VIT®

UNIVERSITY

(Estd. u/s 3 of UGC Act 1956)

ITE3007- Cloud Computing and Virtualization

FACULTY:PRIYA.V

PROJECT TITLE: Using cloud-computing technology to apply machine-learning algorithm to the given data set and find the prediction of target variable based on trained model

Name: Garlapati saiteja
Reg.no: 17BIT0217

Name: Sri Harshitha Palla
Reg.no:17BIT0330

Name:Arun Nuli
Reg.no:17BIT0210

ABSTRACT :
Cloud provides many range of resources to the users over the internet , so here we are going to use this resources to run the machine learning algorithms and to create a model based on particular pattern by the help of given dataset, which will be converted to the data source for our algorithm.

EXPECTED RESULT :
We will be taking a dataset to predict that whether customer is going to buy a particular product or not by training the data model. We will be using Binary Classification to get the decision of customer. Cloud will be taking care of all complexity by running all process over the provided resources instead on the user's system.

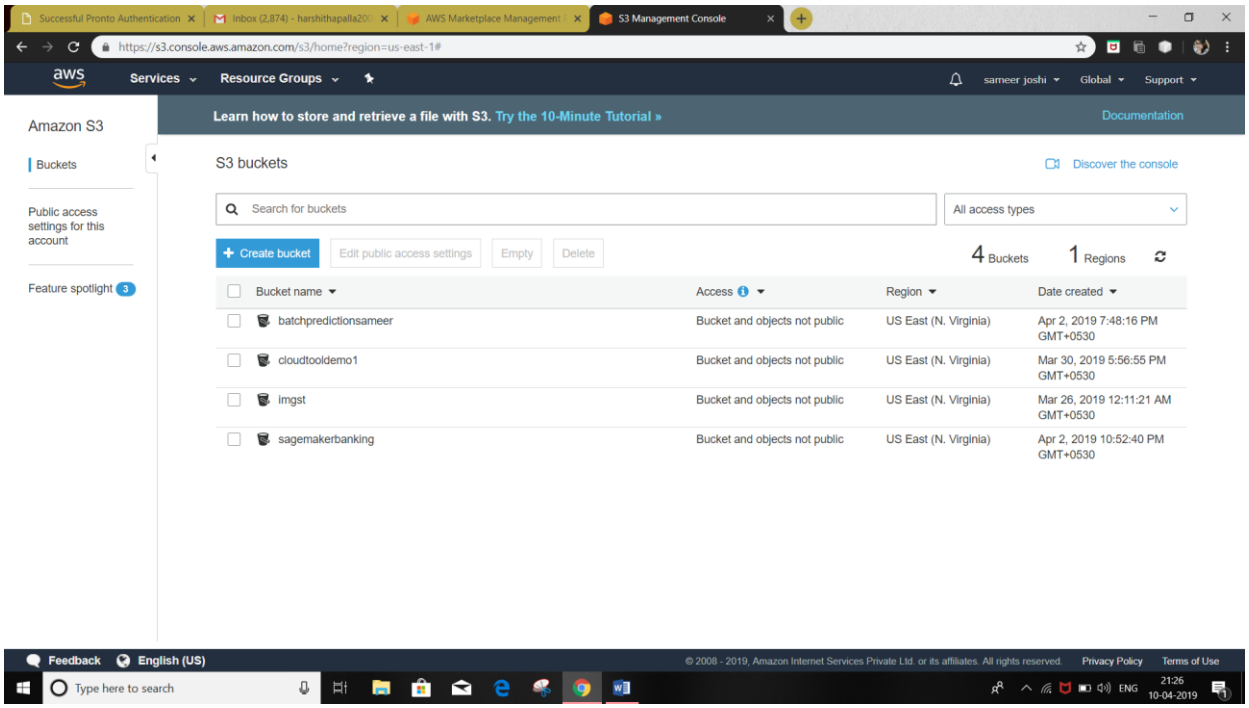
Objective of the project :
Find predictions with the machine-learning algorithm over the cloud.

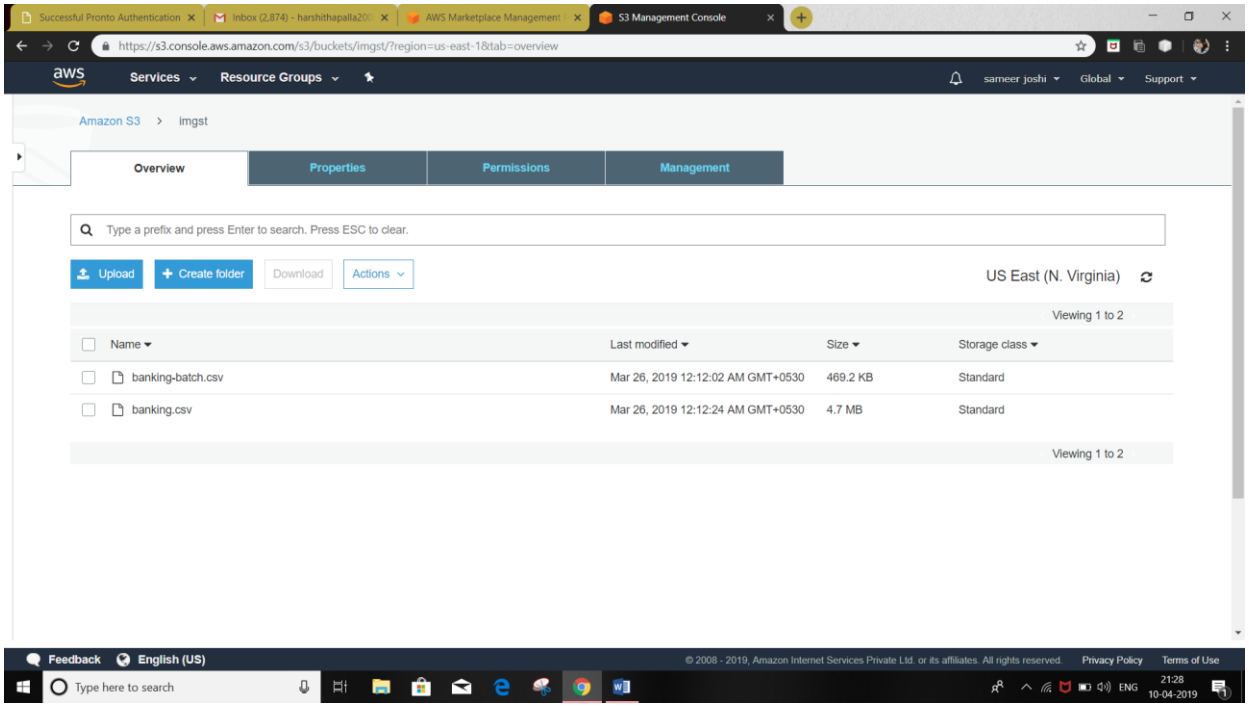
Innovation component in the project :
Using cloud computing over the programming languages to create ML model for a dataset.

Technologies : AWS

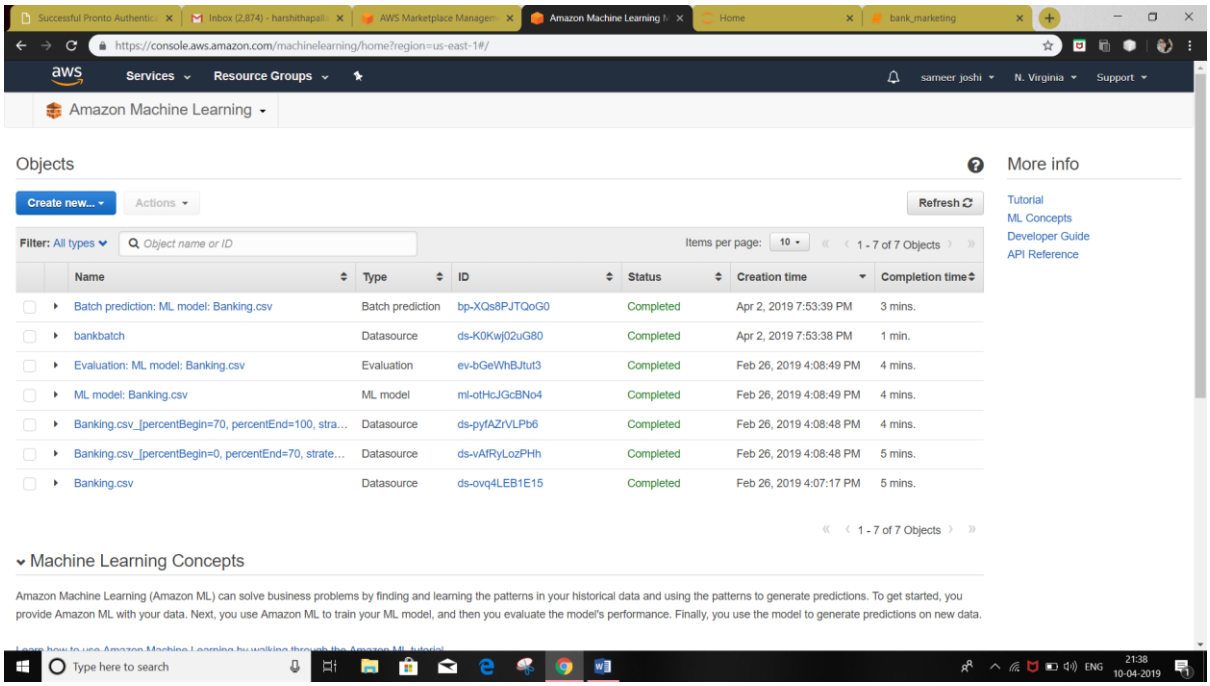
Features from amazon to be used in the project :
S3 management,Amazon Machine learning,Amazon Sagemaker

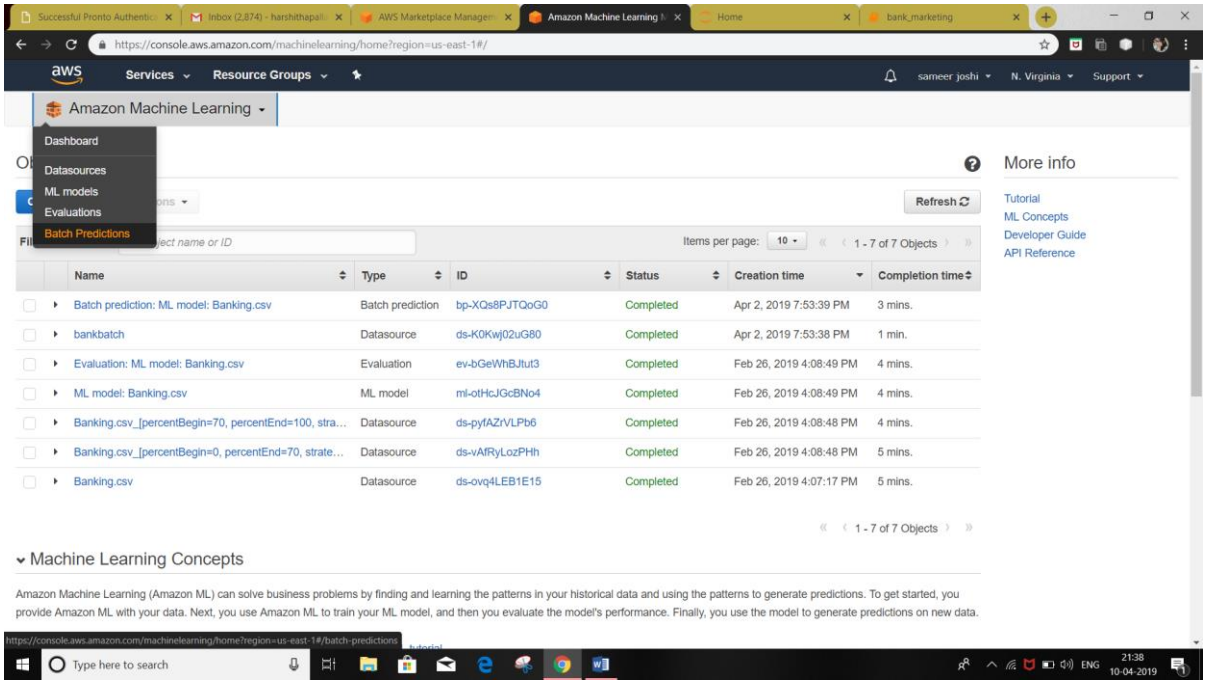
S3:

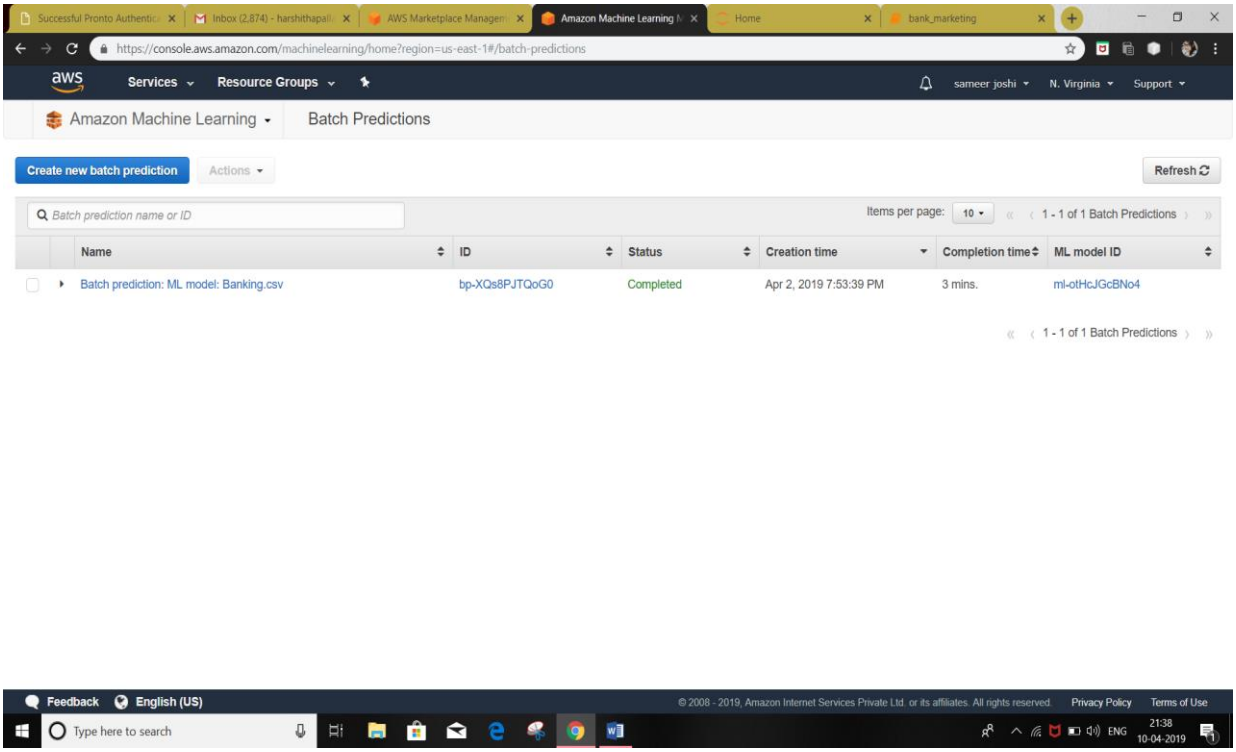

S3 Management Console
https://s3.console.aws.amazon.com/s3/home?region=us-east-1#
Services
Resource Groups
sameer joshi
Global
Support
Amazon S3
Learn how to store and retrieve a file with S3. Try the 10-Minute Tutorial »
Documentation
Buckets
Public access settings for this account
Feature spotlight
S3 buckets
Discover the console
Search for buckets
All access types
Create bucket
Edit public access settings
Empty
Delete
4 Buckets
1 Regions
Bucket name
Access
Region
Date created
batchpredictionsameer
Bucket and objects not public
US East (N. Virginia)
Apr 2, 2019 7:48:16 PM GMT+0530
cloudtooldemo1
Bucket and objects not public
US East (N. Virginia)
Mar 30, 2019 5:56:55 PM GMT+0530
imgst
Bucket and objects not public
US East (N. Virginia)
Mar 26, 2019 12:11:21 AM GMT+0530
sagemakerbanking
Bucket and objects not public
US East (N. Virginia)
Apr 2, 2019 10:52:40 PM GMT+0530
Feedback
English (US)
© 2008 - 2019, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Type here to search
21:26
10-04-2019


S3 Management Console
https://s3.console.aws.amazon.com/s3/buckets/imgst/?region=us-east-1&tab=overview
Services
Resource Groups
sameer joshi
Global
Support
Amazon S3
imgst
Overview
Properties
Permissions
Management
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
Download
Actions
US East (N. Virginia)
Viewing 1 to 2
Name
Last modified
Size
Storage class
banking-batch.csv
Mar 26, 2019 12:12:02 AM GMT+0530
469.2 KB
Standard
banking.csv
Mar 26, 2019 12:12:24 AM GMT+0530
4.7 MB
Standard
Viewing 1 to 2
Feedback
English (US)
© 2008 - 2019, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Type here to search
21:28
10-04-2019

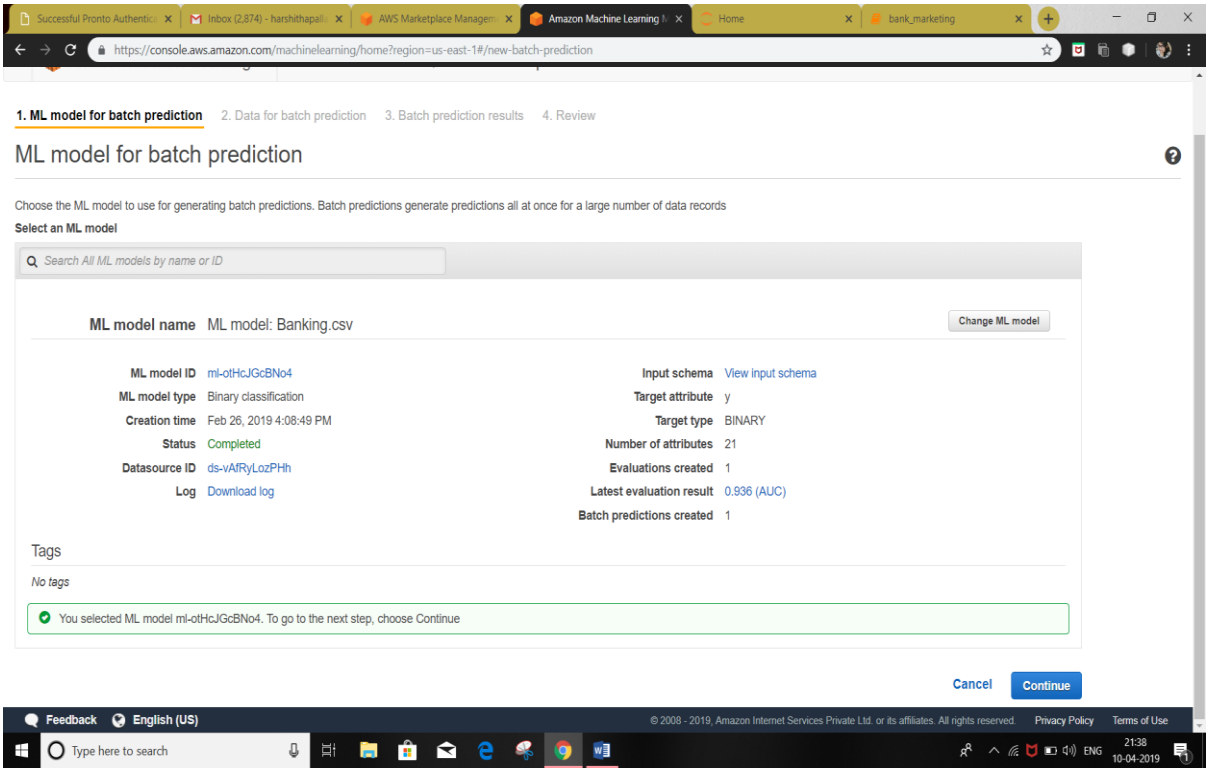
Amazon ML:

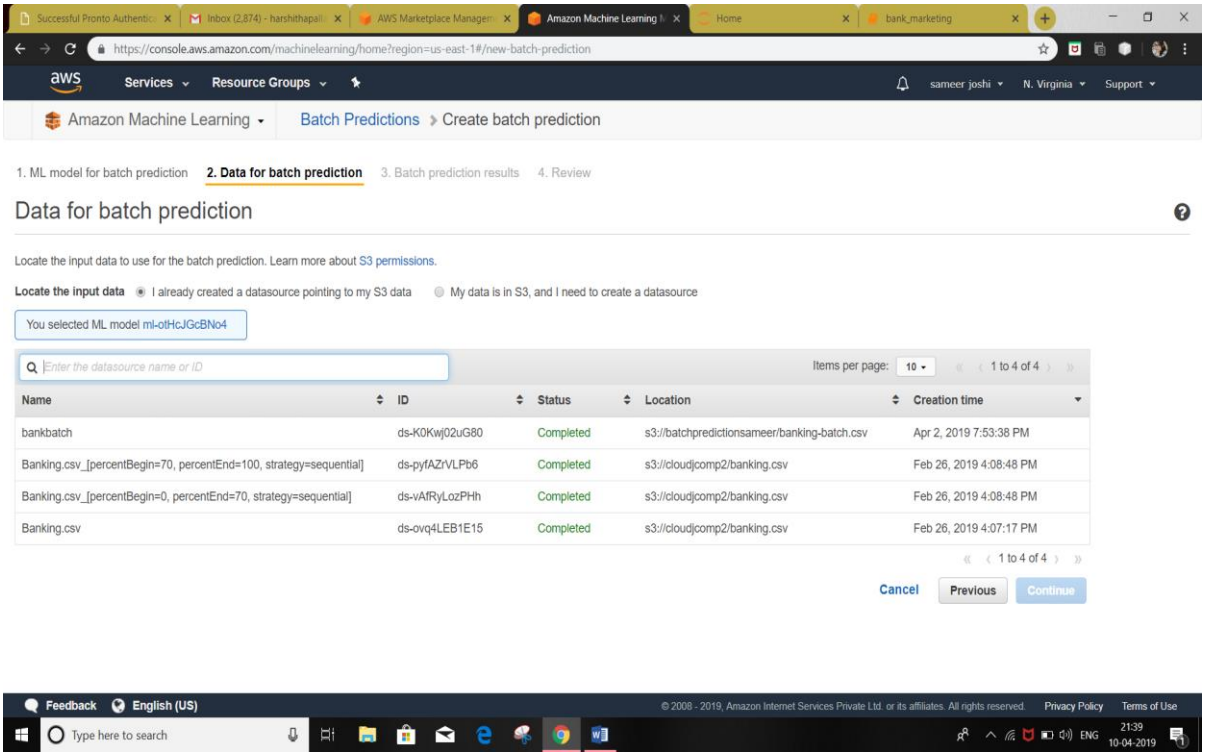

https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/
Services
Resource Groups
sameer joshi
N. Virginia
Support
Amazon Machine Learning
Objects
Create new...
Actions
Refresh
More info
Tutorial
ML Concepts
Developer Guide
API Reference
Filter: All types
Object name or ID
Items per page: 10
1 - 7 of 7 Objects
Name
Type
ID
Status
Creation time
Completion time
Batch prediction: ML model: Banking.csv
Batch prediction
bp-XQs8PJTQoG0
Completed
Apr 2, 2019 7:53:39 PM
3 mins.
bankbatch
Datasource
ds-K0Kwj02uG80
Completed
Apr 2, 2019 7:53:38 PM
1 min.
Evaluation: ML model: Banking.csv
Evaluation
ev-bGeWhBJtut3
Completed
Feb 26, 2019 4:08:49 PM
4 mins.
ML model: Banking.csv
ML model
ml-otHcJGcBNo4
Completed
Feb 26, 2019 4:08:49 PM
4 mins.
Banking.csv_[percentBegin=70, percentEnd=100, stra...
Datasource
ds-pyfAZrVLPb6
Completed
Feb 26, 2019 4:08:48 PM
4 mins.
Banking.csv_[percentBegin=0, percentEnd=70, strate...
Datasource
ds-vAfRyLozPHh
Completed
Feb 26, 2019 4:08:48 PM
5 mins.
Banking.csv
Datasource
ds-ovq4LEB1E15
Completed
Feb 26, 2019 4:07:17 PM
5 mins.
Machine Learning Concepts
Amazon Machine Learning (Amazon ML) can solve business problems by finding and learning the patterns in your historical data and using the patterns to generate predictions. To get started, you provide Amazon ML with your data. Next, you use Amazon ML to train your ML model, and then you evaluate the model's performance. Finally, you use the model to generate predictions on new data.
Type here to search
ENG
21:38
10-04-2019


Amazon Machine Learning
Dashboard
Datasources
ML models
Evaluations
Batch Predictions
Refresh
More info
Tutorial
ML Concepts
Developer Guide
API Reference
Items per page: 10
1 - 7 of 7 Objects
Name
Type
ID
Status
Creation time
Completion time
Batch prediction: ML model: Banking.csv
Batch prediction
bp-XQs8PJTQoG0
Completed
Apr 2, 2019 7:53:39 PM
3 mins.
bankbatch
Datasource
ds-K0Kwj02uG80
Completed
Apr 2, 2019 7:53:38 PM
1 min.
Evaluation: ML model: Banking.csv
Evaluation
ev-bGeWhBJtut3
Completed
Feb 26, 2019 4:08:49 PM
4 mins.
ML model: Banking.csv
ML model
ml-otHcJGcBNo4
Completed
Feb 26, 2019 4:08:49 PM
4 mins.
Banking.csv_[percentBegin=70, percentEnd=100, stra...
Datasource
ds-pyfAZrVLPb6
Completed
Feb 26, 2019 4:08:48 PM
4 mins.
Banking.csv_[percentBegin=0, percentEnd=70, strate...
Datasource
ds-vAfRyLozPHh
Completed
Feb 26, 2019 4:08:48 PM
5 mins.
Banking.csv
Datasource
ds-ovq4LEB1E15
Completed
Feb 26, 2019 4:07:17 PM
5 mins.
Machine Learning Concepts
Amazon Machine Learning (Amazon ML) can solve business problems by finding and learning the patterns in your historical data and using the patterns to generate predictions. To get started, you provide Amazon ML with your data. Next, you use Amazon ML to train your ML model, and then you evaluate the model's performance. Finally, you use the model to generate predictions on new data.
https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/batch-predictions
Type here to search
ENG
21:38
10-04-2019

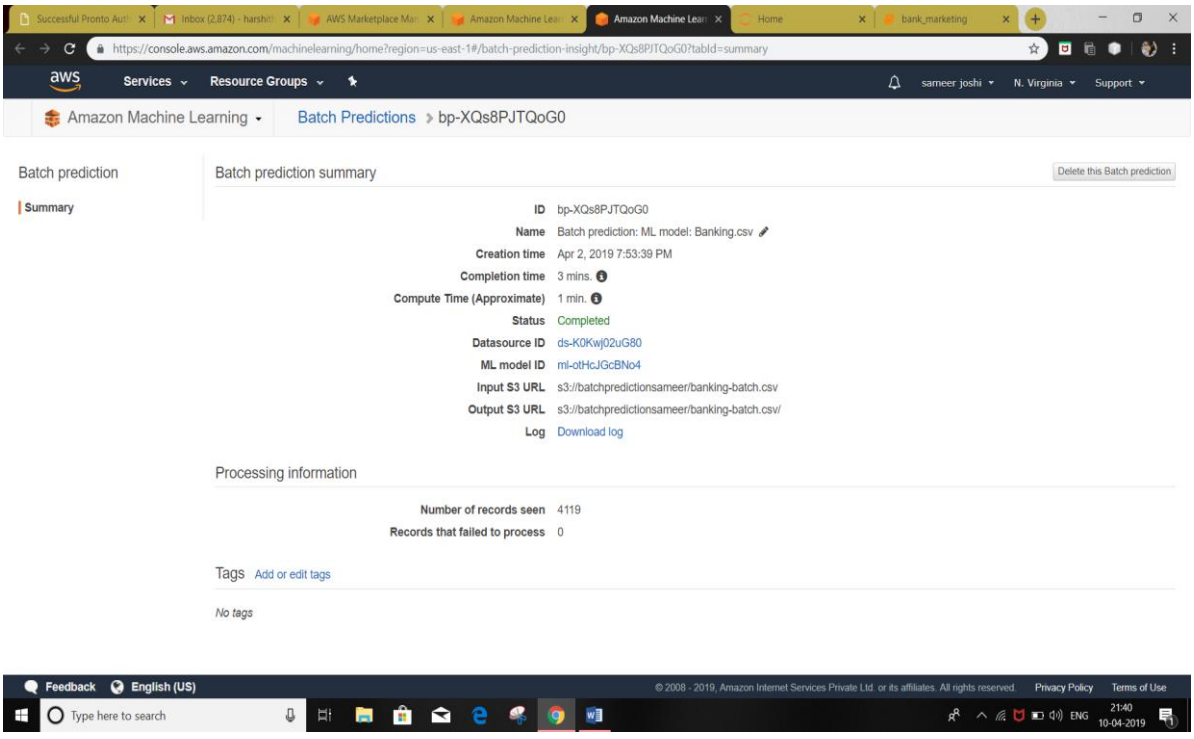

https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/batch-predictions
Services
Resource Groups
sameer joshi
N. Virginia
Support
Amazon Machine Learning
Batch Predictions
Create new batch prediction
Actions
Refresh
Batch prediction name or ID
Items per page: 10
1 - 1 of 1 Batch Predictions
Name
ID
Status
Creation time
Completion time
ML model ID
Batch prediction: ML model: Banking.csv
bp-XQs8PJTQoG0
Completed
Apr 2, 2019 7:53:39 PM
3 mins.
ml-otHcJGcBNo4
1 - 1 of 1 Batch Predictions

Feedback
English (US)
© 2008 - 2019, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Type here to search
ENG
21:38
10-04-2019


https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/new-batch-prediction
1. ML model for batch prediction
2. Data for batch prediction
3. Batch prediction results
4. Review
ML model for batch prediction
Choose the ML model to use for generating batch predictions. Batch predictions generate predictions all at once for a large number of data records
Select an ML model
Search All ML models by name or ID
ML model name ML model: Banking.csv
Change ML model
ML model ID ml-otHcJGcBNo4
ML model type Binary classification
Creation time Feb 26, 2019 4:08:49 PM
Status Completed
Datasource ID ds-vAfRyLozPHh
Log Download log
Input schema View input schema
Target attribute y
Target type BINARY
Number of attributes 21
Evaluations created 1
Latest evaluation result 0.936 (AUC)
Batch predictions created 1
Tags
No tags
You selected ML model ml-otHcJGcBNo4. To go to the next step, choose Continue
Cancel
Continue
Feedback
English (US)
© 2008 - 2019, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Type here to search
ENG
21:38
10-04-2019

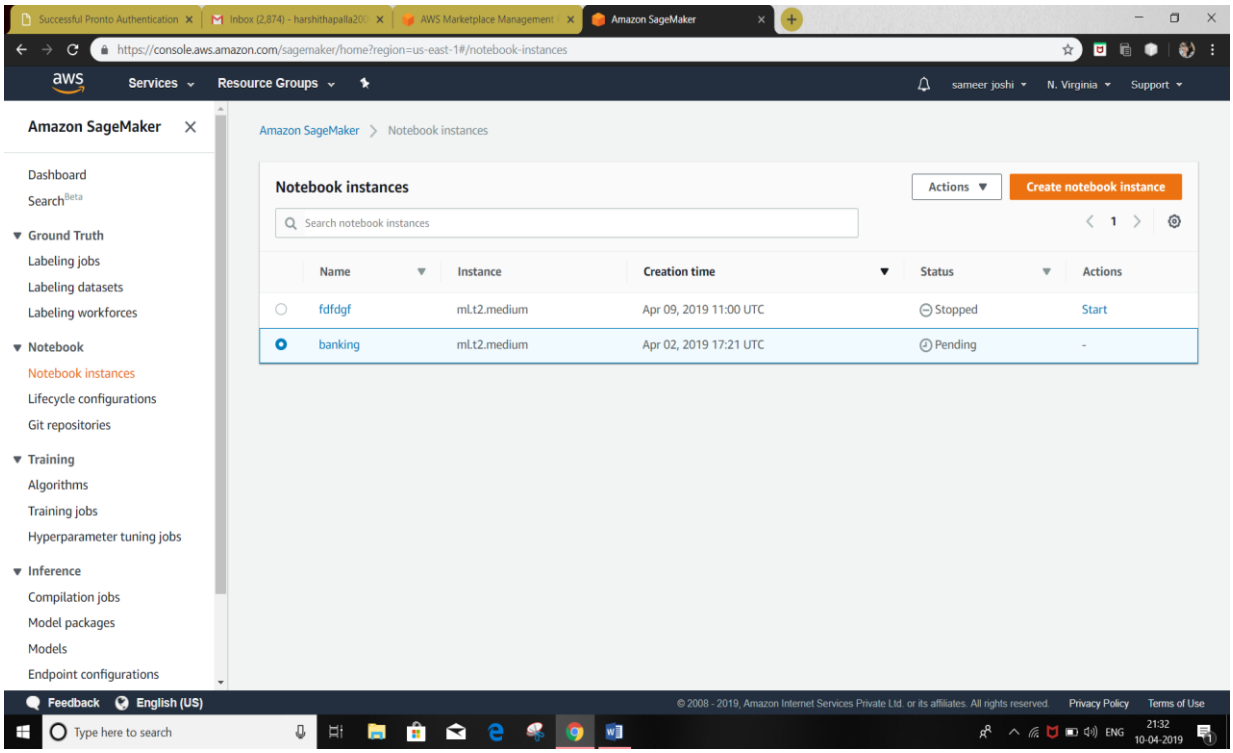

Successful Pronto Authentic
Inbox (2,874) - harshithapall
AWS Marketplace Manager
Amazon Machine Learning
Home
bank_marketing
https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/new-batch-prediction
aws
Services
Resource Groups
sameer joshi
N. Virginia
Support
Amazon Machine Learning
Batch Predictions
Create batch prediction
1. ML model for batch prediction
2. Data for batch prediction
3. Batch prediction results
4. Review
Data for batch prediction
Locate the input data to use for the batch prediction. Learn more about S3 permissions.
Locate the input data
I already created a datasource pointing to my S3 data
My data is in S3, and I need to create a datasource
You selected ML model ml-otHcJGcBNo4
Enter the datasource name or ID
Items per page: 10
1 to 4 of 4
Name
ID
Status
Location
Creation time
bankbatch
ds-K0Kwj02uG80
Completed
s3://batchpredictionsameer/banking-batch.csv
Apr 2, 2019 7:53:38 PM
Banking.csv_[percentBegin=70, percentEnd=100, strategy=sequential]
ds-pyfAZrVLPb6
Completed
s3://cloudjcomp2/banking.csv
Feb 26, 2019 4:08:48 PM
Banking.csv_[percentBegin=0, percentEnd=70, strategy=sequential]
ds-vAfRyLozPHh
Completed
s3://cloudjcomp2/banking.csv
Feb 26, 2019 4:08:48 PM
Banking.csv
ds-ovq4LEB1E15
Completed
s3://cloudjcomp2/banking.csv
Feb 26, 2019 4:07:17 PM
1 to 4 of 4
Cancel
Previous
Continue

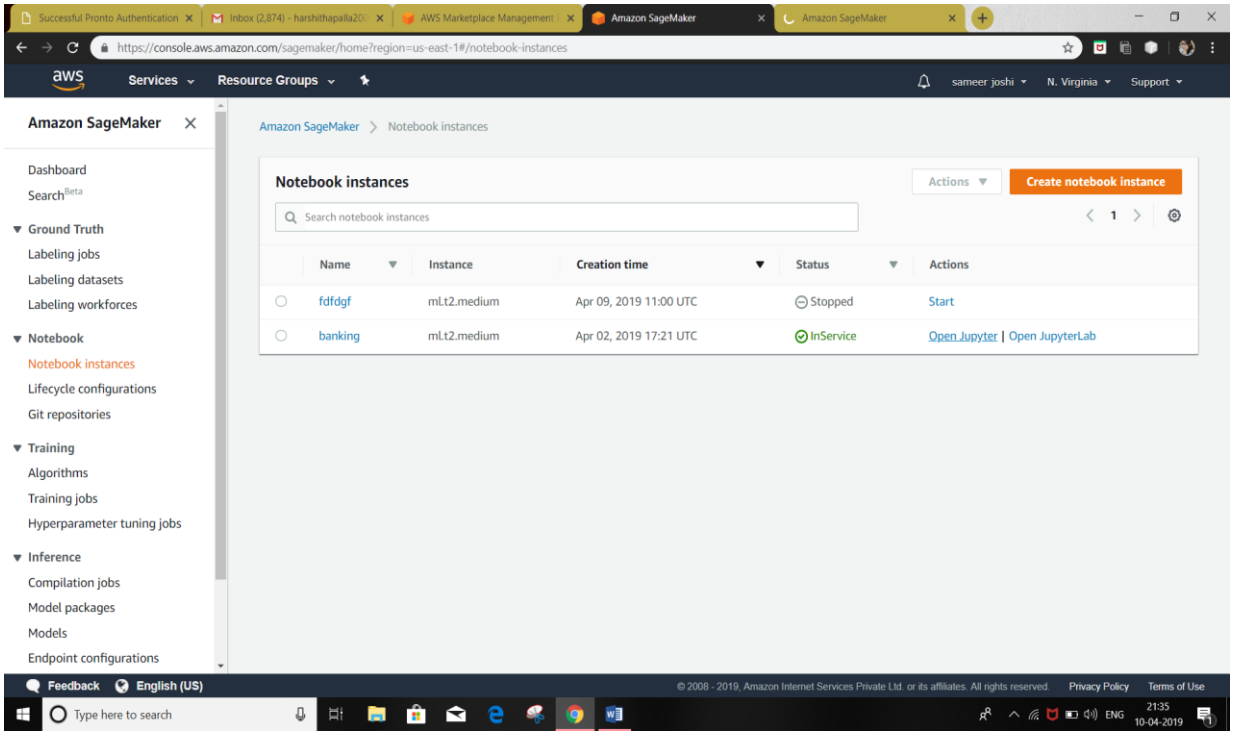
Feedback
English (US)
© 2008 - 2019, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Type here to search
ENG
21:39
10-04-2019

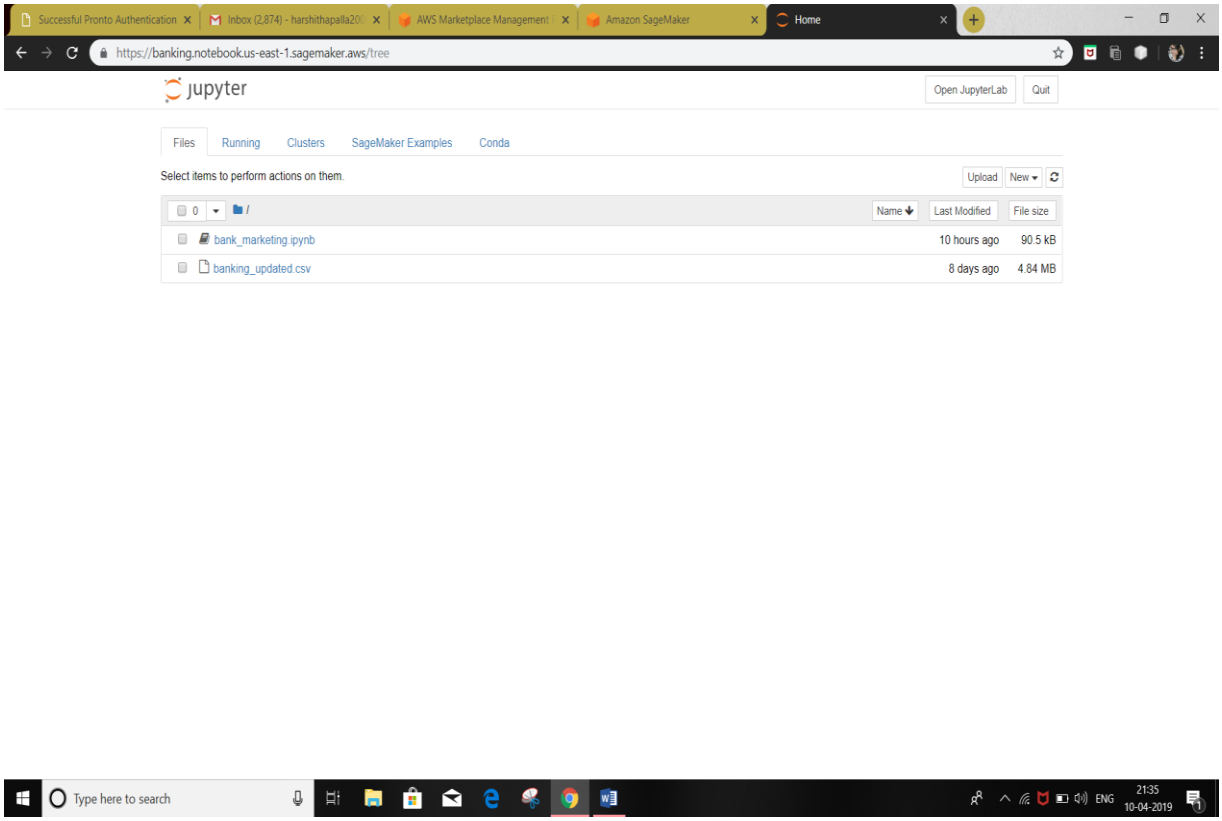

https://console.aws.amazon.com/machinelearning/home?region=us-east-1#/batch-prediction-insight/bp-XQs8PJTQoG0?tabId=summary
aws
Services
Resource Groups
sameer joshi
N. Virginia
Support
Amazon Machine Learning
Batch Predictions
bp-XQs8PJTQoG0
Batch prediction
Summary
Batch prediction summary
Delete this Batch prediction
ID bp-XQs8PJTQoG0
Name Batch prediction: ML model: Banking.csv
Creation time Apr 2, 2019 7:53:39 PM
Completion time 3 mins.
Compute Time (Approximate) 1 min.
Status Completed
Datasource ID ds-K0Kwj02uG80
ML model ID ml-otHcJGcBNo4
Input S3 URL s3://batchpredictionsameer/banking-batch.csv
Output S3 URL s3://batchpredictionsameer/banking-batch.csv/
Log Download log
Processing information
Number of records seen 4119
Records that failed to process 0
Tags Add or edit tags
No tags

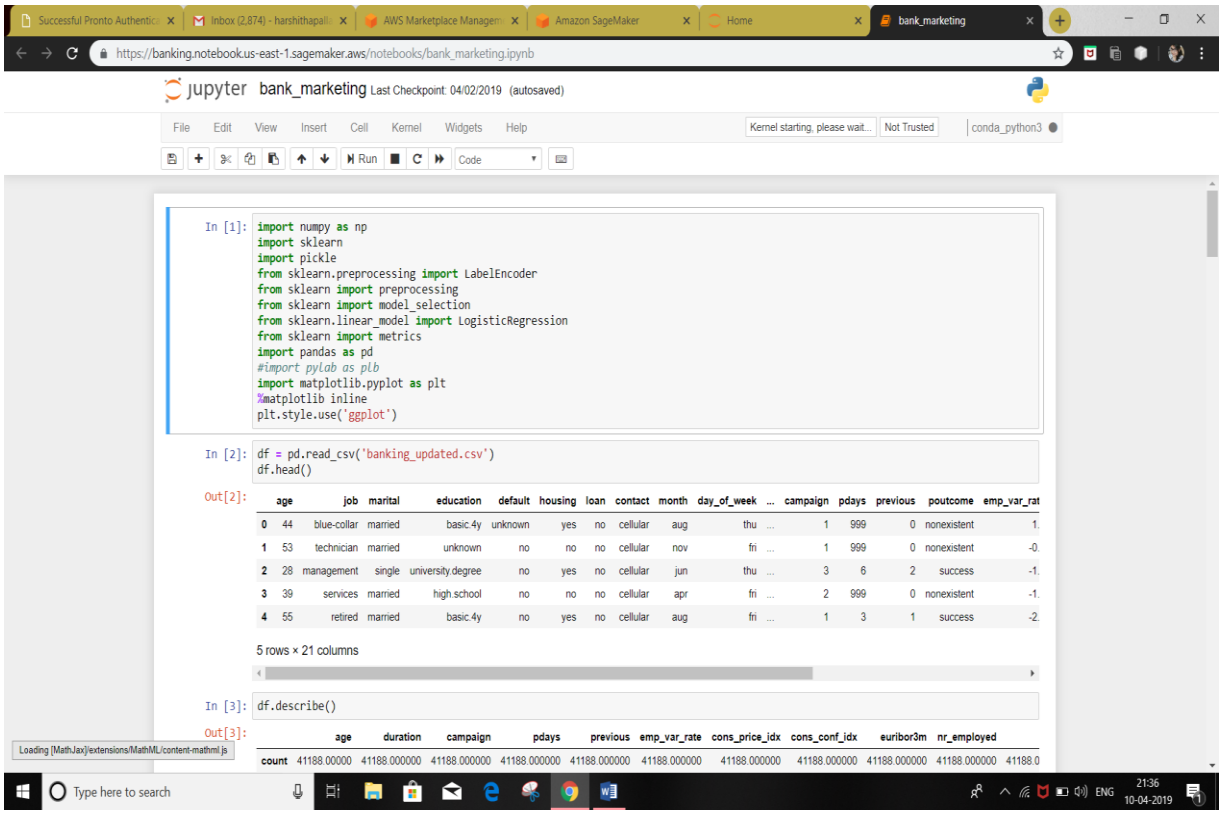
Feedback
English (US)
© 2008 - 2019, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use
Type here to search
ENG
21:40
10-04-2019

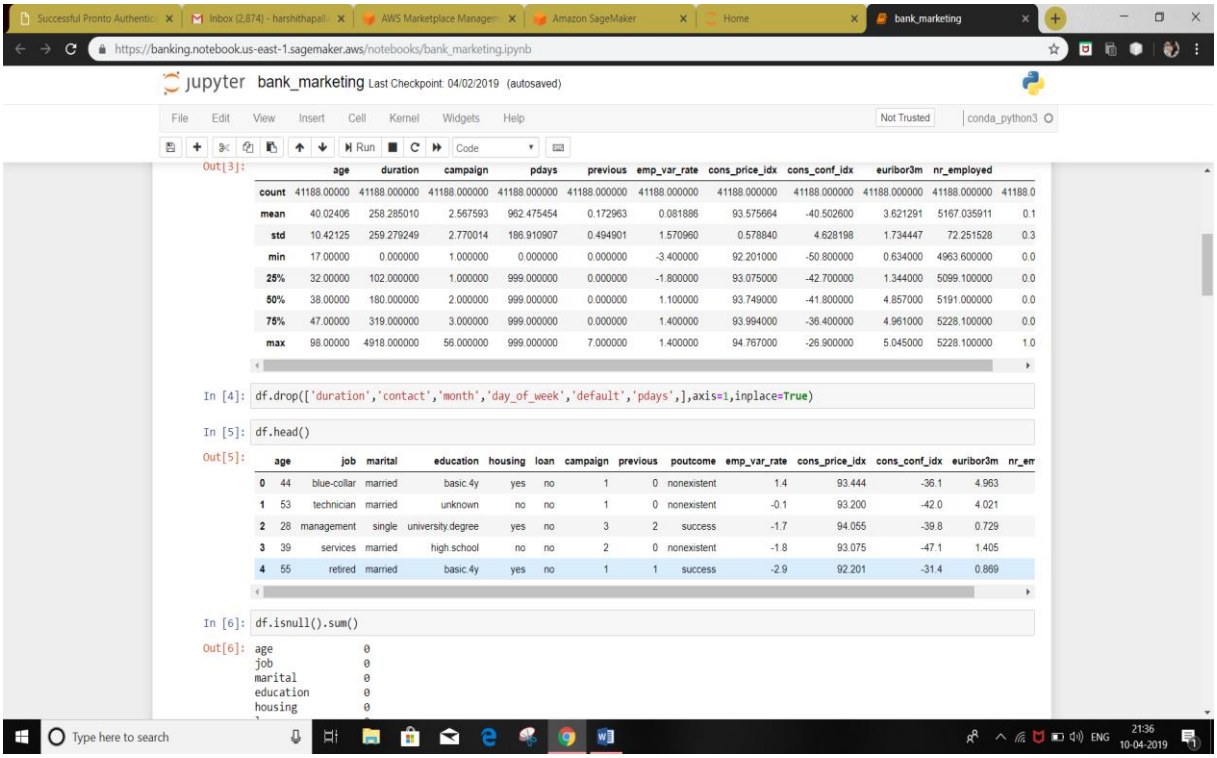
Amazon SageMaker:

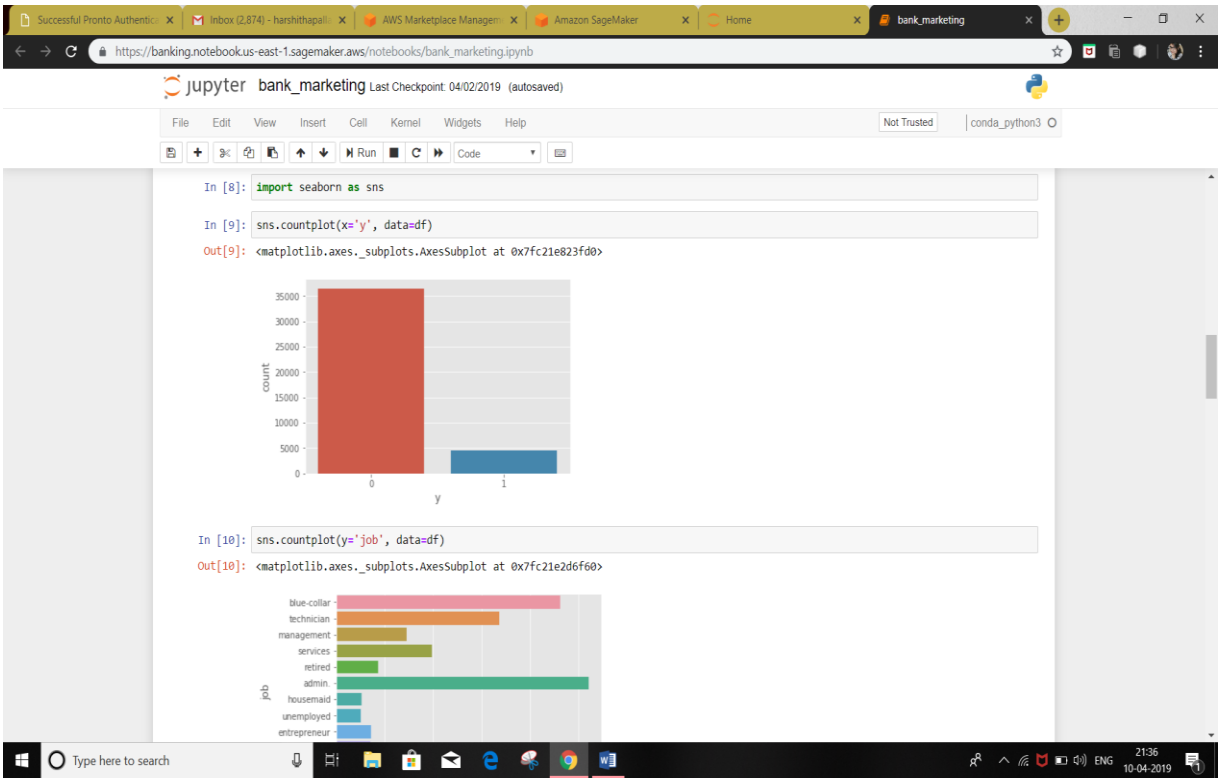


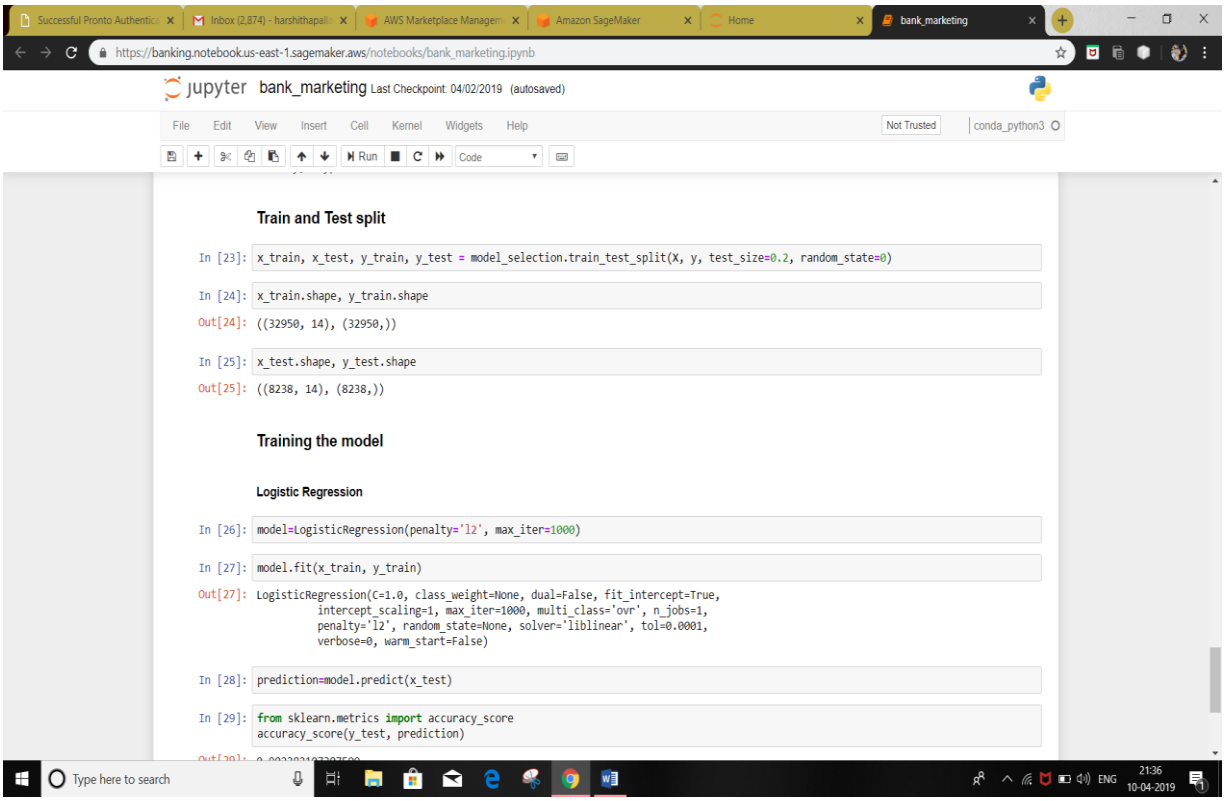


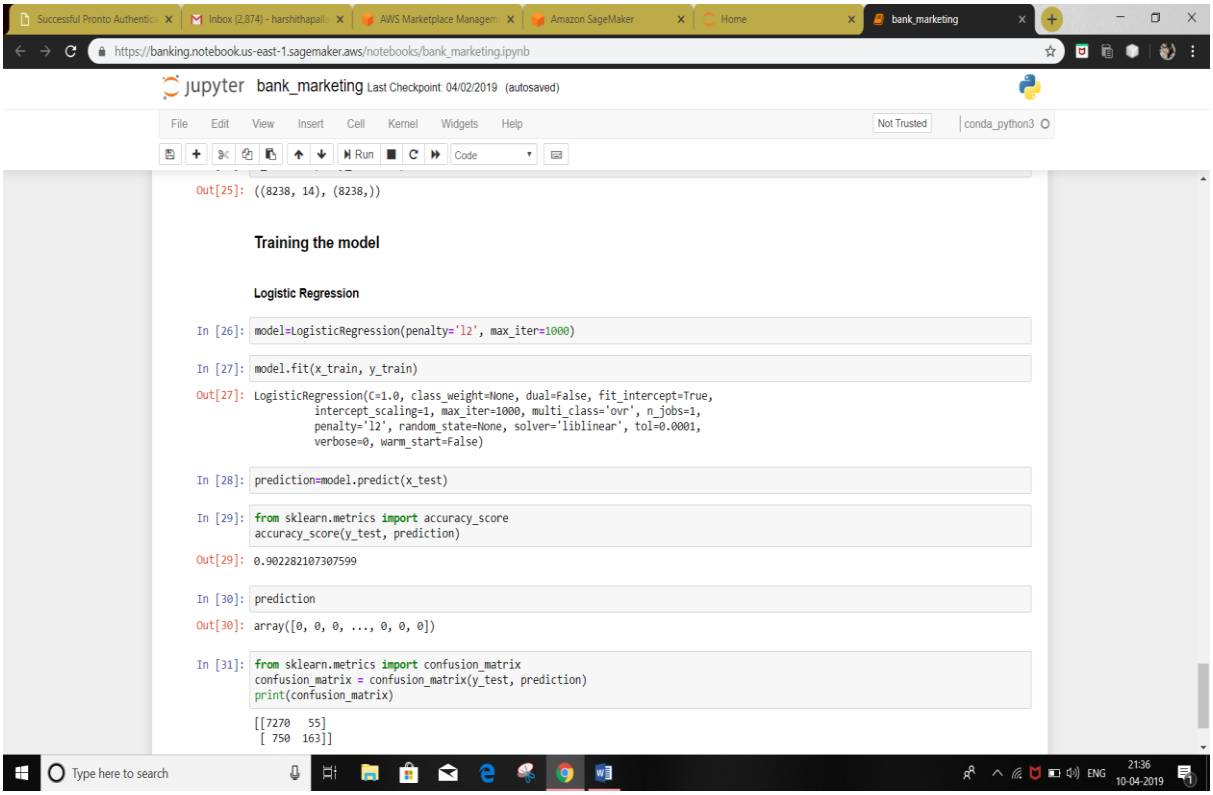

Successful Pronto Authentication
Inbox (2,874) - harshithapalla20
AWS Marketplace Management
Amazon SageMaker
Home
https://banking.notebook.us-east-1.sagemaker.aws/tree
jupyter
Open JupyterLab
Quit
Files
Running
Clusters
SageMaker Examples
Conda
Select items to perform actions on them.
Upload
New
Name
Last Modified
File size
bank_marketing.ipynb
10 hours ago
90.5 kB
banking_updated.csv
8 days ago
4.84 MB
Type here to search
ENG
21:35
10-04-2019


bank_marketing
https://banking.notebook.us-east-1.sagemaker.aws/notebooks/bank_marketing.ipynb
jupyter bank_marketing Last Checkpoint: 04/02/2019 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Kernel starting, please wait...
Not Trusted
conda_python3
Run
Code
In [1]:
import numpy as np
import sklearn
import pickle
from sklearn.preprocessing import LabelEncoder
from sklearn import preprocessing
from sklearn import model_selection
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
import pandas as pd
#import pylab as plb
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')
In [2]:
df = pd.read_csv('banking_updated.csv')
df.head()
Out[2]:
age job marital education default housing loan contact month day_of_week ... campaign pdays previous poutcome emp_var_rat
0 44 blue-collar married basic.4y unknown yes no cellular aug thu ... 1 999 0 nonexistent 1.
1 53 technician married unknown no no no cellular nov fri ... 1 999 0 nonexistent -0.
2 28 management single university.degree no yes no cellular jun thu ... 3 6 2 success -1.
3 39 services married high.school no no no cellular apr fri ... 2 999 0 nonexistent -1.
4 55 retired married basic.4y no yes no cellular aug fri ... 1 3 1 success -2.
5 rows × 21 columns
In [3]:
df.describe()
Out[3]:
age duration campaign pdays previous emp_var_rate cons_price_idx cons_conf_idx euribor3m nr_employed
count 41188.00000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.0
Loading [MathJax]/extensions/MathML/content-mathml.js
Type here to search
ENG
21:36
10-04-2019


jupyter bank_marketing Last Checkpoint: 04/02/2019 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
conda_python3
Out[3]:
age duration campaign pdays previous emp_var_rate cons_price_idx cons_conf_idx euribor3m nr_employed
count 41188.00000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000 41188.000000
mean 40.02406 258.285010 2.567593 962.475454 0.172963 0.081886 93.575664 -40.502600 3.621291 5167.035911
std 10.42125 259.279249 2.770014 186.910907 0.494901 1.570960 0.578840 4.628198 1.734447 72.251528
min 17.00000 0.000000 1.000000 0.000000 0.000000 -3.400000 92.201000 -50.800000 0.634000 4963.600000
25% 32.00000 102.000000 1.000000 999.000000 0.000000 -1.800000 93.075000 -42.700000 1.344000 5099.100000
50% 38.00000 180.000000 2.000000 999.000000 0.000000 1.100000 93.749000 -41.800000 4.857000 5191.000000
75% 47.00000 319.000000 3.000000 999.000000 0.000000 1.400000 93.994000 -36.400000 4.961000 5228.100000
max 98.00000 4918.000000 56.000000 999.000000 7.000000 1.400000 94.767000 -26.900000 5.045000 5228.100000
In [4]: df.drop(['duration','contact','month','day_of_week','default','pdays',],axis=1,inplace=True)
In [5]: df.head()
Out[5]:
age job marital education housing loan campaign previous poutcome emp_var_rate cons_price_idx cons_conf_idx euribor3m nr_em
0 44 blue-collar married basic.4y yes no 1 0 nonexistent 1.4 93.444 -36.1 4.963
1 53 technician married unknown no no 1 0 nonexistent -0.1 93.200 -42.0 4.021
2 28 management single university.degree yes no 3 2 success -1.7 94.055 -39.8 0.729
3 39 services married high.school no no 2 0 nonexistent -1.8 93.075 -47.1 1.405
4 55 retired married basic.4y yes no 1 1 success -2.9 92.201 -31.4 0.869
In [6]: df.isnull().sum()
Out[6]:
age 0
job 0
marital 0
education 0
housing 0
Type here to search
ENG 21:36 10-04-2019


jupyter bank_marketing Last Checkpoint: 04/02/2019 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
conda_python3
In [8]: import seaborn as sns
In [9]: sns.countplot(x='y', data=df)
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc21e823fd0>
count
y
In [10]: sns.countplot(y='job', data=df)
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc21e2d6f60>
job
blue-collar
technician
management
services
retired
admin.
housemaid
unemployed
entrepreneur
Type here to search
ENG 21:36 10-04-2019


bank_marketing
https://banking.notebook.us-east-1.sagemaker.aws/notebooks/bank_marketing.ipynb
jupyter bank_marketing Last Checkpoint: 04/02/2019 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
conda_python3
Run
Code
Train and Test split
In [23]: x_train, x_test, y_train, y_test = model_selection.train_test_split(X, y, test_size=0.2, random_state=0)
In [24]: x_train.shape, y_train.shape
Out[24]: ((32950, 14), (32950,))
In [25]: x_test.shape, y_test.shape
Out[25]: ((8238, 14), (8238,))
Training the model
Logistic Regression
In [26]: model=LogisticRegression(penalty='l2', max_iter=1000)
In [27]: model.fit(x_train, y_train)
Out[27]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=1000, multi_class='ovr', n_jobs=1,
penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
verbose=0, warm_start=False)
In [28]: prediction=model.predict(x_test)
In [29]: from sklearn.metrics import accuracy_score
accuracy_score(y_test, prediction)
Type here to search
ENG
21:36
10-04-2019


bank_marketing
https://banking.notebook.us-east-1.sagemaker.aws/notebooks/bank_marketing.ipynb
jupyter bank_marketing Last Checkpoint: 04/02/2019 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
conda_python3
Run
Code
Out[25]: ((8238, 14), (8238,))
Training the model
Logistic Regression
In [26]: model=LogisticRegression(penalty='l2', max_iter=1000)
In [27]: model.fit(x_train, y_train)
Out[27]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=1000, multi_class='ovr', n_jobs=1,
penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
verbose=0, warm_start=False)
In [28]: prediction=model.predict(x_test)
In [29]: from sklearn.metrics import accuracy_score
accuracy_score(y_test, prediction)
Out[29]: 0.902282107307599
In [30]: prediction
Out[30]: array([0, 0, 0, ..., 0, 0, 0])
In [31]: from sklearn.metrics import confusion_matrix
confusion_matrix = confusion_matrix(y_test, prediction)
print(confusion_matrix)
[[7270 55]
[750 163]]
Type here to search
ENG
21:36
10-04-2019

Algorithm used in Sagemaker:

Code:

```
import numpy as np
import sklearn
import pickle
from sklearn.preprocessing import LabelEncoder
from sklearn import preprocessing
from sklearn import model_selection
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
import pandas as pd
#import pylab as plb
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')

df = pd.read_csv('banking_updated.csv')
df.head()

df.describe()

df.drop(['duration','contact','month','day_of_week','default','pd
ays',],axis=1,inplace=True)

df.head()

df.isnull().sum()

df.replace(['basic.6y','basic.4y', 'basic.9y'], 'basic', inplace=True)

import seaborn as sns

sns.countplot(x='y', data=df)

sns.countplot(y='job', data=df)
```

```
sns.countplot(x='marital', data=df)

df.education.value_counts()

sns.countplot(y='education', data=df)

df.head()

le = preprocessing.LabelEncoder()

df.job = le.fit_transform(df.job)

df.marital = le.fit_transform(df.marital)

df.education = le.fit_transform(df.education)
df.housing = le.fit_transform(df.housing)
df.loan = le.fit_transform(df.loan)
df.poutcome = le.fit_transform(df.poutcome)

df.head()

df.shape

X = df.iloc[:,0:14]
X[0:10]

y = df.iloc[:,14]
y[0:10]

x_train, x_test, y_train, y_test =
model_selection.train_test_split(X, y, test_size=0.2,
random_state=0)

x_train.shape, y_train.shape
```

```
x_test.shape, y_test.shape

model=LogisticRegression(penalty='l2', max_iter=1000)

model.fit(x_train, y_train)

prediction=model.predict(x_test)

from sklearn.metrics import accuracy_score
accuracy_score(y_test, prediction)

prediction

from sklearn.metrics import confusion_matrix
confusion_matrix = confusion_matrix(y_test, prediction)
print(confusion_matrix)
```