

# AWS Athena & AWS SageMaker

Dr. Villanes

## – APPLICATIONS – ENTERPRISE

**APPLICATIONS - ENTERPRISE**

SALES	MARKETING - B2B	MARKETING - B2C	CUSTOMER EXPERIENCE / SERVICE	HUMAN CAPITAL

The diagram illustrates 100+ startups categorized into 10 sectors. The sectors and their associated startups are as follows:

- LEGAL**: RAVEL, DISICO, kira, QSeal, CoConstructive, R. SS
- REGTECH & COMPLIANCE**: Digital text, IQ, TESSIAN, Smarter Knowledge, QoneTrust, ETHNYA, SECURITY.ai
- FINANCE**: Anaplan, ZUORO, SOWSAHANA, TRADESHIFF, SCALE FACTOR, botkeeper, pilot, DIGITS, appzenn
- AUTOMATION & RPA**: UiPath, bluewin, VIDADO, idenitivation, workato, Catalytic, KEYWAY, EVER READY, ALKYPI, FORTIMO, ZINGTANAGIC, Memento, CELONIS
- SECURITY**: TANJUNG, battery spark, Zscaler, StackPath, thumby, FIRETR, feedzoo, DATAVISIO, pingree, sift science, riskrcon, ANOMALI, Recorded Future, SHIELD AI, sparkognition, CYBERPOST, SAKINTECH, Trend Micro, VECTRA, SECURE, GLENN LABS, exbeam, AURA, SANSI, SentinelOne, FORTIMO, BLUEZENEXAGON, XENOVIS, eSolutions, APTHEORIE, CyberAngel, PRIVATAN, netskope, modified

The diagram illustrates the relationship between various industries and the 'APPLICATIONS - INDUSTRY' sector. The industries are arranged in a grid around the central sector, with lines connecting them to the center.

ADVERTISING	EDUCATION	REAL ESTATE	APPLICATIONS - INDUSTRY	COMMERCE	FINANCE - LENDING	INSURANCE
AppNexus MediaMath Criteo IAS Interactive Oracle Ad Manager Bluebeam Albert GumGum Appier The Trade Desk Tapad	Lililishuo 猿辅导 Knewton Geclara 百度AI Korbit	Redfin VTS Opender Orchard Reonomy Skyline SpaceMaker Geophy	<b>GOVT &amp; INTELLIGENCE</b> Palantir OpenGov Dataminr Mark43 Anduril FiscalNote Quid PRIMER	<b>COMMERCE</b> FAIR Stitch Fix HowGood Standard  <b>FINANCE - INVESTING</b> Synapse AYASDI Kenshco Addepar Numerai	Affirm Moneta Tala Zest Upgrade Lendix Bank Gula Upstart 100Credit Reform AI	Root Metromile Amenade Equidware Shift Technology Aflac Cape Capital Evolution O Zelus Zesty AI

The figure displays a grid of startup logos, categorized into six sectors. Each sector has a header and a collection of logos representing companies in that space.

- HEALTHCARE:** Flatiron, Kyrus, Metabiot, babyjohn, 30Med, PathAI, Tempus, patientslike, AI Cure, LinkDoc, sofia, Olive, magn, KomenHealth, Paige, health, innovaccer, lumina, springhealth, THE HUMANS DIGITAL PROJECT, eLAI, eGenic, ezra, Caption Health, IMAGEN, zenbra.
- LIFE SCIENCES:** Beamte, color, Caribio, Beamte, DNAnexus, verily, BEIGS, Z, Genentech, inato, freemove, Alkermes, Inspire, SOPHAs, QWIKIN, insitro, Exscientia, REGENERON, ConceptAI, healx, Verana Health.
- TRANSPORTATION:** Uber, Tesla, WATCO, ZEV, Cruise, Nuro, Aptiv, CARMA, Aurora, nauto, AMOTIVE, arco, PILOT, NIO, Oxtaris, G7, nesco, next, PERCEPTIVE AUTOMATA, moovit, Maser, GHST.
- AGRICULTURE:** FARMERS, Granular, John Deere, BLUE RIVER, Taranis, prospero, semios, Alisha, BUILT, MESA.
- INDUSTRIAL:** AVEVA, Siemens, Uptake, PTC, KUKUKI, SCORTEK, TAGHYUS, ALISE, duoeto, Lightricks.
- OTHER:** stem, Amper, ByteDance, STREPSACE, VERIDIGIS.

- OPEN SOURCE

**DATA SOURCES & APIs**

- DATA MARKETPLACES & DISCOVERY**
  - aws Data Exchange
  - DAWEX
  - data.world
  - snowflake
  - EXPLORIUM
  - narrative
- FINANCIAL & ECONOMIC DATA**
  - Bloomberg
  - THOMSON REUTERS
  - DOW JONES
  - Quandl
  - S&P CAPITAL IQ
  - CIBRIGHTS
  - PLAID
  - SECOND MILEAGE
  - ENVESTNET YODOLITE
  - STOCKTWEETS
  - StockTwits
  - xignite
  - Thomson
  - earnest
  - predata
  - MX
  - thk
  - SmartNews
- AIR / SPACE / SEA**
  - Orbital insight
  - AIRBOTICS
  - spire
  - Windward
  - MarineTraffic
  - ExoLabs
  - Synspective
- PEOPLE / ENTITIES**
  - zoominfo
  - axciom
  - experian
  - EPSILON
  - InsideView
  - Brandwatch
  - Quantcast
  - BASIS
  - SAFEGRAPH
  - Demyst
  - melissa
- LOCATION INTELLIGENCE**
  - FOURSQUARE
  - mapbox
  - sense360
  - pluribus
  - PlaceIQ
  - esri
  - Carto
  - Radar
  - Mapillary
  - cuebiq
  - OpenStreetMap
- OTHER**
  - DATA.GOV
  - IMAGENET
  - LABONE
  - StarKey
  - CRUX

**DATA RESOURCES**

- DATA SERVICES**
  - QUANTUMBLACK
  - Booz | Allen | Hamilton
  - Electrifi
  - fractal
  - EXL
  - DataKind
  - innopLexus
- INCUBATORS & SCHOOLS**
  - FLORALBENT
  - GENERAL ASSEMBLY
  - DataCamp
  - DataLife
  - yalvanze
  - metis
  - INSIGHT
  - The Data Incubator
- RESEARCH**
  - OpenAI
  - facebook research
  - MIRI
  - VECTOR INSTITUTE
  - AI2



## - APPLICATIONS – ENTERPRISE

**APPLICATIONS – ENTERPRISE**

SALES	MARKETING - B2B	MARKETING - B2C	CUSTOMER EXPERIENCE / SERVICE	HUMAN CAPITAL	
salesforce chorus aviso cluwrk fuse	app annie lattice sense tubular NGAGIO	tealium simon sendgrid mparticle bluecore persado	actioniq segment attentive active campaign contxt zeta amperity bloomreach invoca braze	qualtrics surveymonkey usertesting clarabridge medallia zendesk kustomer freshdesk gainsight pendo heap amplitude watson assistant teagogo dialogflow liveperson intercom drift asap adala afniti motomi cafédesk aircall eureka talia frame.ai	hirevue symmetry360 textio gigamonks mya allyo wade & wendt credentia

**LEGAL**

- RAVLEF
- BIDISCO
- QSeal
- Conseal
- kira
- RSS

**REGTECH & COMPLIANCE**

- Digit
- text IQ
- TESSIAN
- SmartGate
- OneTrust
- ETHYCA
- SECURITY.ai

**FINANCE**

- Anaplan
- SA HANA
- TRADESHIFT
- SCALEFACTOR
- workday
- WORKSAP
- ALYPI
- Electric
- colonis

**AUTOMATION & RPA**

- UiPath
- Blue Prism
- workato
- Catalyze
- PRIMAVERA
- Bluewin
- INTEGRADGE
- Homevivo

**SECURITY**

- TANIMU
- Redberry Spark
- feedr
- DATAVISOR
- StackPath
- shift science
- riskrecon
- ANOMALI
- Recorded Future
- SHIELD AI
- springognition
- cyberx
- SANITACE
- Securix
- NETECTRA
- SECURE
- GUARANTEA
- exabeam
- SIEM
- Securix
- SentinelOne
- PRIVATIX
- BLUENIXAGON
- ARXIS
- guardian
- Avastore
- Cyberangel
- netstaps
- ruddid

**PARTNERSHIPS**

- INFOSUM
- GRANDSTREAM
- DATA FUSION

[illegible][illegible]

APS — Al / MA

The diagram illustrates a comprehensive ecosystem of data sources, organized into several key categories:

- Data Marketplaces & Discovery:** Includes platforms like AWS Data Exchange, Snowflake Data Exchange, DAWEX, Explorium, Data.world, and Narrative.
- Financial & Economic Data:** Features Bloomberg, Thomson Reuters, Dow Jones, Quandl, S&P Capital IQ, CB Insights, PLAID, Second Step, Envestnet Yodanis, Stockpile, Xignite, The Motley Fool, Predata, MX, tinkr, and SmartNews.
- Air / Space / Sea:** Lists Orbital Insight, AeroBotics, Spire, Windward, MarineTraffic, and ExoLeads.
- People / Entities:** Includes ZoomInfo, Acxiom, Experian, Epsilon, InsideView, Brandwatch, Quantcast, Basis, SafeGraph, Demyst, and Melissa.
- Location Intelligence:** Comprises Foursquare, Mapbox, Sense360, Phylway, PlaceIQ, Esri, Radar, Mapillary, Cuebiq, and OpenStreetMap.
- Other:** Mentions Data.gov, Imagenet, and ApolloScale.
- Data Resources:** Features QuantumBlack, Booz | Allen | Hamilton, Kaggle, ElectrifAI, Fractalx, ExL, DataKind, Innoplexus, and others.
- Incubators & Schools:** Includes General Assembly, DataCamp, DataCity, Galvanize, Metis, The Data Incubator, and others.
- Research:** Lists OpenAI, Facebook Research, MIR, Vector Institute, and others.



## – APPLICATIONS – ENTERPRISE

**APPLICATIONS - ENTERPRISE**

SALES	MARKETING - B2B	MARKETING - B2C	CUSTOMER EXPERIENCE / SERVICE	HUMAN CAPITAL	
salesbox chorus aviso peoplea cloutier tact.ai fuse/machines	app Annie sense tubular ENGAGIO	tealium simon SendGrid perceptics Bluecore [ PERSADO ]	ACTIONIQ Segment Active Campaign CONTENT SQUARE ZETA bloomreach INVOCAL inze	qualtrics SurveyMonkey UserTesting MEDALLIA zendesk Gainsight pendo HEAP Amplitude Dialogflow ASAPP ada Eureka CLARABRIDGE Kustomer freshdesk Watson Assistant INTERCOM Drift Cahdesk aircall frame ai	HireVue textio mya Allyo Wade&Wendy cere

The diagram illustrates 100+ startups categorized into 10 sectors. The sectors and their associated startups are:

- LEGAL**: RAVEL, DISICO, kira, QSeal, CoConstructive, R.ROSS
- REGTECH & COMPLIANCE**: Digital text IQ, TESSIAN Smarter Knowledge, QoneTrust, ETHNYA, SECURITY.ai
- FINANCE**: Anaplan, ZUORO, SOWSAHANA, TRADESHIFF, SCALE FACTOR, botkeeper, pilot, DIGITS, appzenn
- AUTOMATION & RPA**: UiPath, bluewin, VIDADO, idenfylation, workato, Catalytic, KEYVIN, EVER READY, ALKYPI, FORTIMO, ZINGTANGAO, Memento, CELONIS
- SECURITY**: TANJUNG, battery spark, Zscaler, StackPath, thumby, FIRETR, feedzai, DATAVISIO, pingree, sift science, riskrcon, ANOMALI, Recorded Future, SHIELD AI, sparkognition, CYBERST, SAKINTEK, Trend Micro, VECTRA, SECURE, GLENN LABS, exabeam, APTA, PROSECURE, SentinelOne, FORTIFY, BLUEGRASSION, XENOVIS, eSondran, APTHEORIE, CyberAngel, PRIVATAN, netskope, redified

The diagram illustrates the relationship between various industries and the 'APPLICATIONS - INDUSTRY' sector. The industries are arranged in a grid around the central sector, with lines connecting them to the center.

ADVERTISING	EDUCATION	REAL ESTATE	APPLICATIONS - INDUSTRY	COMMERCE	FINANCE - LENDING	INSURANCE
AppNexus MediaMath Criteo IAS Interactive Oracle Ad Manager Bluebeam Albert GumGum Appier The Trade Desk Tapad	Lililishuo 猿辅导 Knewton Geclara 启源 AI Korbit	Redfin VTS Opender Orchard Reonomy Skyline SpaceMaker Geophy	Palantir OpenGov Dataminr Mark43 Anduril FiscalNote Quid Primer	Fairst Stitch Fix HowGood Standard Finance - Investing Ayasdi Kenshc Addepar Numerai	Affirm Monedo Tala Zest Upgrade Easibanc Gula Upstart 100Credit Reform Ai	Root Metromile Amendea Eudware Shift Technology Ape Cape Evolution Zelus Zesty Ai

The figure displays a grid of startup logos, categorized into six sectors. Each sector has a header and a collection of logos representing companies in that space.

- HEALTHCARE:** Flatiron, Kyrus, Metabiot, babyjohn, 30Med, PathAI, Tempus, patientslike, AI Cure, LinkDoc, sofia, Olive, magn, Khorosha, Paige, health, innovaccer, lumina, springhealth, THE HUMANS DIGITAL PROJECT, eLAI, eGenic, ezra, Caption Health, IMAGEN, zenbra.
- LIFE SCIENCES:** bioRxiv, color, Carotest, Biomercantia, DNAnexus, verily, 10X, Z, Genentech, inato, freemove, Alameda, 10xgenomics, SODiAC, OXENIN, insitro, Exscientia, REGENERON, ConceptAI, healx, Verana Health.
- TRANSPORTATION:** Uber, Tesla, WATCO, ZEV, Cruise, Nuro, Aptiv, CARMA, Aurora, nauto, AMOVITE, arco, PILOT, NIO, Oxtaris, G7, nesco, nvidia, moovit, Maser, GHST, PERCEPTIVE AUTOMATA, semios.
- AGRICULTURE:** FARMERS, Granular, John Deere, BLUE RIVER, TARANIS, prospero, semios.
- INDUSTRIAL:** AVEVA, Siemens, Uptake, Uptime, KONEC, SCORTEC, TAGHYUS, ALISE, BUILT, MESA.
- OTHER:** stem, Amper, ByteDance, STATSPACE, VERIDIGIS, duoeto, Lightbridge.

## - DATA SOURCES & APIs

**DATA SOURCES & APIs**

- DATA MARKETPLACES & DISCOVERY**
  - aws Data Exchange
  - snowflake Data Exchange
  - DAWEX
  - EXPLORUM
  - data.world
  - narrative
- FINANCIAL & ECONOMIC DATA**
  - Bloomberg
  - THOMSON REUTERS
  - DOW JONES
  - Quandl
  - RSP CAPITAL IQ
  - CB Insights
  - PLAID
  - SECOND MILEAGE
  - INVESTMENT YODLEE
  - STOCKTWEETS
  - xignite
  - Twitter
  - earnest
  - predata
  - MX
  - thnk
  - SmartNews
- AIR / SPACE / SEA**
  - Orbital Insight
  - AIRBOTICS
  - spire
  - Windward
  - MarineTraffic
  - EXOLEBS
  - Synapsive
- PEOPLE / ENTITIES**
  - zoominfo
  - axiom
  - experian
  - EPILION
  - InsideView
  - Brandwatch
  - Quantcast
  - BASIS
  - SAFEGRAPH
  - Demyst
  - melissa
- LOCATION INTELLIGENCE**
  - FOURSQUARE
  - mapbox
  - sense360
  - plinytowns
  - PlaceIQ
  - esri
  - Radars
  - Mapillary
  - cuebiq
  - OpenStreetMap
- OTHER**
  - DATA.GOV
  - IMAGENET
  - LABELED
  - MARKETING INTELLIGENCE
  - APOLLOSCALE
  - CRUX
- DATA SERVICES**
  - QUANTUMBLACK
  - Booz | Allen | Hamilton
  - kaggle
  - ElectrifAI
  - fractale
  - EXL
  - DataKind
  - innoprox
- INCUBATORS & SCHOOLS**
  - FLORALBENT
  - GENERAL ASSEMBLY
  - DataCamp
  - DataElio
  - yalvanize
  - WETIS
  - INSIGHT
  - The Data Incubator
- RESEARCH**
  - OpenAI
  - facebook research
  - MIRI
  - VECTOR INSTITUTE
  - AIZ

# What is it?

- Interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL
- Athena is serverless: no infrastructure to manage
- Pay only for the queries that you run
- Benefit: the process is easy for anyone with SQL skills to quickly analyze large-scale datasets.



**Amazon Athena**

# Steps needed



## Select a data set

Identify where your data is located in S3. Athena allows you to query data in CSV, TSV, JSON, Parquet, and ORC formats.



## Create a table

Use the Create Table Wizard or write your own DDL (Data Definition Language) statements using Hive.



## Query data

Run queries on your data. Amazon Athena supports ANSI SQL queries.

# Benefits

- Amazon Athena uses ANSI SQL
- Works with a variety of standard data formats, including CSV, JSON, ORC, Avro, and Parquet
- Athena is ideal for quick, ad-hoc querying
- It can also handle complex analysis, including large joins, and functions
- Amazon Athena automatically executes queries in parallel, so most results come back within seconds.

# Demo AWS Athena



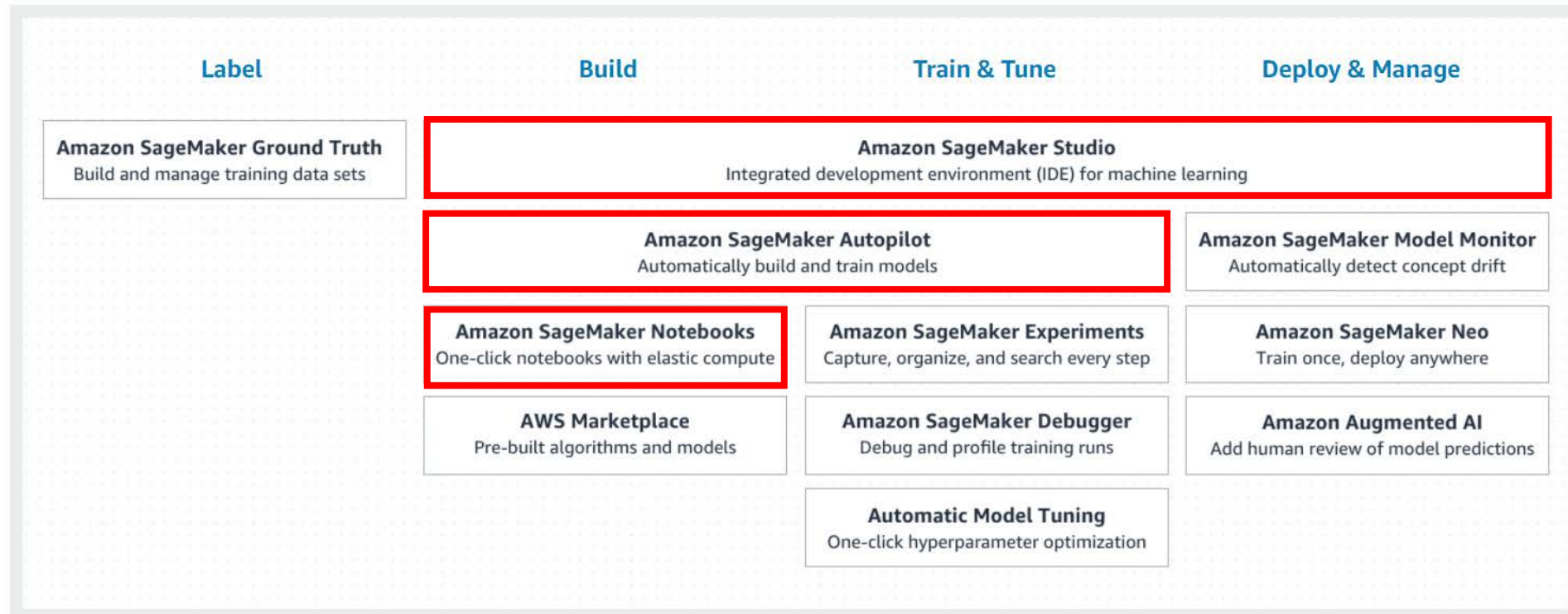
# AWS SageMaker



Amazon SageMaker

# What is it?

- Fully managed service that provides ability to build, train, and deploy machine learning models quickly.



# How to load data from S3?

- conda install -c conda-forge s3fs==0.3.0

```
conda install -c conda-forge s3fs==0.3.0
```

- import pandas as pd

```
train_set = pd.read_csv('s3://sagemaker1109/MLProject21_train.csv')
```

```
import pandas as pd  
train_set = pd.read_csv('s3://sagemaker1109/MLProject21_train.csv')
```

Amazon SageMaker Studio

FileEditViewRunKernelGitTabsSettingsHelp

Feedback

half a minute ago

/ Testing-aws-auto-ml-job /

TRIALS

0 rows selected 0/20 filters

Search column name to start

Clear all

Name	Last modified
tuning-job-1-3bb...	9 seconds ago
tuning-job-1-3bb...	19 seconds ago
tuning-job-1-3bb...	1 minute ago
tuning-job-1-3bb...	1 minute ago
tuning-job-1-3bb...	3 minutes ago
tuning-job-1-3bb...	6 minutes ago
tuning-job-1-3bb...	6 minutes ago
tuning-job-1-3bb...	7 minutes ago
tuning-job-1-3bb...	7 minutes ago
tuning-job-1-3bb...	8 minutes ago
tuning-job-1-3bb...	8 minutes ago
tuning-job-1-3bb...	9 minutes ago
tuning-job-1-3bb...	9 minutes ago
tuning-job-1-3bb...	10 minutes ago
tuning-job-1-3bb...	10 minutes ago
tuning-job-1-3bb...	10 minutes ago
tuning-job-1-3bb...	11 minutes ago
tuning-job-1-3bb...	12 minutes ago

Amazon SageMaker Studio

testing

SageMakerAutopilotCandidat

Untitled.ipynb

Kernel: CPU: 99.80% MEM: 15.69% 4 vCPU + 16 GiB + 1 GPU Python 3 (Data Science) Share

Select Instance

Running notebook

Untitled.ipynb

Current instance type

4 vCPU + 16 GiB + 1 GPU **mLg4dn.xlarge** Fast Launch

If you change your instance, existing settings for this notebook will be lost, and installed packages will not be carried over.

Instances 4 of 29

Fast launch only

Instance Type	Instance Category	vCPU	GPU	Memory	Fast Launch
<input type="radio"/> mlt3.medium	General purpose	2	0	4 GiB	✓
<input checked="" type="radio"/> mLg4dn.xlarge	Accelerated computing	4	1	16 GiB	✓
<input type="radio"/> mLm5.large	General purpose	2	0	8 GiB	✓
<input type="radio"/> mLc5.large	Compute optimized	2	0	4 GiB	✓

Cancel

Save and continue

```
train
[6]: test
[7]: test
[8]: # Lit
test
# New
# use
test
# Sav
featu
# Cov
test
[*]: # Imp
from
# Inp
rf
# Trn
rf.fi
[*]: # Use
predi
# Coe
error
# Pre
print
[*]: conda install -c conda-forge s3fs==0.3.0
```



Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Feedback

[\*]: # Import the model we are using

Tags in Active Cell

Add Tag

Other Tags in Notebook

Select All Cells with this Tag

Rename Tag for All Cells

Delete Tag from All Cells

Slide Type

Raw NBConvert Format

Advanced Tools

Amazon SageMaker Studio

testing

SageMakerAutopilotCandidat

Untitled.ipynb

less than 20 seconds ago

EXPERIMENT: TESTING

Problem typeBinaryClassification

Open candidate generation notebook

Open data exploration notebook

Pre-processing

Candidate Definitions Generated

Feature Engineering

Model Tuning

Amazon SageMaker Autopilot is tuning the model.

At the default setting of creating 250 trials, this process can take hours to complete.

You can always return to this page later by choosing this experiment on the Experiments tab in the navigation panel.

If experiment is taking too long to run, you can stop the experiment

Trials

Job profile

Your trials will display here as they are generated.