

# DATABRICKS\_CODEBASICS

## PROJECT CHALLENGE

*By*

*Dr. Sridhar C*

Ongoing

AI and Data Science Courses

# Build With Databricks: Hands-On Project Challenge



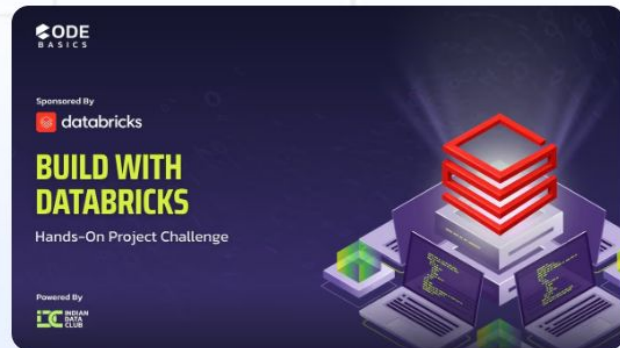
Deadline

Feb 01, 2026 11:59 PM IST



Prize Pool

₹ 2,00,000



## Overview

In this project challenge, the focus is on building a real-world solution using Databricks. The project is self-directed: you will define the problem, select the dataset, and architect an end-to-end solution on the Databricks Lakehouse to showcase your learning in a portfolio-ready project.

Enroll for Free



₹ 2,00,000 Prize Pool



633 Participants



Difficulty: Intermediate

# INTRODUCTION

- Patients with chronic disease often fails to follow up.
- They endup in advanced disease condition without proper care.
- It will be useful if we find the model that predicts these patients, So we will have rigorous follow up.

# DATASET



NUDRAT ABBAS · UPDATED 11 DAYS AGO

▲ 30

<> Code

Download



## Patient Churn Prediction Dataset for Healthcare

Predict patient attrition using satisfaction scores, engagement metrics.



Data Card

Code (8)

Discussion (0)

Suggestions (0)

### About Dataset

#### Context

Patient churn (attrition) is a critical challenge in healthcare, costing providers billions in lost revenue and disrupting continuity of care. Studies show:

- Acquiring a new patient costs 5-25x more than retaining an existing one
- 20-30% of patients switch providers annually
- Low satisfaction and poor engagement are primary drivers of churn

#### Usability ⓘ

10.00

#### License

CC0: Public Domain

#### Expected update frequency

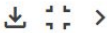
Quarterly

#### Tags

Classification

# DATASET

patient\_churn\_dataset.csv (193.99 kB)




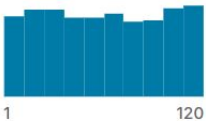
Detail Compact Column

10 of 21 columns

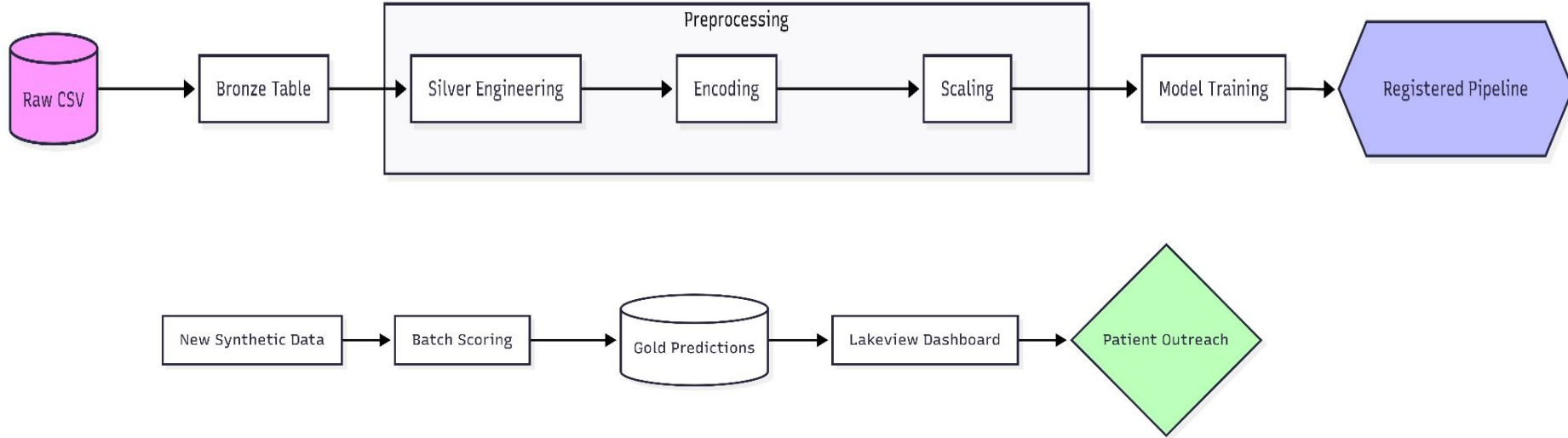
### About this file

Suggest Edits

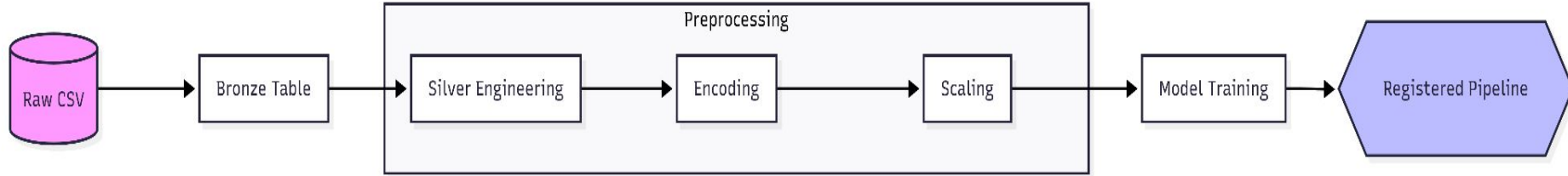
This file contains complete patient behavioral, satisfaction, and engagement data for binary classification (churn prediction). The dataset includes realistic churn patterns with approximately 15-25% churn rate, reflecting industry benchmarks. No missing values in critical fields.

▲ PatientID	# Age	▲ Gender	▲ State	# Tenure_Months	▲ Specialty	▲ Insurance		
Unique patient identifier	Patient age in years							
2000 unique values		Male	51%	NC	13%	General Practice	15%	Self-Pay
		Female	49%	IL	11%	Orthopedics	14%	Medicare
			Other (1527)	76%		Other (1409)	70%	Other (961)
								
C20000	41	Female	PA	62	Pediatrics	Medicaid		
C20001	43	Female	GA	44	Internal Medicine	Self-Pay		
C20002	21	Male	MI	120	Internal Medicine	Medicaid		
C20003	65	Male	FL	118	General Practice	Private		
C20004	18	Female	CA	70	Cardiology	Medicaid		
C20005	65	Male	MI	82	Orthopedics	Self-Pay		
C20006	82	Female	FL	56	Pediatrics	Medicare		
C20007	66	Male	MI	22	General Practice	Private		
C20008	40	Female	PA	60	Cardiology	Medicaid		

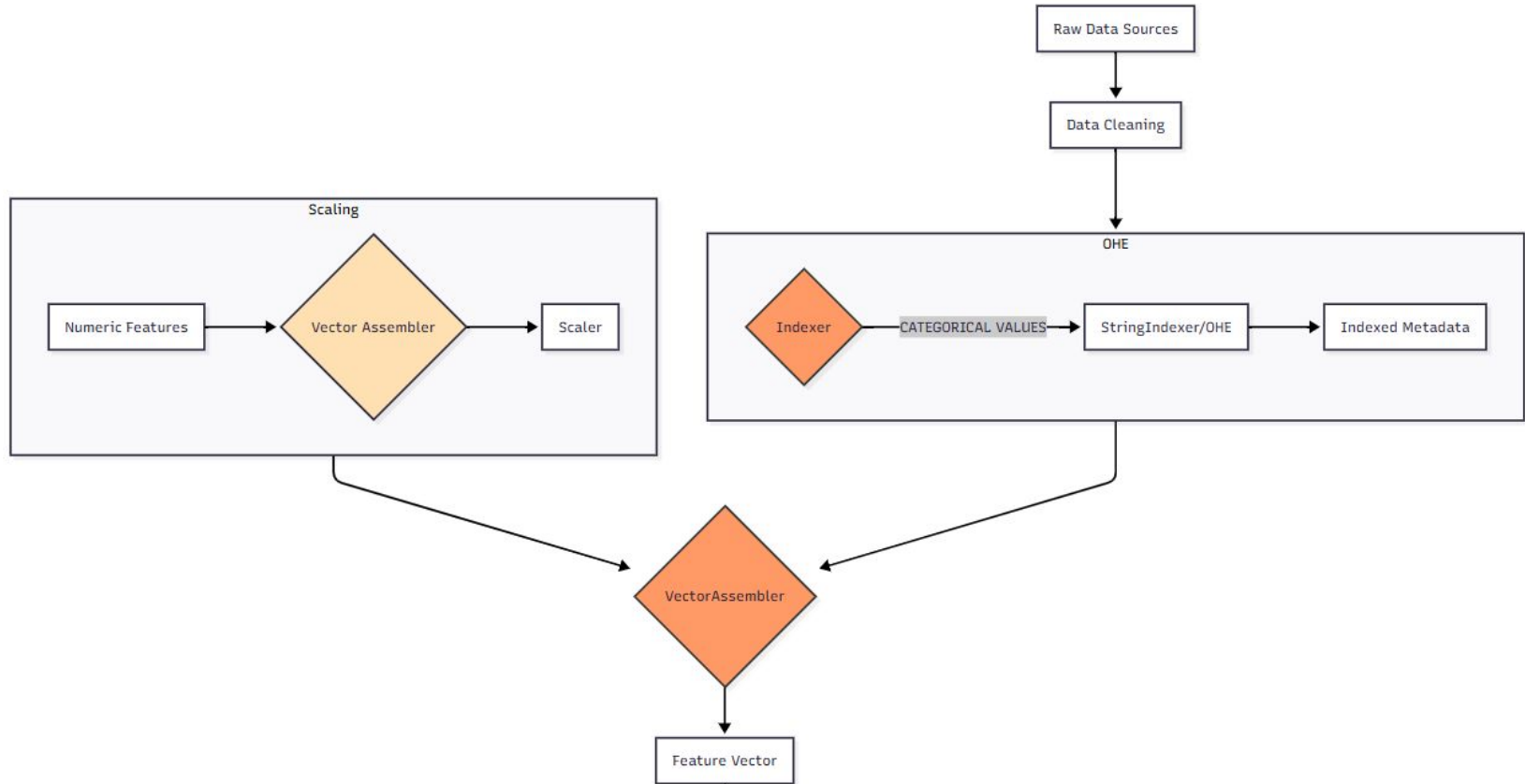
# WORKFLOW



# INITIAL ML TRAINING

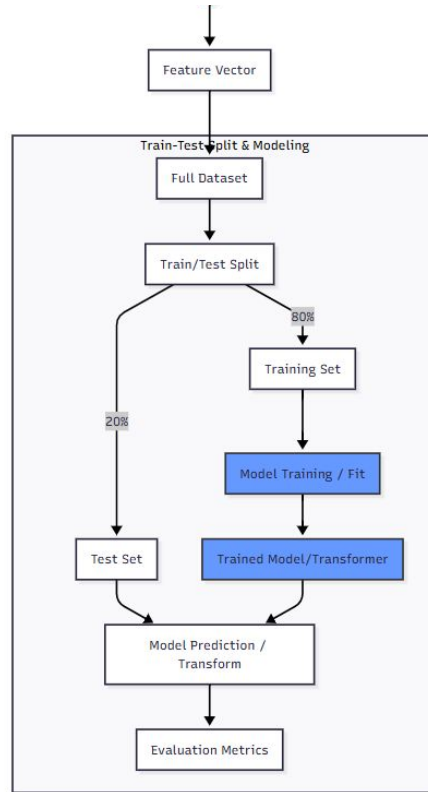


# MODEL TRAINING WITH ML FLOW





# MODEL TRAINING WITH ML FLOW



# EXPERIMENTS

← patient\_churn\_analysis Machine learning ⓘ

Share

View docs

Runs

Models

Traces

MLflow 3 is available!

Featuring unified ML and GenAI experiment tracking, improved model logging, prompt versioning, enhanced LLM judges, advanced tracing for end-to-end agent observability, and beyond. [Learn more](#)

metrics.rmse < 1 and params.model = "tree"

Time created

State: Active

Datasets

+ New run

Sort: Created

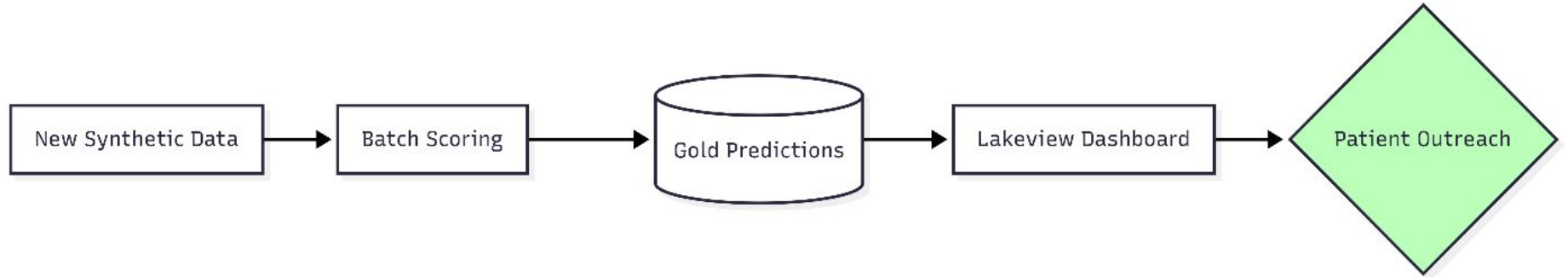
Columns

Group by

<input type="checkbox"/>	Run Name	Created ↕	Dataset	Duration	Source	Models ⓘ
<input type="checkbox"/>	log_regression	5 hours ago	-	5.0s	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	random_forest_classifier	5 hours ago	-	1.0min	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	log_regression	5 hours ago	-	1.2min	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	log_regression	9 hours ago	-	51.8s	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	random_forest_classifier	9 hours ago	-	58.0s	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	log_regression	10 hours ago	-	1.0min	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	agreeable-crane-245	17 hours ago	-	23.2s	ml_train	workspace.default.patient_...
<input type="checkbox"/>	masked-slug-428	17 hours ago	-	1.2min	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	delicate-penguin-212	17 hours ago	-	54.9s	ml_train	patient_churn_prediction.d...
<input type="checkbox"/>	bright-koi-411	17 hours ago	-	25.8s	ml_train	workspace.default.patient_...

13 matching runs

# NEW SYNTHETIC DATA INGESTION



# REGISTERED\_MODEL

Catalog

Serverless Starter Ware... Serverless 2XS

Type to search...

For you All

My organization

> workspace

> system

> ecommerce

> health\_care

> patient\_churn\_prediction

> default

patient\_churn\_logreg

patient\_churn\_logreg\_inc\_thre

patient\_churn\_rf

> information\_schema

> log\_model

> patient\_data

> source\_data

> transportation

> Delta Shares Received

Catalog Explorer > patient\_churn\_prediction >

default

Use with BI tools

Share

Create

Overview

Details

Permissions

Policies

Description

Default schema (auto-created)

Filter models

Tables 0

Volumes 0

Models 3

Functions 3

Sort

Name	Owner	Created at
patient_churn_logreg	csridhar.mbbs.ms.mch@gmail...	Jan 30, 2026, 11:54 PM
patient_churn_logreg_inc_thre	csridhar.mbbs.ms.mch@gmail...	Jan 31, 2026, 07:19 AM
patient_churn_rf	csridhar.mbbs.ms.mch@gmail...	Jan 30, 2026, 11:57 PM

About this schema

Owner csridhar.mbbs.ms.mch@gmail.com

Tags

Add tags

Policies

New policy

# GOLD\_TABLE

Table									
	A <sup>B</sup> <sub>C</sub> PatientID	A <sup>B</sup> <sub>C</sub> Gender	1 <sup>2</sup> <sub>3</sub> Age	A <sup>B</sup> <sub>C</sub> State	A <sup>B</sup> <sub>C</sub> Specialty	1.2 Overall_...	1.2 Churn_...	1.2 Churn_...	A <sup>B</sup> <sub>C</sub> Risk_Segment
1	NEW_40000	Male	56	MI	Orthopedics	1.6	0.7538	1	High Risk
2	NEW_40001	Male	61	FL	Cardiology	4	0.6897	1	Moderate Risk
3	NEW_40002	Female	67	NY	> General ...	3.7	0.7509	1	High Risk
4	NEW_40003	Male	43	IL	> Internal ...	5	0.7164	1	Moderate Risk
5	NEW_40004	Male	58	IL	> General ...	1	0.6752	1	Moderate Risk
6	NEW_40005	Female	51	GA	> Internal ...	1.3	0.8754	1	High Risk
7	NEW_40006	Female	38	IL	Cardiology	2.4	0.6706	1	Moderate Risk
8	NEW_40007	Female	44	CA	Cardiology	1	0.8978	1	High Risk
9	NEW_40008	Female	28	OH	> General ...	1.2	0.8498	1	High Risk
10	NEW_40009	Female	76	TX	> General ...	1.1	0.6485	1	Moderate Risk
11	NEW_40010	Female	58	GA	> Internal ...	2	0.8772	1	High Risk
...	...	...	...	...	...	...	...	...	...
1,000 rows   1.38s runtime									
Refreshed 5 hours ago									

# DASHBOARD

patient\_churn\_dashboard



Edit Draft



5h ago

Schedule



Share

Untitled page

Untitled page 1

Total Patients

1K

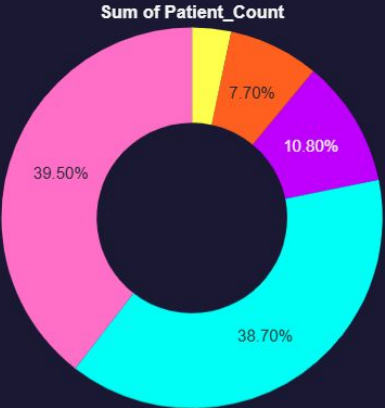
Total Very High Risk Patients

33

Average Churn Risk

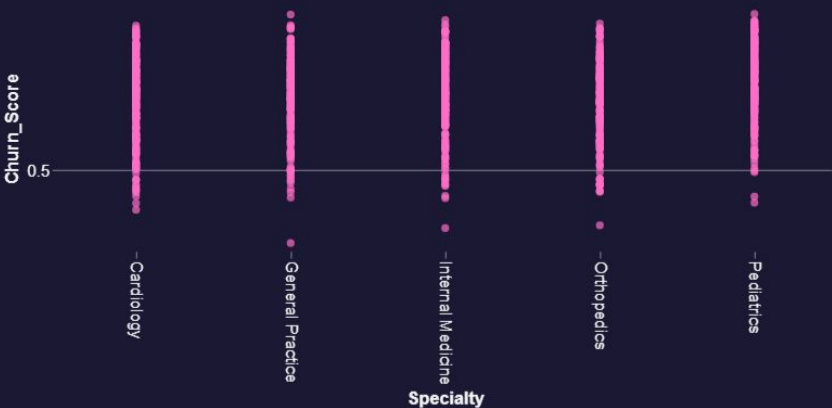
0.71

Patient Count per Risk Segment






Risk\_Segment





- High Risk
- Moderate Risk
- Very Low Risk
- Low Risk
- Very High Risk






# GOVERNANCE




Catalog Explorer > patient\_churn\_prediction > patient\_data >

 **gold\_patients**  

 Open in a dashboard  Share  Create 

Overview Sample Data Details **Permissions** Policies History Lineage Insights Quality

Grant Revoke Privileges  Inherited  Type to filter by principal 

Principal	Privilege	Object
All account users	APPLY TAG	 patient_churn_prediction.patient_data.gold_patients
All account users	MODIFY	 patient_churn_prediction.patient_data.gold_patients
All account users	SELECT	 patient_churn_prediction.patient_data.gold_patients

# CONCLUSION

## MODEL

- Requires more data or better feature engineering to train the model.
- Need to monitor new data to retrain the model



# CONCLUSION

## DATABRICKS

- Best unified platform where we can do end to end project.
- It Provides Unity Catalog Governance.
- It provides seamless connectivity to various other platform