رواد مصر الرقمية

وزارة الاتصـــــالات
وتكنولوجيا المعلومات

DEPI Graduation Project SHR1_AIS3_M1e

HealthCare Predictive Analytics

Group Members & Assigned Tasks:-

- **Ahmed Magdy Ahmed (Leader)**
  - Data Collection
  - ML-Flow
- **Khaled Tarek Mohamed**
  - Data Exploration
  - Data Analysis
  - Data Visualization
- **Modather Abdelmohsen Abdelmawgood**
  - Data Prediction

Contents:-

- **HealthCare Predictive Analytics**
  - **Data Collection**
  - **Data Exploration**
  - **Data Analysis**
  - **Data Visualization**
  - **Data Prediction**
  - **Data ML-Flow**

HealthCare Predictive Analytics

Data Collection

Data Set:-

Name	Age	Gender	Blood Type	Medical C...
Bobby JacksOn	30	Male	B-	Cancer
LesLie TErRy	62	Male	A+	Obesity
DaNnY sMitH	76	Female	A-	Obesity
andrEw waTtS	28	Female	O+	Diabetes
adrIENNE bEll	43	Female	AB+	Cancer
EMILY JOHNSOn	36	Male	A+	Asthma
edwArD EDWaRDs	21	Female	AB-	Diabetes

Data Set:-

Date of A...	Doctor	Hospital	Insurance...	# Billing Am...
2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18856.281305978155
2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327286577885
2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096078842456
2020-11-18	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.78240987528
2022-09-19	Kathleen Hanna	White-White	Aetna	14238.317813937623
2023-12-20	Taylor Newton	Nunez-Humphrey	UnitedHealthcare	48145.11095104189
2020-11-03	Kelly Olson	Group Middleton	Medicare	19580.87234486093
2021-12-28	Suzanne Thomas	Powell Robinson and Valdez,	Cigna	45820.46272159459

Data Set:-

# Room Nu...	Admissio...	Discharge...	Medication	Test Resu...
328	Urgent	2024-02-02	Paracetamol	Normal
265	Emergency	2019-08-26	Ibuprofen	Inconclusive
205	Emergency	2022-10-07	Aspirin	Normal
450	Elective	2020-12-18	Ibuprofen	Abnormal
458	Urgent	2022-10-09	Penicillin	Abnormal
389	Urgent	2023-12-24	Ibuprofen	Normal
389	Emergency	2020-11-15	Paracetamol	Inconclusive
277	Emergency	2022-01-07	Paracetamol	Inconclusive
316	Elective	2020-07-14	Aspirin	Abnormal

HealthCare Predictive Analytics

Data Exploration

About The Data:-

❖Shape

- 55500 Rows
- 15 Columns

❖Size (RxC)

- 832500

❖Data Types

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55500 entries, 0 to 55499
Data columns (total 15 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Name                55500 non-null  object
 1   Age                 55500 non-null  int64
 2   Gender              55500 non-null  object
 3   Blood Type          55500 non-null  object
 4   Medical Condition   55500 non-null  object
 5   Date of Admission   55500 non-null  object
 6   Doctor              55500 non-null  object
 7   Hospital            55500 non-null  object
 8   Insurance Provider  55500 non-null  object
 9   Billing Amount      55500 non-null  float64
 10  Room Number         55500 non-null  int64
 11  Admission Type      55500 non-null  object
 12  Discharge Date      55500 non-null  object
 13  Medication          55500 non-null  object
 14  Test Results        55500 non-null  object
dtypes: float64(1), int64(2), object(12)
memory usage: 6.4+ MB
```

About The Data:-

❖Description

	Age	Billing Amount	Room Number
count	55500.000000	55500.000000	55500.000000
mean	51.539459	25539.316097	301.134829
std	19.602454	14211.454431	115.243069
min	13.000000	-2008.492140	101.000000
25%	35.000000	13241.224652	202.000000
50%	52.000000	25538.069376	302.000000
75%	68.000000	37820.508436	401.000000
max	89.000000	52764.276736	500.000000

About The Data:-

❖Description

	Age	Billing Amount	Room Number
count	55500.000000	55500.000000	55500.000000
mean	51.539459	25539.316097	301.134829
std	19.602454	14211.454431	115.243069
min	13.000000	-2008.492140	101.000000
25%	35.000000	13241.224652	202.000000
50%	52.000000	25538.069376	302.000000
75%	68.000000	37820.508436	401.000000
max	89.000000	52764.276736	500.000000

HealthCare Predictive Analytics

Data Analysis

About The Data:-

- **Cleanings**
  - Duplicates
  - Missing Values

- **Medical Condition**
  - Arthritis
  - Diabetes
  - Hypertension
  - Obesity
  - Cancer
  - Asthma

About The Data:-

❖Medication

- Lipitor
- Ibuprofen
- Aspirin
- Paracetamol
- Penicillin

❖Test Results

- Normal
- Abnormal
- Inconclusive

About The Data:-

❖Insurance Providers

- Aetna
- Blue Cross
- UnitedHealthcare
- Medicare
- Cigna

❖Admission Types

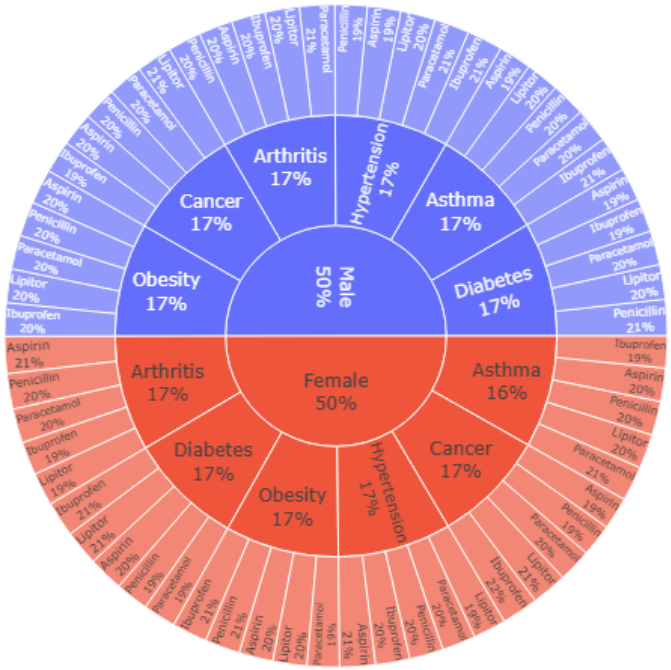
- Elective
- Urgent
- Emergency

HealthCare Predictive Analytics

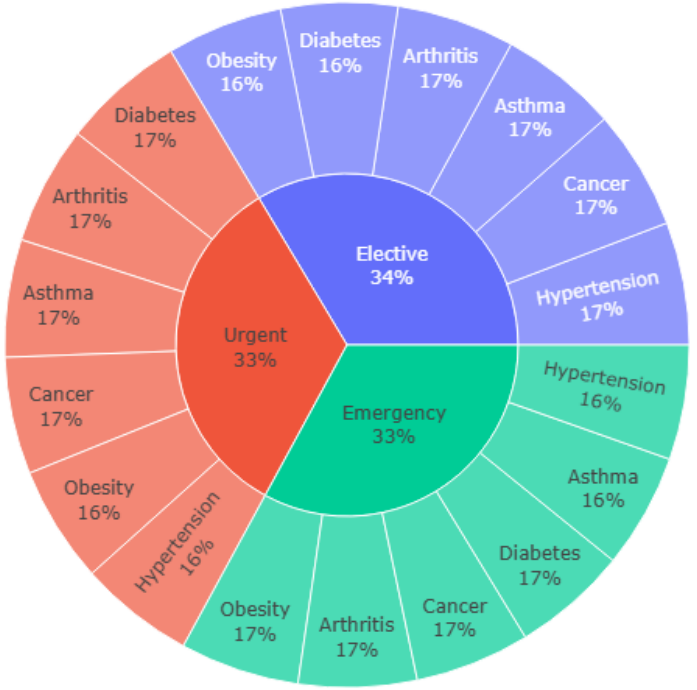
Data Visualization

Visualizations:-

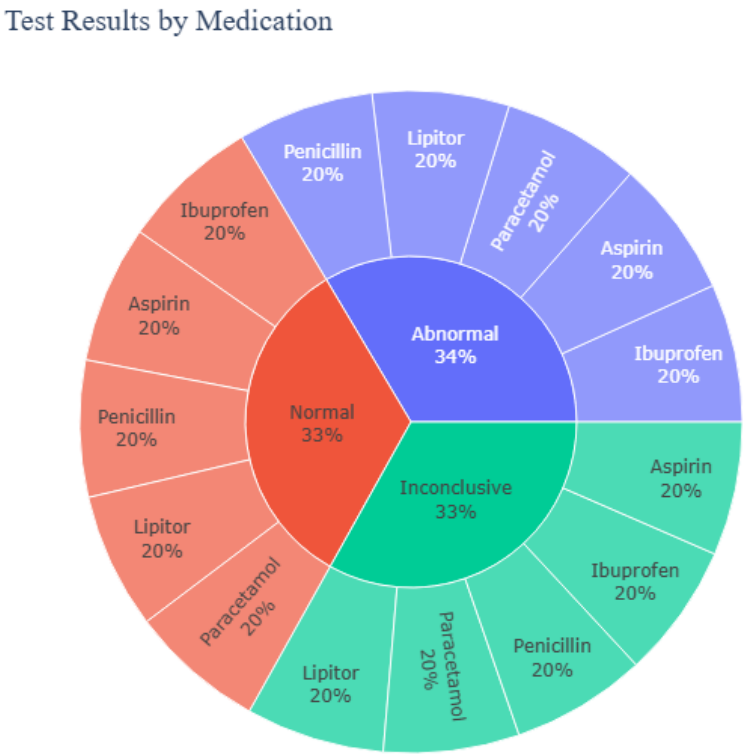
Patient's Status

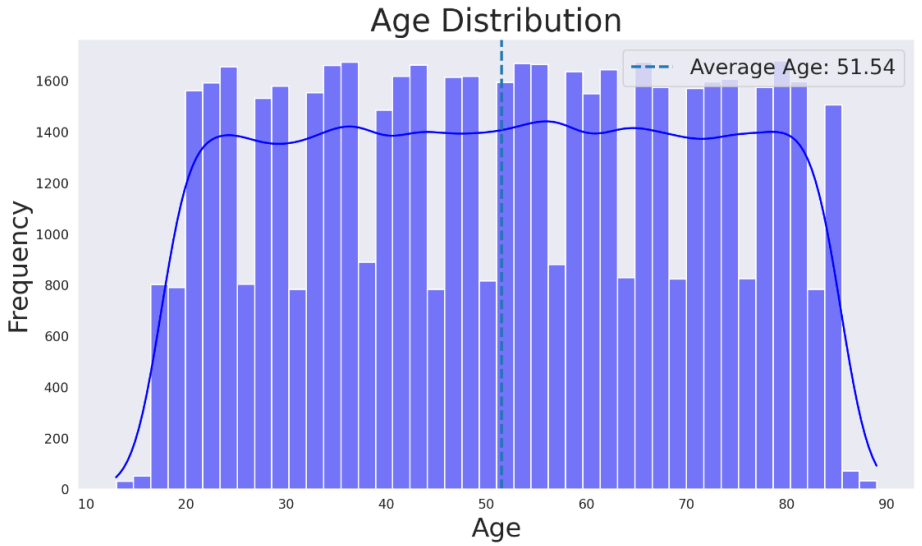


Admission Type by Medical Condition



Visualizations:-





HealthCare Predictive Analytics

Data Prediction

Prediction:-

- **Prediction Target**
  - Test Results
- **Prediction Categories**
  - All Data Except (Test Results)
- **Main Method**
  - Decision Tree Classifier (DTC)
- **Optimized Method**
  - Random Forest Classifier (RFC)

Prediction:-

❖Main Method Prediction Results

- Accuracy: *44.90%*
- Precision Score: *0.4493153322783424*7
- Recall Score: *0.449*
- F1 Score: *0.449129234867101*

❖Optimized Method Prediction Results

- Accuracy: *48.65%*
- Precision Score: *0.4870509927393602*6
- Recall Score: *0.4865*
- F1 Score: *0.4865200362489088*

Prediction:-

❖Prediction Issues

- String Data Types

Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
Bobby JacksOn	30	Male	B-	Cancer	2024-01-31	Matthew Smith	Sons and Miller	Blue Cross	18856.281306	328	Urgent	2024-02-02	Paracetamol	Normal
LesLie TErRy	62	Male	A+	Obesity	2019-08-20	Samantha Davies	Kim Inc	Medicare	33643.327287	265	Emergency	2019-08-26	Ibuprofen	Inconclusive
DaNnY sMitH	76	Female	A-	Obesity	2022-09-22	Tiffany Mitchell	Cook PLC	Aetna	27955.096079	205	Emergency	2022-10-07	Aspirin	Normal
andrEw waTtS	28	Female	O+	Diabetes	2020-11-18	Kevin Wells	Hernandez Rogers and Vang,	Medicare	37909.782410	450	Elective	2020-12-18	Ibuprofen	Abnormal
adrIENNE bEll	43	Female	AB+	Cancer	2022-09-19	Kathleen Hanna	White-White	Aetna	14238.317814	458	Urgent	2022-10-09	Penicillin	Abnormal

Prediction:-

❖Issue Handling

- Label Encoding

Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount	Room Number	Admission Type	Discharge Date	Medication	Test Results
3068	30	1	5	2	1729	26612	29933	1	18856.281306	328	2	1730	3	2
15211	62	1	0	5	104	33648	16012	3	33643.327287	265	1	109	1	1
6476	76	0	1	5	1233	37828	5473	0	27955.096079	205	1	1247	0	2
26935	28	0	6	3	560	22511	12317	3	37909.782410	450	0	589	1	0
26241	43	0	2	2	1230	21259	33598	0	14238.317814	458	2	1249	4	0

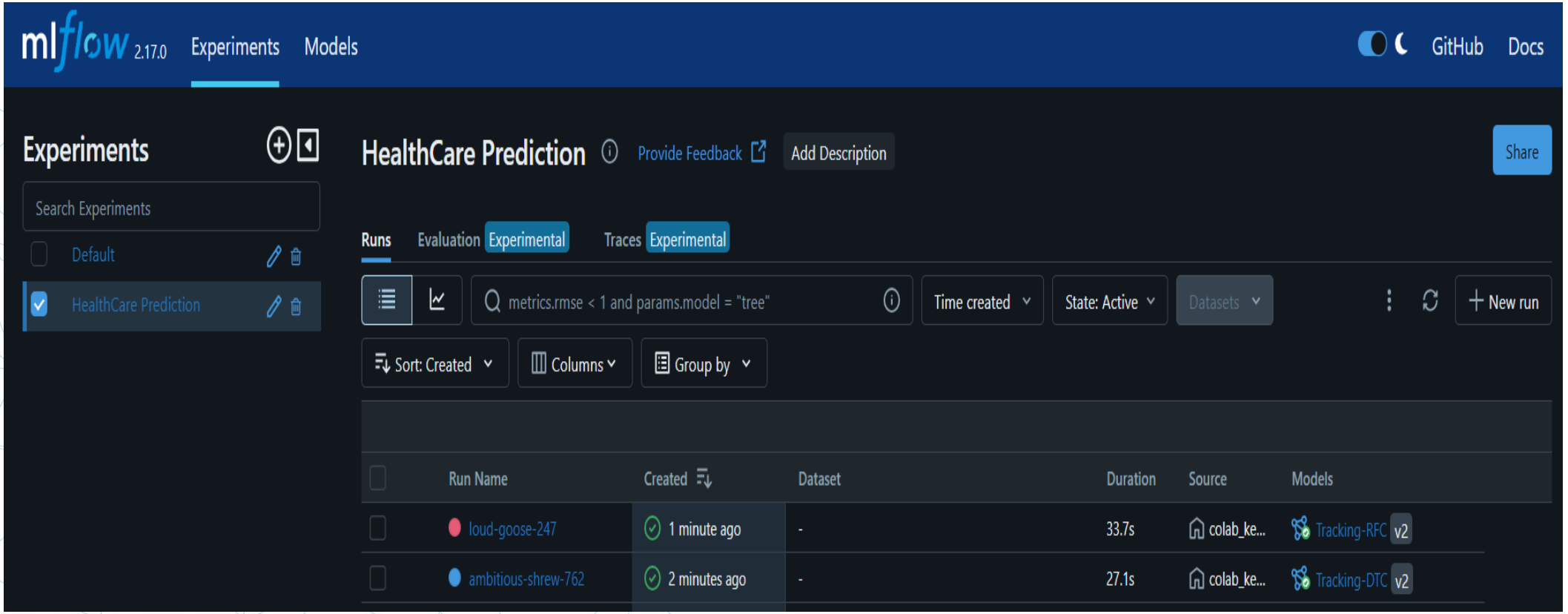
HealthCare Predictive Analytics

Data ML-Flow

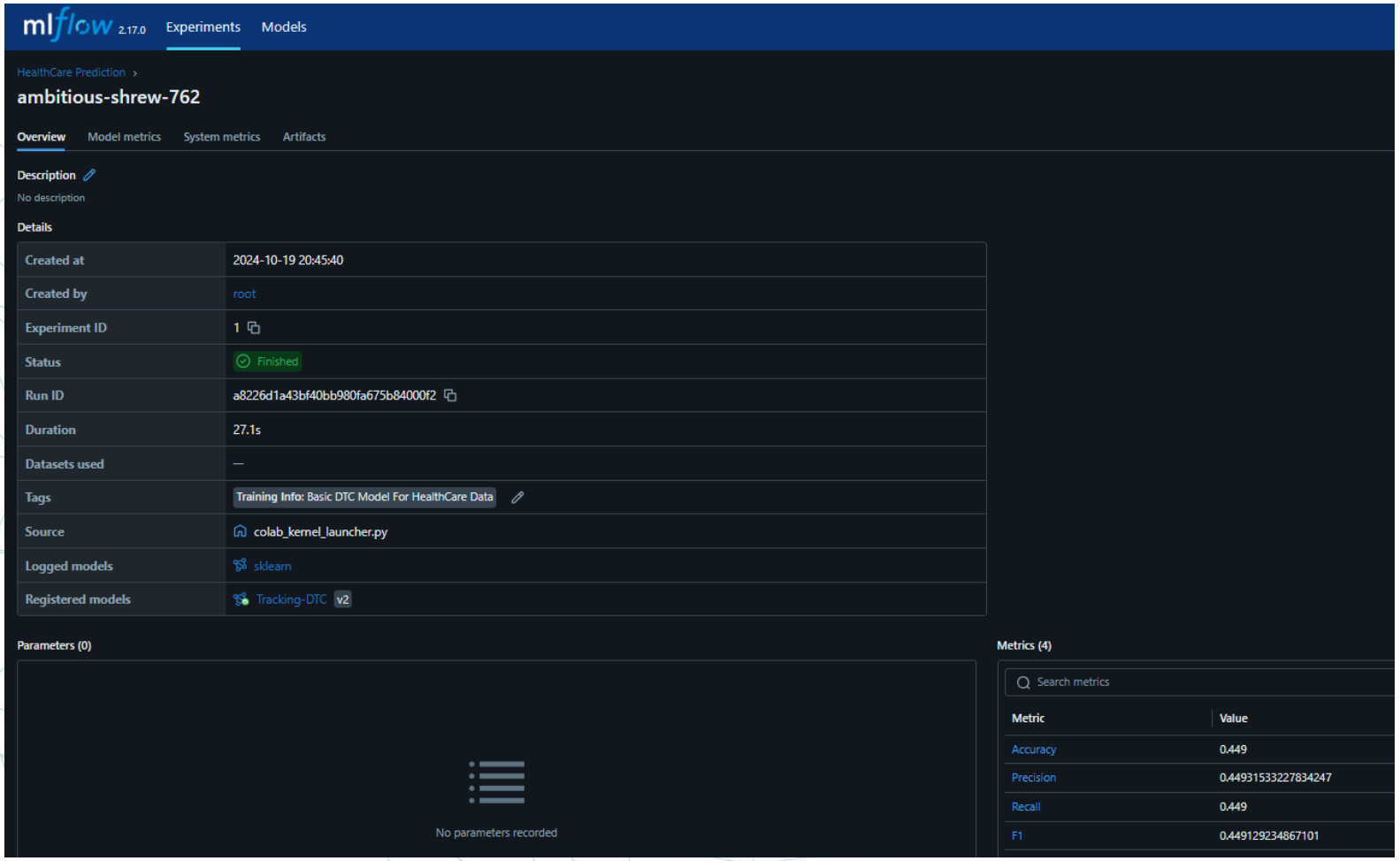
ML-Flow:-

- **Experiment Name**
  - HealthCare Prediction
- **Tracking**
  - Decision Tree Classifier (DTC)
  - Random Forest Classifier (RFC)
- **Metrices**
  - Accuracy
  - Precision
  - Recall
  - F1

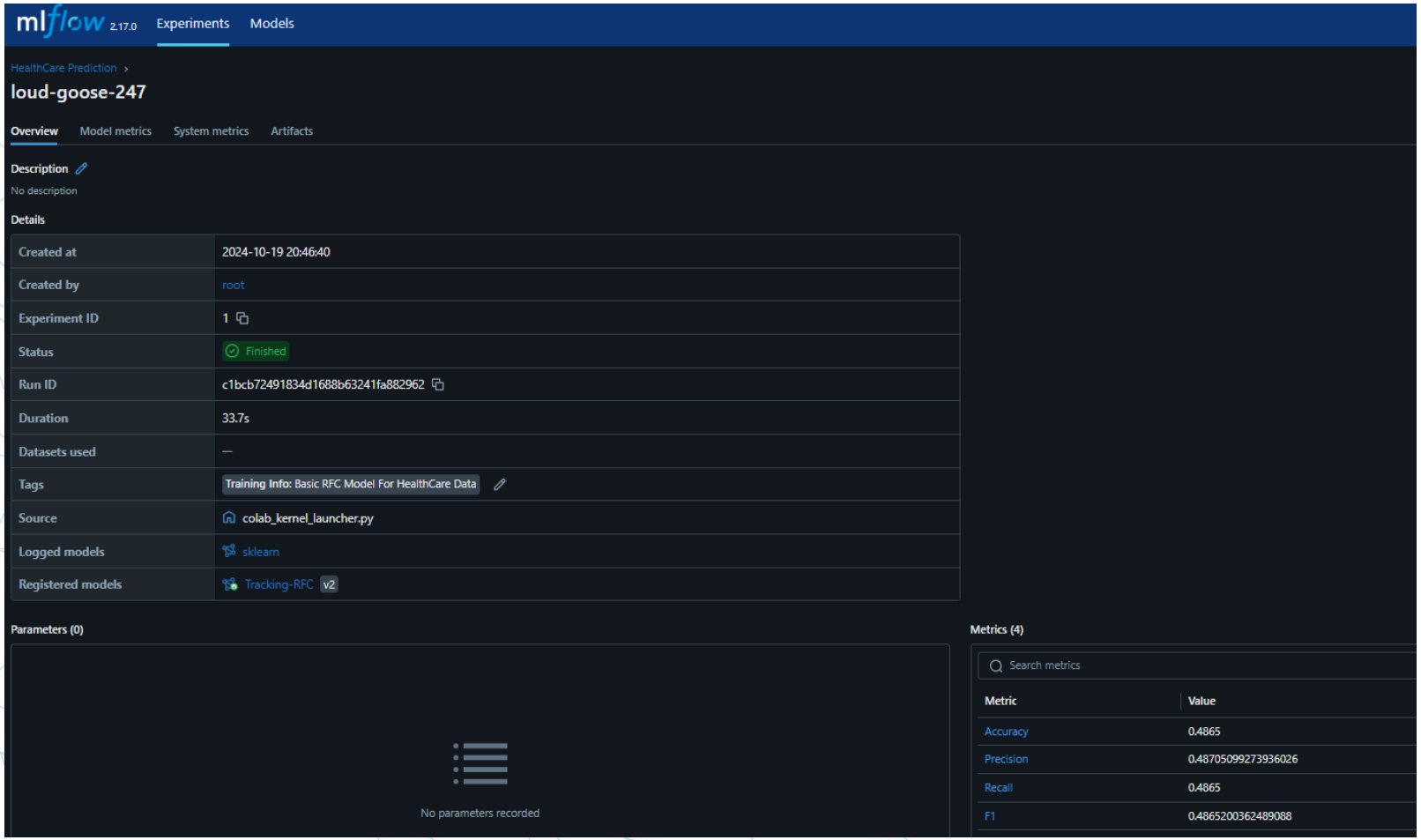
ML-Flow Server:-



ML-Flow Server:-



ML-Flow Server:-



Thank You !