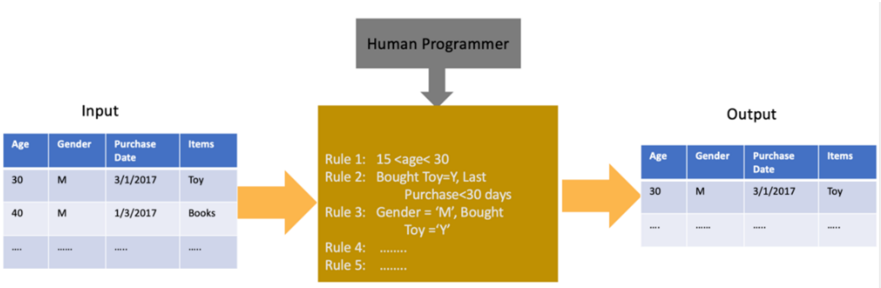
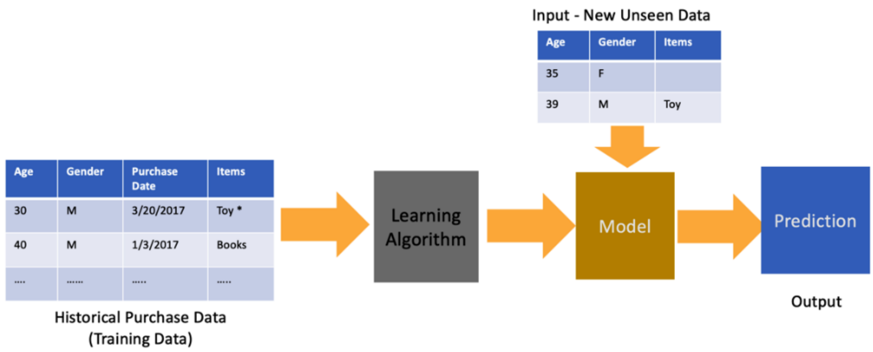
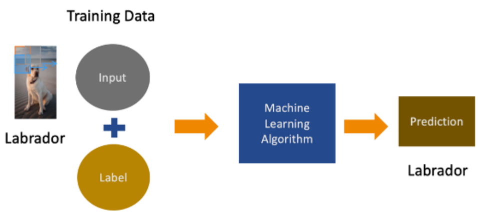
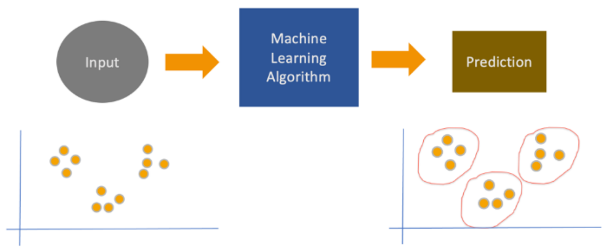
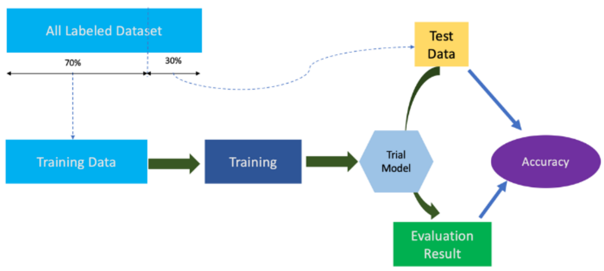
Chapter 1: Introducing Machine Learning and the AWS Machine Learning Stack

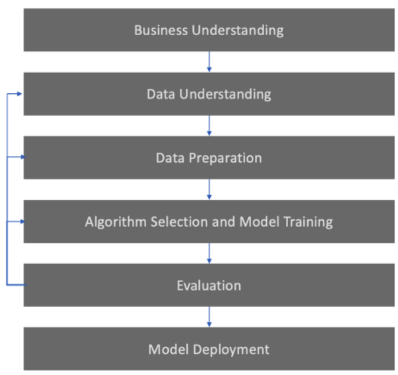


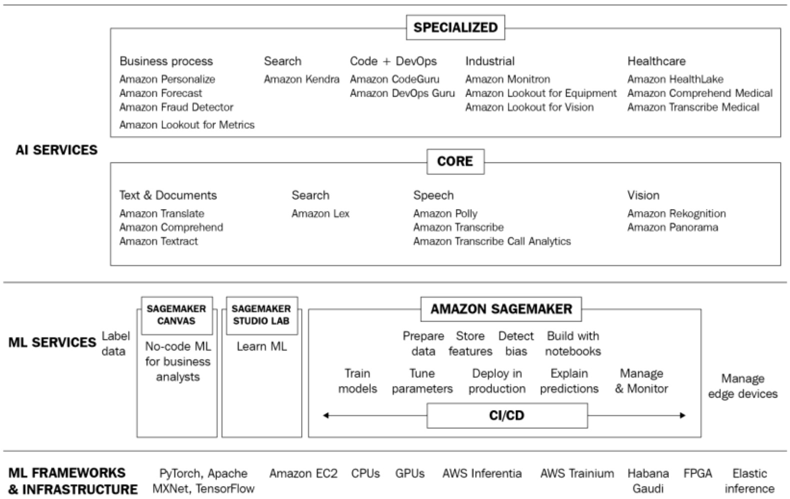




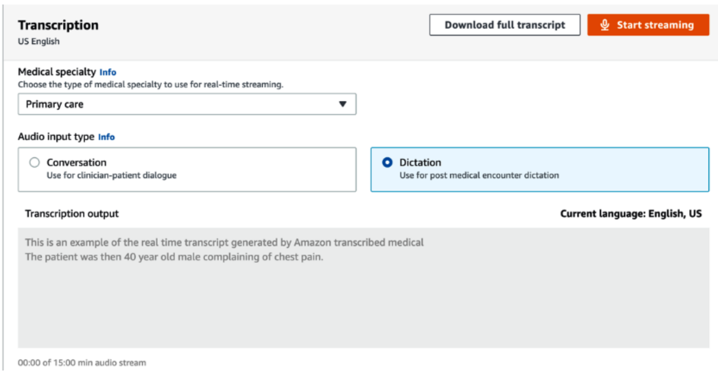


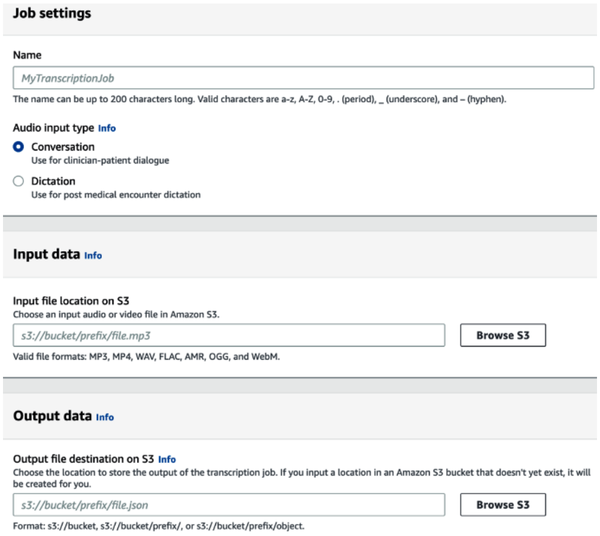


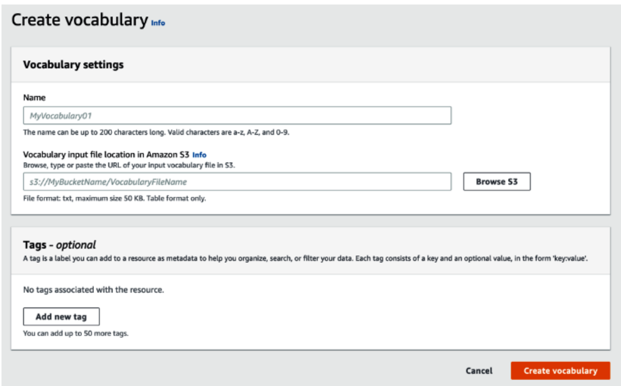


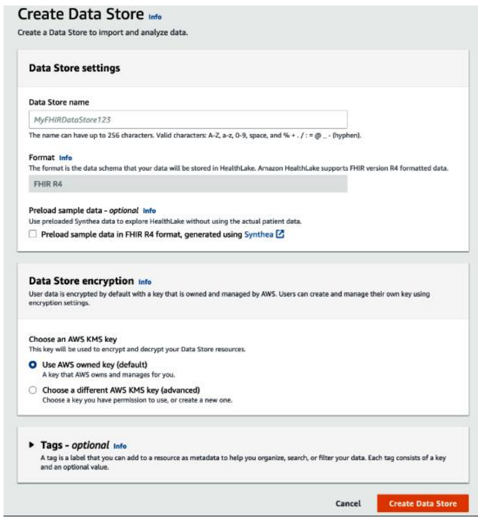

AI SERVICES
SPECIALIZED
Business process
Amazon Personalize
Amazon Forecast
Amazon Fraud Detector
Amazon Lookout for Metrics
Search
Amazon Kendra
Code + DevOps
Amazon CodeGuru
Amazon DevOps Guru
Industrial
Amazon Monitron
Amazon Lookout for Equipment
Amazon Lookout for Vision
Healthcare
Amazon HealthLake
Amazon Comprehend Medical
Amazon Transcribe Medical
CORE
Text & Documents
Amazon Translate
Amazon Comprehend
Amazon Textract
Search
Amazon Lex
Speech
Amazon Polly
Amazon Transcribe
Amazon Transcribe Call Analytics
Vision
Amazon Rekognition
Amazon Panorama
ML SERVICES
Label data
SAGEMAKER CANVAS
No-code ML for business analysts
SAGEMAKER STUDIO LAB
Learn ML
AMAZON SAGEMAKER
Prepare data
Store features
Detect bias
Build with notebooks
Train models
Tune parameters
Deploy in production
Explain predictions
Manage & Monitor
CI/CD
Manage edge devices
ML FRAMEWORKS & INFRASTRUCTURE
PyTorch, Apache MXNet, TensorFlow
Amazon EC2
CPUs
GPUs
AWS Inferentia
AWS Trainium
Habana Gaudi
FPGA
Elastic inference

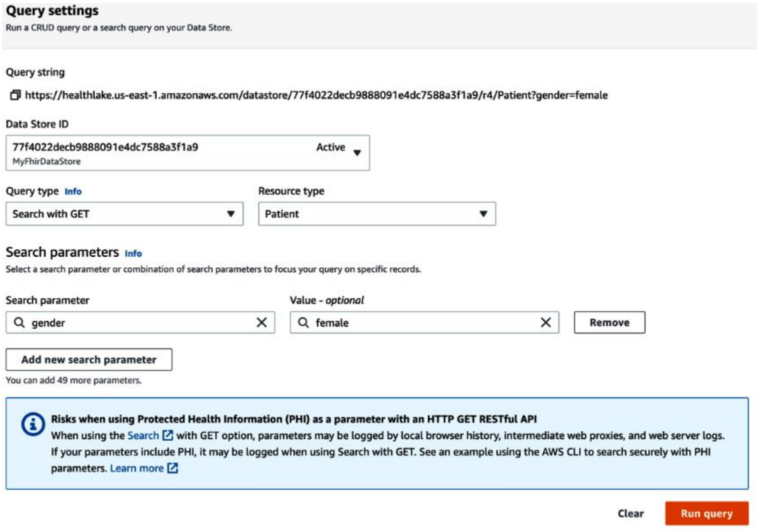
Chapter 2: Exploring Key AWS Machine Learning Services for Healthcare and Life Sciences

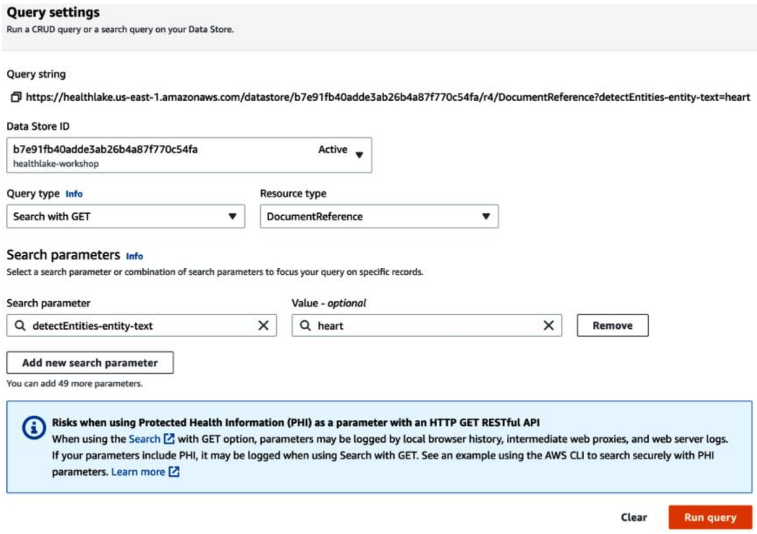




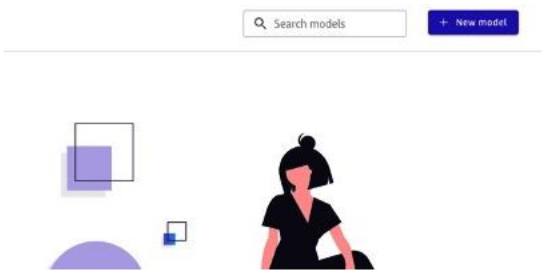


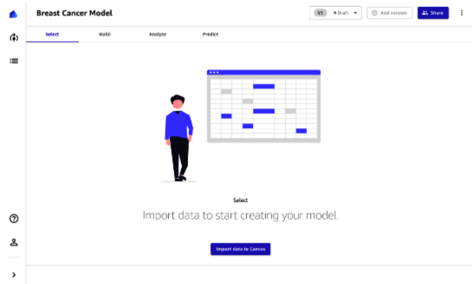


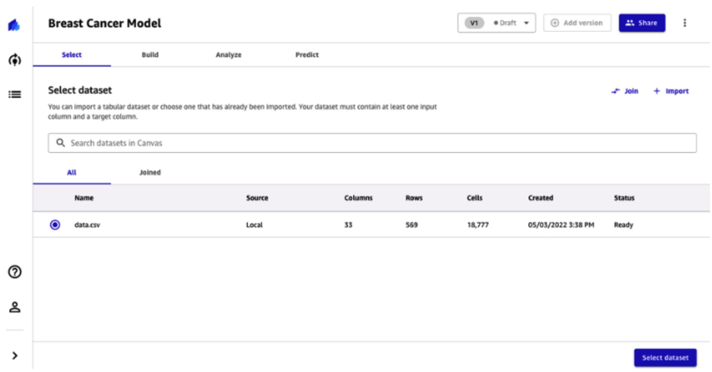

Query settings
Run a CRUD query or a search query on your Data Store.
Query string
https://healthlake.us-east-1.amazonaws.com/datastore/77f4022decb9888091e4dc7588a3f1a9/r4/Patient?gender=female
Data Store ID
77f4022decb9888091e4dc7588a3f1a9
MyFhirDataStore
Active
Query type Info
Search with GET
Resource type
Patient
Search parameters Info
Select a search parameter or combination of search parameters to focus your query on specific records.
Search parameter
gender
Value - optional
female
Remove
Add new search parameter
You can add 49 more parameters.
Risks when using Protected Health Information (PHI) as a parameter with an HTTP GET RESTful API
When using the Search with GET option, parameters may be logged by local browser history, intermediate web proxies, and web server logs. If your parameters include PHI, it may be logged when using Search with GET. See an example using the AWS CLI to search securely with PHI parameters. Learn more
Clear
Run query

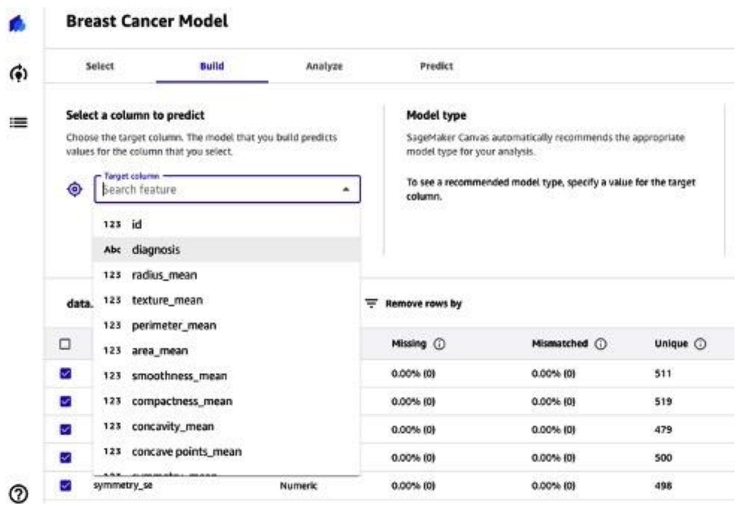

Query settings
Run a CRUD query or a search query on your Data Store.
Query string
https://healthlake.us-east-1.amazonaws.com/datastore/b7e91fb40adde3ab26b4a87f770c54fa/r4/DocumentReference?detectEntities-entity-text=heart
Data Store ID
b7e91fb40adde3ab26b4a87f770c54fa
healthlake-workshop
Active
Query type Info
Search with GET
Resource type
DocumentReference
Search parameters Info
Select a search parameter or combination of search parameters to focus your query on specific records.
Search parameter
detectEntities-entity-text
Value - optional
heart
Remove
Add new search parameter
You can add 49 more parameters.
Risks when using Protected Health Information (PHI) as a parameter with an HTTP GET RESTful API
When using the Search with GET option, parameters may be logged by local browser history, intermediate web proxies, and web server logs. If your parameters include PHI, it may be logged when using Search with GET. See an example using the AWS CLI to search securely with PHI parameters. Learn more
Clear
Run query

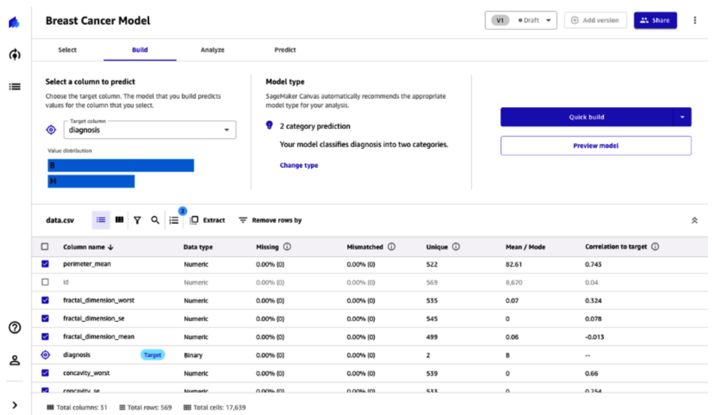
Chapter 3: Machine Learning for Patient Risk Stratification

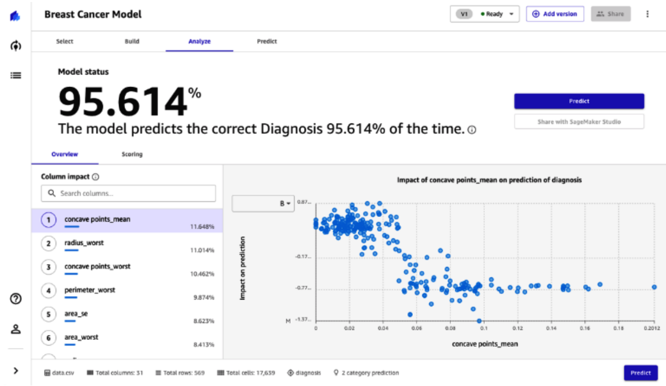


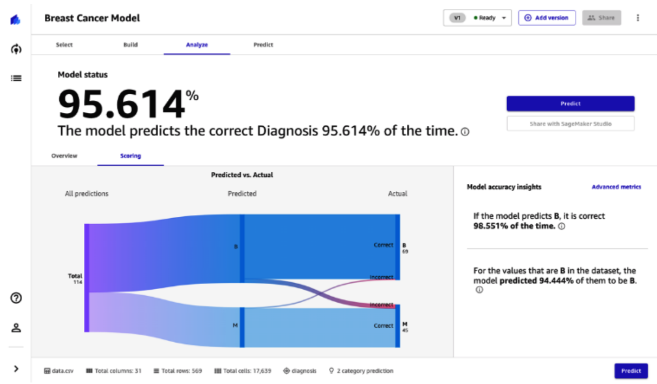




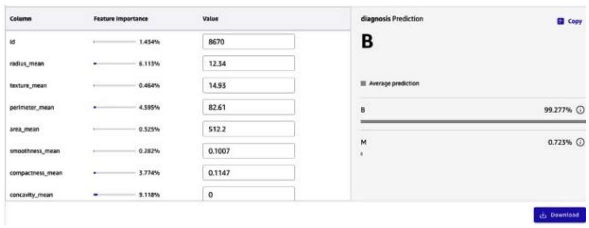




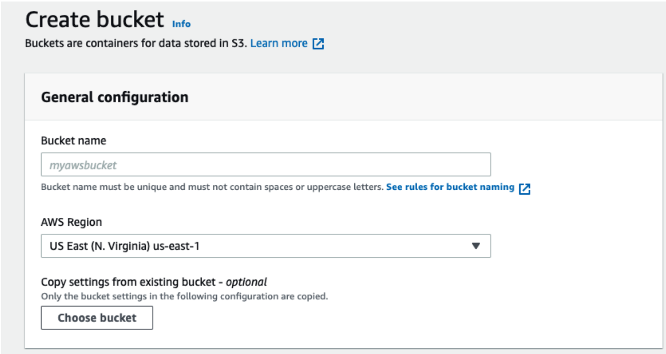

Breast Cancer Model
V1
Ready
Add version
Share
Select
Build
Analyze
Predict
Model status
95.614%
The model predicts the correct Diagnosis 95.614% of the time.
Predict
Share with SageMaker Studio
Overview
Scoring
Column impact
Search columns...
1 concave points_mean 11.648%
2 radius_worst 11.014%
3 concave points_worst 10.462%
4 perimeter_worst 9.874%
5 area_se 8.623%
6 area_worst 8.413%
Impact of concave points_mean on prediction of diagnosis
Impact on prediction
concave points_mean
data.csv
Total columns: 31
Total rows: 569
Total cells: 17,639
diagnosis
2 category prediction
Predict

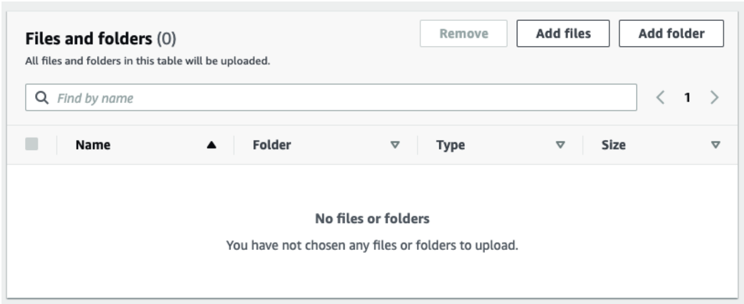

Breast Cancer Model
V1
Ready
Add version
Share
Select
Build
Analyze
Predict
Model status
95.614%
The model predicts the correct Diagnosis 95.614% of the time.
Predict
Share with SageMaker Studio
Overview
Scoring
Predicted vs. Actual
All predictions
Predicted
Actual
Total
Correct
Incorrect
Model accuracy insights
Advanced metrics
If the model predicts B, it is correct 98.551% of the time.
For the values that are B in the dataset, the model predicted 94.444% of them to be B.
data.csv
Total columns: 31
Total rows: 569
Total cells: 17,639
diagnosis
2 category prediction
Predict

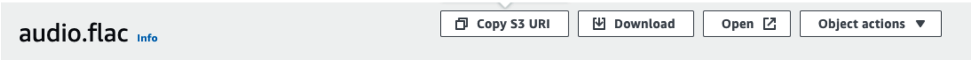

Advanced metrics
Positive Class
B
M
F1
96.454%
Accuracy
95.614%
Precision
98.551%
Recall
94.444%
AUC
0.989%
Model performance
Predicted values
Actual values
True positive (TP)
68
59.6% of predicted results
False negative (FN)
4
3.5% of predicted results
False positive (FP)
1
0.9% of predicted results
True negative (TN)
41
36.0% of predicted results
Close
Download


Column
Feature importance
Value
id
1.454%
8670
radius_mean
6.113%
12.34
texture_mean
0.464%
14.93
perimeter_mean
4.395%
82.61
area_mean
0.325%
512.2
smoothness_mean
0.282%
0.1007
compactness_mean
3.774%
0.1147
concavity_mean
9.118%
0
diagnosis Prediction
Copy
B
Average prediction
B
99.277%
M
0.723%
Download

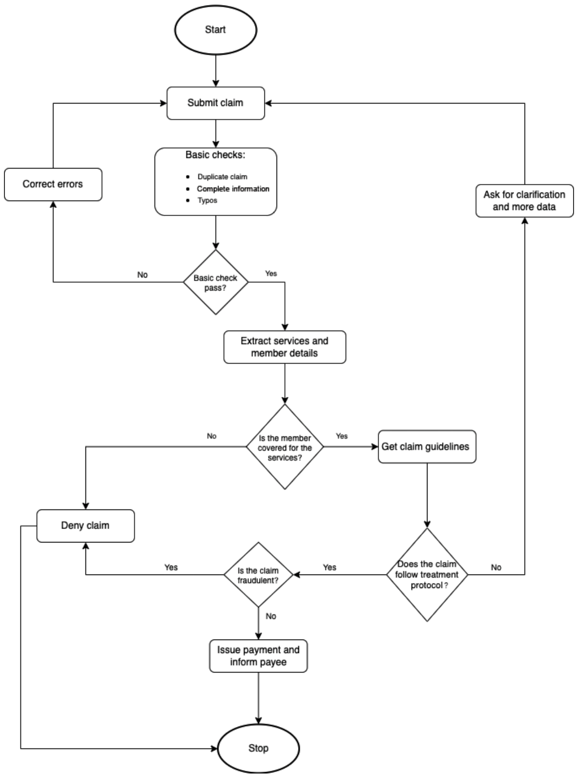
Chapter 4: Using Machine Learning to Improve Operational Efficiency for Healthcare Providers

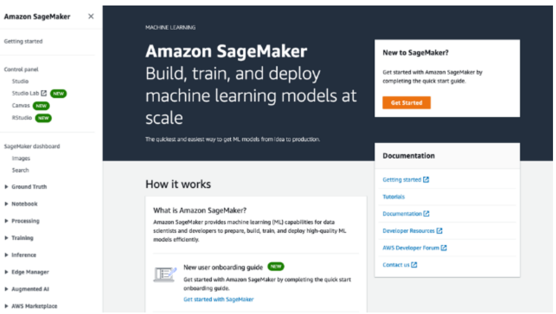


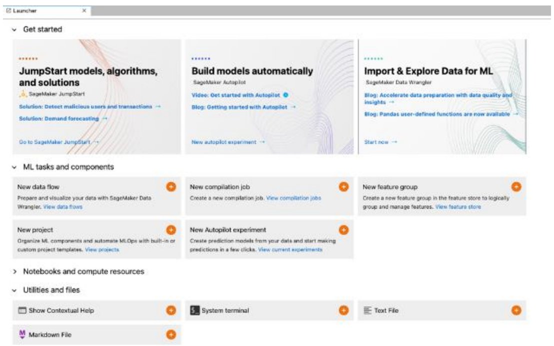


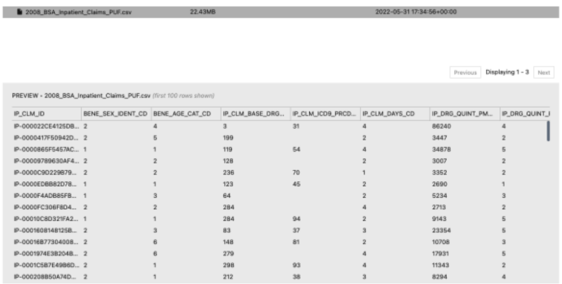


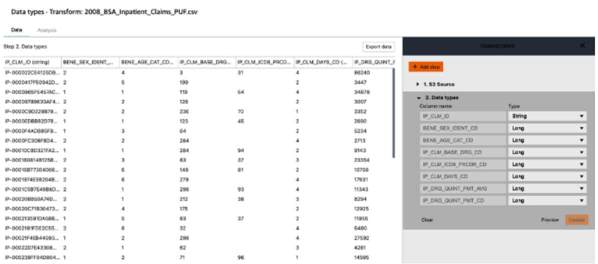
Chapter 5: Implementing Machine Learning for Healthcare Payors

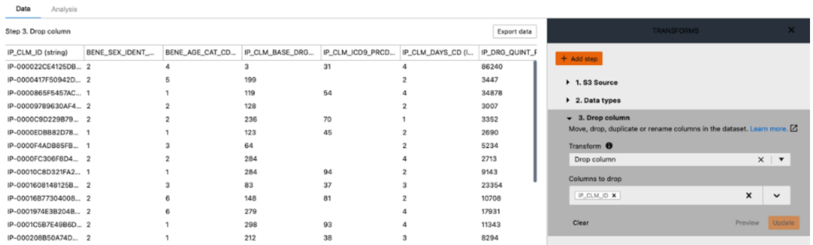


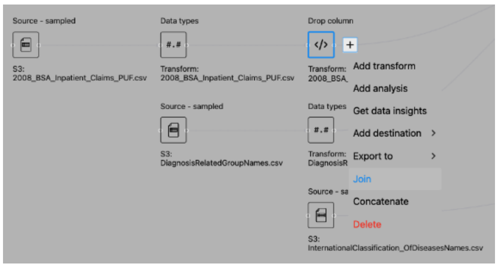


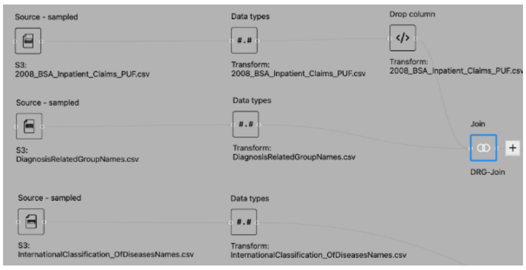




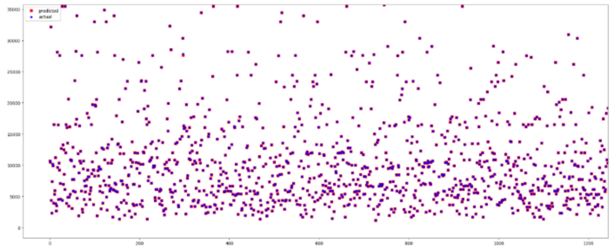



Data
Analysis
Step 3. Drop column
Export data
IP_CLM_ID (string)
BENE_SEX_IDENT_...
BENE_AGE_CAT_CD...
IP_CLM_BASE_DRG...
IP_CLM_ICD9_PRCD...
IP_CLM_DAYS_CD (I...
IP_DRG_QUINT_P
IP-000022CE4125DB... 2 4 3 31 4 86240
IP-0000417F50942D... 2 5 199 2 3447
IP-0000865F5457AC... 1 1 119 54 4 34878
IP-000097B9630AF4... 2 2 128 2 3007
IP-0000C9D229B79... 2 2 236 70 1 3352
IP-0000EDBB82D78... 1 1 123 45 2 2690
IP-0000F4ADB85FB... 1 3 64 2 5234
IP-0000FC306F8D4... 2 2 284 4 2713
IP-00010C8D321FA2... 1 1 284 94 2 9143
IP-0001608148125B... 2 3 83 37 3 23354
IP-00016B7730400B... 2 6 148 81 2 10708
IP-0001974E3B204B... 2 6 279 4 17931
IP-0001C5B7E49B6D... 2 1 298 93 4 11343
IP-000208B50A74D... 2 1 212 38 3 8294
TRANSFORMS
+ Add step
1. S3 Source
2. Data types
3. Drop column
Move, drop, duplicate or rename columns in the dataset. Learn more.
Transform
Drop column
Columns to drop
IP_CLM_ID
Clear
Preview
Update

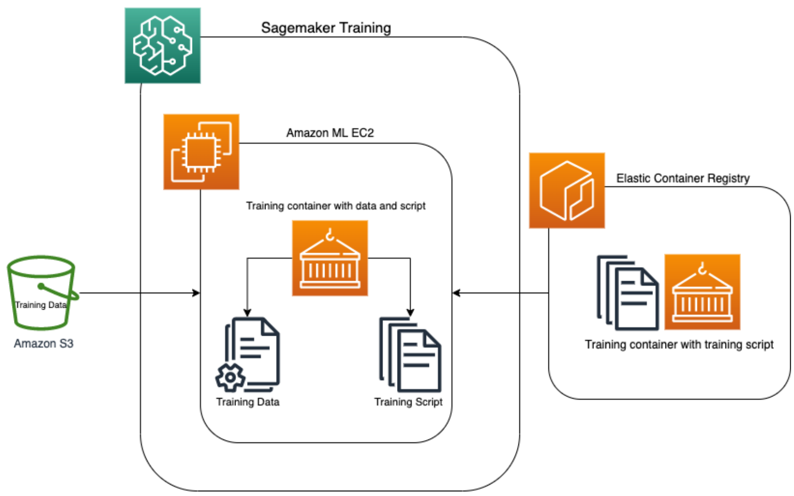

Source - sampled
S3:
2008_BSA_Inpatient_Claims_PUF.csv
Data types
Transform:
2008_BSA_Inpatient_Claims_PUF.csv
Drop column
Transform:
2008_BSA
Source - sampled
S3:
DiagnosisRelatedGroupNames.csv
Data types
Transform:
DiagnosisR
Source - sa
S3:
InternationalClassification_OfDiseasesNames.csv
Add transform
Add analysis
Get data insights
Add destination
Export to
Join
Concatenate
Delete

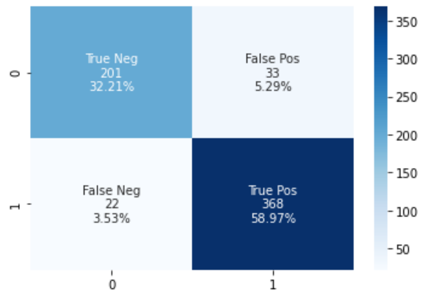

Source - sampled
S3:
2008_BSA_Inpatient_Claims_PUF.csv
Data types
Transform:
2008_BSA_Inpatient_Claims_PUF.csv
Drop column
Transform:
2008_BSA_Inpatient_Claims_PUF.csv
Source - sampled
S3:
DiagnosisRelatedGroupNames.csv
Data types
Transform:
DiagnosisRelatedGroupNames.csv
Join
DRG-Join
Source - sampled
S3:
InternationalClassification_OfDiseasesNames.csv
Data types
Transform:
InternationalClassification_OfDiseasesNames.csv

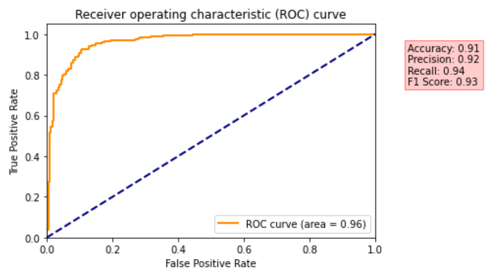

Back to transform
Export data
Export your data in your current session to S3.
Amazon S3 location
s3://sagemaker
Browse
File type
CSV (*.csv)
Delimiter
Comma (,)
Compression
None
KMS key ID or ARN
Optional


predicted
actual

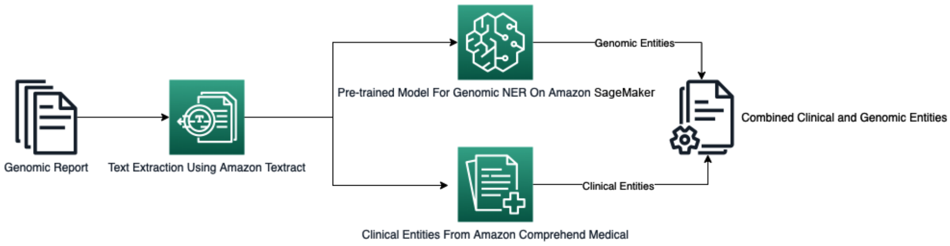
Chapter 6: Implementing Machine Learning for Medical Devices and Radiology Images







Chapter 7: Applying Machine Learning to Genomics

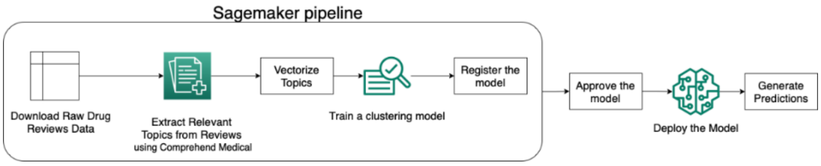


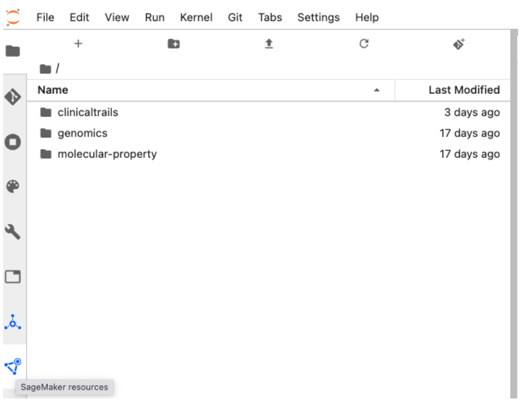
Chapter 8: Applying Machine Learning to Molecular Data

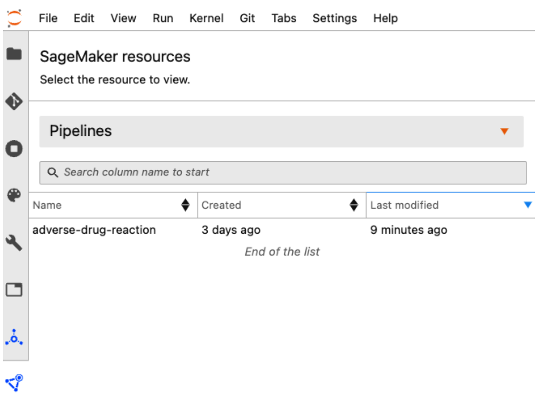
```
! ls models/
```

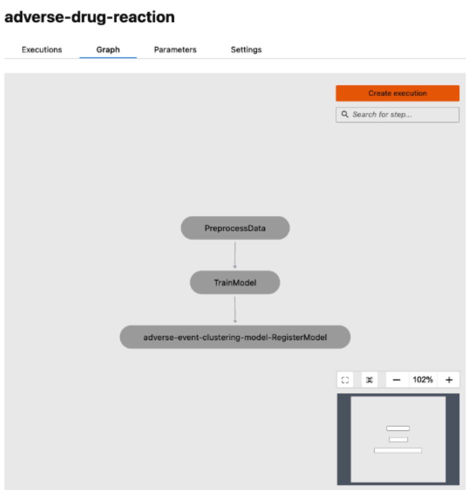
```
bbb_martins_model                        cyp3a4_substrate_carbonmangels_model
bioavailability_ma_model                 cyp3a4_veith_model
caco2_wang_model                         half_life_obach_model
clearance_hepatocyte_az_model            hia_hou_model
clearance_microsome_az_model             hydrationfreeenergy_freesolv_model
cyp1a2_veith_model                       lipophilicity_astrazeneca_model
cyp2c19_veith_model                      pgp_broccatelli_model
cyp2c9_substrate_carbonmangels_model     ppbr_az_model
cyp2c9_veith_model                       solubility_aqsoldb_model
cyp2d6_substrate_carbonmangels_model     vdss_lombardo_model
cyp2d6_veith_model
```

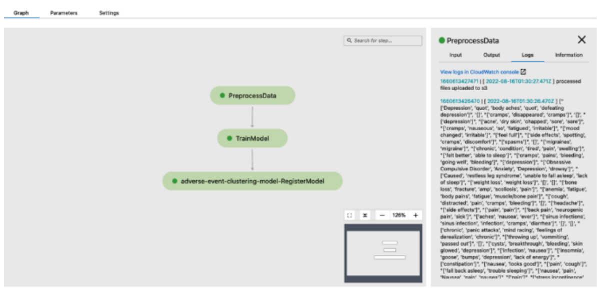
Chapter 9: Applying Machine Learning to Clinical Trials and Pharmacovigilance

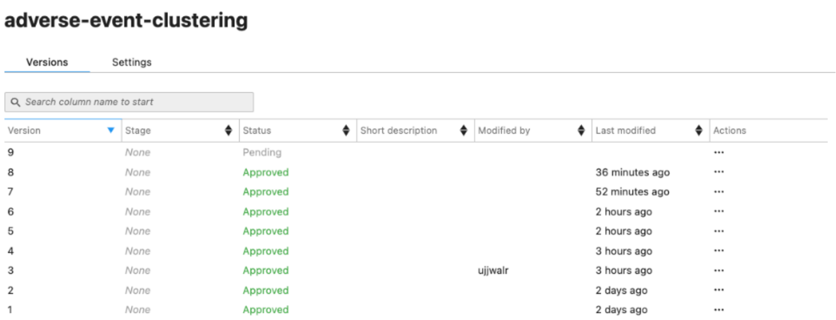




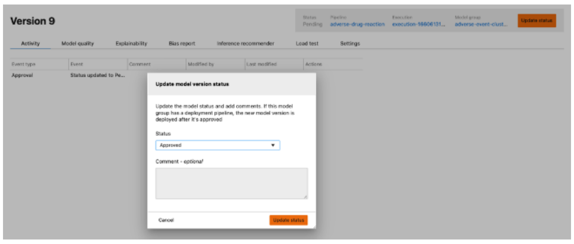


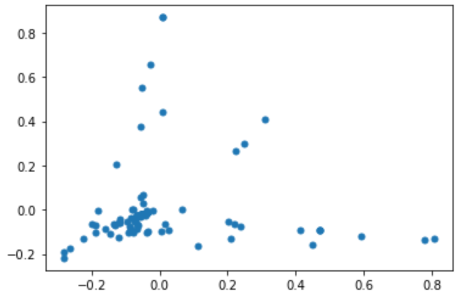




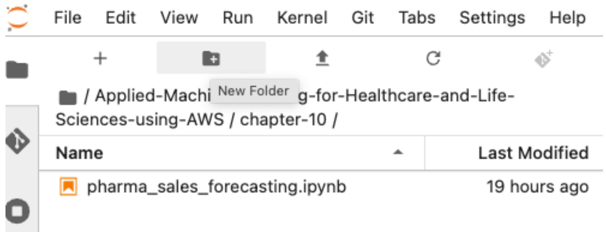


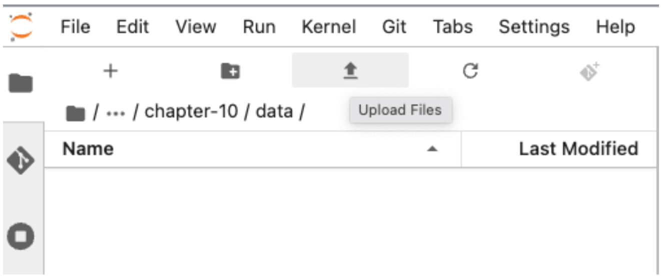

Version 9
Status
Pending
Pipeline
adverse-drug-reaction
Execution
execution-16606131...
Model group
adverse-event-clust...
Update status
Activity
Model quality
Explainability
Bias report
Inference recommender
Load test
Settings
Event type
Event
Comment
Modified by
Last modified
Actions
Approval
Status updated to Pe...
7 minutes ago


Version 9
Update model version status
Approved
Cancel

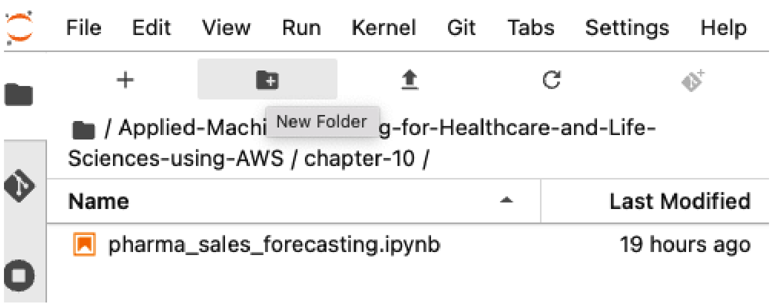

0.8
0.6
0.4
0.2
0.0
−0.2
−0.2
0.0
0.2
0.4
0.6
0.8

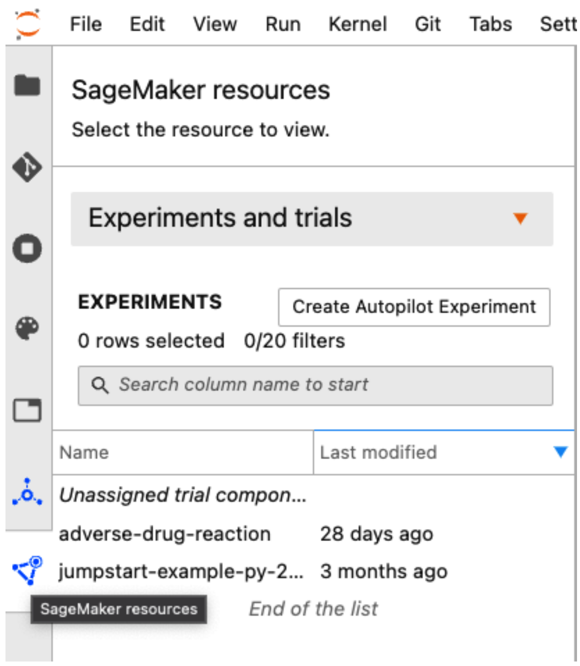
Chapter 10: Utilizing Machine Learning in the Pharmaceutical Supply Chain

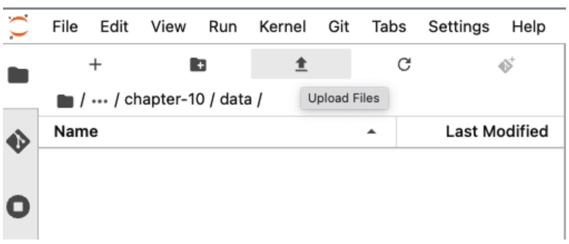


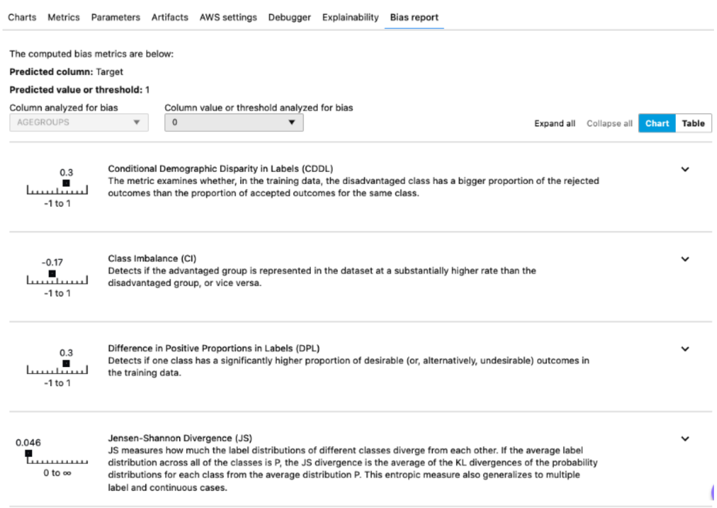


Chapter 11: Understanding Common Industry Challenges and Solutions









Charts Metrics Parameters Artifacts AWS settings Debugger Explainability **Bias report**

The computed bias metrics are below:

Predicted column: Target

Predicted value or threshold: 1

Column analyzed for bias: AGEGROUPS

Column value or threshold analyzed for bias: 1

Expand all Collapse all Chart Table

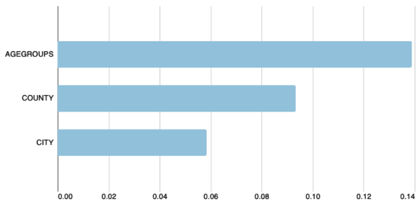
Bias metric	Bias value	Description
Conditional Demographic Disparity in Labels (CDDL)	-0.023	The metric examines whether, in the training data, the disa...
Class Imbalance (CI)	0.97	Detects if the advantaged group is represented in the data...
Difference in Positive Proportions in Labels (DPL)	-0.54	Detects if one class has a significantly higher proportion of...
Jensen-Shannon Divergence (JS)	0.052	JS measures how much the label distributions of different ...
Kullback-Liebler Divergence (KL)	-0.36	In a binary case, a relative entropy measure of how much t...
Kolmogorov-Smirnov Distance (KS)	0.54	This metric is equal to the maximum divergence in a label a...
L-p Norm (LP)	0.54	This measure of distance in label distributions is the norme...
Total Variation Distance (TVD)	0.27	This measure of distance in label distributions is half the H...
Accuracy Difference (AD)	-0.37	This metric examines whether the classification by the mod...
Conditional Demographic Disparity in Predicted Labels (CD...	-0.069	The metric examines whether the model predicted a bigger...
Difference in Acceptance Rates (DAR)	-0.051	The difference in the rates of positive predicted outcomes ...
Difference in Conditional Acceptance (DCA)	3.6	This metric compares the actual labels to the predicted lab...
Difference in Conditional Outcomes (DCR)	-0.6	This metric compares the actual labels to the predicted lab...
Disparate (Adverse) Impact (DI)	10	This metric examines whether the model predicts outcome...
Difference in Positive Proportions in Predicted Labels (DPPL)	-0.9	This metric examines whether the model predicts outcome...
Difference in Rejection Rates (DRR)	-0.6	The difference in the rates of negative predicted outcomes...
Counterfactuals: Fliptest (FT)	1	The fliptest is an approach that looks at each member of th...
Recall Difference (RD)	-0.79	Checks whether there is a difference in recall of the model ...
Treatment Equality (TE)	-71	This is defined as the difference in the ratio of false negativ...

Charts Metrics Parameters Artifacts AWS settings Debugger **Explainability** Bias report

Explaining your model's predictions

Amazon SageMaker Studio helps you understand your machine learning model by portraying the importance of its features in terms of SHAP values. We plot the aggregated SHAP value for each feature across all instances of the dataset.

Predicted column: label0



Export PDF report Download raw data