

# **Movies Recommender by Personalize**

Hands on Lab

**March 2019** 

#### Machine Learning and Artificial Intelligence

# **Overview**

Amazon Personalize is a machine learning service that makes it easy for developers to create individualized recommendations for customers using their applications.

With Amazon Personalize, you provide an activity stream from your application – page views, signups, purchases, and so forth – as well as an inventory of the items you want to recommend, such as articles, products, videos, or music. You can also choose to provide Amazon Personalize with additional demographic information from your users such as age, or geographic location. Amazon Personalize will process and examine the data, identify what is meaningful, select the right algorithms, and train and optimize a personalization model that is customized for your data.

## Machine Learning and Artificial Intelligence

# **Movies Recommender with Personalize**

In this lab you will build a movies recommender with Amazon Personalize. The problem statement is, given a historical rating of movies by users, the system should be able to recommend movies that a particular user may like.

Amazon Personalize, as a managed recommendation engine service by AWS, will be used to create the data schema, build data set, import data, train the engine, deploy the solution, and host the recommendation inference.

#### Machine Learning and Artificial Intelligence

## **Prepare data**

{

Please set-up an AWS SageMaker Notebook Instance, open Jupyter Notebook, and complete Step PS1 until Step PS3 in <a href="https://github.com/yudho/machine-learning-workshop/blob/master/recommendation-engine-with-personalize/personalize sample notebook.ipynb">https://github.com/yudho/machine-learning-workshop/blob/master/recommendation-engine-with-personalize/personalize sample notebook.ipynb</a>

## **Prepare S3 bucket access**

- 1. Open S3 https://s3.console.aws.amazon.com/s3/home?region=us-east-1
- 2. Take note of the S3 bucket name that ends with "-product-recommendation-personalize". We will use it later.
- 3. Click the S3 bucket name
- 4. Go to "Permissions" tab, and click "Bucket Policy".
- 5. Paste this policy into the Bucket policy editor, change the bolded "<s3 bucket name>" to the name of this S3 bucket and click Save:

#### Machine Learning and Artificial Intelligence

## Create dataset group, dataset, schema, and import job

- 1. Open Amazon Personalize console in N.Virginia region by following <a href="https://console.aws.amazon.com/personalize/home?region=us-east-1">https://console.aws.amazon.com/personalize/home?region=us-east-1</a>
- 2. If it displays a homepage (since you do not have any Amazon Personalize resources in that region), then click "Get Started"
- 3. The Create dataset group page should be displayed
- 4. Enter this detail and click Next
  - a. Dataset group name: movie100-dataset-group
  - b. Input data: upload user-item interaction data
- 5. This page below will be displayed. Enter detail:
  - a. Dataset name: ratings
  - b. Schema selection: Create new schema
  - c. New schema name: movie100-schema

#### Machine Learning and Artificial Intelligence

## **Dataset details**

#### Dataset name

The text you type here appears in the Dataset dashboard and detail page. It can help you distinguish this dataset from others.

#### ratings

The dataset name must have 2 to 256 characters. Valid characters: a-z, A-Z, 0-9, and .: + = @ % - (hyphen).

# Schema details

#### Schema selection Info



Use existing schema Choose an existing schema that matches your dataset.



## Create new schema

Create a new schema to match your dataset

#### New schema name

The text you type here appears in the Schema dashboard. It can help you distinguish this schema from others.

#### movie100-schema

The schema name must have 2 to 256 characters. Valid characters: a-z, A-Z, 0-9, and .: + = @ % - (hyphen).

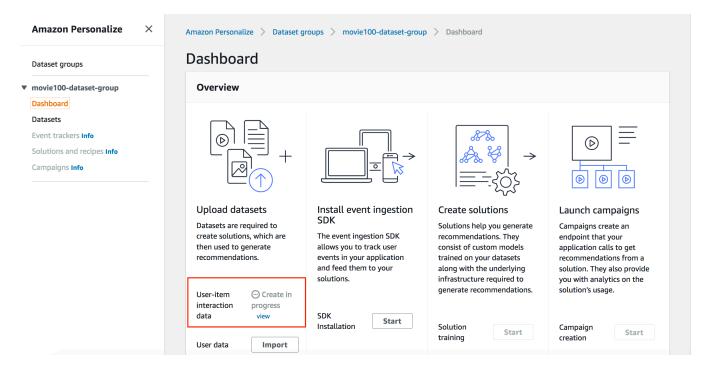
6. For Schema definition, enter this JSON document below. Then click Next

```
{
"type": "record",
"name": "Interactions",
"namespace": "com.amazonaws.personalize.schema",
"fields": [
  {
    "name": "USER_ID",
    "type": "string"
```

#### Machine Learning and Artificial Intelligence

- 7. The Import user-item interaction data should be displayed. Otherwise, an error might happen. In that case just wait for a while and repeat step 1-6, except that, instead of creating new schema, select an existing one.
- 8. Fill in this detail and click Next:
  - a. Dataset import job name: movie100-dataset-import-job
  - b. Automatic dataset import jobs: Off
  - c. IAM service role:
    - i. click 'Create a new role' and when prompted, select 'Specific S3 buckets' under 'S3 buckets you specify' dialog, and for the value, paste the bucket name that you have taken note in step 2 in "Prepare S3 bucket access" section above
  - d. Data location: s3://<bucket name>/DEMO-movie-lens-100k.csv
    - i. Please substitute <bucket\_name> with the name of your S3 bucket from step 2 in "Prepare S3 bucket access" section above
- 9. You will be directed to the movie100-dataset-group Dashboard page. Please wait until dataset finish importing

#### Machine Learning and Artificial Intelligence



#### **Create solution**

After we have all the data-related stuff, now we need to train the recommender engine with those ratings data we have. In Amazon Personalize, creating solution means training the engine.

In this step, we can choose either to train with specific recipe/algorithm (e.g. HRNN), or do AutoML and let the Personalize organize the training with multiple recipes/algorithms and choose the best one.

- 1. On movie100-dataset-group Dashboard page, click "Start" on Solution training
- 2. For Solution name, enter movie100-solution
- 3. For Recipe selection, choose **Automatic (AutoML)**.
- 4. For Event type and Minimum transactions per second, leave the default values.
- 5. Click Finish

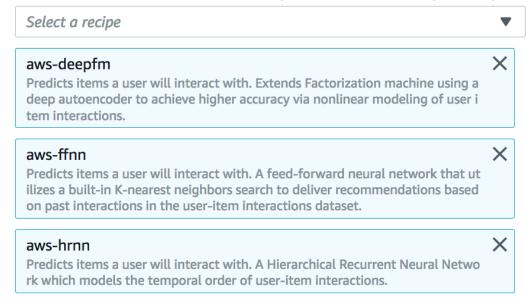
## Machine Learning and Artificial Intelligence

Amazon Personalize > Dataset groups > movie100-dataset-group > Solutions > Create solution Create solution Info You create a solution using a recipe that is tailored to a specific use case. **Solution configuration** Solution name The solution name that you type here can help you distinguish this solution from other solutions. movie100-solution The solution name must have 2 to 256 characters. Valid characters: a-z, A-Z, 0-9, and .: + = @ % - (hyphen). Minimum provisioned transactions per second Info The minimum amount of throughput in transactions per second (TPS) that is provisioned for this solution. Type in a value between 1 and 500. Recipe selection Info A recipe consists of recommendation algorithms and data processing steps needed to create a solution. Amazon Personalize has several prebuilt recipes. Manual Choose the recipe manually. Automatic (AutoML) Amazon Personalize will find the best recipe for your dataset. AutoML recipe list Select candidate recipes that you would like Amazon Personalize to examine and determine the best recipe. By default Amazon Personalize will train on DeepEM FENN and HRNN recipes when performing AutoMI

#### Machine Learning and Artificial Intelligence

#### AutoML recipe list

Select candidate recipes that you would like Amazon Personalize to examine and determine the best recipe. If default, Amazon Personalize will train on DeepFM, FFNN and HRNN recipes when performing AutoML.



## Launch campaign

After we have trained the engine, we should deploy our recommender engine to a server/container. In Amazon Personalize, we can launch campaign to actually deploy the trained solution.

- 1. On movie100-dataset-group Dashboard page, click 'Start' on Campaign creation
- 2. For Campaign name, enter movie100-campaign
- 3. For Solution, select the solution that we created in the section 'Create solution' above
- 4. Click Create campaign

#### **Get recommendation**

Now, it is time to test our recommendation engine by making inference.

- 1. Go to https://docs.aws.amazon.com/personalize/latest/dg/getting-started-console.html
- 2. Expand the section 'Step 4: Get Recommendations'
- 3. Follow the instructions in there to test the recommendation
- 4. Good, we see some IDs, but what movies are they? Let's follow the steps below to get commendation from SageMaker Notebook Instance where we can define the movie titles mappings
- 5. From AWS Personalize console, locate the campaign under movie100-dataset-group.
- 6. Identify the ARN of the campaign and note it down
- 7. Go back to the SageMaker Notebook that we used earlier in this lab and go straight to Step PS11.

## Machine Learning and Artificial Intelligence

- 8. Run 'Select a user' step
- 9. Under 'Call GetRecomendations' section, change the campaignArn value to the ARN you noted in step 6 above, enclosed in single quote to make it a string
- 10. Run that section
- 11. You should see the list of recommended movies for that user