description

Model description

300

Machine Learning Service

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

Spark Service

DSX-Spark

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** .

Machine Learning

Existing New

Machine Learning

IBM Watson Machine Learning is a full-service Bluemix offering that makes it easy for developers and data scientists to work together to integrate predictive capabilities with their applications. The Machine Learning service is a set of REST APIs that you can call from any programming language to develop applications that make smarter decisions, solve tough problems, and improve user outcomes.

Features

SPSS analytics platform features
SPSS streams management and deployment with realtime scoring and batch processing options.

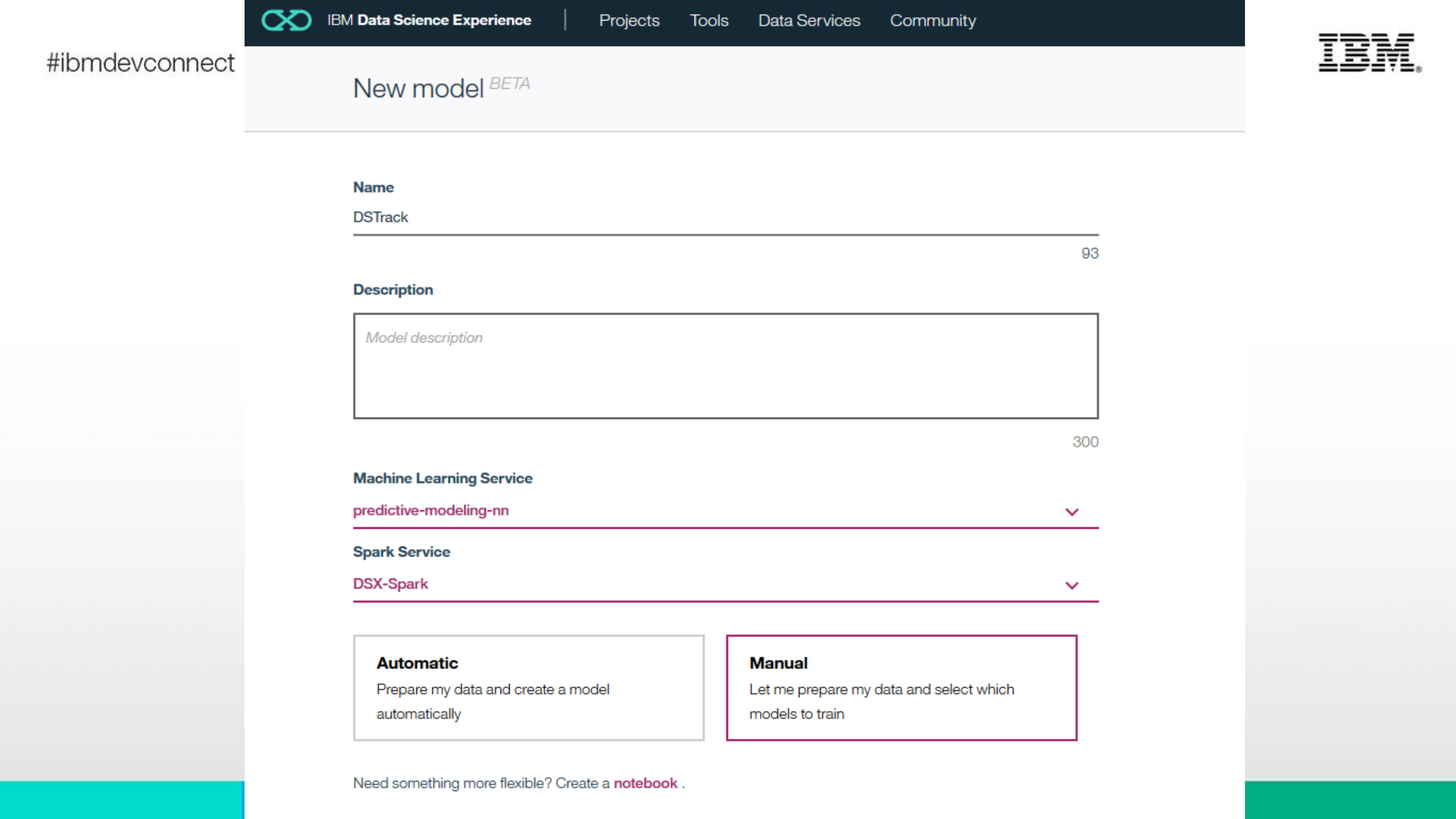
Spark and Python Machine Learning features
Take advantage of Spark MLlib and scikit-learn machine learning models management and deployment - online, batch and streaming.

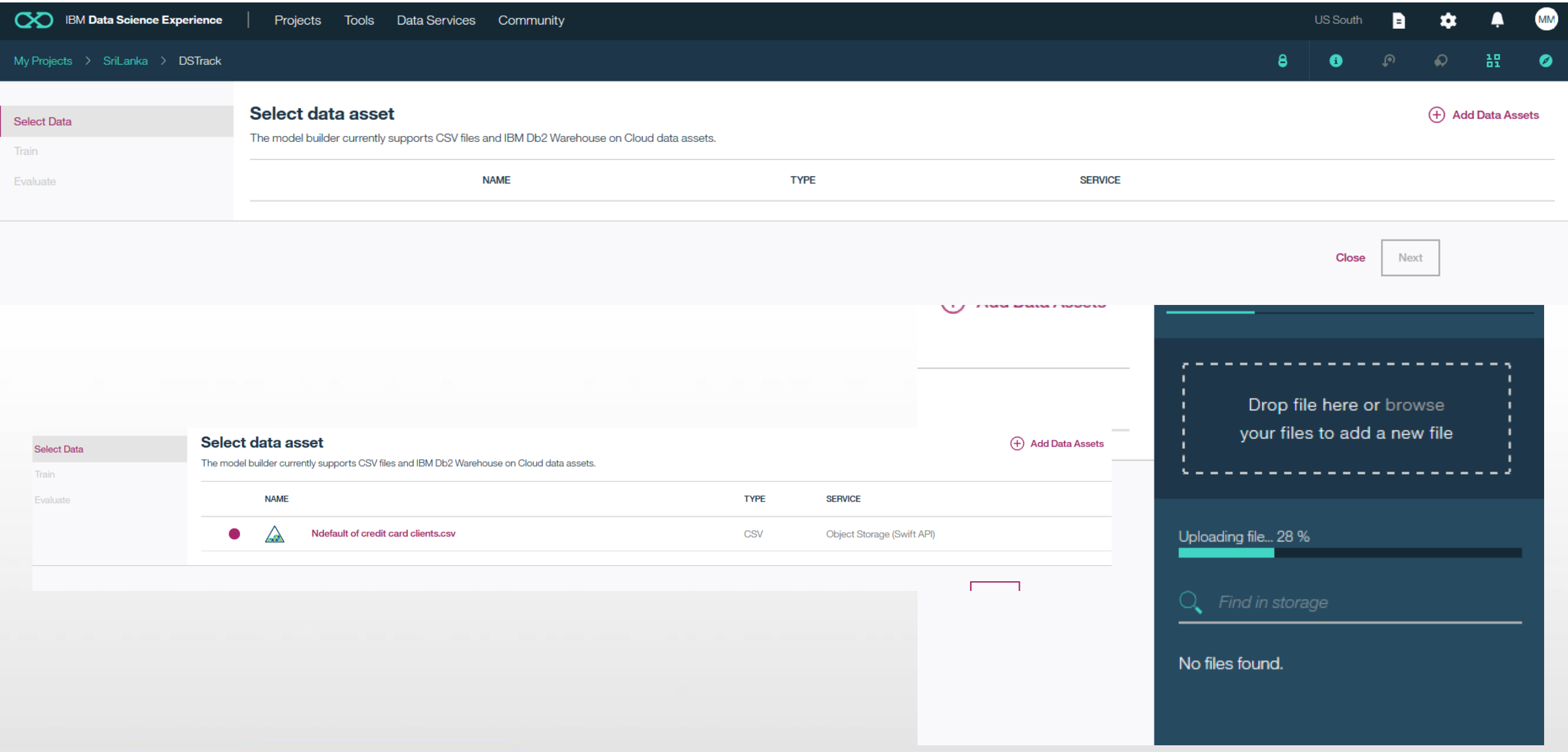
Integration with Data Science Experience
Visit http://datascience.ibm.com. Create and train predictive analytics models with the best tools and the latest expertise in a social environment built by data scientists.

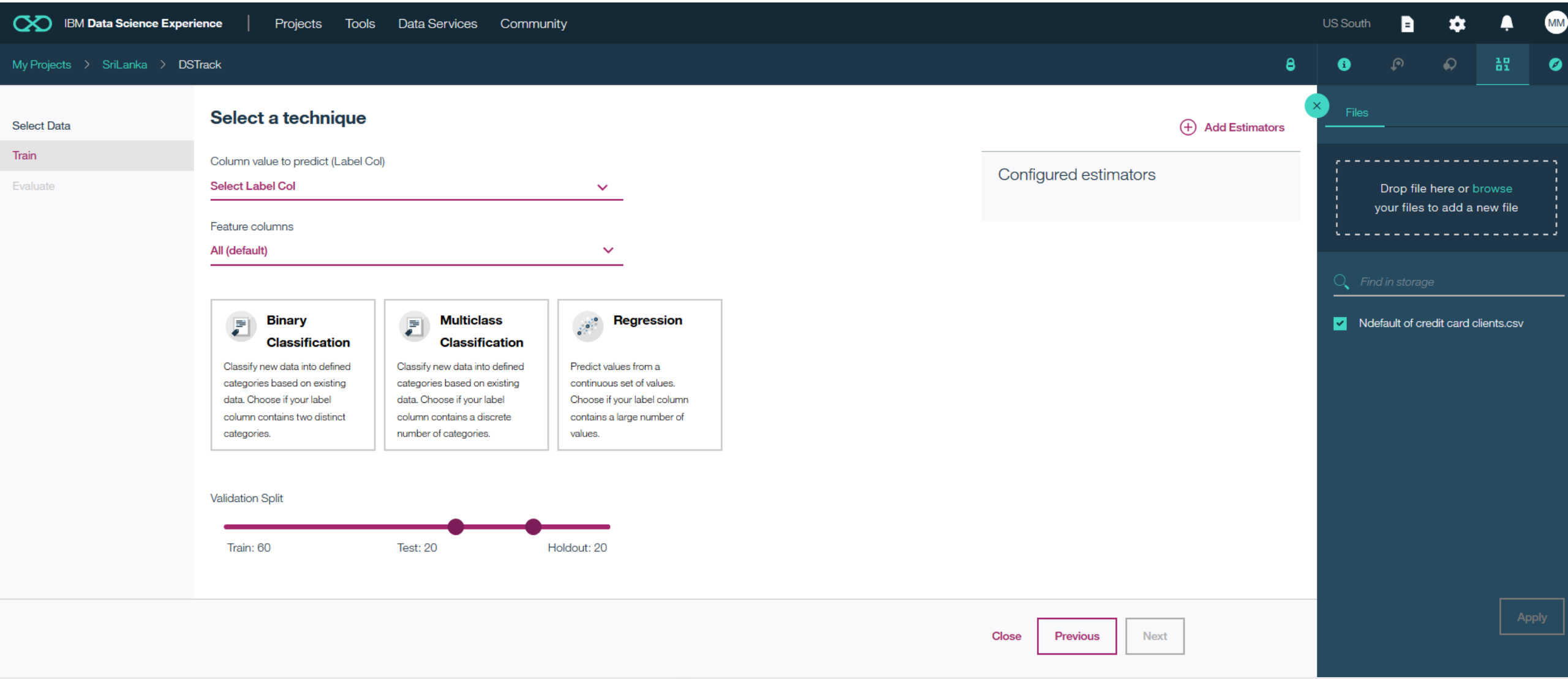
Pricing Plan: Monthly Process shown above reflect the: **United States**

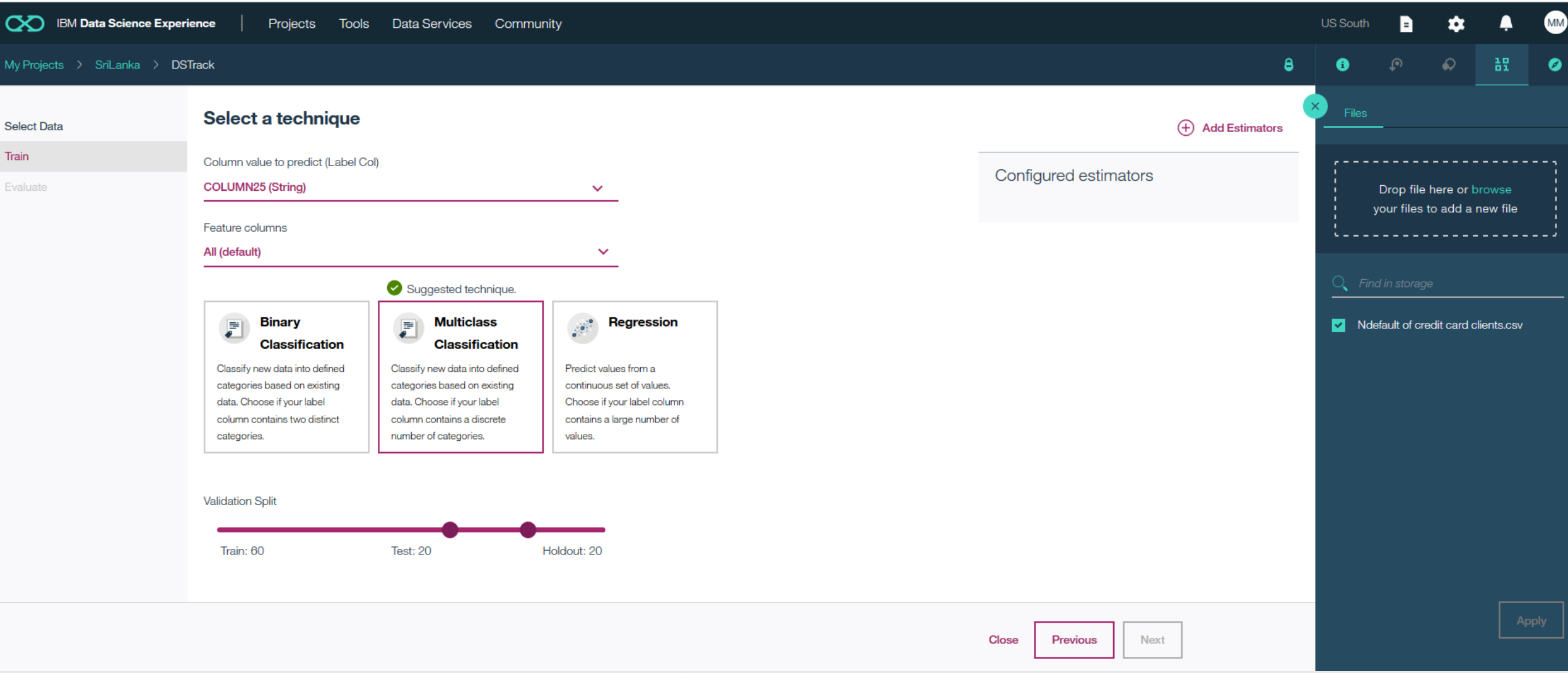
Plan	Features	Pricing
○ Lite	Service instance (5 models per instance) 5,000 predictions 5 compute hours	Free
○ Standard	Predictions Compute hours	$0.5 USD/1,000 predictions $0.45 USD/hour
○ Professional	Service instance 2,000,000 predictions included and then billed per 1,000 predictions 1,000 compute hours included and then billed per compute hour	$1,000 USD/Instance $0.4 USD/1,000 predictions $0.4 USD/hour

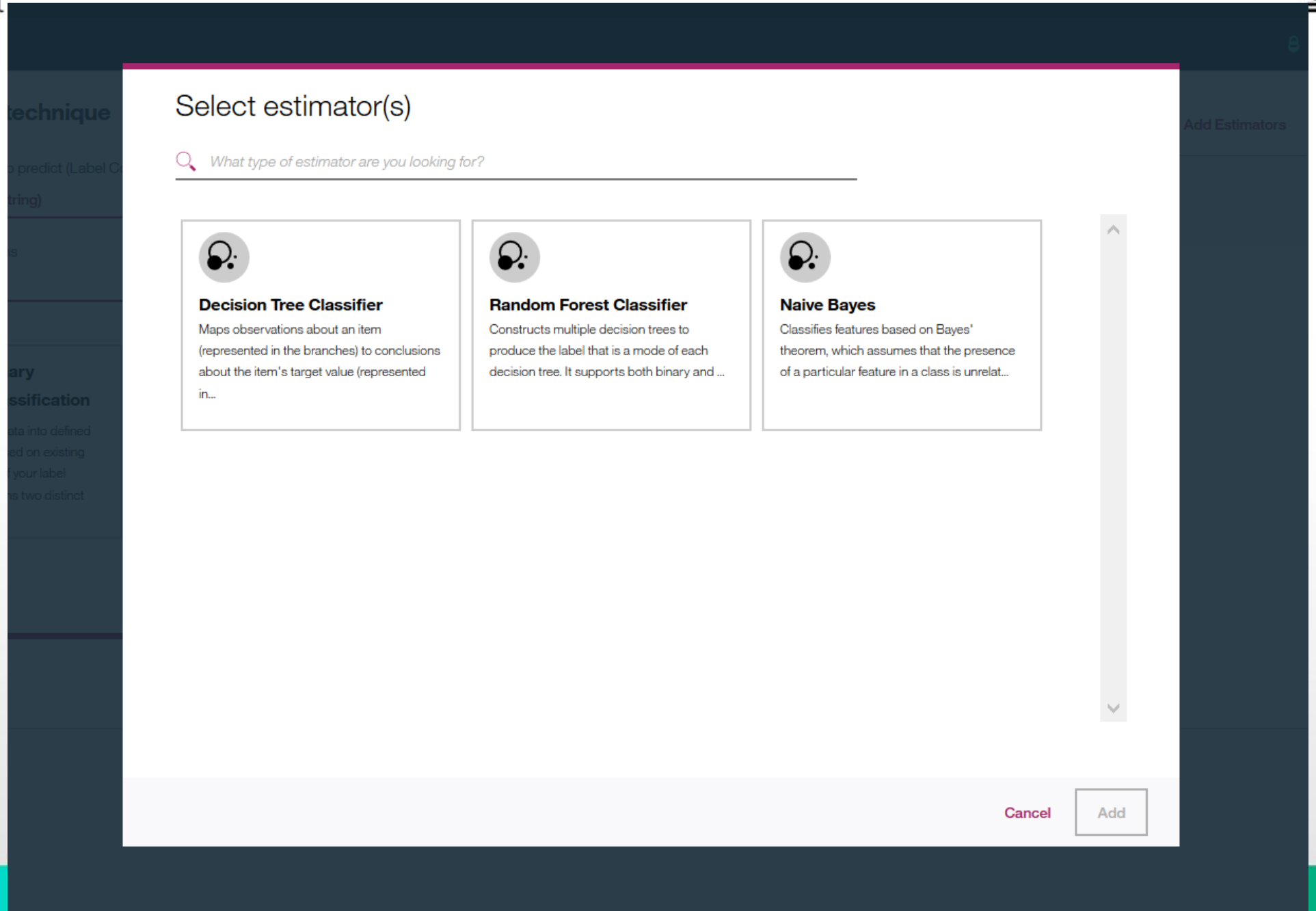
Terms


IBM Data Science Experience
Projects
Tools
Data Services
Community
New model BETA
Name
DSTrack
93
Description
Model description
300
Machine Learning Service
predictive-modeling-nn
Spark Service
DSX-Spark
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 .


IBM Data Science Experience
Projects
Tools
Data Services
Community
US South
MM
My Projects > SriLanka > DSTrack
Select Data
Train
Evaluate
Select data asset
The model builder currently supports CSV files and IBM Db2 Warehouse on Cloud data assets.
Add Data Assets
NAME
TYPE
SERVICE
Close
Next
Drop file here or browse your files to add a new file
Uploading file... 28 %
Find in storage
No files found.
Ndefault of credit card clients.csv
CSV
Object Storage (Swift API)


IBM Data Science Experience
Projects
Tools
Data Services
Community
US South
MM
My Projects > SriLanka > DSTrack
Select Data
Train
Evaluate
Select a technique
Add Estimators
Column value to predict (Label Col)
Select Label Col
Feature columns
All (default)
Configured estimators
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
Files
Drop file here or browse your files to add a new file
Find in storage
Ndefault of credit card clients.csv
Apply
Close
Previous
Next


IBM Data Science Experience
Projects
Tools
Data Services
Community
US South
MM
My Projects
SriLanka
DSTrack
Select Data
Train
Evaluate
Select a technique
Column value to predict (Label Col)
COLUMN25 (String)
Feature columns
All (default)
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
Add Estimators
Configured estimators
Files
Drop file here or browse your files to add a new file
Find in storage
Ndefault of credit card clients.csv
Apply
Close
Previous
Next


Select estimator(s)
What type of estimator are you looking for?
Decision Tree Classifier
Maps observations about an item (represented in the branches) to conclusions about the item's target value (represented in...
Random Forest Classifier
Constructs multiple decision trees to produce the label that is a mode of each decision tree. It supports both binary and ...
Naive Bayes
Classifies features based on Bayes' theorem, which assumes that the presence of a particular feature in a class is unrelat...
Cancel
Add

Select Data
Train
Evaluate

Select model

	ESTIMATOR TYPE	STATUS	PERFORMANCE	WEIGHTED TRUE POSITIVE RATE	WEIGHTED FALSE POSITIVE RATE	WEIGHTED PRECISION	WEIGHTED F MEASURE	WEIGHTED RECALL	LAST EVALUATION	ACTIONS
●	RandomForestClassifier	Training...	...	...	...	...	...	...	...	...

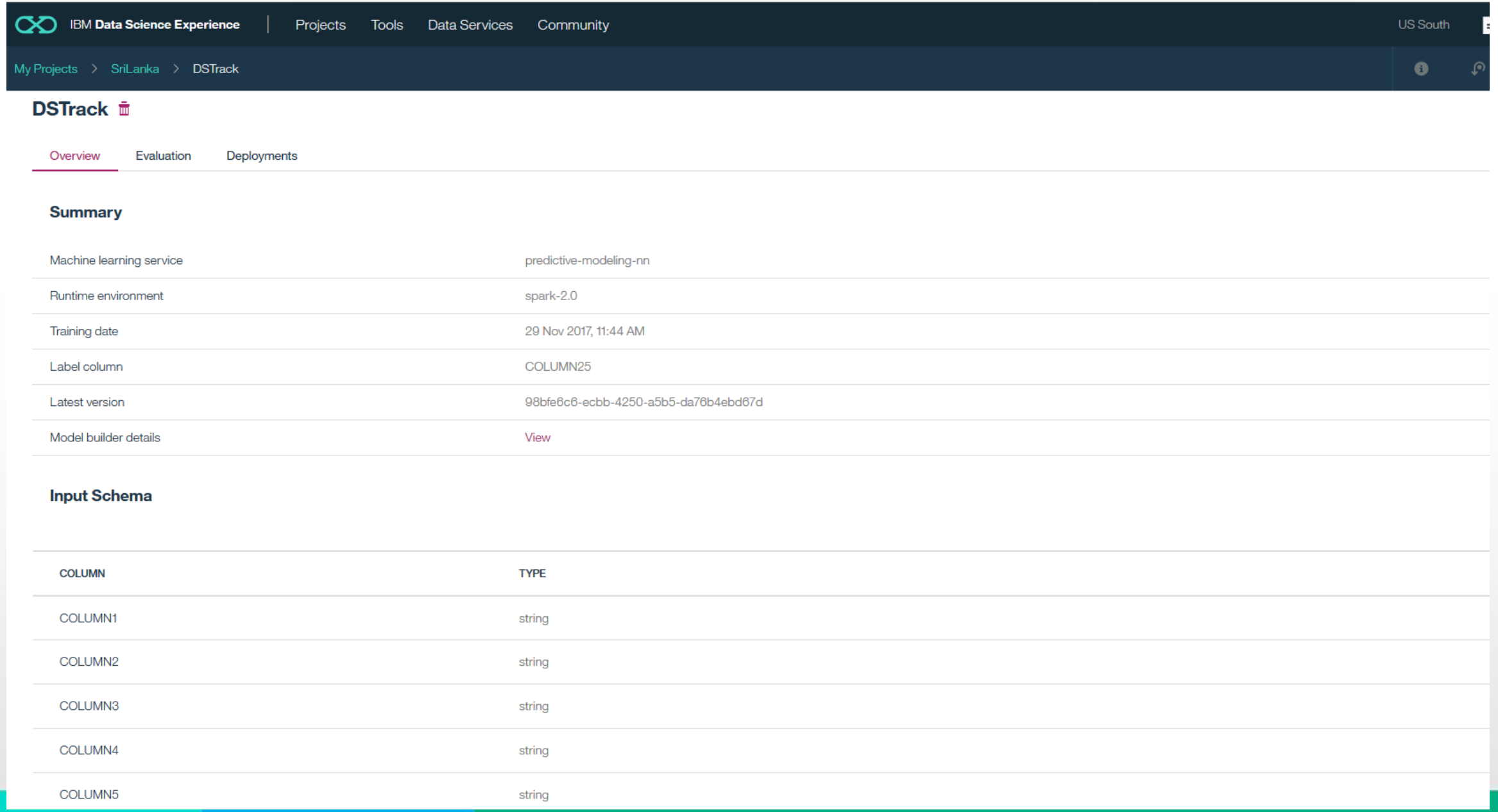
Close Previous Save

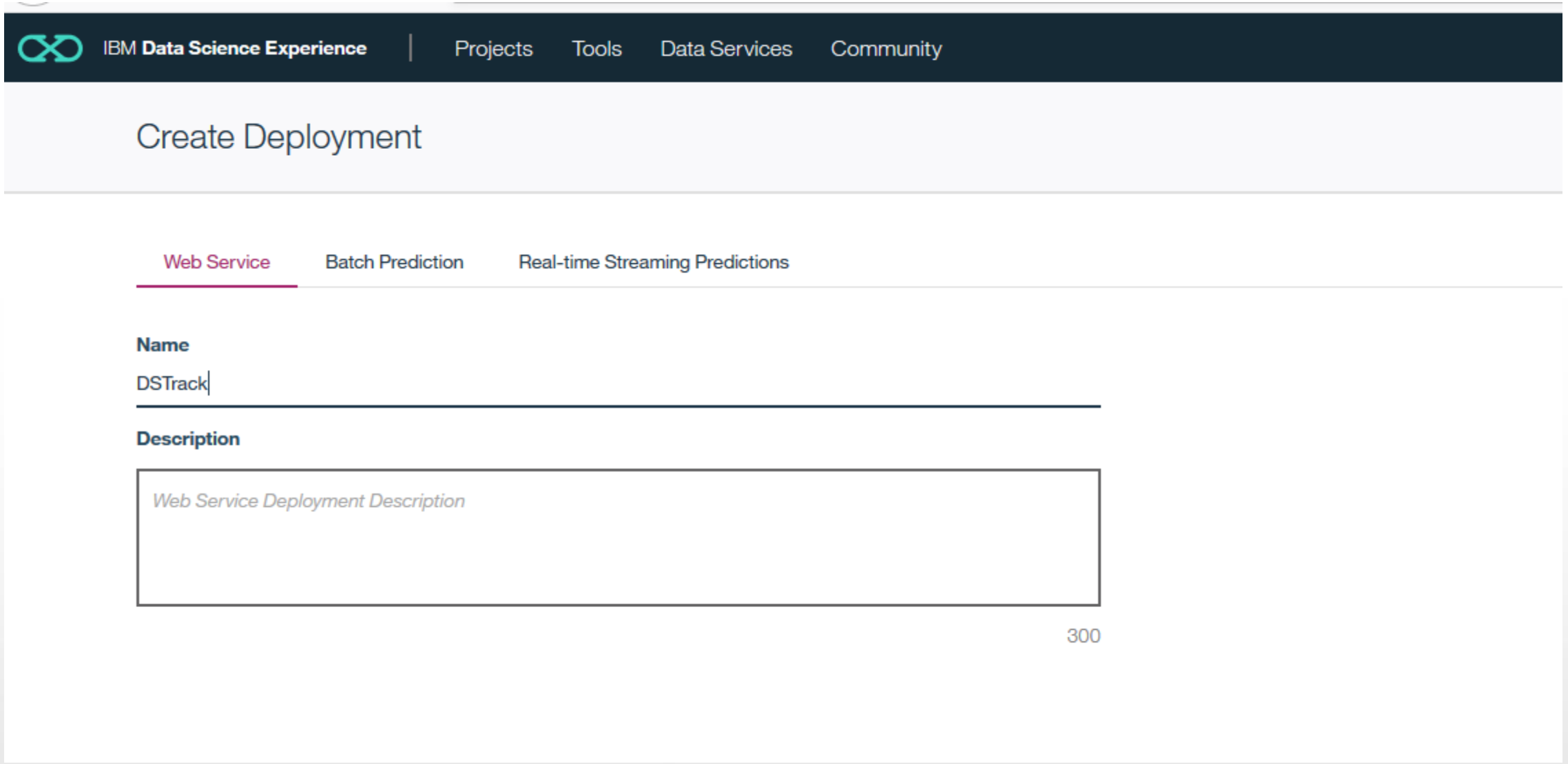
Select Data
Train
Evaluate

Select model

	ESTIMATOR TYPE	STATUS	PERFORMANCE	WEIGHTED TRUE POSITIVE RATE	WEIGHTED FALSE POSITIVE RATE	WEIGHTED PRECISION	WEIGHTED F MEASURE	WEIGHTED RECALL	LAST EVALUATION	ACTIONS
●	RandomForestClassifier	Trained & Evaluated	Good	0.81885	0.5096	0.80207	0.7966	0.81885	29 Nov 2017, 10:43 AM	⋮

Close Previous Save


IBM Data Science Experience
Projects
Tools
Data Services
Community
US South
My Projects
SriLanka
DSTrack
DSTrack
Overview
Evaluation
Deployments
Summary
Machine learning service
predictive-modeling-nn
Runtime environment
spark-2.0
Training date
29 Nov 2017, 11:44 AM
Label column
COLUMN25
Latest version
98bfe6c6-ecbb-4250-a5b5-da76b4ebd67d
Model builder details
View
Input Schema
COLUMN
TYPE
COLUMN1
string
COLUMN2
string
COLUMN3
string
COLUMN4
string
COLUMN5
string


IBM Data Science Experience
Projects
Tools
Data Services
Community
Create Deployment
Web Service
Batch Prediction
Real-time Streaming Predictions
Name
DSTrack
Description
Web Service Deployment Description
300

DSTrack

Overview | Implementation | Test

Deployment

Name	DSTrack
Type	Web Service
Deployment ID	06e01496-93a5-4caf-bfe1-0264c6db82fb
Status	ACTIVE
Machine learning service	predictive-modeling-nn
Created	29 Nov 2017 11:52am
Last modified	29 Nov 2017 11:53am

Model

Name	DSTrack
Model ID	5f6489d9-5189-41b4-8947-52e29e2abbe8
Version ID	98bfe6c6-ecbb-4250-a5b5-da76b4ebd67d
URL	https://ibm-watson-ml.mybluemix.net/v2/artifacts/models/5f6489d9-5189-41b4-8947-52e29e2abbe8/versions/98bfe6c6-ecbb-4250-a5b5-da76b4ebd67d

Overview | Implementation | Test

Input data

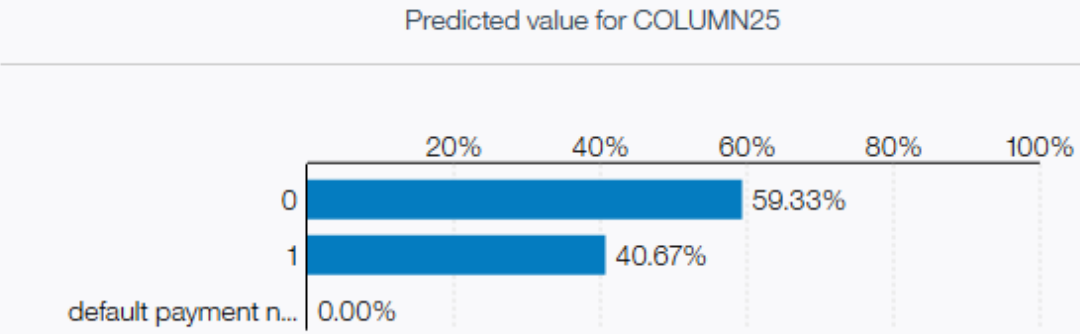
COLUMN4
2

COLUMN5
2

COLUMN6
2

COLUMN7
X6

Predict



Introduction to Jupyter Notebooks

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more

Language of choice

The Notebook has support for over 40 programming languages, including Python, R, Julia, and Scala.

Interactive output

Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.

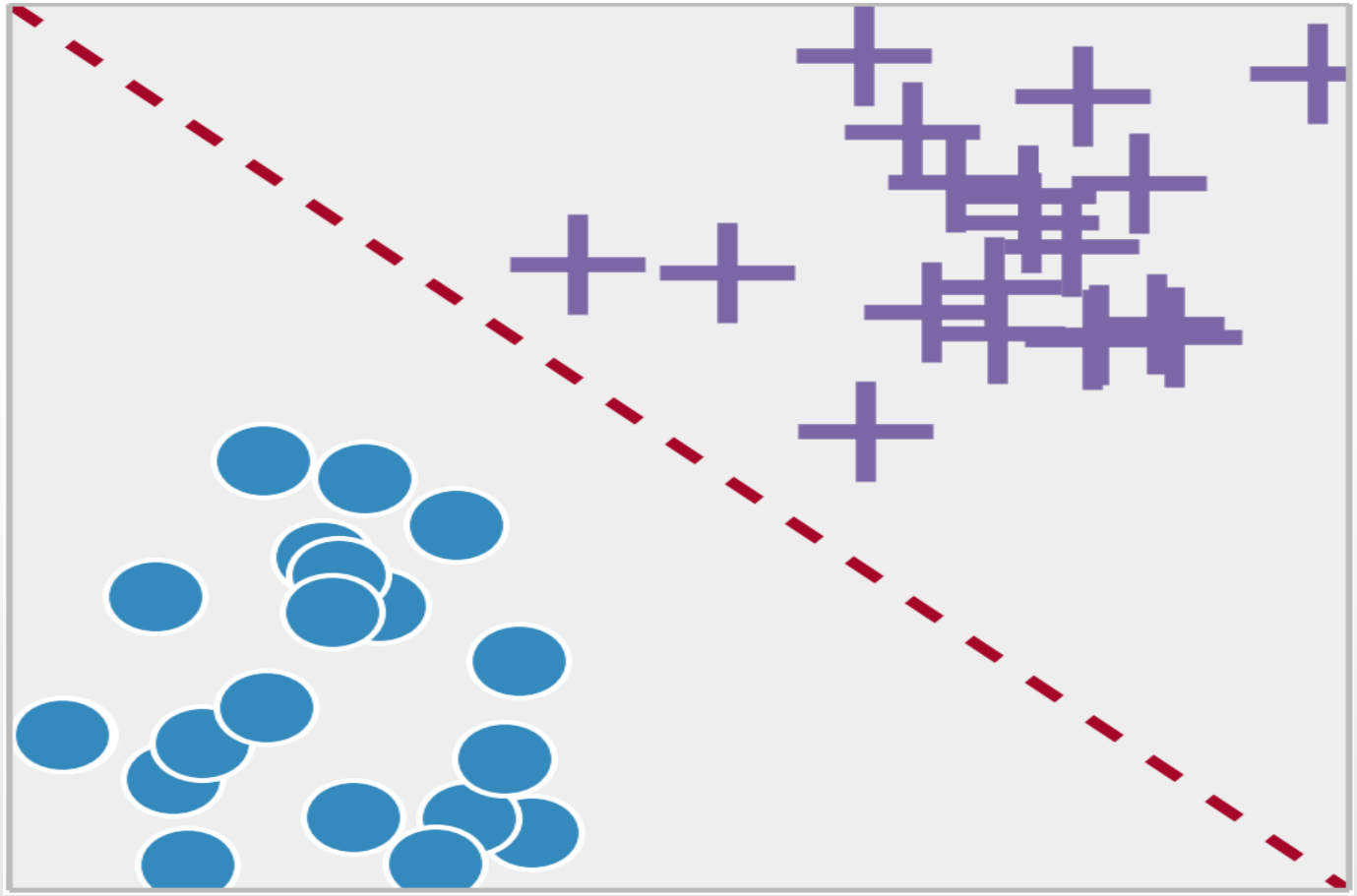
Share notebooks

Notebooks can be shared with others using email, Dropbox, GitHub and the Jupyter Notebook Viewer.

Big data integration

Leverage big data tools, such as Apache Spark, from Python, R and Scala. Explore that same data with pandas, scikit-learn, ggplot2, TensorFlow.

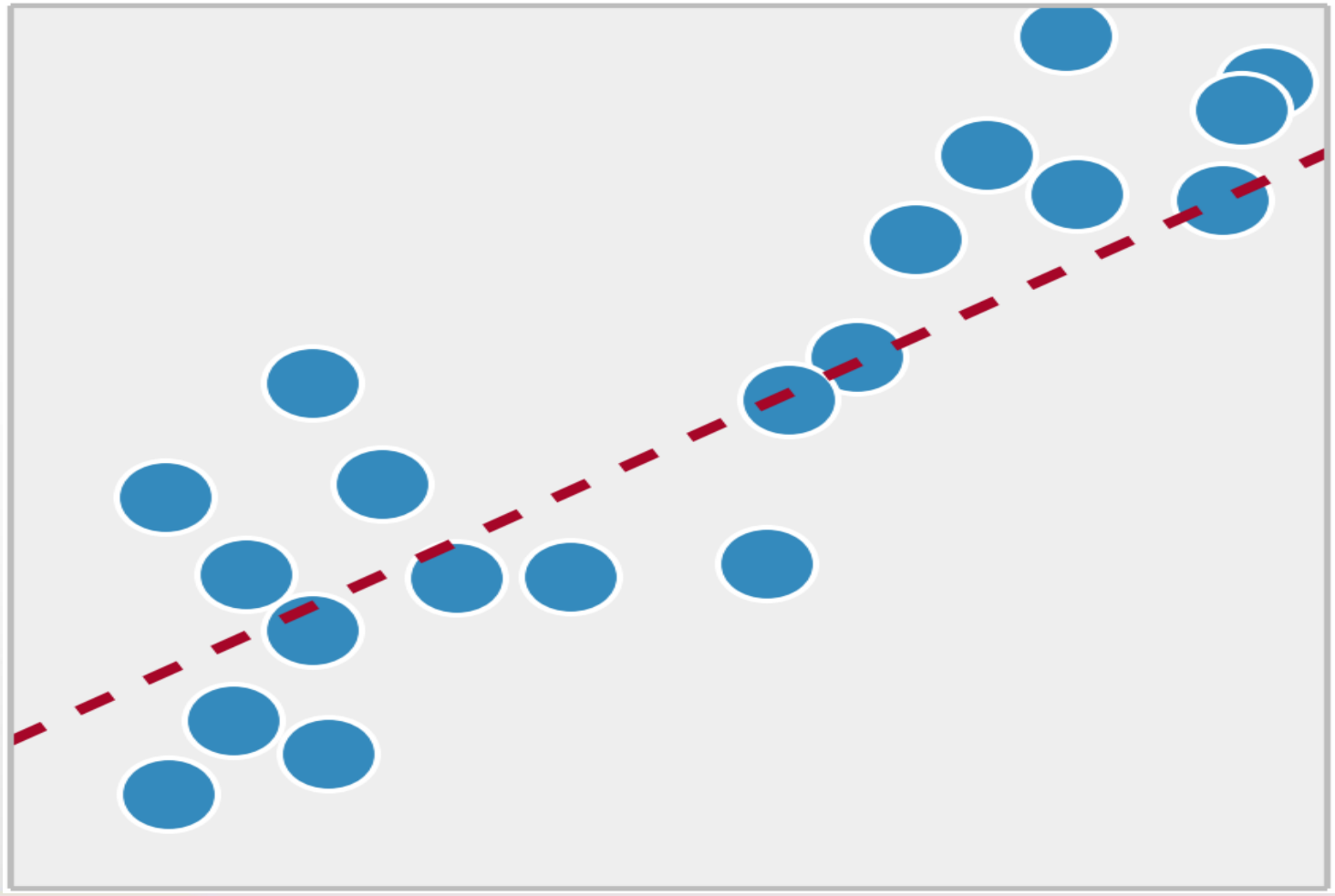
Classification: Given a known relationship, identify the class that the data belongs.



Applications of Classification

- Spam Filtering
- Detecting Cancer
- OCR (Optical Character Recognition)
- Email Sorting
- Recommender Systems

Regression: Given a set of data, find the best relationship that represent the set of data.

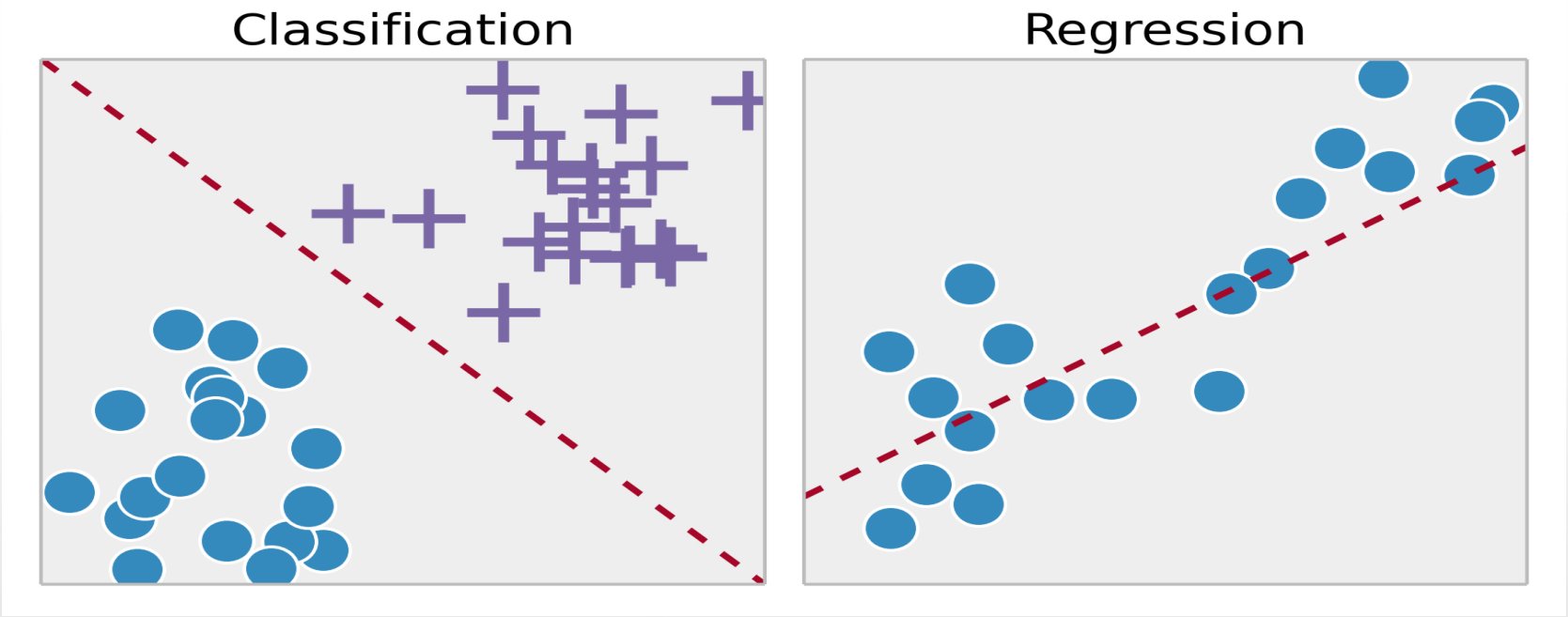


Applications of Regression

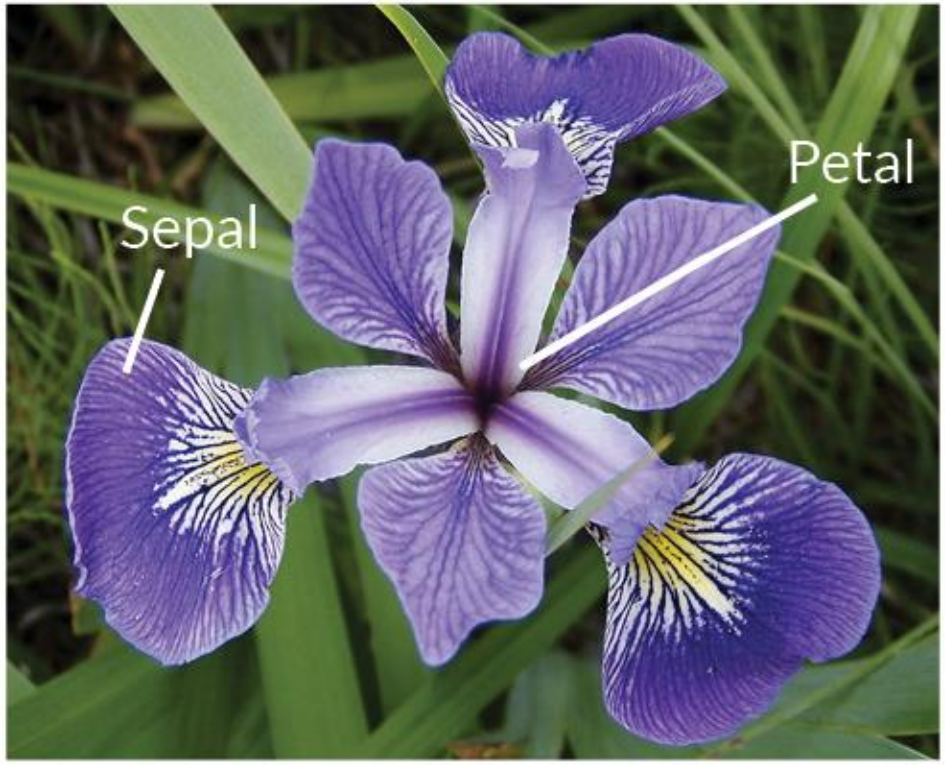
- Predicting Stock Prices
- Estimating Real Estate price boom
- Optimizing machine running times in factories
- Predicting Weather
- User Preference learning

In summary

Classification: The output variable takes class labels.
Regression: The output variable takes continuous values.



IRIS Classifier



Iris Versicolor

Iris Setosa

Iris Virginica

IRIS Dataset

- This data sets consists of 3 different types of irises' (Setosa, Versicolour and Virginica) petal and sepal length, stored in a 150x4 Table
- The rows being the samples and the columns being: Sepal Length, Sepal Width, Petal Length and Petal Width

Get in touch

Please leave your valuable feedback here – http://devconfeedback.mybluemix.net/

Access the resources on - https://github.com/IBMDevConnect17/colombo17

Get connected with your fellow developers @ https://developer.ibm.com/in/

Join our Slack team @ https://ibmdevconnect.slack.com .

Send in your request to - http://ibm.biz/slackrequest

You can also join our Meetup group

Mumbai : https://www.meetup.com/Cloud-Mumbai-Meetup/

Hyderabad: https://www.meetup.com/Hyderabad-Cognitive-with-Cloud

Bangalore : https://www.meetup.com/IBMDevConnect-Bangalore

IBM Code repository- https://developer.ibm.com/code/patterns/