Train a model with Amazon SageMaker Autopilot

Introduction

In this lab, you will use Amazon Sagemaker Autopilot to train a BERT-based natural language processing (NLP) model. The model will analyze customer feedback and classify the messages into positive (1), neutral (0) and negative (-1) sentiment.

Table of Contents

Amazon SageMaker Autopilot automatically trains and tunes the best machine learning models for classification or regression, based on your data while allowing to maintain full control and visibility.

SageMaker Autopilot will inspect the raw dataset, apply feature processors, pick the best set of algorithms, train and tune multiple models, and then rank the models based on performance - all with just a few clicks. Autopilot transparently generates a set of Python scripts and notebooks for a complete end-to-end pipeline including data analysis, candidate generation, feature engineering, and model training/tuning.

SageMaker Autopilot job consists of the following high-level steps:

- *Data analysis* where the data is summarized and analyzed to determine which feature engineering techniques, hyper-parameters, and models to explore.
- *Feature engineering* where the data is scrubbed, balanced, combined, and split into train and validation.
- *Model training and tuning* where the top performing features, hyper-parameters, and models are selected and trained.

These re-usable scripts and notebooks give us full visibility into how the model candidates were created. Since Autopilot integrates natively with SageMaker Studio, we can visually explore the different models generated by SageMaker Autopilot.

SageMaker Autopilot can be used by people without machine learning experience to automatically train a model from a dataset. Additionally, experienced developers can use Autopilot to train a baseline model from which they can iterate and manually improve.

Autopilot is available through the SageMaker Studio UI and AWS Python SDK. In this notebook, you will use the AWS Python SDK to train a series of text-classification models and deploy the model with the highest accuracy.

For more details on Autopilot, have a look at this **Amazon Science Publication** (https://www.amazon.science/publications/amazon-sagemaker-autopilot-a-white-box-automl-solution-at-scale).

Use case: analyze customer sentiment

Customer feedback appears across many channels including social media and partner websites. As a company, you want to capture this valuable product feedback to spot negative trends and improve the situation, if needed. Here you will train a model to classify the feedback messages into positive (1), neutral (0) and negative (-1) sentiment.

First, let's install and import required modules.

In [2]:

```
# please ignore warning messages during the installation
!pip install --disable-pip-version-check -q sagemaker==2.35.0
```

```
WARNING: Running pip as the 'root' user can result in broken permission
s and conflicting behaviour with the system package manager. It is reco
mmended to use a virtual environment instead: https://pip.pypa.io/warni
ngs/venv
```

In [3]:

```
import boto3
import sagemaker
import pandas as pd
import numpy as np
import botocore
import time
import json

config = botocore.config.Config(user_agent_extra='dlai-pds/c1/w3')

# low-level service client of the boto3 session
sm = boto3.client(service_name='sagemaker',
                  config=config)

sm_runtime = boto3.client('sagemaker-runtime',
                          config=config)

sess = sagemaker.Session(sagemaker_client=sm,
                         sagemaker_runtime_client=sm_runtime)

bucket = sess.default_bucket()
role = sagemaker.get_execution_role()
region = sess.boto_region_name
```

In [4]:

```
import matplotlib.pyplot as plt
%matplotlib inline
%config InlineBackend.figure_format='retina'
```

1. Review transformed dataset

Let's transform the dataset into a format that Autopilot recognizes. Specifically, a comma-separated file of `label,features` as shown here:

```
sentiment,review_body
-1,"this is bad"
0,"this is ok"
1,"this is great"
...
```

Sentiment is one of three classes: negative (-1), neutral (0), or positive (1). Autopilot requires that the target variable, `sentiment` is first and the set of features, just `review_body` in this case, come next.

In [5]:

```
!aws s3 cp 's3://dlai-practical-data-science/data/balanced/womens_clothing_ecommerce_reviews_balanced.csv' ./
```

```
download: s3://dlai-practical-data-science/data/balanced/womens_clothing_ecommerce_reviews_balanced.csv to ./womens_clothing_ecommerce_reviews_balanced.csv
```

In [6]:

```
path = './womens_clothing_ecommerce_reviews_balanced.csv'

df = pd.read_csv(path, delimiter=',')
df.head()
```

Out[6]:

	sentiment	review_body	product_category
0	-1	This suit did nothing for me. the top has zero...	Swim
1	-1	Like other reviewers i saw this dress on the ...	Dresses
2	-1	I wish i had read the reviews before purchasin...	Knits
3	-1	I ordered these pants in my usual size (xl) an...	Legwear
4	-1	I noticed this top on one of the sales associa...	Knits

In [7]:

```
path_autopilot = './womens_clothing_ecommerce_reviews_balanced_for_autopilot.csv'

df[['sentiment', 'review_body']].to_csv(path_autopilot,
                                        sep=',',
                                        index=False)
```

2. Configure the Autopilot job

2.1. Upload data to S3 bucket

In [8]:

```
autopilot_train_s3_uri = sess.upload_data(bucket=bucket, key_prefix='autopilot/dat
a', path=path_autopilot)
autopilot_train_s3_uri
```

Out[8]:

```
's3://sagemaker-us-east-1-822315549952/autopilot/data/womens_clothing_e
commerce_reviews_balanced_for_autopilot.csv'
```

Check the existence of the dataset in this S3 bucket folder:

In [9]:

```
!aws s3 ls $autopilot_train_s3_uri
```

```
2022-08-21 23:59:04    2253749 womens_clothing_ecommerce_reviews_balanc
ed_for_autopilot.csv
```

2.2. S3 output for generated assets

Set the S3 output path for the Autopilot outputs. This includes Jupyter notebooks (analysis), Python scripts (feature engineering), and trained models.

In [10]:

```
model_output_s3_uri = 's3://{}/autopilot'.format(bucket)

print(model_output_s3_uri)
```

```
s3://sagemaker-us-east-1-822315549952/autopilot
```

2.3. Configure the Autopilot job

Create the Autopilot job name.

In [11]:

```
import time

timestamp = int(time.time())

auto_ml_job_name = 'automl-dm-{}'.format(timestamp)
```

When configuring our Autopilot job, you need to specify the maximum number of candidates, `max_candidates`, to explore as well as the input/output S3 locations and target column to predict. In this case, you want to predict `sentiment` from the review text.

Exercise 1

Configure the Autopilot job.

Instructions: Create an instance of the `sagemaker.automl.automl.AutoML` estimator class passing the required configuration parameters. Target attribute for predictions here is `sentiment`.

```
automl = sagemaker.automl.automl.AutoML(
    target_attribute_name='...', # the name of the target attribute for pred
ictions
    base_job_name=..., # Autopilot job name
    output_path=..., # output data path
    max_candidates=..., # maximum number of candidates
    sagemaker_session=sess,
    role=role,
    max_runtime_per_training_job_in_seconds=1200,
    total_job_runtime_in_seconds=7200
)
```

In [12]:

```
max_candidates = 3

automl = sagemaker.automl.automl.AutoML(
    ### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
    target_attribute_name='sentiment', # Replace None
    base_job_name=auto_ml_job_name, # Replace None
    output_path=model_output_s3_uri, # Replace None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
    max_candidates=max_candidates,
    sagemaker_session=sess,
    role=role,
    max_runtime_per_training_job_in_seconds=1200,
    total_job_runtime_in_seconds=7200
)
```

3. Launch the Autopilot job

Exercise 2

Launch the Autopilot job.

Instructions: Call `fit` function of the configured estimator passing the S3 bucket input data path and the Autopilot job name.

```
automl.fit(
    ..., # input data path
    job_name=auto_ml_job_name, # Autopilot job name
    wait=False,
    logs=False
)
```

In [13]:

```
automl.fit(
    ### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
    autopilot_train_s3_uri, # Replace None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
    job_name=auto_ml_job_name,
    wait=False,
    logs=False
)
```

4. Track Autopilot job progress

Once the Autopilot job has been launched, you can track the job progress directly from the notebook using the SDK capabilities.

4.1. Autopilot job description

Function `describe_auto_ml_job` of the Amazon SageMaker service returns the information about the AutoML job in dictionary format. You can review the response syntax and response elements in the **documentation** (https://docs.aws.amazon.com/sagemaker/latest/APIReference/API_DescribeAutoMLJob.html).

In [14]:

```
job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name)
```

4.2. Autopilot job status

To track the job progress you can use two response elements: `AutoMLJobStatus` and `AutoMLJobSecondaryStatus` , which correspond to the primary (Completed | InProgress | Failed | Stopped | Stopping) and secondary (AnalyzingData | FeatureEngineering | ModelTuning etc.) job states respectively. To see if the AutoML job has started, you can check the existence of the `AutoMLJobStatus` and `AutoMLJobSecondaryStatus` elements in the job description response.

In this notebook, you will use the following scheme to track the job progress:

```
    # check if the job is still at certain stage
    while [check 'AutoMLJobStatus' and 'AutoMLJobSecondaryStatus'] in job_descri
    ption_response:
        # update the job description response
        job_description_response = automl.describe_auto_ml_job(AutoMLJobName=aut
    o_ml_job_name)
        # print the message the Autopilot job is in the stage ...
        print([message])
        # git a time step to check the status again
        sleep(15)
    print("Autopilot job complete...")
```

In [15]:

```
while 'AutoMLJobStatus' not in job_description_response.keys() and 'AutoMLJobSecond
aryStatus' not in job_description_response.keys():
    job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_nam
e)
    print('[INFO] Autopilot job has not yet started. Please wait. ')
    # function `json.dumps` encodes JSON string for printing.
    print(json.dumps(job_description_response, indent=4, sort_keys=True, default=st
r))
    print('[INFO] Waiting for Autopilot job to start...')
    sleep(15)

print('[OK] AutoML job started.')
```

```
[OK] AutoML job started.
```

4.3. Review the SageMaker processing jobs

The Autopilot creates required SageMaker processing jobs during the run:

- First processing job (data splitter) checks the data sanity, performs stratified shuffling and splits the data into training and validation.
- Second processing job (candidate generator) first streams through the data to compute statistics for the dataset. Then, uses these statistics to identify the problem type, and possible types of every column-predictor: numeric, categorical, natural language, etc.

In [16]:

```
from IPython.core.display import display, HTML

display(HTML('<b>Review <a target="blank" href="https://console.aws.amazon.com/sagemaker/home?region={}#/processing-jobs/">processing jobs</a></b>'.format(region)))
```

Review processing jobs (https://console.aws.amazon.com/sagemaker/home?region=us-east-1#/processing-jobs/)

You can review the updates on that page during the run of the Autopilot job.

4.4. Wait for the data analysis step to finish

Here you will use the same scheme as above to check the completion of the data analysis step. This step can be identified with the (primary) job status value `InProgress` and secondary job status values `Starting` and then `AnalyzingData`.

This cell will take approximately 10 minutes to run.

In [17]:

```
%%time

job_status = job_description_response['AutoMLJobStatus']
job_sec_status = job_description_response['AutoMLJobSecondaryStatus']

if job_status not in ('Stopped', 'Failed'):
    while job_status in ('InProgress') and job_sec_status in ('Starting', 'Analyzin
gData'):
        job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job
_name)
        job_status = job_description_response['AutoMLJobStatus']
        job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
        print(job_status, job_sec_status)
        time.sleep(15)
    print('[OK] Data analysis phase completed.\n')

print(json.dumps(job_description_response, indent=4, sort_keys=True, default=str))
```

```
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress AnalyzingData
InProgress FeatureEngineering
[OK] Data analysis phase completed.

{
    "AutoMLJobArn": "arn:aws:sagemaker:us-east-1:822315549952:automl-jo
b/automl-dm-1661126447",
    "AutoMLJobArtifacts": {
        "CandidateDefinitionNotebookLocation": "s3://sagemaker-us-east-
1-822315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidat
es/automl-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/note
books/SageMakerAutopilotCandidateDefinitionNotebook.ipynb",
        "DataExplorationNotebookLocation": "s3://sagemaker-us-east-1-82
2315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidates/a
utoml-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/notebook
s/SageMakerAutopilotDataExplorationNotebook.ipynb"
    },
    "AutoMLJobConfig": {
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": 7200,
            "MaxCandidates": 3,
            "MaxRuntimePerTrainingJobInSeconds": 1200
        },
        "SecurityConfig": {
            "EnableInterContainerTrafficEncryption": false
        }
    },
    "AutoMLJobName": "automl-dm-1661126447",
    "AutoMLJobSecondaryStatus": "FeatureEngineering",
    "AutoMLJobStatus": "InProgress",
    "CreationTime": "2022-08-22 00:05:00.600000+00:00",
    "GenerateCandidateDefinitionsOnly": false,
    "InputDataConfig": [
        {
```

```
            "ChannelType": "training",
            "ContentType": "text/csv;header=present",
            "DataSource": {
                "S3DataSource": {
                    "S3DataType": "S3Prefix",
                    "S3Uri": "s3://sagemaker-us-east-1-822315549952/aut
opilot/data/womens_clothing_ecommerce_reviews_balanced_for_autopilot.cs
v"
                }
            },
            "TargetAttributeName": "sentiment"
        }
    ],
    "LastModifiedTime": "2022-08-22 00:14:12.851000+00:00",
    "OutputDataConfig": {
        "S3OutputPath": "s3://sagemaker-us-east-1-822315549952/autopilo
t"
    },
    "ResolvedAttributes": {
        "AutoMLJobObjective": {
            "MetricName": "Accuracy"
        },
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": 7200,
            "MaxCandidates": 3,
            "MaxRuntimePerTrainingJobInSeconds": 1200
        },
        "ProblemType": "MulticlassClassification"
    },
    "ResponseMetadata": {
        "HTTPHeaders": {
            "content-length": "1809",
            "content-type": "application/x-amz-json-1.1",
            "date": "Mon, 22 Aug 2022 00:14:15 GMT",
            "x-amzn-requestid": "95e38d31-e2ce-4e60-aa9d-f72fd4aa4618"
        },
        "HTTPStatusCode": 200,
        "RequestId": "95e38d31-e2ce-4e60-aa9d-f72fd4aa4618",
        "RetryAttempts": 0
    },
    "RoleArn": "arn:aws:iam::822315549952:role/sagemaker-studio-vpc-fir
ewall-us-east-1-sagemaker-execution-role"
}
CPU times: user 321 ms, sys: 23.8 ms, total: 344 ms
Wall time: 6min 19s
```


Wait for Autopilot to finish generating the notebooks.

4.5. View generated notebooks

Once data analysis is complete, SageMaker AutoPilot generates two notebooks:

- Data exploration
- Candidate definition

Notebooks are included in the AutoML job artifacts generated during the run. Before checking the existence of the notebooks, you can check if the artifacts have been generated.

Exercise 3

Check if the Autopilot job artifacts have been generated.

Instructions: Use status check scheme described above. The generation of artifacts can be identified by existence of `AutoMLJobArtifacts` element in the keys of the job description response.

In [19]:

```
### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
# get the information about the running Autopilot job
job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name) # Replace None

# keep in the while loop until the Autopilot job artifacts will be generated
while 'AutoMLJobArtifacts' not in job_description_response: # Replace all None
    # update the information about the running Autopilot job
    job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name) # Replace None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
    print('[INFO] Autopilot job has not yet generated the artifacts. Please wait. ')
    print(json.dumps(job_description_response, indent=4, sort_keys=True, default=str))
    print('[INFO] Waiting for AutoMLJobArtifacts...')
    time.sleep(15)

print('[OK] AutoMLJobArtifacts generated.')
```

```
[OK] AutoMLJobArtifacts generated.
```

Wait for Autopilot to make the notebooks available.

Exercise 4

Check if the notebooks have been created.

Instructions: Use status check scheme described above. Notebooks creation can be identified by existence of `DataExplorationNotebookLocation` element in the keys of the `job_description_response['AutoMLJobArtifacts']` dictionary.

In [20]:

```
### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
# get the information about the running Autopilot job
job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name) # Replace None

# keep in the while loop until the notebooks will be created
while 'DataExplorationNotebookLocation' not in job_description_response['AutoMLJobArtifacts']: # Replace all None
    # update the information about the running Autopilot job
    job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name) # Replace None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
    print('[INFO] Autopilot job has not yet generated the notebooks. Please wait. ')
    print(json.dumps(job_description_response, indent=4, sort_keys=True, default=str))
    print('[INFO] Waiting for DataExplorationNotebookLocation...')
    time.sleep(15)

print('[OK] DataExplorationNotebookLocation found.')
```

```
[OK] DataExplorationNotebookLocation found.
```

Review the generated resources in S3 directly. Following the link, you can find the notebooks in the folder `notebooks` and download them by clicking on object `Actions` / `Object actions -> Download as` / `Download` .

In [21]:

```
from IPython.core.display import display, HTML

generated_resources = job_description_response['AutoMLJobArtifacts']['DataExplorationNotebookLocation']
download_path = generated_resources.rsplit('/notebooks/SageMakerAutopilotDataExplorationNotebook.ipynb')[0]
job_id = download_path.rsplit('/', 1)[-1]

if not job_id:
    print('No AutoMLJobArtifacts found.')
else:
    display(HTML('<b>Review <a target="blank" href="https://s3.console.aws.amazon.com/s3/buckets/{}/autopilot/{}/sagemaker-automl-candidates/{}/">generated notebooks</a> in S3 bucket</b>'.format(bucket, auto_ml_job_name, job_id)))
```

Review generated notebooks (https://s3.console.aws.amazon.com/s3/buckets/sagemaker-us-east-1-822315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidates/automl-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/) in S3 bucket

5. Feature engineering

Exercise 5

Check the completion of the feature engineering step.

Instructions: Use status check scheme described above. Feature engineering step can be identified with the (primary) job status value `InProgress` and secondary job status value `FeatureEngineering`.

This cell will take approximately 10 minutes to run.

In [22]:

```
%%time

job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name)
job_status = job_description_response['AutoMLJobStatus']
job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
print(job_status)
print(job_sec_status)
if job_status not in ('Stopped', 'Failed'):
    ### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
    while job_status == 'InProgress' and job_sec_status == 'FeatureEngineering': #
 Replace all None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
        job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job
_name)
        job_status = job_description_response['AutoMLJobStatus']
        job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
        print(job_status, job_sec_status)
        time.sleep(5)
    print('[OK] Feature engineering phase completed.\n')

print(json.dumps(job_description_response, indent=4, sort_keys=True, default=str))
```

```
InProgress
GeneratingExplainabilityReport
[OK] Feature engineering phase completed.

{
    "AutoMLJobArn": "arn:aws:sagemaker:us-east-1:822315549952:automl-jo
b/automl-dm-1661126447",
    "AutoMLJobArtifacts": {
        "CandidateDefinitionNotebookLocation": "s3://sagemaker-us-east-
1-822315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidat
es/automl-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/note
books/SageMakerAutopilotCandidateDefinitionNotebook.ipynb",
        "DataExplorationNotebookLocation": "s3://sagemaker-us-east-1-82
2315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidates/a
utoml-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/notebook
s/SageMakerAutopilotDataExplorationNotebook.ipynb"
    },
    "AutoMLJobConfig": {
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": 7200,
            "MaxCandidates": 3,
            "MaxRuntimePerTrainingJobInSeconds": 1200
        },
        "SecurityConfig": {
            "EnableInterContainerTrafficEncryption": false
        }
    },
    "AutoMLJobName": "automl-dm-1661126447",
    "AutoMLJobSecondaryStatus": "GeneratingExplainabilityReport",
    "AutoMLJobStatus": "InProgress",
    "BestCandidate": {
        "CandidateName": "automl-dm-1661126447GqsLp9je4jVG-003-724e794
e",
        "CandidateProperties": {
            "CandidateMetrics": [
                {
                    "MetricName": "Accuracy",
                    "Set": "Validation",
                    "StandardMetricName": "Accuracy",
                    "Value": 0.6161699891090393
                },
                {
                    "MetricName": "PrecisionMacro",
                    "Set": "Validation",
                    "StandardMetricName": "PrecisionMacro",
                    "Value": 0.6182199716567993
                },
                {
                    "MetricName": "BalancedAccuracy",
                    "Set": "Validation",
                    "StandardMetricName": "BalancedAccuracy",
                    "Value": 0.6161699891090393
                },
                {
                    "MetricName": "F1macro",
                    "Set": "Validation",
                    "StandardMetricName": "F1macro",
```

```
                    "Value": 0.6169800162315369
                },
                {
                    "MetricName": "RecallMacro",
                    "Set": "Validation",
                    "StandardMetricName": "RecallMacro",
                    "Value": 0.6161699891090393
                }
            ]
        },
        "CandidateStatus": "Completed",
        "CandidateSteps": [
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:processing-job/automl-dm-1661126447-db-1-848d0ed3feb448c9ab6491e
9f701fe5b10178",
                "CandidateStepName": "automl-dm-1661126447-db-1-848d0ed
3feb448c9ab6491e9f701fe5b10178",
                "CandidateStepType": "AWS::SageMaker::ProcessingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447-dpp0-1-472c0d3c133b4b138860983
2bf05be317b1",
                "CandidateStepName": "automl-dm-1661126447-dpp0-1-472c0
d3c133b4b1388609832bf05be317b1",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:transform-job/automl-dm-1661126447-dpp0-rpb-1-59e569b7f6ef495d92
15d86e33748a5",
                "CandidateStepName": "automl-dm-1661126447-dpp0-rpb-1-5
9e569b7f6ef495d9215d86e33748a5",
                "CandidateStepType": "AWS::SageMaker::TransformJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447gqslp9je4jvg-003-724e794e",
                "CandidateStepName": "automl-dm-1661126447GqsLp9je4jVG-
003-724e794e",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            }
        ],
        "CreationTime": "2022-08-22 00:24:54+00:00",
        "EndTime": "2022-08-22 00:28:42+00:00",
        "FinalAutoMLJobObjectiveMetric": {
            "MetricName": "validation:accuracy",
            "Value": 0.6161699891090393
        },
        "InferenceContainers": [
            {
                "Environment": {
                    "AUTOML_SPARSE_ENCODE_RECORDIO_PROTOBUF": "1",
                    "AUTOML_TRANSFORM_MODE": "feature-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "applicatio
n/x-recordio-protobuf",
```

```
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            },
            {
                "Environment": {
                    "MAX_CONTENT_LENGTH": "20971520",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,probabilities"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-xgboost:1.3-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/tuning/automl-dm--dpp0-xgb/automl-dm-166
1126447GqsLp9je4jVG-003-724e794e/output/model.tar.gz"
            },
            {
                "Environment": {
                    "AUTOML_TRANSFORM_MODE": "inverse-label-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_INPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,labels,probabilities",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            }
        ],
        "LastModifiedTime": "2022-08-22 00:29:35.467000+00:00",
        "ObjectiveStatus": "Succeeded"
    },
    "CreationTime": "2022-08-22 00:05:00.600000+00:00",
    "GenerateCandidateDefinitionsOnly": false,
    "InputDataConfig": [
        {
            "ChannelType": "training",
            "ContentType": "text/csv;header=present",
            "DataSource": {
                "S3DataSource": {
                    "S3DataType": "S3Prefix",
                    "S3Uri": "s3://sagemaker-us-east-1-822315549952/aut
opilot/data/womens_clothing_ecommerce_reviews_balanced_for_autopilot.cs
v"
                }
```

```
            },
            "TargetAttributeName": "sentiment"
        }
    ],
    "LastModifiedTime": "2022-08-22 00:33:51.810000+00:00",
    "OutputDataConfig": {
        "S3OutputPath": "s3://sagemaker-us-east-1-822315549952/autopilo
t"
    },
    "ResolvedAttributes": {
        "AutoMLJobObjective": {
            "MetricName": "Accuracy"
        },
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": 7200,
            "MaxCandidates": 3,
            "MaxRuntimePerTrainingJobInSeconds": 1200
        },
        "ProblemType": "MulticlassClassification"
    },
    "ResponseMetadata": {
        "HTTPHeaders": {
            "content-length": "5633",
            "content-type": "application/x-amz-json-1.1",
            "date": "Mon, 22 Aug 2022 00:33:59 GMT",
            "x-amzn-requestid": "47bfdfbe-60a6-42f3-a8c8-6faa547d5ac7"
        },
        "HTTPStatusCode": 200,
        "RequestId": "47bfdfbe-60a6-42f3-a8c8-6faa547d5ac7",
        "RetryAttempts": 0
    },
    "RoleArn": "arn:aws:iam::822315549952:role/sagemaker-studio-vpc-fir
ewall-us-east-1-sagemaker-execution-role"
}
CPU times: user 20.8 ms, sys: 784 µs, total: 21.6 ms
Wall time: 152 ms
```


6. Model training and tuning

When you launched the Autopilot job, you requested that 3 model candidates are generated and compared. Therefore, you should see three (3) SageMaker training jobs below.

In [23]:

```
from IPython.core.display import display, HTML

display(HTML('<b>Review <a target="blank" href="https://console.aws.amazon.com/sage
maker/home?region={}#/hyper-tuning-jobs/">hyper-parameter tuning jobs</a></b>'.form
at(region)))
```

Review [hyper-parameter tuning jobs (https://console.aws.amazon.com/sagemaker/home?region=us-east-1#/hyper-tuning-jobs/)]

6.1. Wait for training and tuning

Exercise 6

Check the completion of the model tuning step.

Instructions: Use status check scheme described above. Model tuning step can be identified with the (primary) job status value `InProgress` and secondary job status value `ModelTuning`.

This cell will take approximately 5-10 minutes to run.

In [24]:

```
%%time

job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name)
job_status = job_description_response['AutoMLJobStatus']
job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
print(job_status)
print(job_sec_status)
if job_status not in ('Stopped', 'Failed'):
    ### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
    while job_status == 'InProgress' and job_sec_status == 'ModelTuning': # Replace
all None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
        job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job
_name)
        job_status = job_description_response['AutoMLJobStatus']
        job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
        print(job_status, job_sec_status)
        time.sleep(5)
    print('[OK] Model tuning phase completed.\n')

print(json.dumps(job_description_response, indent=4, sort_keys=True, default=str))
```

```
InProgress
GeneratingExplainabilityReport
[OK] Model tuning phase completed.

{
    "AutoMLJobArn": "arn:aws:sagemaker:us-east-1:822315549952:automl-jo
b/automl-dm-1661126447",
    "AutoMLJobArtifacts": {
        "CandidateDefinitionNotebookLocation": "s3://sagemaker-us-east-
1-822315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidat
es/automl-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/note
books/SageMakerAutopilotCandidateDefinitionNotebook.ipynb",
        "DataExplorationNotebookLocation": "s3://sagemaker-us-east-1-82
2315549952/autopilot/automl-dm-1661126447/sagemaker-automl-candidates/a
utoml-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea32a8e69/notebook
s/SageMakerAutopilotDataExplorationNotebook.ipynb"
    },
    "AutoMLJobConfig": {
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": 7200,
            "MaxCandidates": 3,
            "MaxRuntimePerTrainingJobInSeconds": 1200
        },
        "SecurityConfig": {
            "EnableInterContainerTrafficEncryption": false
        }
    },
    "AutoMLJobName": "automl-dm-1661126447",
    "AutoMLJobSecondaryStatus": "GeneratingExplainabilityReport",
    "AutoMLJobStatus": "InProgress",
    "BestCandidate": {
        "CandidateName": "automl-dm-1661126447GqsLp9je4jVG-003-724e794
e",
        "CandidateProperties": {
            "CandidateMetrics": [
                {
                    "MetricName": "Accuracy",
                    "Set": "Validation",
                    "StandardMetricName": "Accuracy",
                    "Value": 0.6161699891090393
                },
                {
                    "MetricName": "PrecisionMacro",
                    "Set": "Validation",
                    "StandardMetricName": "PrecisionMacro",
                    "Value": 0.6182199716567993
                },
                {
                    "MetricName": "BalancedAccuracy",
                    "Set": "Validation",
                    "StandardMetricName": "BalancedAccuracy",
                    "Value": 0.6161699891090393
                },
                {
                    "MetricName": "F1macro",
                    "Set": "Validation",
                    "StandardMetricName": "F1macro",
```

```
                    "Value": 0.6169800162315369
                },
                {
                    "MetricName": "RecallMacro",
                    "Set": "Validation",
                    "StandardMetricName": "RecallMacro",
                    "Value": 0.6161699891090393
                }
            ]
        },
        "CandidateStatus": "Completed",
        "CandidateSteps": [
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:processing-job/automl-dm-1661126447-db-1-848d0ed3feb448c9ab6491e
9f701fe5b10178",
                "CandidateStepName": "automl-dm-1661126447-db-1-848d0ed
3feb448c9ab6491e9f701fe5b10178",
                "CandidateStepType": "AWS::SageMaker::ProcessingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447-dpp0-1-472c0d3c133b4b138860983
2bf05be317b1",
                "CandidateStepName": "automl-dm-1661126447-dpp0-1-472c0
d3c133b4b1388609832bf05be317b1",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:transform-job/automl-dm-1661126447-dpp0-rpb-1-59e569b7f6ef495d92
15d86e33748a5",
                "CandidateStepName": "automl-dm-1661126447-dpp0-rpb-1-5
9e569b7f6ef495d9215d86e33748a5",
                "CandidateStepType": "AWS::SageMaker::TransformJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447gqslp9je4jvg-003-724e794e",
                "CandidateStepName": "automl-dm-1661126447GqsLp9je4jVG-
003-724e794e",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            }
        ],
        "CreationTime": "2022-08-22 00:24:54+00:00",
        "EndTime": "2022-08-22 00:28:42+00:00",
        "FinalAutoMLJobObjectiveMetric": {
            "MetricName": "validation:accuracy",
            "Value": 0.6161699891090393
        },
        "InferenceContainers": [
            {
                "Environment": {
                    "AUTOML_SPARSE_ENCODE_RECORDIO_PROTOBUF": "1",
                    "AUTOML_TRANSFORM_MODE": "feature-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "applicatio
n/x-recordio-protobuf",
```

```
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            },
            {
                "Environment": {
                    "MAX_CONTENT_LENGTH": "20971520",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,probabilities"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-xgboost:1.3-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/tuning/automl-dm--dpp0-xgb/automl-dm-166
1126447GqsLp9je4jVG-003-724e794e/output/model.tar.gz"
            },
            {
                "Environment": {
                    "AUTOML_TRANSFORM_MODE": "inverse-label-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_INPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,labels,probabilities",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            }
        ],
        "LastModifiedTime": "2022-08-22 00:29:35.467000+00:00",
        "ObjectiveStatus": "Succeeded"
    },
    "CreationTime": "2022-08-22 00:05:00.600000+00:00",
    "GenerateCandidateDefinitionsOnly": false,
    "InputDataConfig": [
        {
            "ChannelType": "training",
            "ContentType": "text/csv;header=present",
            "DataSource": {
                "S3DataSource": {
                    "S3DataType": "S3Prefix",
                    "S3Uri": "s3://sagemaker-us-east-1-822315549952/aut
opilot/data/womens_clothing_ecommerce_reviews_balanced_for_autopilot.cs
v"
                }
```

```
            },
            "TargetAttributeName": "sentiment"
        }
    ],
    "LastModifiedTime": "2022-08-22 00:34:42.451000+00:00",
    "OutputDataConfig": {
        "S3OutputPath": "s3://sagemaker-us-east-1-822315549952/autopilo
t"
    },
    "ResolvedAttributes": {
        "AutoMLJobObjective": {
            "MetricName": "Accuracy"
        },
        "CompletionCriteria": {
            "MaxAutoMLJobRuntimeInSeconds": 7200,
            "MaxCandidates": 3,
            "MaxRuntimePerTrainingJobInSeconds": 1200
        },
        "ProblemType": "MulticlassClassification"
    },
    "ResponseMetadata": {
        "HTTPHeaders": {
            "content-length": "5634",
            "content-type": "application/x-amz-json-1.1",
            "date": "Mon, 22 Aug 2022 00:34:50 GMT",
            "x-amzn-requestid": "b3f91fdb-e3c1-4f14-b1eb-03be756f5149"
        },
        "HTTPStatusCode": 200,
        "RequestId": "b3f91fdb-e3c1-4f14-b1eb-03be756f5149",
        "RetryAttempts": 0
    },
    "RoleArn": "arn:aws:iam::822315549952:role/sagemaker-studio-vpc-fir
ewall-us-east-1-sagemaker-execution-role"
}
CPU times: user 17.4 ms, sys: 3.76 ms, total: 21.2 ms
Wall time: 158 ms
```


Please wait until ^^ Autopilot ^^ completes above

Finally, you can check the completion of the Autopilot job looking for the `Completed` job status.

In [25]:

```
%%time

from pprint import pprint

job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job_name)
pprint(job_description_response)
job_status = job_description_response['AutoMLJobStatus']
job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
print('Job status:  {}'.format(job_status))
print('Secondary job status:  {}'.format(job_sec_status))
if job_status not in ('Stopped', 'Failed'):
    while job_status not in ('Completed'):
        job_description_response = automl.describe_auto_ml_job(job_name=auto_ml_job
_name)
        job_status = job_description_response['AutoMLJobStatus']
        job_sec_status = job_description_response['AutoMLJobSecondaryStatus']
        print('Job status:  {}'.format(job_status))
        print('Secondary job status:  {}'.format(job_sec_status))
        time.sleep(10)
    print('[OK] Autopilot job completed.\n')
else:
    print('Job status: {}'.format(job_status))
    print('Secondary job status: {}'.format(job_status))
```

```
{'AutoMLJobArn': 'arn:aws:sagemaker:us-east-1:822315549952:automl-job/a
utoml-dm-1661126447',
 'AutoMLJobArtifacts': {'CandidateDefinitionNotebookLocation': 's3://sa
gemaker-us-east-1-822315549952/autopilot/automl-dm-1661126447/sagemaker
-automl-candidates/automl-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d
32ea32a8e69/notebooks/SageMakerAutopilotCandidateDefinitionNotebook.ipy
nb',
                        'DataExplorationNotebookLocation': 's3://sagema
ker-us-east-1-822315549952/autopilot/automl-dm-1661126447/sagemaker-aut
oml-candidates/automl-dm-1661126447-pr-1-2cc3465ca86f49c69ede49797d32ea
32a8e69/notebooks/SageMakerAutopilotDataExplorationNotebook.ipynb'},
 'AutoMLJobConfig': {'CompletionCriteria': {'MaxAutoMLJobRuntimeInSecon
ds': 7200,
                                            'MaxCandidates': 3,
                                            'MaxRuntimePerTrainingJobIn
Seconds': 1200},
                     'SecurityConfig': {'EnableInterContainerTrafficEnc
ryption': False}},
 'AutoMLJobName': 'automl-dm-1661126447',
 'AutoMLJobSecondaryStatus': 'GeneratingExplainabilityReport',
 'AutoMLJobStatus': 'InProgress',
 'BestCandidate': {'CandidateName': 'automl-dm-1661126447GqsLp9je4jVG-0
03-724e794e',
                   'CandidateProperties': {'CandidateMetrics': [{'Metri
cName': 'Accuracy',
                                                                 'Set':
'Validation',
                                                                 'Stand
ardMetricName': 'Accuracy',
                                                                 'Valu
e': 0.6161699891090393},
                                                                {'Metri
cName': 'PrecisionMacro',
                                                                 'Set':
'Validation',
                                                                 'Stand
ardMetricName': 'PrecisionMacro',
                                                                 'Valu
e': 0.6182199716567993},
                                                                {'Metri
cName': 'BalancedAccuracy',
                                                                 'Set':
'Validation',
                                                                 'Stand
ardMetricName': 'BalancedAccuracy',
                                                                 'Valu
e': 0.6161699891090393},
                                                                {'Metri
cName': 'F1macro',
                                                                 'Set':
'Validation',
                                                                 'Stand
ardMetricName': 'F1macro',
                                                                 'Valu
e': 0.6169800162315369},
                                                                {'Metri
cName': 'RecallMacro',
```

```
                                                                    'Set':
'Validation',
                                                                    'Stand
ardMetricName': 'RecallMacro',
                                                                    'Valu
e': 0.6161699891090393}]},
                    'CandidateStatus': 'Completed',
                    'CandidateSteps': [{'CandidateStepArn': 'arn:aws:sag
emaker:us-east-1:822315549952:processing-job/automl-dm-1661126447-db-1-
848d0ed3feb448c9ab6491e9f701fe5b10178',
                                        'CandidateStepName': 'automl-dm-
1661126447-db-1-848d0ed3feb448c9ab6491e9f701fe5b10178',
                                        'CandidateStepType': 'AWS::SageM
aker::ProcessingJob'},
                                       {'CandidateStepArn': 'arn:aws:sag
emaker:us-east-1:822315549952:training-job/automl-dm-1661126447-dpp0-1-
472c0d3c133b4b1388609832bf05be317b1',
                                        'CandidateStepName': 'automl-dm-
1661126447-dpp0-1-472c0d3c133b4b1388609832bf05be317b1',
                                        'CandidateStepType': 'AWS::SageM
aker::TrainingJob'},
                                       {'CandidateStepArn': 'arn:aws:sag
emaker:us-east-1:822315549952:transform-job/automl-dm-1661126447-dpp0-r
pb-1-59e569b7f6ef495d9215d86e33748a5',
                                        'CandidateStepName': 'automl-dm-
1661126447-dpp0-rpb-1-59e569b7f6ef495d9215d86e33748a5',
                                        'CandidateStepType': 'AWS::SageM
aker::TransformJob'},
                                       {'CandidateStepArn': 'arn:aws:sag
emaker:us-east-1:822315549952:training-job/automl-dm-1661126447gqslp9je
4jvg-003-724e794e',
                                        'CandidateStepName': 'automl-dm-
1661126447GqsLp9je4jVG-003-724e794e',
                                        'CandidateStepType': 'AWS::SageM
aker::TrainingJob'}],
                    'CreationTime': datetime.datetime(2022, 8, 22, 0, 2
4, 54, tzinfo=tzlocal()),
                    'EndTime': datetime.datetime(2022, 8, 22, 0, 28, 42,
tzinfo=tzlocal()),
                    'FinalAutoMLJobObjectiveMetric': {'MetricName': 'val
idation:accuracy',
                                                      'Value': 0.6161699
891090393},
                    'InferenceContainers': [{'Environment': {'AUTOML_SPA
RSE_ENCODE_RECORDIO_PROTOBUF': '1',
                                                             'AUTOML_TRA
NSFORM_MODE': 'feature-transform',
                                                             'SAGEMAKER_
DEFAULT_INVOCATIONS_ACCEPT': 'application/x-recordio-protobuf',
                                                             'SAGEMAKER_
PROGRAM': 'sagemaker_serve',
                                                             'SAGEMAKER_
SUBMIT_DIRECTORY': '/opt/ml/model/code'},
                                                'Image': '683313688378.dkr.
ecr.us-east-1.amazonaws.com/sagemaker-sklearn-automl:2.5-1-cpu-py3',
                                                'ModelDataUrl': 's3://sagem
aker-us-east-1-822315549952/autopilot/automl-dm-1661126447/data-process
```

```
or-models/automl-dm-1661126447-dpp0-1-472c0d3c133b4b1388609832bf05be317
b1/output/model.tar.gz'},
                                                {'Environment': {'MAX_CONTEN
T_LENGTH': '20971520',
                                                                 'SAGEMAKER_
DEFAULT_INVOCATIONS_ACCEPT': 'text/csv',
                                                                 'SAGEMAKER_
INFERENCE_OUTPUT': 'predicted_label',
                                                                 'SAGEMAKER_
INFERENCE_SUPPORTED': 'predicted_label,probability,probabilities'},
                                                 'Image': '683313688378.dkr.
ecr.us-east-1.amazonaws.com/sagemaker-xgboost:1.3-1-cpu-py3',
                                                 'ModelDataUrl': 's3://sagem
aker-us-east-1-822315549952/autopilot/automl-dm-1661126447/tuning/autom
l-dm--dpp0-xgb/automl-dm-1661126447GqsLp9je4jVG-003-724e794e/output/mod
el.tar.gz'},
                                                {'Environment': {'AUTOML_TRA
NSFORM_MODE': 'inverse-label-transform',
                                                                 'SAGEMAKER_
DEFAULT_INVOCATIONS_ACCEPT': 'text/csv',
                                                                 'SAGEMAKER_
INFERENCE_INPUT': 'predicted_label',
                                                                 'SAGEMAKER_
INFERENCE_OUTPUT': 'predicted_label',
                                                                 'SAGEMAKER_
INFERENCE_SUPPORTED': 'predicted_label,probability,labels,probabilitie
s',
                                                                 'SAGEMAKER_
PROGRAM': 'sagemaker_serve',
                                                                 'SAGEMAKER_
SUBMIT_DIRECTORY': '/opt/ml/model/code'},
                                                 'Image': '683313688378.dkr.
ecr.us-east-1.amazonaws.com/sagemaker-sklearn-automl:2.5-1-cpu-py3',
                                                 'ModelDataUrl': 's3://sagem
aker-us-east-1-822315549952/autopilot/automl-dm-1661126447/data-process
or-models/automl-dm-1661126447-dpp0-1-472c0d3c133b4b1388609832bf05be317
b1/output/model.tar.gz'}],
                    'LastModifiedTime': datetime.datetime(2022, 8, 22,
0, 29, 35, 467000, tzinfo=tzlocal()),
                    'ObjectiveStatus': 'Succeeded'},
 'CreationTime': datetime.datetime(2022, 8, 22, 0, 5, 0, 600000, tzinfo
=tzlocal()),
 'GenerateCandidateDefinitionsOnly': False,
 'InputDataConfig': [{'ChannelType': 'training',
                      'ContentType': 'text/csv;header=present',
                      'DataSource': {'S3DataSource': {'S3DataType': 'S3
Prefix',
                                                      'S3Uri': 's3://sa
gemaker-us-east-1-822315549952/autopilot/data/womens_clothing_ecommerce
_reviews_balanced_for_autopilot.csv'}},
                      'TargetAttributeName': 'sentiment'}],
 'LastModifiedTime': datetime.datetime(2022, 8, 22, 0, 35, 16, 174000,
tzinfo=tzlocal()),
 'OutputDataConfig': {'S3OutputPath': 's3://sagemaker-us-east-1-8223155
49952/autopilot'},
 'ResolvedAttributes': {'AutoMLJobObjective': {'MetricName': 'Accurac
y'},
```

```
                        'CompletionCriteria': {'MaxAutoMLJobRuntimeInSe
conds': 7200,
                                               'MaxCandidates': 3,
                                               'MaxRuntimePerTrainingJo
bInSeconds': 1200},
                        'ProblemType': 'MulticlassClassification'},
 'ResponseMetadata': {'HTTPHeaders': {'content-length': '5634',
                                      'content-type': 'application/x-am
z-json-1.1',
                                      'date': 'Mon, 22 Aug 2022 00:35:1
6 GMT',
                                      'x-amzn-requestid': '4171a79e-739
4-4656-b949-2d48cc69d1da'},
                      'HTTPStatusCode': 200,
                      'RequestId': '4171a79e-7394-4656-b949-2d48cc69d1d
a',
                      'RetryAttempts': 0},
 'RoleArn': 'arn:aws:iam::822315549952:role/sagemaker-studio-vpc-firewa
ll-us-east-1-sagemaker-execution-role'}
```


```
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
```

```
Job status:  InProgress
Secondary job status:  GeneratingExplainabilityReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
```

```
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  InProgress
Secondary job status:  GeneratingModelInsightsReport
Job status:  Completed
Secondary job status:  Completed
[OK] Autopilot job completed.

CPU times: user 882 ms, sys: 98.6 ms, total: 980 ms
Wall time: 11min 22s
```

Before moving to the next section make sure the status above indicates `Autopilot job completed`.

6.2. Compare model candidates

Once model tuning is complete, you can view all the candidates (pipeline evaluations with different hyperparameter combinations) that were explored by AutoML and sort them by their final performance metric.

Exercise 7

List candidates generated by Autopilot sorted by accuracy from highest to lowest.

Instructions: Use `list_candidates` function passing the Autopilot job name `auto_ml_job_name` with the accuracy field `FinalObjectiveMetricValue` . It returns the list of candidates with the information about them.

```
candidates = automl.list_candidates(
    job_name=..., # Autopilot job name
    sort_by='...' # accuracy field name
)
```

In [26]:

```
candidates = automl.list_candidates(
    ### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
    job_name=auto_ml_job_name, # Replace None
    sort_by='FinalObjectiveMetricValue' # Replace None
    ### END SOLUTION - DO NOT delete this comment for grading purposes
)
```

You can review the response syntax and response elements of the function `list_candidates` in the **documentation** (https://docs.aws.amazon.com/sagemaker/latest/APIReference/API_AutoMLCandidate.html). Now let's put the candidate existence check into the loop:

In [27]:

```
while candidates == []:
    candidates = automl.list_candidates(job_name=auto_ml_job_name)
    print('[INFO] Autopilot job is generating the candidates. Please wait.')
    time.sleep(10)

print('[OK] Candidates generated.')
```

```
[OK] Candidates generated.
```

The information about each of the candidates is in the dictionary with the following keys:

In [28]:

```
print(candidates[0].keys())
```

```
dict_keys(['CandidateName', 'FinalAutoMLJobObjectiveMetric', 'Objective
Status', 'CandidateSteps', 'CandidateStatus', 'InferenceContainers', 'C
reationTime', 'EndTime', 'LastModifiedTime', 'CandidateProperties'])
```

`CandidateName` contains the candidate name and the `FinalAutoMLJobObjectiveMetric` element contains the metric information which can be used to identify the best candidate later. Let's check that they were generated.

In [29]:

```
while 'CandidateName' not in candidates[0]:
    candidates = automl.list_candidates(job_name=auto_ml_job_name)
    print('[INFO] Autopilot job is generating CandidateName. Please wait. ')
    sleep(10)

print('[OK] CandidateName generated.')
```

```
[OK] CandidateName generated.
```

In [30]:

```
while 'FinalAutoMLJobObjectiveMetric' not in candidates[0]:
    candidates = automl.list_candidates(job_name=auto_ml_job_name)
    print('[INFO] Autopilot job is generating FinalAutoMLJobObjectiveMetric. Please
wait. ')
    sleep(10)

print('[OK] FinalAutoMLJobObjectiveMetric generated.')
```

```
[OK] FinalAutoMLJobObjectiveMetric generated.
```

In [31]:

```
print(json.dumps(candidates, indent=4, sort_keys=True, default=str))
```

```
[
    {
        "CandidateName": "automl-dm-1661126447GqsLp9je4jVG-003-724e794
e",
        "CandidateProperties": {
            "CandidateArtifactLocations": {
                "Explainability": "s3://sagemaker-us-east-1-82231554995
2/autopilot/automl-dm-1661126447/documentation/explainability/output",
                "ModelInsights": "s3://sagemaker-us-east-1-82231554995
2/autopilot/automl-dm-1661126447/documentation/model_monitor/output"
            },
            "CandidateMetrics": [
                {
                    "MetricName": "Accuracy",
                    "Set": "Validation",
                    "StandardMetricName": "Accuracy",
                    "Value": 0.6161699891090393
                },
                {
                    "MetricName": "PrecisionMacro",
                    "Set": "Validation",
                    "StandardMetricName": "PrecisionMacro",
                    "Value": 0.6182199716567993
                },
                {
                    "MetricName": "BalancedAccuracy",
                    "Set": "Validation",
                    "StandardMetricName": "BalancedAccuracy",
                    "Value": 0.6161699891090393
                },
                {
                    "MetricName": "F1macro",
                    "Set": "Validation",
                    "StandardMetricName": "F1macro",
                    "Value": 0.6169800162315369
                },
                {
                    "MetricName": "RecallMacro",
                    "Set": "Validation",
                    "StandardMetricName": "RecallMacro",
                    "Value": 0.6161699891090393
                }
            ]
        },
        "CandidateStatus": "Completed",
        "CandidateSteps": [
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:processing-job/automl-dm-1661126447-db-1-848d0ed3feb448c9ab6491e
9f701fe5b10178",
                "CandidateStepName": "automl-dm-1661126447-db-1-848d0ed
3feb448c9ab6491e9f701fe5b10178",
                "CandidateStepType": "AWS::SageMaker::ProcessingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447-dpp0-1-472c0d3c133b4b138860983
```

```
2bf05be317b1",
                "CandidateStepName": "automl-dm-1661126447-dpp0-1-472c0
d3c133b4b1388609832bf05be317b1",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:transform-job/automl-dm-1661126447-dpp0-rpb-1-59e569b7f6ef495d92
15d86e33748a5",
                "CandidateStepName": "automl-dm-1661126447-dpp0-rpb-1-5
9e569b7f6ef495d9215d86e33748a5",
                "CandidateStepType": "AWS::SageMaker::TransformJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447gqslp9je4jvg-003-724e794e",
                "CandidateStepName": "automl-dm-1661126447GqsLp9je4jVG-
003-724e794e",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            }
        ],
        "CreationTime": "2022-08-22 00:24:54+00:00",
        "EndTime": "2022-08-22 00:28:42+00:00",
        "FinalAutoMLJobObjectiveMetric": {
            "MetricName": "validation:accuracy",
            "Value": 0.6161699891090393
        },
        "InferenceContainers": [
            {
                "Environment": {
                    "AUTOML_SPARSE_ENCODE_RECORDIO_PROTOBUF": "1",
                    "AUTOML_TRANSFORM_MODE": "feature-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "applicatio
n/x-recordio-protobuf",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            },
            {
                "Environment": {
                    "MAX_CONTENT_LENGTH": "20971520",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,probabilities"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-xgboost:1.3-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/tuning/automl-dm--dpp0-xgb/automl-dm-166
1126447GqsLp9je4jVG-003-724e794e/output/model.tar.gz"
            },
```

```
            {
                "Environment": {
                    "AUTOML_TRANSFORM_MODE": "inverse-label-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_INPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,labels,probabilities",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            }
        ],
        "LastModifiedTime": "2022-08-22 00:29:35.467000+00:00",
        "ObjectiveStatus": "Succeeded"
    },
    {
        "CandidateName": "automl-dm-1661126447GqsLp9je4jVG-002-e6d4689
b",
        "CandidateProperties": {
            "CandidateMetrics": [
                {
                    "MetricName": "Accuracy",
                    "Set": "Validation",
                    "StandardMetricName": "Accuracy",
                    "Value": 0.6143500208854675
                },
                {
                    "MetricName": "PrecisionMacro",
                    "Set": "Validation",
                    "StandardMetricName": "PrecisionMacro",
                    "Value": 0.6165099740028381
                },
                {
                    "MetricName": "BalancedAccuracy",
                    "Set": "Validation",
                    "StandardMetricName": "BalancedAccuracy",
                    "Value": 0.6143500208854675
                },
                {
                    "MetricName": "F1macro",
                    "Set": "Validation",
                    "StandardMetricName": "F1macro",
                    "Value": 0.6152899861335754
                },
                {
                    "MetricName": "RecallMacro",
                    "Set": "Validation",
                    "StandardMetricName": "RecallMacro",
                    "Value": 0.6143500208854675
                }
            ]
```

```
        },
        "CandidateStatus": "Completed",
        "CandidateSteps": [
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:processing-job/automl-dm-1661126447-db-1-848d0ed3feb448c9ab6491e
9f701fe5b10178",
                "CandidateStepName": "automl-dm-1661126447-db-1-848d0ed
3feb448c9ab6491e9f701fe5b10178",
                "CandidateStepType": "AWS::SageMaker::ProcessingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447-dpp0-1-472c0d3c133b4b138860983
2bf05be317b1",
                "CandidateStepName": "automl-dm-1661126447-dpp0-1-472c0
d3c133b4b1388609832bf05be317b1",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:transform-job/automl-dm-1661126447-dpp0-rpb-1-59e569b7f6ef495d92
15d86e33748a5",
                "CandidateStepName": "automl-dm-1661126447-dpp0-rpb-1-5
9e569b7f6ef495d9215d86e33748a5",
                "CandidateStepType": "AWS::SageMaker::TransformJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447gqslp9je4jvg-002-e6d4689b",
                "CandidateStepName": "automl-dm-1661126447GqsLp9je4jVG-
002-e6d4689b",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            }
        ],
        "CreationTime": "2022-08-22 00:24:46+00:00",
        "EndTime": "2022-08-22 00:26:59+00:00",
        "FinalAutoMLJobObjectiveMetric": {
            "MetricName": "validation:accuracy",
            "Value": 0.6143500208854675
        },
        "InferenceContainers": [
            {
                "Environment": {
                    "AUTOML_SPARSE_ENCODE_RECORDIO_PROTOBUF": "1",
                    "AUTOML_TRANSFORM_MODE": "feature-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "applicatio
n/x-recordio-protobuf",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            },
```

```
            {
                "Environment": {
                    "MAX_CONTENT_LENGTH": "20971520",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,probabilities"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-xgboost:1.3-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/tuning/automl-dm--dpp0-xgb/automl-dm-166
1126447GqsLp9je4jVG-002-e6d4689b/output/model.tar.gz"
            },
            {
                "Environment": {
                    "AUTOML_TRANSFORM_MODE": "inverse-label-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_INPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,labels,probabilities",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
            }
        ],
        "LastModifiedTime": "2022-08-22 00:29:35.373000+00:00",
        "ObjectiveStatus": "Succeeded"
    },
    {
        "CandidateName": "automl-dm-1661126447GqsLp9je4jVG-001-615e42a
b",
        "CandidateProperties": {
            "CandidateMetrics": [
                {
                    "MetricName": "Accuracy",
                    "Set": "Validation",
                    "StandardMetricName": "Accuracy",
                    "Value": 0.40408000349998474
                },
                {
                    "MetricName": "PrecisionMacro",
                    "Set": "Validation",
                    "StandardMetricName": "PrecisionMacro",
                    "Value": 0.4040899872779846
                },
                {
                    "MetricName": "BalancedAccuracy",
                    "Set": "Validation",
                    "StandardMetricName": "BalancedAccuracy",
                    "Value": 0.40408000349998474
```

```
                },
                {
                    "MetricName": "F1macro",
                    "Set": "Validation",
                    "StandardMetricName": "F1macro",
                    "Value": 0.40386998653411865
                },
                {
                    "MetricName": "RecallMacro",
                    "Set": "Validation",
                    "StandardMetricName": "RecallMacro",
                    "Value": 0.40408000349998474
                }
            ]
        },
        "CandidateStatus": "Completed",
        "CandidateSteps": [
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:processing-job/automl-dm-1661126447-db-1-848d0ed3feb448c9ab6491e
9f701fe5b10178",
                "CandidateStepName": "automl-dm-1661126447-db-1-848d0ed
3feb448c9ab6491e9f701fe5b10178",
                "CandidateStepType": "AWS::SageMaker::ProcessingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447-dpp1-1-dbbe8e8d135e44cb8204674
f11243c3729a",
                "CandidateStepName": "automl-dm-1661126447-dpp1-1-dbbe8
e8d135e44cb8204674f11243c3729a",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:transform-job/automl-dm-1661126447-dpp1-csv-1-bd49999d1904408697
ac04eb2f73e9c",
                "CandidateStepName": "automl-dm-1661126447-dpp1-csv-1-b
d49999d1904408697ac04eb2f73e9c",
                "CandidateStepType": "AWS::SageMaker::TransformJob"
            },
            {
                "CandidateStepArn": "arn:aws:sagemaker:us-east-1:822315
549952:training-job/automl-dm-1661126447gqslp9je4jvg-001-615e42ab",
                "CandidateStepName": "automl-dm-1661126447GqsLp9je4jVG-
001-615e42ab",
                "CandidateStepType": "AWS::SageMaker::TrainingJob"
            }
        ],
        "CreationTime": "2022-08-22 00:24:42+00:00",
        "EndTime": "2022-08-22 00:27:50+00:00",
        "FinalAutoMLJobObjectiveMetric": {
            "MetricName": "validation:accuracy",
            "Value": 0.40408000349998474
        },
        "InferenceContainers": [
            {
```

```
                "Environment": {
                    "AUTOML_TRANSFORM_MODE": "feature-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "applicatio
n/x-recordio-protobuf",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp1-1-dbbe8e8d135e44cb8204674f11243c3729a/output/model.tar.gz"
            },
            {
                "Environment": {
                    "MAX_CONTENT_LENGTH": "20971520",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,probabilities"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-xgboost:1.3-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/tuning/automl-dm--dpp1-xgb/automl-dm-166
1126447GqsLp9je4jVG-001-615e42ab/output/model.tar.gz"
            },
            {
                "Environment": {
                    "AUTOML_TRANSFORM_MODE": "inverse-label-transform",
                    "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                    "SAGEMAKER_INFERENCE_INPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                    "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,p
robability,labels,probabilities",
                    "SAGEMAKER_PROGRAM": "sagemaker_serve",
                    "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
                },
                "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/
sagemaker-sklearn-automl:2.5-1-cpu-py3",
                "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/
autopilot/automl-dm-1661126447/data-processor-models/automl-dm-16611264
47-dpp1-1-dbbe8e8d135e44cb8204674f11243c3729a/output/model.tar.gz"
            }
        ],
        "LastModifiedTime": "2022-08-22 00:29:35.374000+00:00",
        "ObjectiveStatus": "Succeeded"
    }
]
```


You can print the names of the candidates with their metric values:

In [32]:

```
print("metric" + str(candidates[0]['FinalAutoMLJobObjectiveMetric']['MetricName']))

for index, candidate in enumerate(candidates):
    print(str(index) + "  "
        + candidate['CandidateName'] + "  "
        + str(candidate['FinalAutoMLJobObjectiveMetric']['Value']))
```

```
metric validation:accuracy
0  automl-dm-1661126447GqsLp9je4jVG-003-724e794e  0.6161699891090393
1  automl-dm-1661126447GqsLp9je4jVG-002-e6d4689b  0.6143500208854675
2  automl-dm-1661126447GqsLp9je4jVG-001-615e42ab  0.40408000349998474
```

6.3. Review best candidate

Now that you have successfully completed the Autopilot job on the dataset and visualized the trials, you can get the information about the best candidate model and review it.

Exercise 8

Get the information about the generated best candidate job.

Instructions: Use `best_candidate` function passing the Autopilot job name. This function will give an error if candidates have not been generated.

In [33]:

```
candidates = automl.list_candidates(job_name=auto_ml_job_name)

if candidates != []:
    best_candidate = automl.best_candidate(
        ### BEGIN SOLUTION - DO NOT delete this comment for grading purposes
        job_name=auto_ml_job_name # Replace None
        ### END SOLUTION - DO NOT delete this comment for grading purposes
    )
    print(json.dumps(best_candidate, indent=4, sort_keys=True, default=str))
```

```
{
    "CandidateName": "automl-dm-1661126447GqsLp9je4jVG-003-724e794e",
    "CandidateProperties": {
        "CandidateArtifactLocations": {
            "Explainability": "s3://sagemaker-us-east-1-822315549952/au
topilot/automl-dm-1661126447/documentation/explainability/output",
            "ModelInsights": "s3://sagemaker-us-east-1-822315549952/aut
opilot/automl-dm-1661126447/documentation/model_monitor/output"
        },
        "CandidateMetrics": [
            {
                "MetricName": "Accuracy",
                "Set": "Validation",
                "StandardMetricName": "Accuracy",
                "Value": 0.6161699891090393
            },
            {
                "MetricName": "PrecisionMacro",
                "Set": "Validation",
                "StandardMetricName": "PrecisionMacro",
                "Value": 0.6182199716567993
            },
            {
                "MetricName": "BalancedAccuracy",
                "Set": "Validation",
                "StandardMetricName": "BalancedAccuracy",
                "Value": 0.6161699891090393
            },
            {
                "MetricName": "F1macro",
                "Set": "Validation",
                "StandardMetricName": "F1macro",
                "Value": 0.6169800162315369
            },
            {
                "MetricName": "RecallMacro",
                "Set": "Validation",
                "StandardMetricName": "RecallMacro",
                "Value": 0.6161699891090393
            }
        ]
    },
    "CandidateStatus": "Completed",
    "CandidateSteps": [
        {
            "CandidateStepArn": "arn:aws:sagemaker:us-east-1:8223155499
52:processing-job/automl-dm-1661126447-db-1-848d0ed3feb448c9ab6491e9f70
1fe5b10178",
            "CandidateStepName": "automl-dm-1661126447-db-1-848d0ed3feb
448c9ab6491e9f701fe5b10178",
            "CandidateStepType": "AWS::SageMaker::ProcessingJob"
        },
        {
            "CandidateStepArn": "arn:aws:sagemaker:us-east-1:8223155499
52:training-job/automl-dm-1661126447-dpp0-1-472c0d3c133b4b1388609832bf0
5be317b1",
            "CandidateStepName": "automl-dm-1661126447-dpp0-1-472c0d3c1
```

```
33b4b1388609832bf05be317b1",
            "CandidateStepType": "AWS::SageMaker::TrainingJob"
        },
        {
            "CandidateStepArn": "arn:aws:sagemaker:us-east-1:8223155499
52:transform-job/automl-dm-1661126447-dpp0-rpb-1-59e569b7f6ef495d9215d8
6e33748a5",
            "CandidateStepName": "automl-dm-1661126447-dpp0-rpb-1-59e56
9b7f6ef495d9215d86e33748a5",
            "CandidateStepType": "AWS::SageMaker::TransformJob"
        },
        {
            "CandidateStepArn": "arn:aws:sagemaker:us-east-1:8223155499
52:training-job/automl-dm-1661126447gqslp9je4jvg-003-724e794e",
            "CandidateStepName": "automl-dm-1661126447GqsLp9je4jVG-003-
724e794e",
            "CandidateStepType": "AWS::SageMaker::TrainingJob"
        }
    ],
    "CreationTime": "2022-08-22 00:24:54+00:00",
    "EndTime": "2022-08-22 00:28:42+00:00",
    "FinalAutoMLJobObjectiveMetric": {
        "MetricName": "validation:accuracy",
        "Value": 0.6161699891090393
    },
    "InferenceContainers": [
        {
            "Environment": {
                "AUTOML_SPARSE_ENCODE_RECORDIO_PROTOBUF": "1",
                "AUTOML_TRANSFORM_MODE": "feature-transform",
                "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "application/x-
recordio-protobuf",
                "SAGEMAKER_PROGRAM": "sagemaker_serve",
                "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
            },
            "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/sage
maker-sklearn-automl:2.5-1-cpu-py3",
            "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/auto
pilot/automl-dm-1661126447/data-processor-models/automl-dm-1661126447-d
pp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
        },
        {
            "Environment": {
                "MAX_CONTENT_LENGTH": "20971520",
                "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,proba
bility,probabilities"
            },
            "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/sage
maker-xgboost:1.3-1-cpu-py3",
            "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/auto
pilot/automl-dm-1661126447/tuning/automl-dm--dpp0-xgb/automl-dm-1661126
447GqsLp9je4jVG-003-724e794e/output/model.tar.gz"
        },
        {
            "Environment": {
```

```
                "AUTOML_TRANSFORM_MODE": "inverse-label-transform",
                "SAGEMAKER_DEFAULT_INVOCATIONS_ACCEPT": "text/csv",
                "SAGEMAKER_INFERENCE_INPUT": "predicted_label",
                "SAGEMAKER_INFERENCE_OUTPUT": "predicted_label",
                "SAGEMAKER_INFERENCE_SUPPORTED": "predicted_label,proba
bility,labels,probabilities",
                "SAGEMAKER_PROGRAM": "sagemaker_serve",
                "SAGEMAKER_SUBMIT_DIRECTORY": "/opt/ml/model/code"
            },
            "Image": "683313688378.dkr.ecr.us-east-1.amazonaws.com/sage
maker-sklearn-automl:2.5-1-cpu-py3",
            "ModelDataUrl": "s3://sagemaker-us-east-1-822315549952/auto
pilot/automl-dm-1661126447/data-processor-models/automl-dm-1661126447-d
pp0-1-472c0d3c133b4b1388609832bf05be317b1/output/model.tar.gz"
        }
    ],
    "LastModifiedTime": "2022-08-22 00:29:35.467000+00:00",
    "ObjectiveStatus": "Succeeded"
}
```

Check the existence of the candidate name for the best candidate.

In [34]:

```
while 'CandidateName' not in best_candidate:
    best_candidate = automl.best_candidate(job_name=auto_ml_job_name)
    print('[INFO] Autopilot Job is generating BestCandidate CandidateName. Please w
ait. ')
    print(json.dumps(best_candidate, indent=4, sort_keys=True, default=str))
    sleep(10)

print('[OK] BestCandidate CandidateName generated.')
```

```
[OK] BestCandidate CandidateName generated.
```

Check the existence of the metric value for the best candidate.

In [35]:

```
while 'FinalAutoMLJobObjectiveMetric' not in best_candidate:
    best_candidate = automl.best_candidate(job_name=auto_ml_job_name)
    print('[INFO] Autopilot Job is generating BestCandidate FinalAutoMLJobObjective
Metric. Please wait. ')
    print(json.dumps(best_candidate, indent=4, sort_keys=True, default=str))
    sleep(10)

print('[OK] BestCandidate FinalAutoMLJobObjectiveMetric generated.')
```

```
[OK] BestCandidate FinalAutoMLJobObjectiveMetric generated.
```

Print the information about the best candidate:

In [36]:

```
best_candidate_identifier = best_candidate['CandidateName']
print("Candidate name: " + best_candidate_identifier)
print("Metric name: " + best_candidate['FinalAutoMLJobObjectiveMetric']['MetricNam
e'])
print("Metric value: " + str(best_candidate['FinalAutoMLJobObjectiveMetric']['Valu
e']))
```

```
Candidate name: automl-dm-1661126447GqsLp9je4jVG-003-724e794e
Metric name: validation:accuracy
Metric value: 0.6161699891090393
```

7. Review all output in S3 bucket

You will see the artifacts generated by Autopilot including the following:

```
   data-processor-models/        # "models" learned to transform raw data into
    features
   documentation/                # explainability and other documentation about
   your model
   preprocessed-data/            # data for train and validation
   sagemaker-automl-candidates/  # candidate models which autopilot compares
   transformed-data/             # candidate-specific data for train and valida
   tion
   tuning/                       # candidate-specific tuning results
   validations/                  # validation results
```

In [37]:

```
from IPython.core.display import display, HTML

display(
    HTML(
        '<b>Review all <a target="blank" href="https://s3.console.aws.amazon.com/s
3/buckets/{}?region={}&prefix=autopilot/{}/">output in S3</a></b>'.format(
            bucket, region, auto_ml_job_name
        )
    )
)
```

Review all [output in S3 (https://s3.console.aws.amazon.com/s3/buckets/sagemaker-us-east-1-822315549952?region=us-east-1&prefix=autopilot/automl-dm-1661126447/)](https://s3.console.aws.amazon.com/s3/buckets/sagemaker-us-east-1-822315549952?region=us-east-1&prefix=autopilot/automl-dm-1661126447/)

8. Deploy and test best candidate model

8.1. Deploy best candidate model

While batch transformations are supported, you will deploy our model as a REST Endpoint in this example.

First, you need to customize the inference response. The inference containers generated by SageMaker Autopilot allow you to select the response content for predictions. By default the inference containers are configured to generate the `predicted_label`. But you can add `probability` into the list of inference response keys.

In [38]:

```
inference_response_keys = ['predicted_label', 'probability']
```

Now you will create a SageMaker endpoint from the best candidate generated by Autopilot. Wait for SageMaker to deploy the endpoint.

This cell will take approximately 5-10 minutes to run.

In [39]:

```
autopilot_model = automl.deploy(
    initial_instance_count=1,
    instance_type='ml.m5.large',
    candidate=best_candidate,
    inference_response_keys=inference_response_keys,
    predictor_cls=sagemaker.predictor.Predictor,
    serializer=sagemaker.serializers.JSONSerializer(),
    deserializer=sagemaker.deserializers.JSONDeserializer()
)

print('\nEndpoint name:  {}'.format(autopilot_model.endpoint_name))
```

```
---------------!
Endpoint name:  sagemaker-sklearn-automl-2022-08-22-00-49-09-018
```

Please wait until the ^^ endpoint ^^ is deployed.

Review the SageMaker endpoint in the AWS console.

In [40]:

```
from IPython.core.display import display, HTML

display(HTML('<b>Review <a target="blank" href="https://console.aws.amazon.com/sagemaker/home?region={}#/endpoints/{}">SageMaker REST endpoint</a></b>'.format(region, autopilot_model.endpoint_name)))
```

Review [SageMaker REST endpoint (https://console.aws.amazon.com/sagemaker/home?region=us-east-1#/endpoints/sagemaker-sklearn-automl-2022-08-22-00-49-09-018)]

Invoke a few predictions for the actual reviews using the deployed endpoint.

In [41]:

```
#sm_runtime = boto3.client('sagemaker-runtime')

review_list = ['This product is great!',
               'OK, but not great.',
               'This is not the right product.']

for review in review_list:

    # remove commas from the review since we're passing the inputs as a CSV
    review = review.replace(",", "")

    response = sm_runtime.invoke_endpoint(
        EndpointName=autopilot_model.endpoint_name, # endpoint name
        ContentType='text/csv', # type of input data
        Accept='text/csv', # type of the inference in the response
        Body=review # review text
        )

    response_body=response['Body'].read().decode('utf-8').strip().split(',')

    print('Review: ', review, ' Predicated class: {}'.format(response_body[0]))

print("(-1 = Negative, 0=Neutral, 1=Positive)")
```

```
Review:  This product is great!  Predicated class: 1
Review:  OK but not great.  Predicated class: 0
Review:  This is not the right product.  Predicated class: -1
(-1 = Negative, 0=Neutral, 1=Positive)
```

You used Amazon SageMaker Autopilot to automatically find the best model, hyper-parameters, and feature-engineering scripts for our dataset. Autopilot uses a uniquely-transparent approach to AutoML by generating re-usable Python scripts and notebooks.

Upload the notebook into S3 bucket for grading purposes.

Note: you may need to click on "Save" button before the upload.

In [42]:

```
!aws s3 cp ./C1_W3_Assignment.ipynb s3://$bucket/C1_W3_Assignment_Learner.ipynb
```

```
upload: ./C1_W3_Assignment.ipynb to s3://sagemaker-us-east-1-8223155499
52/C1_W3_Assignment_Learner.ipynb
```

Please go to the main lab window and click on `Submit` button (see the `Finish the lab` section of the instructions).

In [ ]: