

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.

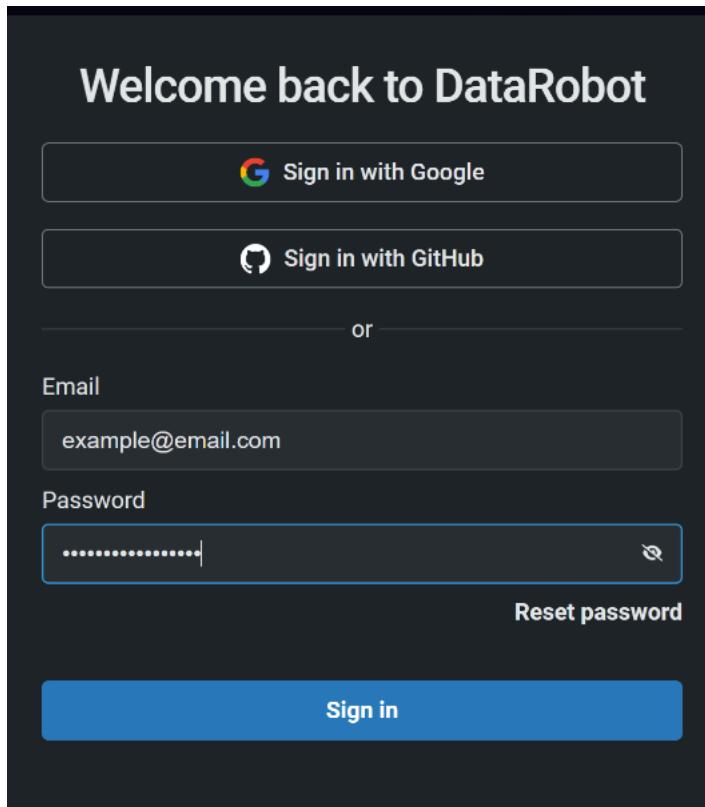
The screenshot shows the DataRobot homepage. At the top, there is a navigation bar with links for Platform, Solutions, Customers, Partners, Resources, and Company. On the right side of the header, there are buttons for 'Contact Us' and 'Book a Demo'. Below the header, the main content area features a large banner with the text 'Experience the DataRobot AI Platform' and 'Less Friction, More AI. Get Started Today With a Free 30-Day Trial.' To the right of the banner is a 'Start For Free' form. The form contains fields for First Name, Last Name, Business Email, Phone, Company, Job Title, and Country. There is also a checkbox for 'Yes! Please email me news and' followed by a dropdown menu. The entire page has a light blue background with some subtle shadows.

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.

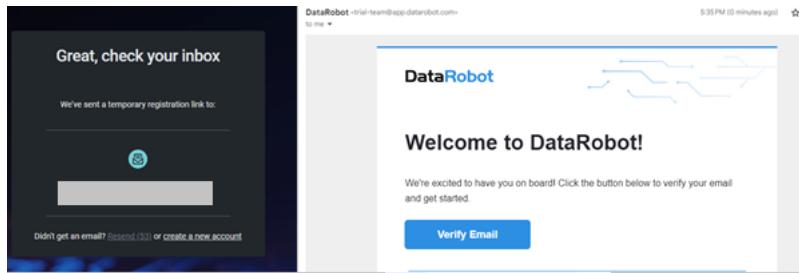
For step-by-step guidance on creating a GitHub account, please refer to the following link:

[GitHub Account Setup Guide](#)

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

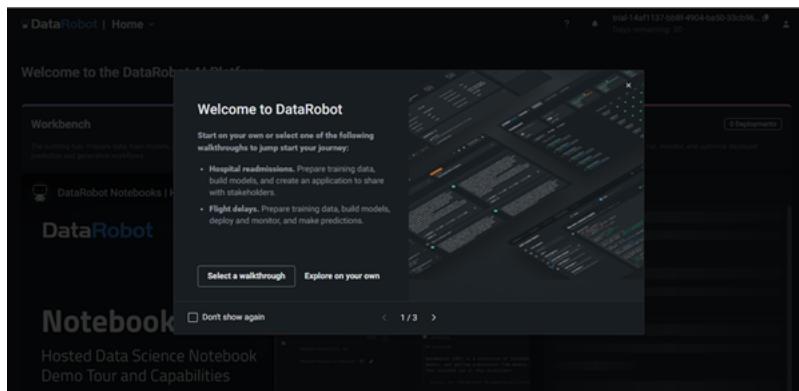


Step 5: Confirm your email by clicking **Verify Email** in your inbox.



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**.

Welcome to the DataRobot AI Platform

Workbench 3 User Cases Registry 0 Registered models Console 0 Deployments

DataRobot Notebooks | Hosted Data Science Notebook Demo Tour and Capabilities

Notebooks

Hosted Data Science Notebook Demo Tour and Capabilities

Step 8: Click **Create Use Case**.

Use Case directory

+ Create Use Case

Name	Datasets	Vector databases...	Experiments	Playgrounds	Apps	Notebooks	Created by
F. Walkthrough	1	0	0	0	0	0	P P
H. Walkthrough	2	0	0	0	0	0	P P

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

USE CASE
Getting_Hands_on

RECENTS Data Vector databases... Experiments Playgrounds Applications Notebooks

Add...

DATA

DESCRIPTION Add a description

CREATED November 19, 2023

LAST MODIFIED just now

OWNER P P

TEAM None

Manage members

YOUR ROLE Owner

Use Case assets

There's nothing here yet. Click "Add" to start creating or adding assets.

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*.

Add data

Data Registry

Search

<input type="checkbox"/>	Name	Created on	Type	Last modified
<input type="checkbox"/>	HOSPITAL_READMISSION_DATA	1d ago	Snapshot	1d ago

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

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**.

Add data

HOSPITAL_READMISSION_DATA

Data Registry / HOSPITAL_READMISSION_...

Features Data preview

race	gender	age	weight	admission_type_
Categorical	Categorical	Categorical	Categorical	Categorical
Caucasian	Male	[50-60)		Emergency
Caucasian	Female	[70-80)		Urgent
Caucasian	Male	[60-70)		Elective
Caucasian	Male	[60-70)		
Caucasian	Female	[40-50)		Not Available
Caucasian	Male	[90-100)		Emergency
Caucasian	Male	[70-80)		Urgent
Caucasian	Female	[80-90)		Urgent

Snapshot sample 51 features | 1,511 rows

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.

USE CASE
Untitled_4

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.

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

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

Category	Count
All Features	51
Informative Features	40
Raw Features	51

race Categorical

age Categorical

weight Categorical

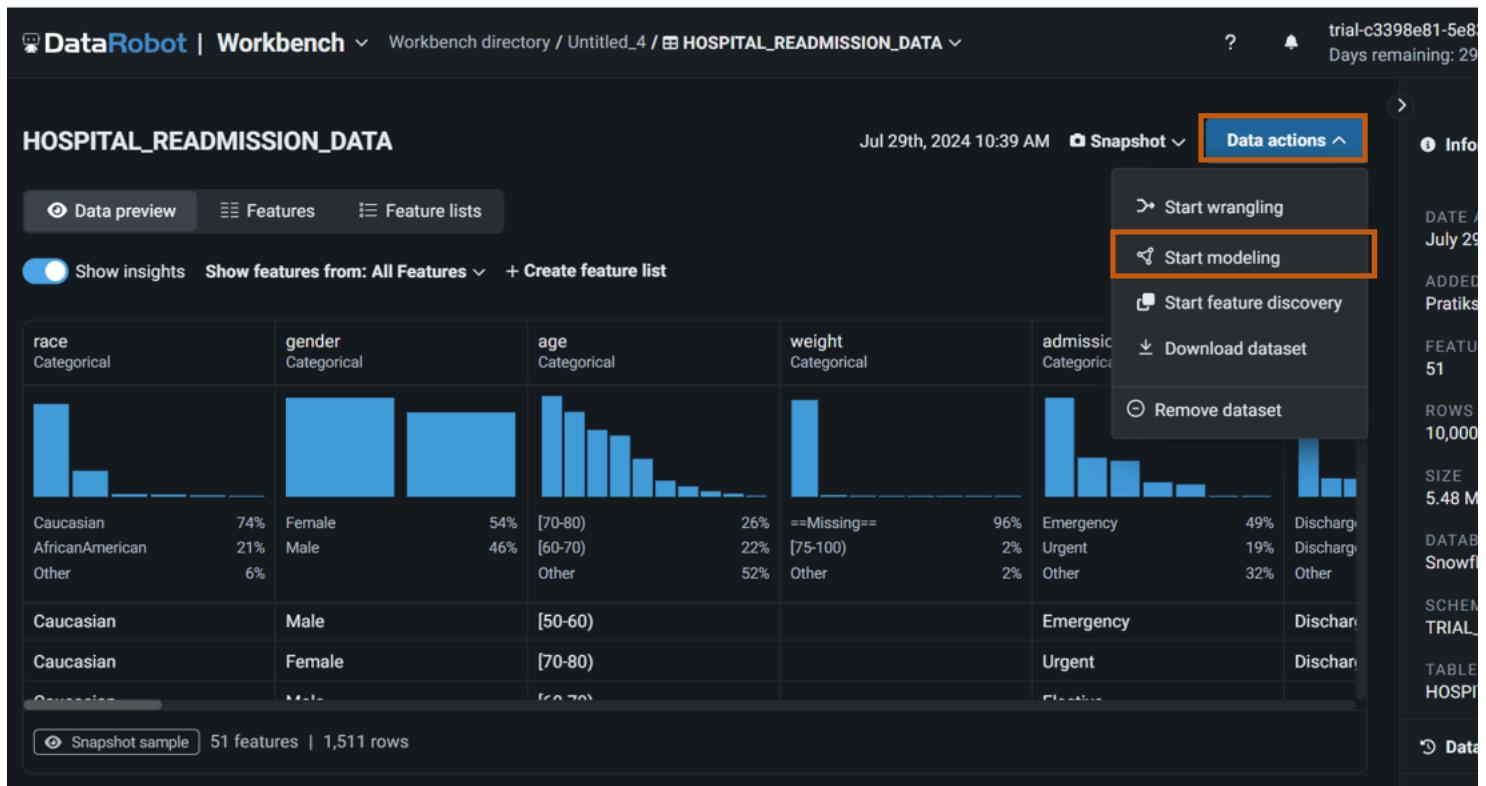
admission_type_id Categorical

discharge Categorical

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.



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

Set up new experiment

Dataset Target Additional settings

Target feature

Select the feature to make predictions on.

			Uniq...	Missi...	Mean	Std Dev
payer_code						
pioglitazone						
race						
readmitted			7	9592	-	-
repaglinide						
troglitazone	35	Categorical	1	0	-	-
tolbutamide	30	Categorical	2	0	-	-
tolazamide	36	Categorical	2	0	-	-
time_in_hospital	7	Numeric	14	0	4.43	3.021

Experiment summary
HOSPITAL_READMISSION_DATA 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-c3398e81-5e6 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

Experiment summary

- HOSPITAL_READIN
- 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-c3398e81-5e6 Days remaining: 2

Set up new experiment

Dataset — **Target** — **Additional settings** — **Exit**

Data partitioning — **Time series modeling** Preview — **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_READIN
- 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.

The screenshot shows the DataRobot Workbench interface. The top navigation bar displays "DataRobot | Workbench", the current directory "Workbench directory / Untitled_4 / HOSPITAL_READMISSION_DATA - 2024-07-30 11:23:57", and a trial identifier "trial-c3398e81-5e8". The sidebar on the left includes tabs for "Experiment" (selected) and "Comparison", and sections for "View experiment info", "Filter", "Validation", "Search", and "MODELS 0/0". The main area lists three models:

- Keras Slim Residual Neural Network Classifier using... (Building...)
- Elastic-Net Classifier (L2 / Binomial Deviance) (Building...)
- Keras Slim Residual Neural Network Classifier using... (Building...)

A large circular progress bar in the center indicates that "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**.

The screenshot shows the DataRobot Workbench interface with the "Model Overview" section highlighted by an orange box. The "Experiment" tab is selected. The main area displays the "Model Overview" for the first model listed in the sidebar:

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

Training scores: LogLoss

Validation	0.6089
Cross-validation	Score
Holdout	0.6111

Training settings

Training feature list	Informative Features
Training sample size	64% (6,400 rows)

Blueprint

Feature Impact

Feature Effects

Individual Prediction Explanations

ROC Curve

The sidebar on the left shows the same three models as in the previous screenshot, with the first one highlighted.

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

The screenshot shows the Model Overview page with two main sections highlighted:

- Blueprint**: A flowchart illustrating the data processing pipeline. It starts with 'Data' (split into Categorical, Numeric, and Text variables), followed by 'One-Hot Encoding' for categorical, 'Missing Values Imputed' for numeric, and 'Standardize' for all. These then feed into an 'Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance)' block, which outputs 'Prediction'.
- Feature Impact**: A horizontal bar chart showing the impact of various features on the model's output. The top 15 features listed are:
 - number_inpatient
 - number_diagnoses
 - medical_specialty
 - admission_type_id
 - discharge_disposition_id
 - admission_source_id
 - diag_3
 - diag_1
 - age
 - diag_2
 - num_lab_procedures
 - number_emergency
 - diabetesMed
 - race
 - insulin
 - payer_code
 - change
 - diag_2_desc
 - glucotazone

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

The screenshot shows the Model Overview page with the **Make Predictions** section highlighted. This section includes:

- Training scores: LogLoss**: Validation: 0.6089, Cross-validation: Score 0.6111, Holdout: 0.6111.
- Training settings**: Training feature list, Training sample size: 64% (6,400 rows).
- Blueprint**: A link to the blueprint diagram.
- Make Predictions**: A sub-section with the following options:
 - Make new predictions**: A section for uploading prediction datasets.
 - Prediction dataset**: A note stating "Prediction datasets cannot exceed 5.00 GB." with a file upload input field.
 - Prediction options**:
 - Include additional feature values in pred**: A note explaining it includes specified feature values from the prediction dataset.
 - Add all features** (selected) vs **Add specified features**.
 - Include Prediction Explanations**.
 - Download recent predictions**: A note stating "Predictions are stored and available to download for 48 hours." with a "No recent predictions" message.

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

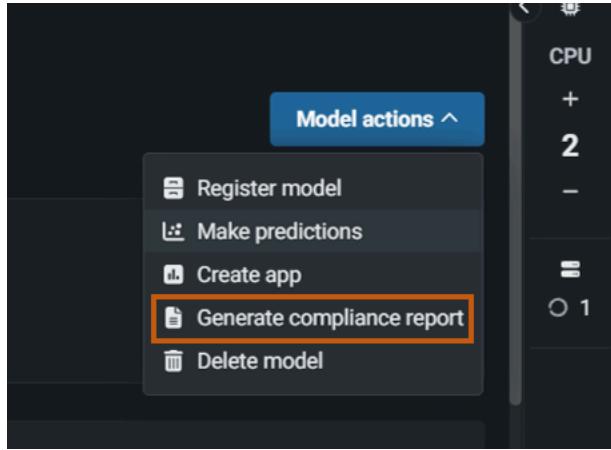


Table of Contents

- 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