

Lab: Explore a Simple Generative Tool

Estimated time needed: 30 minutes

Overview

Generative AI models have revolutionized how you interact with technology, enabling you to create new content, generate realistic images, and translate languages with remarkable accuracy.

In this lab, you will gain hands-on experience with a simple generative AI tool, DataRobot, exploring its capabilities and applications.

Learning Objectives

After completing this lab, you will be able to:

- Sign up in DataRobot
- Add a data set to the use case
- Work on model building

Task 1: Sign-up in DataRobot

Step 1: Click www.datarobot.com

Step 2: Fill in the required information under the "Start for free" section and create an account.

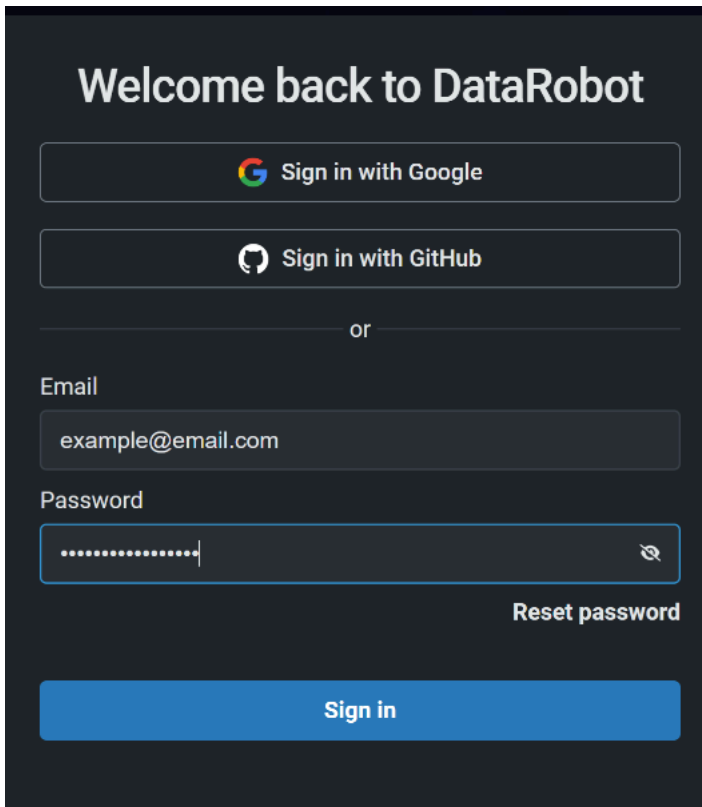
The screenshot shows the DataRobot website's homepage. The header includes the DataRobot logo, navigation links (Platform, Solutions, Customers, Partners, Resources, Company), and links for Support, Log In, Contact Us, and Book a Demo. The main content area features a large heading 'Experience the DataRobot AI Platform' and a subheading 'Less Friction, More AI. Get Started Today With a Free 30-Day Trial.' Below this, a paragraph states: 'Unleash your skills and embark on a new era of AI with a single platform that streamlines your predictive and generative AI workflows.' A section titled 'Start your free 30 day trial to:' lists four bullet points with checkmarks: 'Experience how to fast-track preparing data, running experiments, and testing your models', 'See how you can finally automate all your AI Experimentation and AI Production processes in a single solution', 'Build predictive and generative AI use-cases all in a single platform', and 'Learn how to accelerate your AI plans with hands-on labs, use-case code snippets, and an AI expert community'. On the right side, there is a 'Start For Free' form with input fields for First Name *, Last Name *, Business Email *, Phone, Company *, Job Title *, and Country *. At the bottom of the form, there is a checkbox labeled 'Yes! Please email me news and'.

Note: To access the DataRobot platform, you must sign up using a work email address. If you do not have a relevant work email, an alternative is to create a GitHub account using your Gmail address. Once registered, you can log in to DataRobot using your GitHub credentials.

For step-by-step guidance on creating a GitHub account, please refer to the following link:

[GitHub Account Setup Guide](#)

Step 4: A new window will open; select the relevant option for signing up.



Welcome back to DataRobot

Sign in with Google

Sign in with GitHub

or

Email

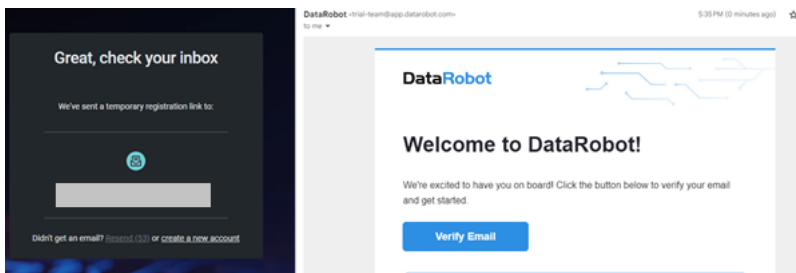
example@email.com

Password

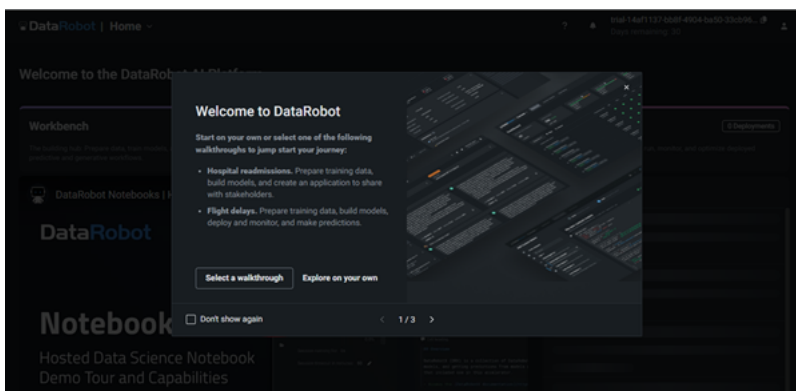
Reset password

Sign in

Step 5: Confirm your email by clicking **Verify Email** in your inbox.

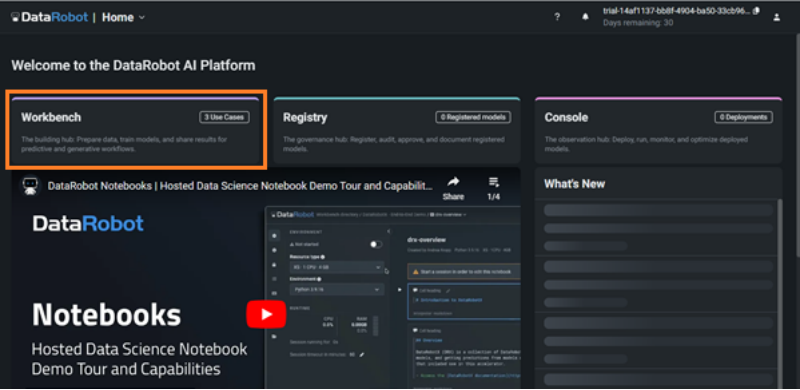


Step 6: Sign up and start your first experience of using the Generative AI tool.
The dashboard will look like the image below. You may like to familiarize yourself with the application by clicking **Select a walkthrough**.

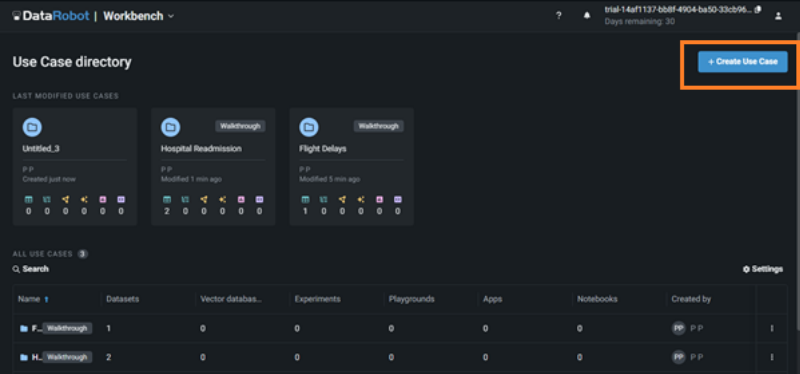


Task 2: Add a data set

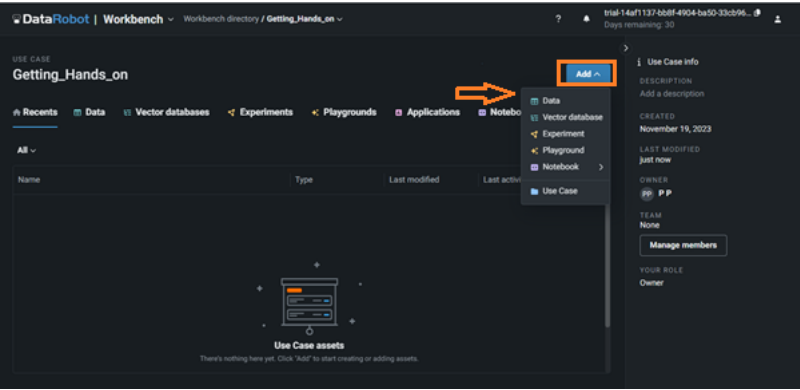
Step 7: The dashboard will appear shortly, and your screen will look as shown below. Click **Workbench**.



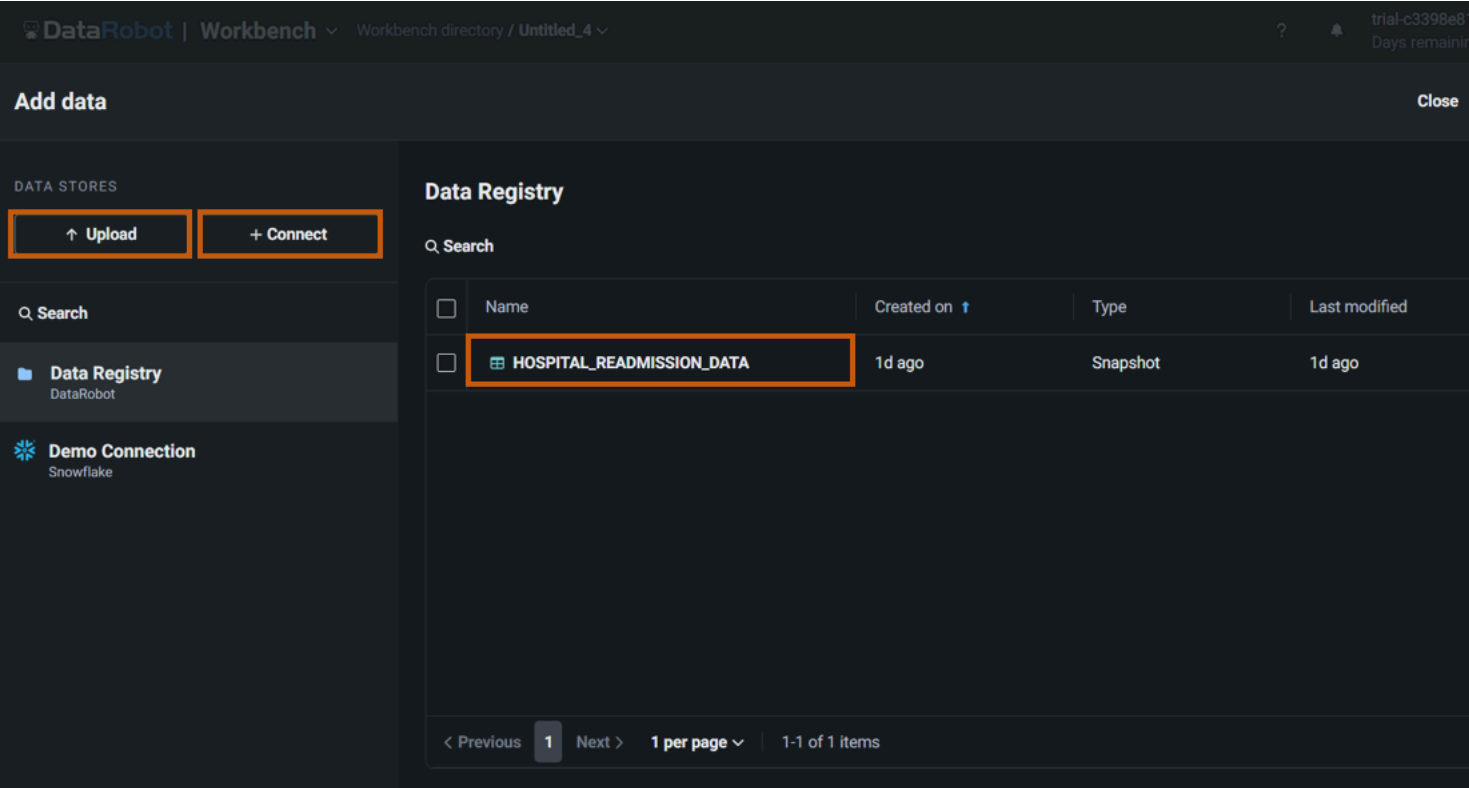
Step 8: Click Create Use Case.



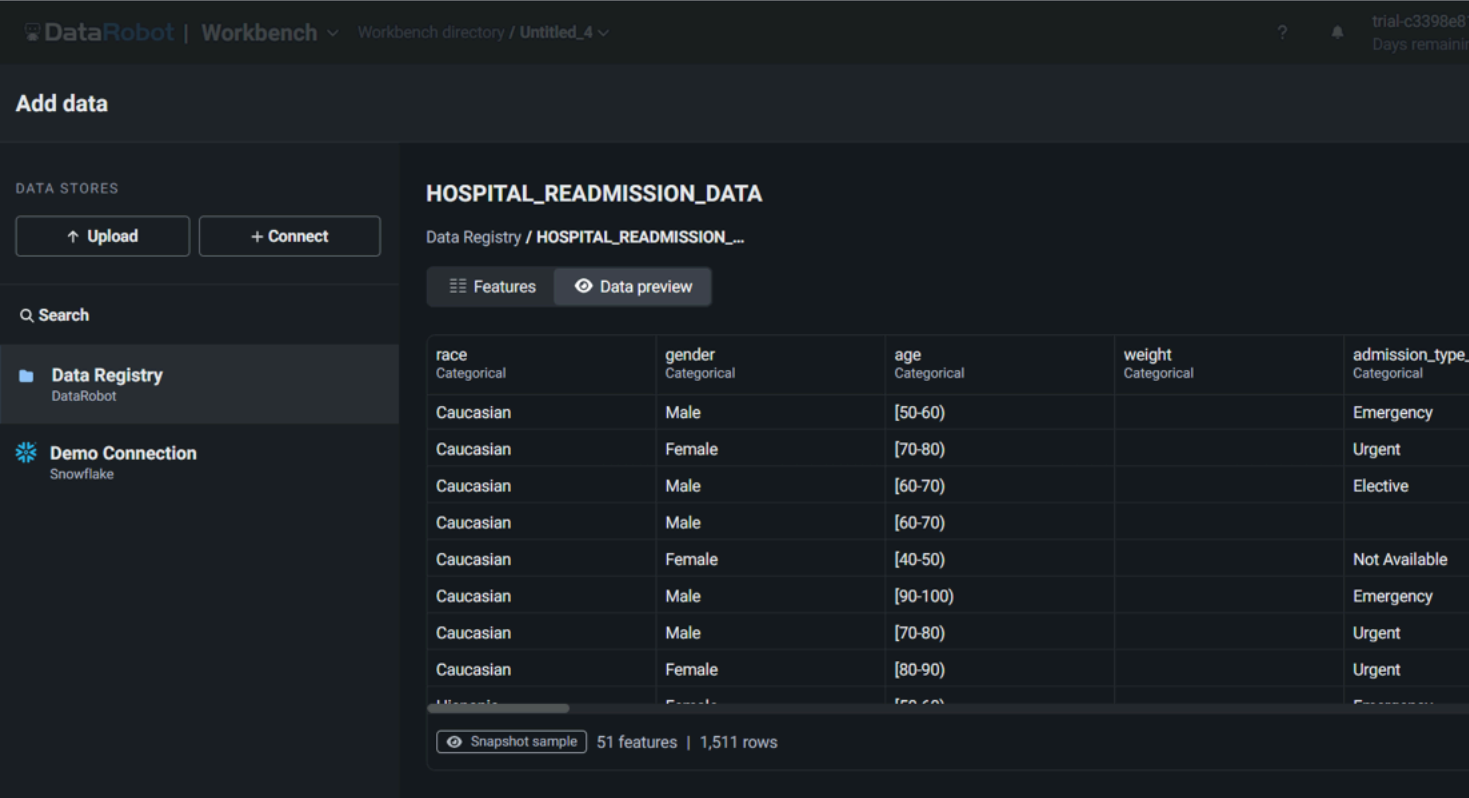
Step 9: Click Add and Data to include the data set in your use case.



Step 10: Upload your data set or Connect to the data source; however, for this lab, you can select an in-built sample data set *HOSPITAL_READMISSION_DATA*.



Step 11: Once you select the data set, you can see a preview of it. You can also view the data set's features, as shown below. Click **Add to Use Case**.



Step 12: After you add the data set to the use case, the workbench will appear as shown below. You can click the data set to see the feature insights.

DataRobot | Workbench Workbench directory / Untitled_4

< Use Case directory

USE CASE
Untitled_4 Add Data

All Data 1 Vector databases Experiments Playgrounds Applications

Type Search Settings

Name	Created By	Last Modified	Type	Source	Rows
HOSPITAL_READMISSION_DATA	PV Pratiksha V.	now	Snapshot	Snowflake	10000

< Previous 1 Next > 1 per page 1-1 of 1 items

Step 13: Explore the **All Features** menu to display specific features.

DataRobot | Workbench Workbench directory / Untitled_4 / HOSPITAL_READMISSION_DATA

HOSPITAL_READMISSION_DATA Jul 29th, 2024 10:39 AM Snapshot Data actions

Data preview **Features** Feature lists

☒ Show insights Show features from: All Features + Create feature list

Search

DATAROBOT FEATURE LISTS

- All Features 51
- Informative Features 40
- Raw Features 51

race	age	weight	admission_type_id	discharge
Caucasian	[70-80]	==Missing==	Emergency	Discharge
AfricanAmerican	[60-70]	[75-100]	Urgent	Discharge
Other	Other	Other	Other	Other

Snapshot sample 51 features | 1,511 rows

Task 3: Work on Data Modeling

Step 14: Click **Start**. You will have options **Modelling** and **Start wrangling**. You can try data wrangling if you want to. For this lab, you will work on model building. Click **Start** and select **Modelling**. It will take a while to prepare a data set for modelling.

DataRobot | Workbench Workbench directory / Untitled_4 / HOSPITAL_READMISSION_DATA

Jul 29th, 2024 10:39 AM Snapshot Data actions

HOSPITAL_READMISSION_DATA

Data preview Features Feature lists

Show insights Show features from: All Features + Create feature list

race	gender	age	weight	admission	discharge
Categorical	Categorical	Categorical	Categorical	Categorical	Categorical
Caucasian	Female	[70-80]	==Missing==	Emergency	Discharge
AfricanAmerican	Male	[60-70]	[75-100]	Urgent	Discharge
Other		Other	Other	Other	Other

Snapshot sample | 51 features | 1,511 rows

Start wrangling
Start modeling
Start feature discovery
Download dataset
Remove dataset

Step 15: Once done, you need to select the **Target feature**. Select **readmitted** as your target feature.

DataRobot | Workbench Workbench directory / Untitled_4

Set up new experiment Dataset Target Additional settings

Target feature
 Select the feature to make predictions on.

payer_code
 pioglitazone
 race
readmitted
 repagalinide
 troglitazone
 tolbutamide
 tolazamide
 time_in_hospital

	Uniq...	Missi...	Mean	Std Dev
readmitted	7	9592	-	-
troglitazone	1	0	-	-
tolbutamide	2	0	-	-
tolazamide	2	0	-	-
time_in_hospital	14	0	4.43	3.021

Experiment summary
 HOSPITAL_READ
 07-30 11:23:57

Dataset
 Name

Rows
 Features

Target
 No target selected

Step 16: The workbench screen will be displayed as shown below. Click **Next**.

DataRobot | Workbench Workbench directory / Untitled_4

trial-c3398e81-5e Days remaining: 2

Set up new experiment

Dataset Target Additional settings

Target feature
Select the feature to make predictions on.

readmitted

Target type: Binary classification

Positive class: ☐ 0 ☒ 1

Modeling mode
Set the mode used for selecting which blueprints to build when training models.

Quick Autopilot

Optimization metric
Set the metric used when training models to evaluate and optimize accuracy.

LogLoss (Accuracy) Recommended

Number of rows

Values of readmitted	Number of rows
False	~6000
True	~4000

Experiment summary
HOSPITAL_READM
07-30 11:23:57

Dataset
Name

Rows
Features

Target
Feature
Target type
Positive class
Modeling mode
Optimization metric
Training feature list

Partitioning

Step 17: You can modify the model setting in **Additional Settings**; once done, click **Next** and then click **Start modelling**.

DataRobot | Workbench Workbench directory / Untitled_4

trial-c3398e81-5e Days remaining: 2

Set up new experiment

Dataset Target Additional settings Exit

Data partitioning **Time series modeling** **Additional settings**

Partitioning method
Select the method for assigning rows to partitions when training models.

Stratified sampling
Rows are assigned to ensure similar target distribution across each partition.

Validation type

☒ **Cross-validation**
Trains models on a specified number of folds, maximizing data use but also increasing run time.

☐ **Training-validation-holdout**
Splits data into three partitions: trains models on the training set, assess performance on the validation set, and evaluates the model on unseen data in the holdout set.

Cross-validation folds
Enter a value from 2 - 50.

Holdout percentage
Set the subset of data that is unavailable during training and validation. Enter a value

Experiment summary
HOSPITAL_READM
07-30 11:23:57

Dataset
Name

Rows
Features

Target
Feature
Target type
Positive class
Modeling mode
Optimization metric
Training feature list

Partitioning

Step 18: Building models will take a while.

DataRobot | Workbench

Workbench directory / Untitled_4 / HOSPITAL_READMISSION_DATA - 2024-07-30 11:23:57

trial-c3398e81-5e8 Days remaining: 29

Experiment Comparison

View experiment info

Filter 0

Validation • LogLoss

Search

MODELS 0/0

Models are building...

Keras Slim Residual Neural Network Classifier using... Building...

Informative Features

64% (6,400 rows)

Elastic-Net Classifier (L2 / Binomial Deviance) Building...

Informative Features

64% (6,400 rows)

Models are building...

Step 19: once the modelling is complete, you can pick a model of your choice, and the DataRobot will show the **Model Overview**.

DataRobot | Workbench

Workbench directory / Untitled_4 / HOSPITAL_READMISSION_DATA - 2024-07-30 11:23:57

trial-c3398e81-5e8 Days remaining: 29

Experiment Comparison

View experiment info

Filter 0

Validation • LogLoss

Search

MODELS 3/3

Models are building...

Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance) 0.6089

Informative Features

64% (6,400 rows)

Elastic-Net Classifier (L2 / Binomial Deviance) 0.6115

Informative Features

64% (6,400 rows)

Keras Slim Residual Neural Network Classifier using... 0.6327

Model Overview

Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance)

Training scores: LogLoss

Validation	0.6089
Cross-validation	Score
Holdout	0.6111

Training settings

Training feature list

Informative Features 64% (6,400 rows)

Training sample size

Blueprint

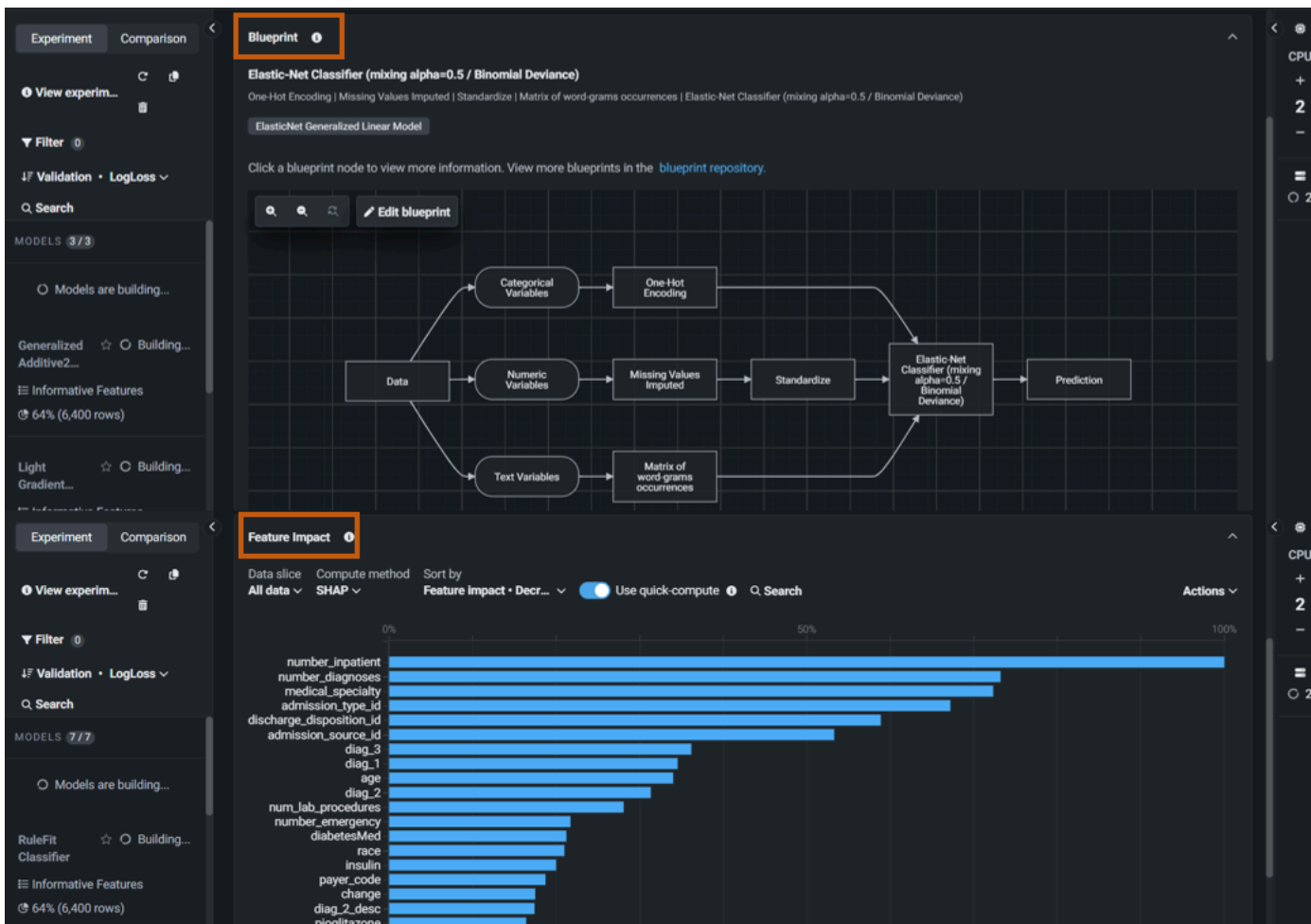
Feature Impact

Feature Effects

Individual Prediction Explanations

ROC Curve

Step 20: You can explore various model overview components like **Blueprint**, **Feature Impact**, and so on.



Step 21: If you have test or unseen data, you can also make predictions by clicking **Make Predictions** under **Model actions**.

The screenshot displays the Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance) interface. The left sidebar shows the 'Model Overview' tab selected, with a search bar and a list of models. The main area shows the 'Model Overview' section, which includes training scores, training settings, and informative features. The 'Training scores' section shows LogLoss, Validation, Cross-validation, and Holdout scores. The 'Training settings' section shows the training feature list and training sample size. The 'Informative Features' section shows the percentage of informative features (64% (6,400 rows)). Below the 'Model Overview' section, the 'Make Predictions' section is visible, showing options to make new predictions and download recent predictions. The 'Make new predictions' section includes a prediction dataset upload area, prediction options (including feature values and explanations), and a 'Compute and download predictions' button. The 'Download recent predictions' section shows that there are no recent predictions available for download.

Step 22: You can also click **Generate compliance report** and **download compliance report** for your use case.

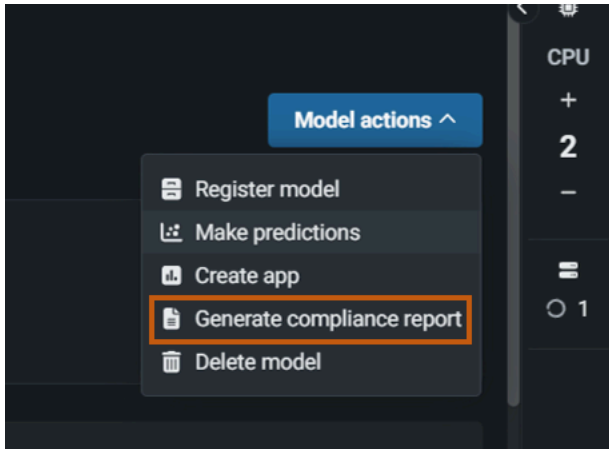


Table of Contents

- 1 How To Use This Document
- 2 DataRobot Model Development Documentation
- 3 Executive Summary and Model Overview
 - 3.1 Model Stakeholders
 - 3.2 Model Development Purpose and Intended Use
 - 3.3 Model Description and Overview
 - 3.4 Overview of Model Results
 - 3.5 Model Interdependencies
- 4 Model Data Overview
 - 4.1 Feature Association
 - 4.2 Data Source Overview and Appropriateness
 - 4.3 Input Data Extraction, Preparation, and Quality & Completeness

Conclusion

In this lab, you have signed up in DataRobot, added a data set in a use case, and worked on data modelling.

Author(s)

[Dr. Pooja](#)



Skills Network