

Watson Studio

*A walk through of the lifecycle of a data scientist in the era of
Cloud Native development*

Justin McCoy

@mccoyjus

justin.mccoy@us.ibm.com

IBM
CODE
—

Agenda

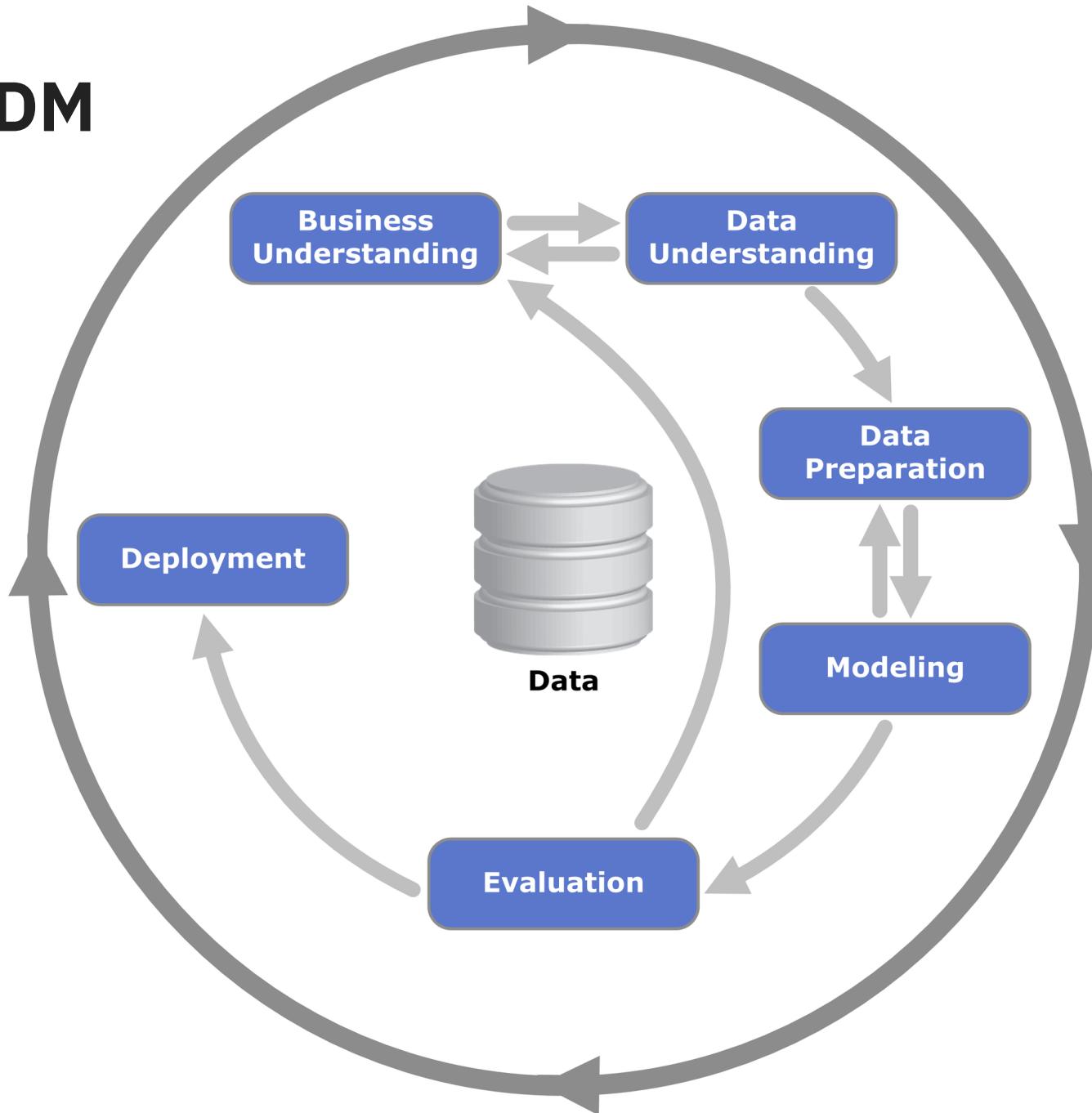
- Challenges faced by Wrangles of data
- Tools and libraries
- Walkthrough: Automatic Model Builder
- Lab: RNN Predicting Oil Prices
- Lab: RNN Predicting Oil Prices Model Deployment



How many people
here consider
themselves Data
Scientists?

What is Data Science and what does it mean to be a Data Scientist?

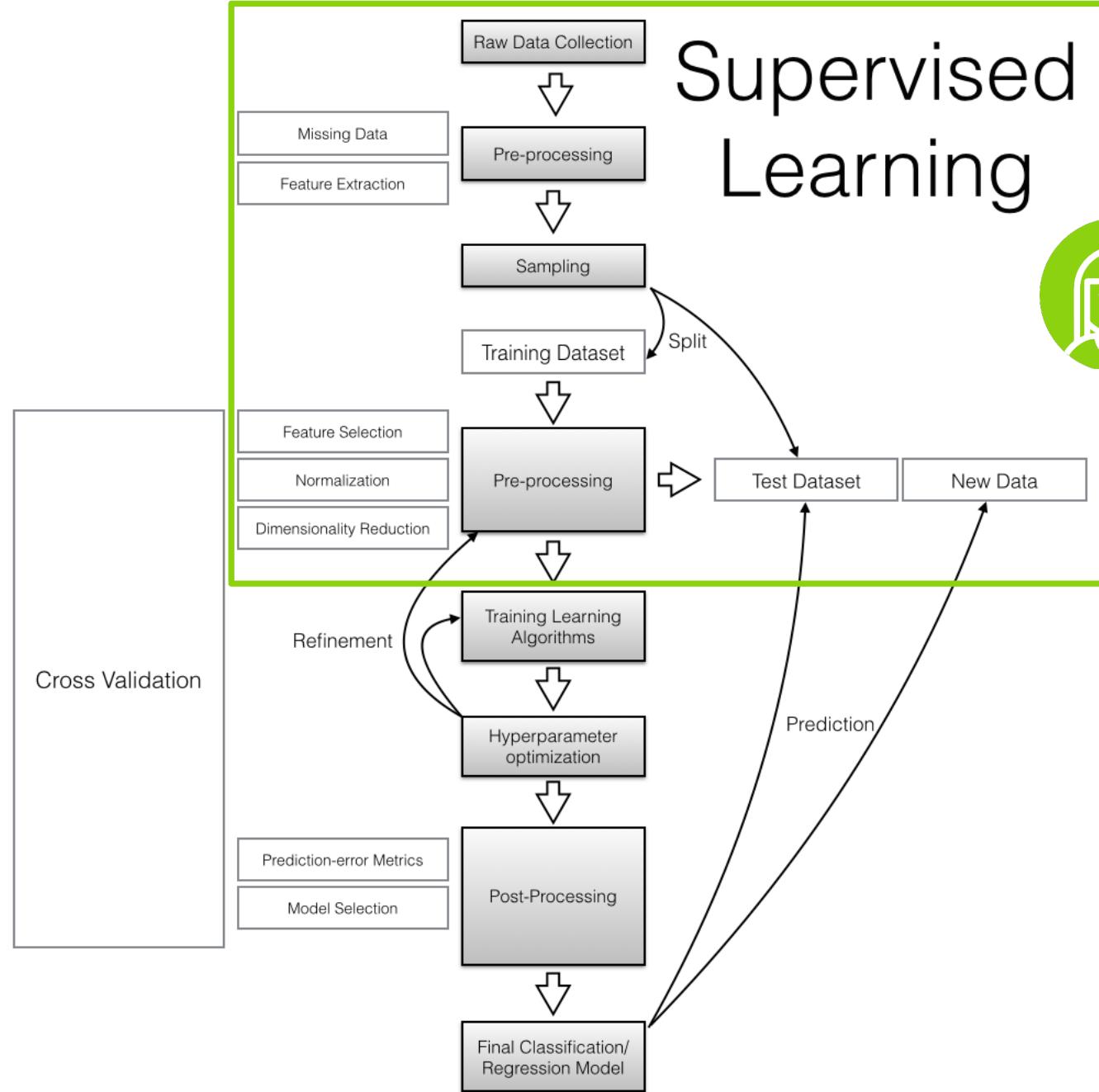
CRISP-DM



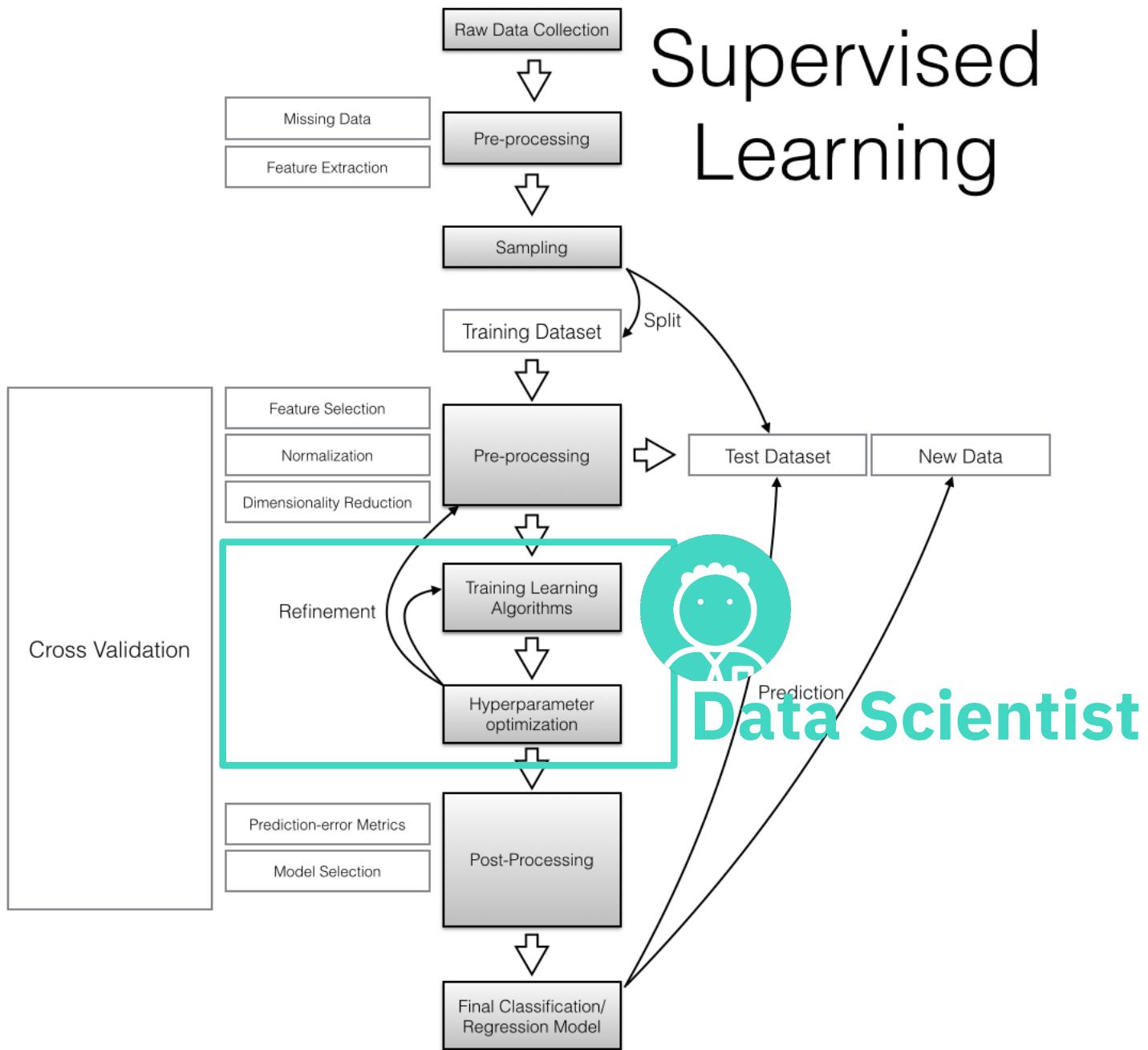
Supervised Learning



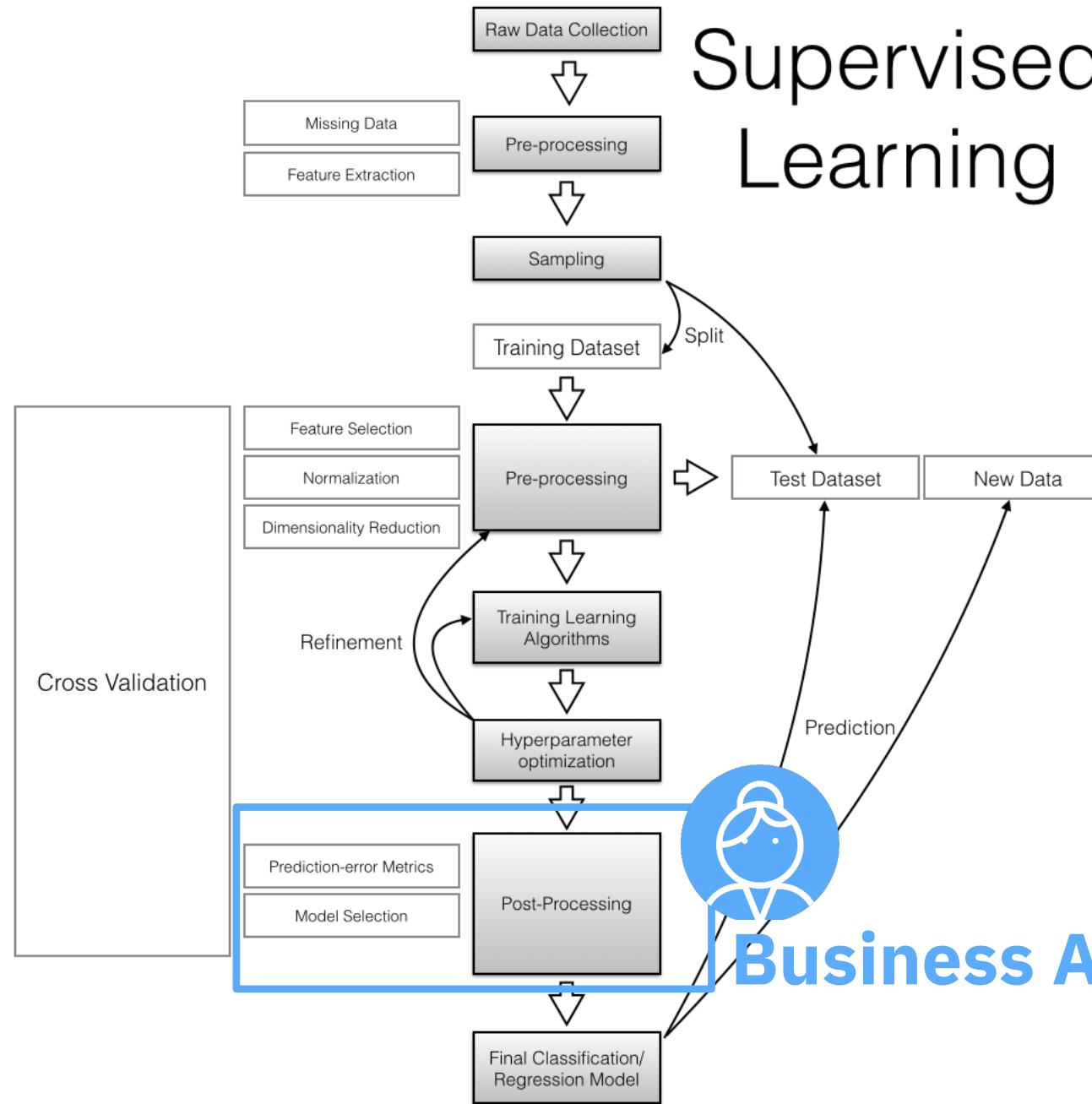
Data Engineer



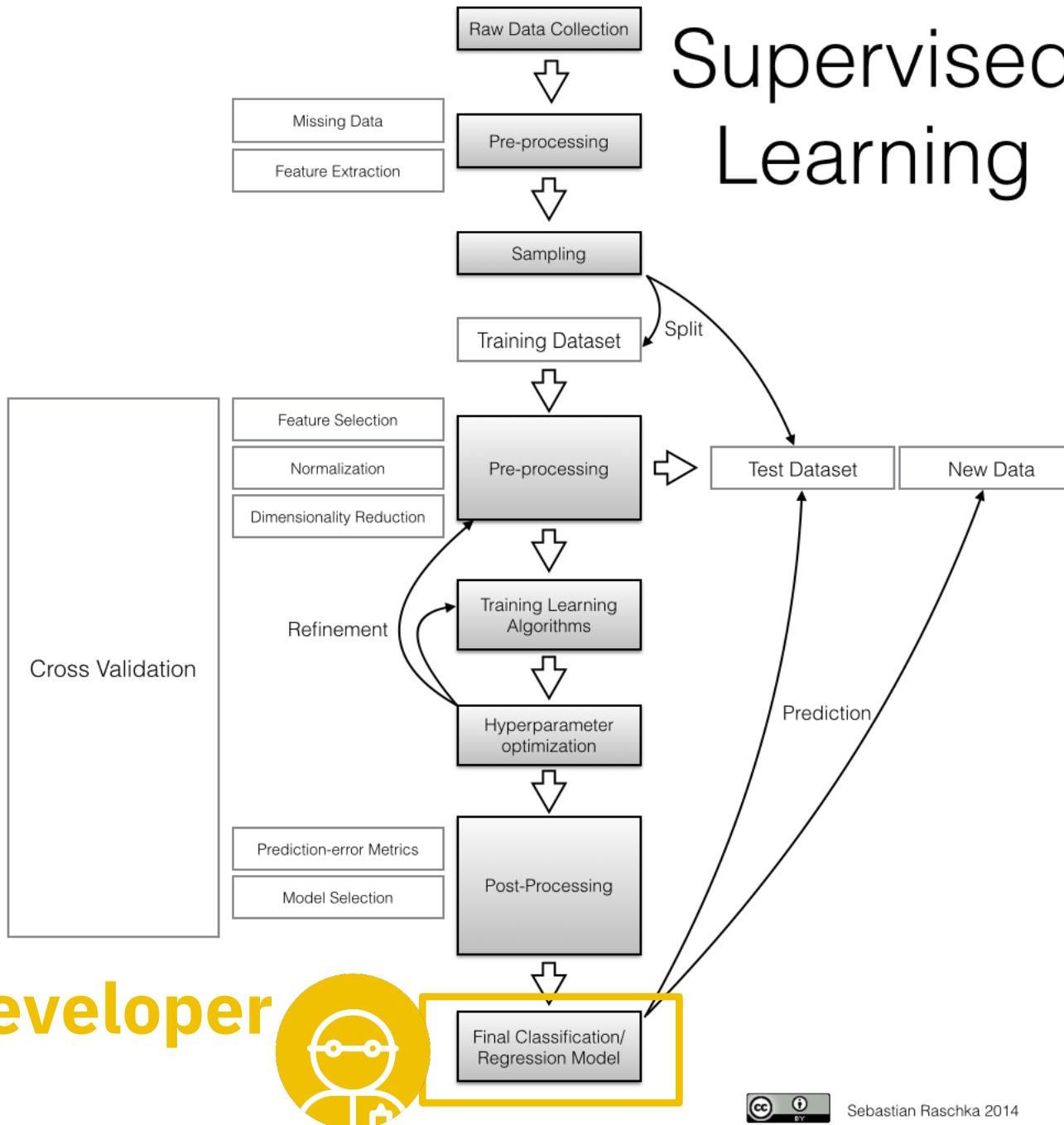
Supervised Learning



Supervised Learning



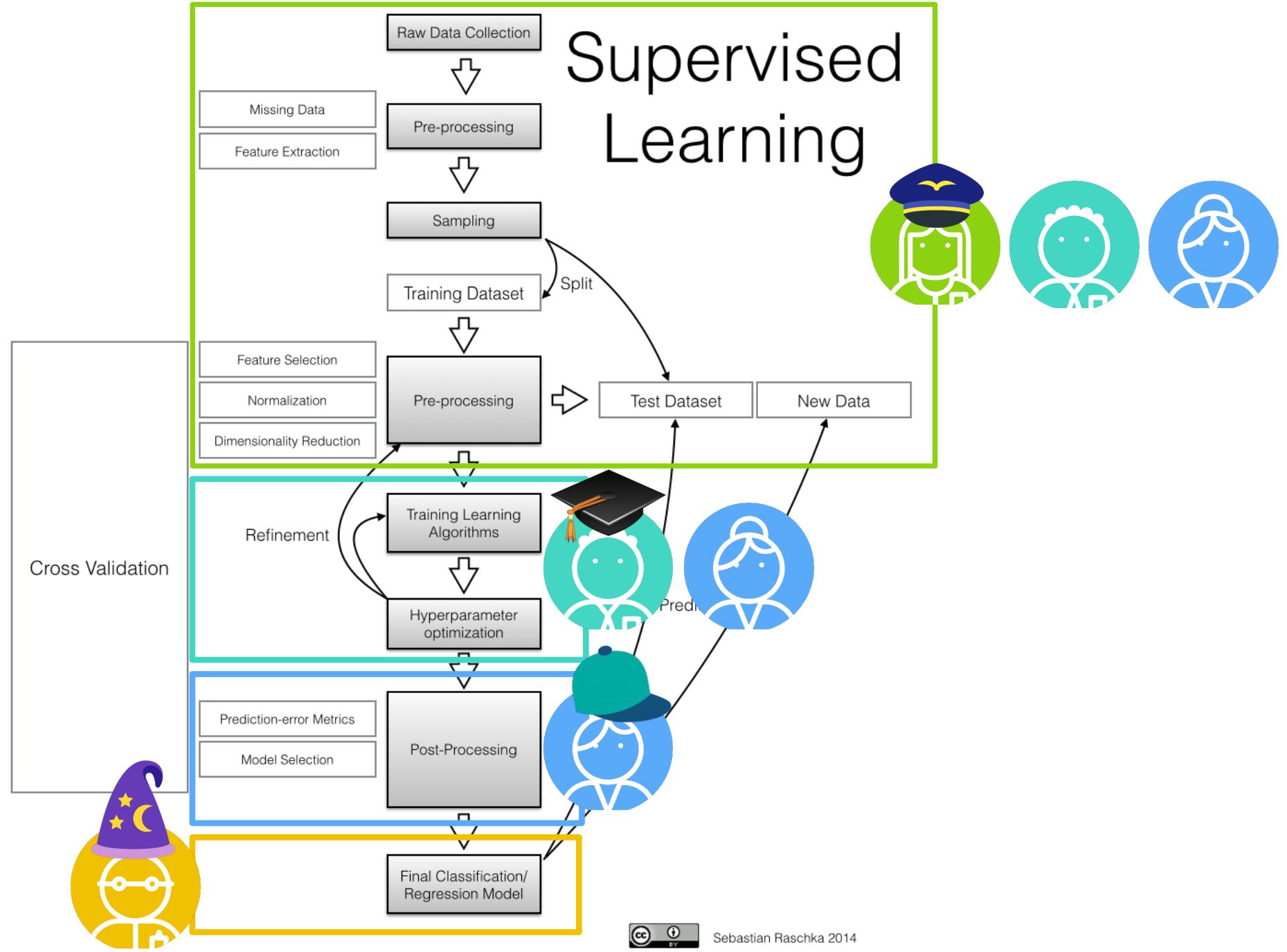
Supervised Learning



App Developer



Supervised Learning



What does
this mean for
you?



Tools of the trade





Welcome Justin!

Watson Studio and Watson Knowledge Catalog are both part of IBM Watson.

Get started with key tasks

[New project](#)[Catalog and find data](#)[Refine data](#)[New notebook](#)[Deep learning](#)[Govern data](#)[Hide ▾](#)

Data, Connections, Catalogs

The screenshot shows the IBM Watson Data Catalog interface. At the top, there's a navigation bar with links for IBM Watson, Projects, Tools, Catalog, Community, Services, US South, and user profile icons. Below the navigation bar, the title "MyCo Data Catalog" is displayed, along with "Add to Catalog" and other account-related buttons.

The main content area has tabs for "Browse Assets", "Usage Statistics", "Access Control", and "Settings". A search bar asks "What assets are you looking for?". Below the search bar, there are three filter categories: "Watson Recommends", "Highly Rated", and "Recently Added".

The "Recently Added" section displays four data assets:

- 2017 U.S. Auto Claims Satisfaction Study J.D. Power**: Owner: Ricardo Buglio, Added: Mar 15, 2018 11:04 AM. Tags: insurance, auto. Rating: 5 stars.
- Vehicle Insurance Documentation**: Owner: Ricardo Buglio, Added: Mar 15, 2018 10:35 AM. Tags: insurance, auto. Rating: 5 stars.
- CompetitorProximity.csv**: Owner: Carmen Ruppach, Added: Mar 13, 2018 9:48 PM. Rating: 5 stars.
- Great Outdoor Customer V**: Owner: Carmen Ruppach, Added: Mar 13, 2018 9:33 PM. Tags: GOSALES. Rating: 5 stars.

A "Collapse" button is located at the top right of this section. To the left of the recently added section is a "Filter" sidebar with an "Asset types" section containing checkboxes for Data Asset (60), Notebook (4), Connection (3), and dashboard (2). There's also a link to "View all available tags (42)".

The bottom half of the screen shows the "Available Assets" section, which lists 69 assets. The table has columns for NAME, OWNER, TAGS, TYPE, and DATE ADDED. Some rows show small profile pictures next to the owner names. On the far right of the table, there are three vertical dots and a grid icon for each row. To the right of the table is a green circular icon with a white message icon and a red notification dot.

NAME	OWNER	TAGS	TYPE	DATE ADDED
2017 Small Business Banking L...	JL Jay Limburn	banking, loan	Data Asset	Feb 19, 2018
2017 U.S. Auto Claims Satisfact...	Ricardo Buglio	insurance, auto	Data Asset	Mar 15, 2018
ANCESTRY	JL Jay Limburn	discovered, SAMPLES	Data Asset	Dec 13, 2017

Environments

The screenshot shows the IBM Watson Environments interface. At the top, there's a navigation bar with links for Projects, Tools, Catalog, Community, Services, Docs, Support, Manage, and a user profile icon. Below the navigation bar, the URL path is shown as My Projects / Test / Environments / My Custom Anaconda Runtime. On the right side of the header, there are several small icons for information, support, manage, and notifications.

The main content area has a title "My Custom Anaconda Runtime". Below it, a message says "There is no description available for this environment." and "Last updated: 16 Apr 2018".

A "Summary" section contains a table with the following data:

Environment	My Custom Anaconda Runtime
Creator	Justin Mccoy
Language	Python 3.5
Hardware configuration	16 vCPU and 64 GB RAM
Software configuration	Default Python 3.5

Below the summary is a "New notebook" button with a plus sign and the text "New notebook".

The interface is divided into two main sections: "Software configuration details" and "Customization".

Software configuration details: This section displays a large block of JSON configuration code:

```
name: DSX-Python35
channels:
- defaults
dependencies:
- alabaster=0.7.10
- anaconda-client=1.6.5
- anaconda-project=0.8.0
- asn1crypto=0.22.0
- astroid=1.5.3
- astropy=2.0.2
- babel=2.5.0
- backports
- backports.shutil_get_terminal_size=1.0.0
- backports.weakref=1.0rc1
- beautifulsoup4=4.6.0
- biopython=1.69
- bitarray=0.8.1
- bkcharts=0.2
- blaze=0.11.3
- bleach=2.0.0
- bokeh=0.12.10
- boto=2.48.0
- boto3=1.14.7
- botocore=1.7.20
- bottleneck=1.2.1
- bz2file=1.0.6
```

Customization: This section contains a text area with placeholder comments and a "Customization" button:

```
# Please add conda channels here
channels:
- defaults

# Please add conda packages here
dependencies:

# Please add pip packages here
# To add pip packages, please comment out the next line
#- pip:
|
```

At the bottom of the customization area are "Apply" and "Cancel" buttons. A note below the customization area states: "With the specified customization you can modify the default software configuration below. The changes are applied on top of the software configuration. Learn more"

Experiments

IBM Watson Projects Tools Community Services Docs Support Manage JM

My Projects / BP Workshops / Oil Price RNN Experiment

Oil Price RNN Experiment Delete Add training runs

Training Runs Compare Runs Overview

Compare runs

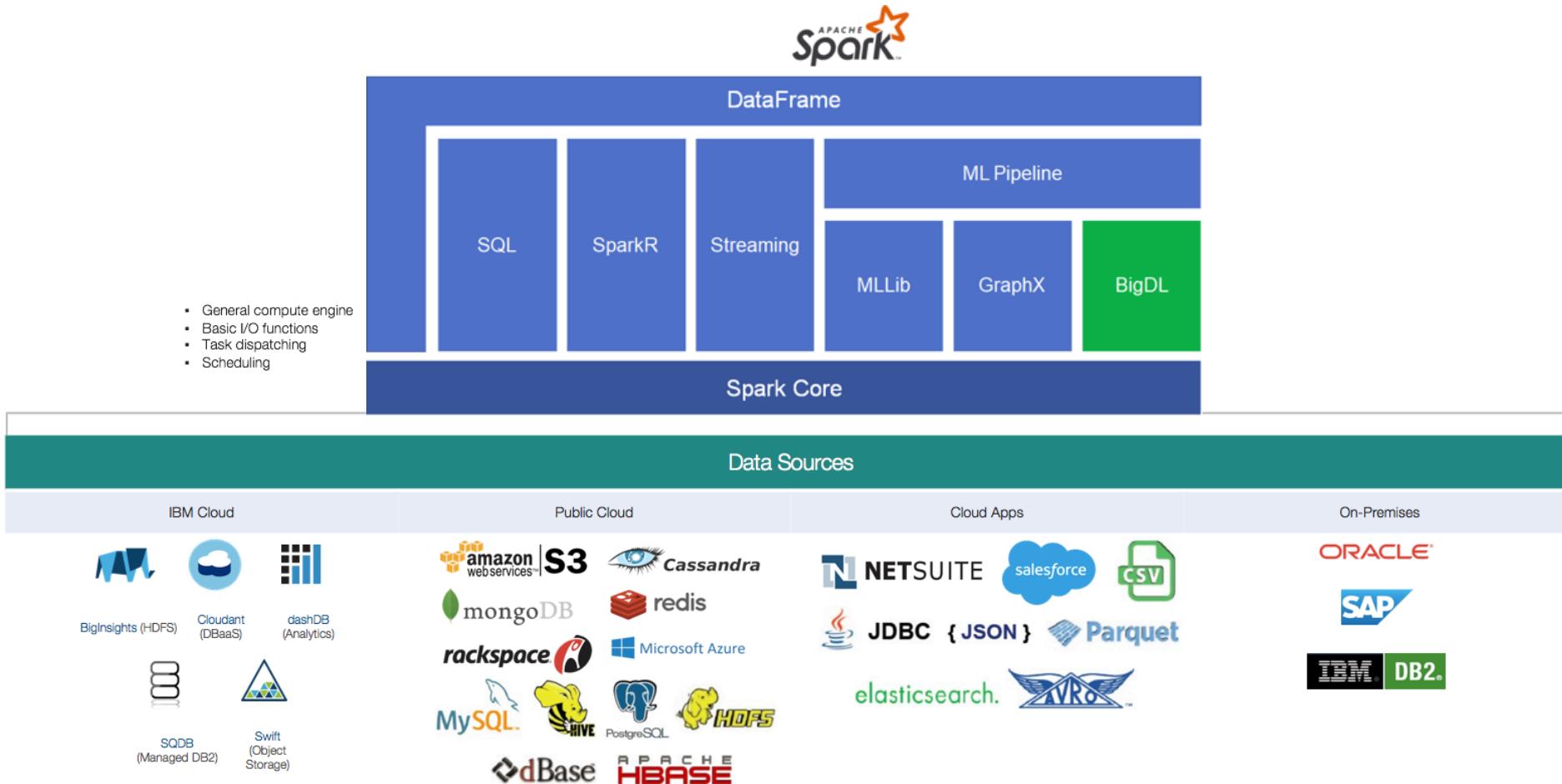
NAME	STATUS	LOSS ▲	MEAN_ABSOLUTE_ERROR	VAL LOSS	VAL_MEAN_ABSOLUTE_ERROR	DROPOUT RATE	PREV_PERIODS
OILPRICE_RNN_0	completed	6.933	1.934			0.5	1
OILPRICE_RNN_5	completed			22.665	3.997	0.1	1
OILPRICE_RNN_4	completed			22.914	3.672	0.1	2
OILPRICE_RNN_3	completed			20.212	3.758	0.1	1
OILPRICE_RNN_2	completed			101.483	8.797	0.3	2
OILPRICE_RNN_1	completed			73.688	7.38	0.3	2

mean_absolute_error

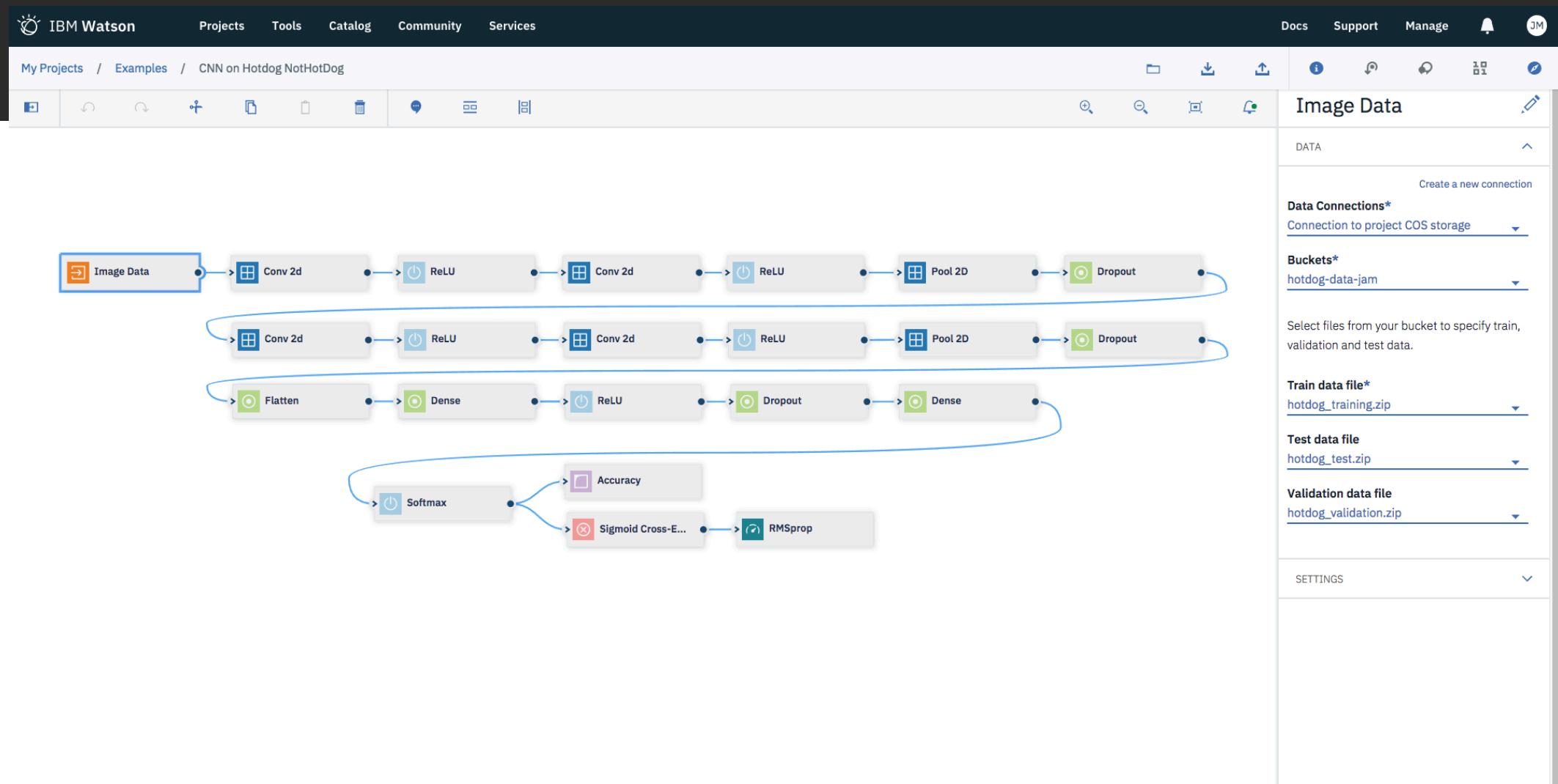
The chart displays the mean absolute error over 50 iterations for five different RNN models. The y-axis represents the mean absolute error, ranging from 0 to 24. The x-axis represents the iteration number, ranging from 0 to 49. All models show a rapid initial decrease in error, followed by a plateau or slight fluctuations at lower values. OILPRICE_RNN_0 consistently shows the lowest error, starting around 20 and plateauing near 2. OILPRICE_RNN_5 starts highest at approximately 22 and plateaus around 4. The other three models (OILPRICE_RNN_2, 3, 4) start between 7 and 10 and plateau between 2 and 4.

Iteration	OILPRICE_RNN_0	OILPRICE_RNN_2	OILPRICE_RNN_3	OILPRICE_RNN_4	OILPRICE_RNN_5
0	20.0	18.0	10.0	10.0	22.0
10	2.0	2.0	2.0	2.0	2.0
20	2.0	2.0	2.0	2.0	2.0
30	2.0	2.0	2.0	2.0	2.0
40	2.0	2.0	2.0	2.0	2.0
50	2.0	2.0	2.0	2.0	2.0

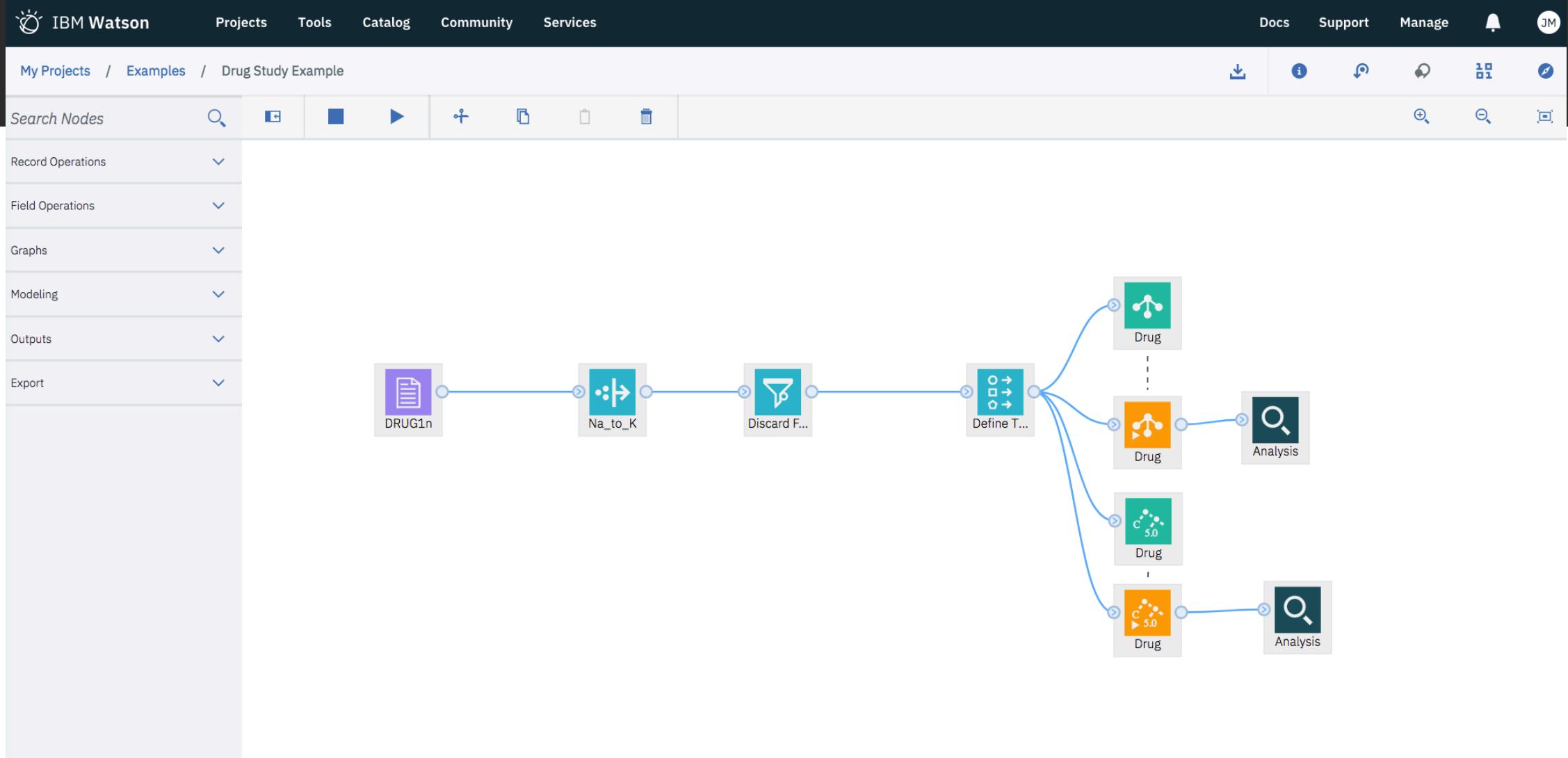
Apache Spark



Neural Network Modeler



SPSS Modeler Flows



Automatic Model Builder

IBM Watson Projects Tools Catalog Community Services Docs Support Manage JM

My Projects / Examples / Heart Failure Prediction

Select Data

Train Evaluate

Select a technique

Column value to predict (Label Col)
HEARTFAILURE (String)

Feature columns
All (default)

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

Model Deployment as Web Service

The screenshot shows the 'Create Deployment' page in the IBM Watson interface. At the top, there's a navigation bar with links for 'IBM Watson', 'Projects', 'Tools', 'Catalog', 'Community', 'Services', 'Docs', 'Support', 'Manage', and a user icon 'JM'. Below the navigation bar, the main title 'Create Deployment' is displayed. Underneath it, there are three tabs: 'Web Service' (which is selected), 'Batch Prediction', and 'Real-time Streaming Predictions'. The 'Name' field contains 'Heart Failure Prediction Model'. The 'Description' field has a placeholder text 'Web Service Deployment Description' and a character count indicator '300'. In the bottom right corner of the form area, there are 'Cancel' and 'Save' buttons, and a circular 'Next Step' button with a blue arrow icon.

IBM Watson

Projects Tools Catalog Community Services

Docs Support Manage JM

Create Deployment

Web Service Batch Prediction Real-time Streaming Predictions

Name
Heart Failure Prediction Model

Description
Web Service Deployment Description

300

Cancel Save

Next Step

Update, deploy, and monitor models w/Watson's ML API

Token Show/Hide | List Operations | Expand Operations

Service Instances Show/Hide | List Operations | Expand Operations

Published Models Show/Hide | List Operations | Expand Operations

Deployments Show/Hide | List Operations | Expand Operations

GET /v3/wml_instances/{instance_id}/published_models/{published_model_id}/deployments Lists all deployments for the particular model

POST /v3/wml_instances/{instance_id}/published_models/{published_model_id}/deployments Create the deployment - online, stream (Beta) , batch (Beta)

DELETE /v3/wml_instances/{instance_id}/published_models/{published_model_id}/deployments/{deployment_id} Deletes the deployment

GET /v3/wml_instances/{instance_id}/published_models/{published_model_id}/deployments/{deployment_id} Details about specific deployment

PATCH /v3/wml_instances/{instance_id}/published_models/{published_model_id}/deployments/{deployment_id} (Beta) Update stream state (start/stop)

POST /v3/wml_instances/{instance_id}/published_models/{published_model_id}/deployments/{deployment_id}/online Make an online prediction

GET /v3/wml_instances/{instance_id}/deployments Lists all deployments

[BASE URL: / , API VERSION: 3.0.0] VALID {...}

Data Refinery

IBM Watson Projects Tools Catalog Community Services Docs Support Manage JM

My Projects / Examples / patientdataV6.csv / Data Refinery

+ Operation Code an operation to cleanse and shape your data

Data Profile Visualizations Steps

AVGHEARTBEATSPERMIN String

PALPITATIONSPERDAY String

CHOLESTEROL String

BMI String

HEARTFAILUR String

FREQUENCY

Value	Frequency
80	31
91	26
83	25
93	24
72	23
89	22
85	21
79	20
84	19
94	18

FREQUENCY

Value	Frequency
25	32
19	31
16	30
2	30
7	29
21	29
11	28
13	28
24	27
36	27

FREQUENCY

Value	Frequency
217	20
185	18
239	17
166	17
169	16
234	16
157	15
159	15
172	15
191	15

FREQUENCY

Value	Frequency
22	81
26	81
29	80
30	80
28	78
27	78
23	75
31	72
25	68
32	67

FREQUENCY

Value	Frequency
N	100
Y	100

STATISTICS

Statistic	Value
Maximum length	3
Minimum length	2
Mean length	2.355
Unique	104

STATISTICS

Statistic	Value
Maximum length	2
Minimum length	1
Mean length	1.775
Unique	46

STATISTICS

Statistic	Value
Maximum length	3
Minimum length	3
Mean length	3
Unique	95

STATISTICS

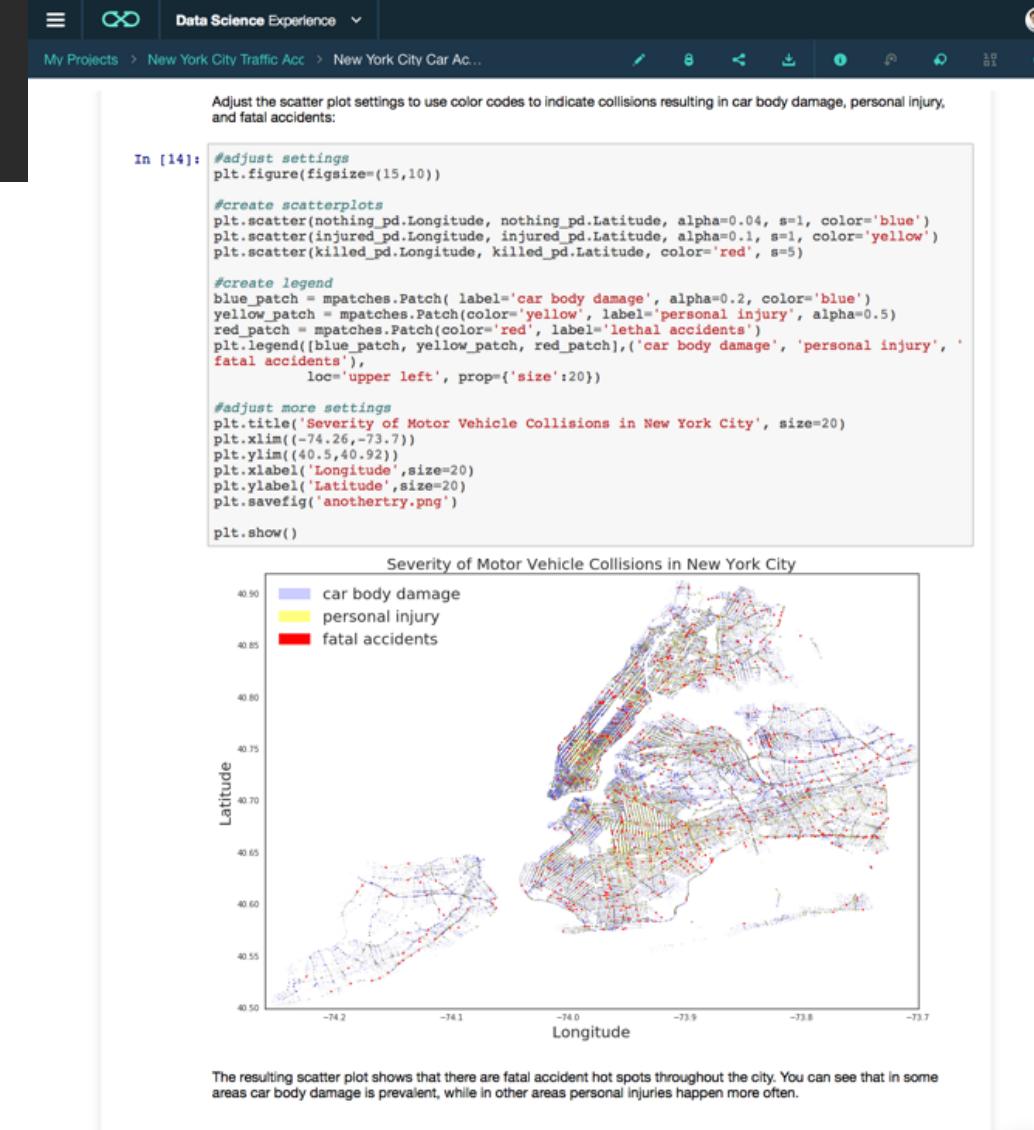
Statistic	Value
Maximum length	2
Minimum length	2
Mean length	2
Unique	15

RStudio

The screenshot shows the RStudio interface with several panels:

- Code Editor:** A large panel on the left containing R code. A green box highlights the code area. The code is part of a shiny application, specifically the server.R script, which filters flight delays by year and month and generates a UI output.
- Console:** Below the code editor, the R console displays the standard R startup message and the command `> library("shiny", lib.loc="/usr/lib64/R/library")`.
- Environment:** A panel at the top right showing disk usage (10%) and a project dropdown.
- File Browser:** A panel on the right with a red border, titled "flights", showing a file structure with files like data, global.R, gomap.js, server.R, styles.css, and ui.R.

Jupyter Notebooks



Streams Designer

IBM Watson Projects Tools Catalog Community Services US South NL

My Projects / Streaming Analytics / Data Historian

Data Historian Status: running

Sample Data Aggregation Aggregation Cloud Objec...

Data Historian Ingest Rate Events per Second

Sample Data

Time	Events per Second
04:45	0 EPS
04:48	16 EPS
04:51	28 EPS
04:54	10 EPS
04:57	2 EPS
04:58	0 EPS

Sample Data Throughput Events per Second

Output

Time	Events per Second
04:45	0 EPS
04:48	16 EPS
04:51	28 EPS
04:54	10 EPS
04:57	2 EPS
04:58	0 EPS

Dynamic Dashboards



Community

IBM Watson Projects Tools Catalog Community Services Docs Support Manage JM + Create new ▾

All filters What are you looking for?

Popular filters: Spark Deep Learning Brunel

Sort by: Featured ▾

Featured

ARTICLE Webinar: April 25 - R and TensorFlow AUTHOR RStudio DATE Apr 12, 2018 TOPIC Data Science FORMAT Web page	TUTORIAL Build Deep Learning Architectures With... AUTHOR developerWorks TV DATE Apr 02, 2018 LEVEL Beginner TOPIC Deep Learning +2	ARTICLE Introducing IBM Watson Studio AUTHOR Armand Ruiz DATE Mar 20, 2018 TOPIC Watson FORMAT Web page	ARTICLE Apple, IBM add machine learning to... AUTHOR TechCrunch DATE Mar 20, 2018 TOPIC Watson FORMAT Web page
--	---	---	--

All content

DATA SET Fashion-MNIST AUTHOR IBM DATE Apr 12, 2018 TOPIC Science & Technology	ARTICLE How Bayesian inference works AUTHOR Brandon Rohrer DATE Apr 12, 2018 TOPIC Data Science FORMAT Web page	ARTICLE Webinar: April 25 - R and TensorFlow AUTHOR RStudio DATE Apr 12, 2018 TOPIC Data Science FORMAT Web page	TUTORIAL How to Handle Imbalanced Classes in Machine... AUTHOR EliteDataScience DATE Apr 11, 2018 LEVEL Beginner TOPIC Machine Learning
ARTICLE Text Classifier Algorithms in Machine... AUTHOR Stats and Bots DATE Apr 11, 2018 TOPIC Machine Learning FORMAT Web page	TUTORIAL Serverless Data Flow Sequencing with Watson... AUTHOR Damian Cummins DATE Apr 09, 2018 LEVEL Intermediate TOPIC Analytics +2	NOTEBOOK Use PMML to predict iris species AUTHOR IBM DATE Apr 06, 2018 TOPIC Science & Technology	NOTEBOOK A TensorFlow regression model to predict... AUTHOR IBM DATE Apr 06, 2018 TOPIC Economy & Business

Collaboration and Sharing

test

Add collaborators

Invite

nilmeier@us.ibm.com X gfilla@us.ibm.com X Add more people... 

Access level

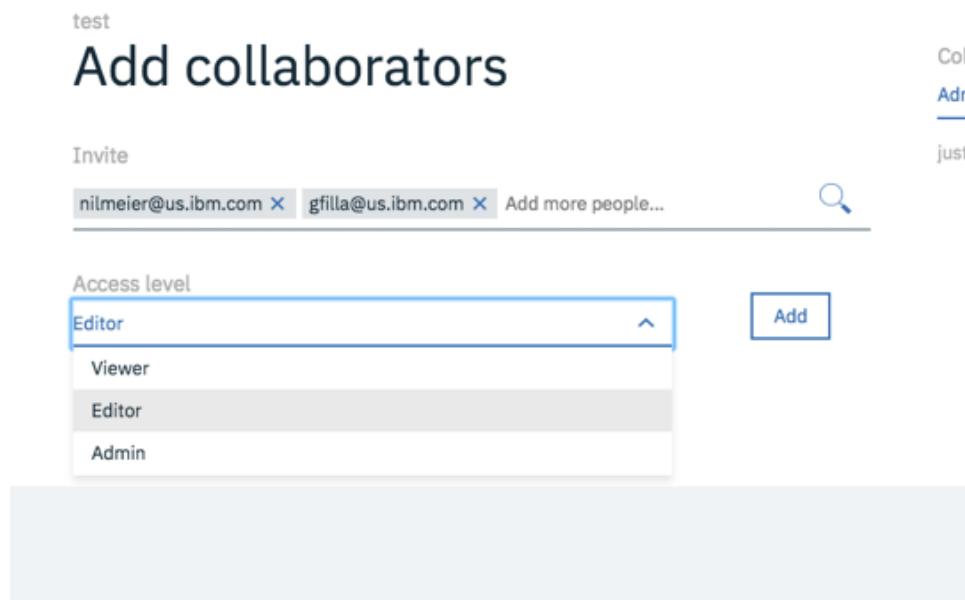
Editor  

Viewer

Editor

Admin

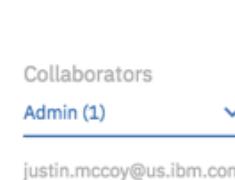
 



Collaborators

Admin (1) 

justin.mccoy@us.ibm.com



Share Migrate and Deploy Model on IBM Cloud

Share a read-only view of this notebook.

Share with anyone who has the link.

Cell content

- Only text and output
- All content excluding sensitive code cells
- All content, including code

 This option allows you to exclude code cells containing sensitive data. You can hide these code cells with a specific comment line in your code. When you share the notebook, for all code cells containing this comment line the code will not be displayed. A version is saved for your notebook. The link always points to the most recent version of the notebook.

Permalink to view notebook

<https://dataplatform.ibm.com/analytics/notebooks/v2/f448ffb2-e1f2-444b-8e3c-ee8d3e4841ee/view?> 

Share on social media





Github Integration

The screenshot shows the IBM Watson Platform interface. At the top, there is a dark header bar with the "IBM Watson" logo, navigation links for "Projects", "Tools", "Catalog", "Community", and "Services", and a user profile icon. Below the header, a secondary navigation bar has tabs for "Profile", "Apps and Services", and "Integrations", with "Integrations" being the active tab. The main content area is titled "GitHub Integration". It contains instructions: "Want to publish notebooks to your GitHub repository? Before you can publish to GitHub, you need to create an access token. Visit GitHub personal access tokens, select repo scope and generate a token. GitHub Enterprise is not currently supported." Below this text is a horizontal dotted line, followed by a "Remove" button in a blue-bordered box. At the bottom, there is a note: "After the access token is saved, a GitHub repository can be connected to a project on the project's Settings page."

GitHub Integration

Want to publish notebooks to your GitHub repository?
Before you can publish to GitHub, you need to create an access token. Visit [GitHub personal access tokens](#), select repo scope and generate a token.
GitHub Enterprise is not currently supported.

.....

0

[Remove](#)

After the access token is saved, a GitHub repository can be connected to a project on the project's Settings page.

Enough!
Let's build
something.



Create and deploy a model to predict Heart Rate Failure



<https://github.com/IBM/predictive-model-on-watson-ml>

Predicting Oil Prices Using an RNN with Watson Studio



<https://github.com/djccarew/timeseries-rnn-lab-part1>

<https://github.com/djccarew/timeseries-rnn-lab-part2>

Cleaning Data with Data Refinery



<https://developer.ibm.com/code/labs/Data-Science-Data-Refinery>

Thank you

 ibm.com/cloud/watson-studio

 twitter.com/ibmdata-science

 developer.ibm.com/code

