Watson Studio – Tweet Analysis Workshop

Prerequisites: IBM Cloud Setup, and a Watson Studio service instance. See the Setup and Installation Guide for more details.

Over the course of this workshop, you will learn how to create an end-to-end machine learning project on IBM Cloud. We will start with data preparation using Jupyter notebooks in IBM Watson Studio. Then move on to data analysis using the Watson Discovery service. For further refinement and modelling we will use Watson Knowledge Studio. Finally, the data visualization will be completed using the dashboard tool in Watson Studio. Through this workshop, you will be using many services with the goal of mapping all the workplace related issues found from Twitter on a world map.

The sample data needed for this tutorial is listed below:

- **cxe_2019_tweets.csv**
- **Tweet Analysis - Workshop.ipynb**
- **wks_dictionary.zip**
- **wks_types.json**
- **wks_documents.zip**

You should be able to complete this workshop in 2 hours 20 minutes, depending on loading and running times. To get started, open the IBM Cloud site at https://cloud.ibm.com and sign in.

Part 1: Setup the project environment.

Step 1 - Create a project.

1. From the homepage of Watson Studio, click **Create a Project**.

Normally, your data for this project would be acquired by creating a Twitter Developer Account and using Twitter's API to pull the data directly from the Jupyter notebooks. However, to simplify the process, we have already downloaded the data from Twitter and saved in into a CSV file. To locate and upload the file, please follow the instructions below.

2. Select the **Create an empty project** option, then name your project and click **Create**.

Note: You can leave the description blank, its optional to fill in

Step 2 - Load the data asset.

1. Click on the **Assets** tab.
2. Drag **cxe_2019_tweets.csv** into the side bar or click browse and select it.

Step 3 - Create a Jupyter Notebook.

1. Click on **Add to project**, then choose the **Notebook** card.
2. In the New notebook page navigate to the **From File** tab.
3. Click **Choose file** and select the "**Tweet Analysis - Workshop.ipynb**" file.
4. Select the runtime as **Default Python 3.6 Free (1 vCPU and 4 GB RAM).**
5. Click on **Create Notebook.**

Part 2: Working with the notebook.

Step 1 - Create and insert project token.

1. Click on the 3 dots (image) symbol and then click **Insert project token**.
2. From the popup, click on the **project name** in the message.
3. You will be redirect to the settings tab, scroll down and locate access token and click on **new token**.
4. On the confirm creation pop make the options listed here:

- Name (E.g Mytoken)
- Access role: Editor

5. Click on **Create**.
6. Navigate back to the notebook, close the popup and click the **3 dots (image) symbol**.
7. A project token has been added to the top of the notebook as a hidden cell.
8. Run the hidden cell to add the project token in the scope of your project.

Step 2 - Importing libraries.

1. Run the cell under section **1.1. Importing Libraries** in the notebook.

We will be skipping sections two and three because we have already downloaded the data form Twitter. However, if you are interested in seeing how the data was acquired then you can view steps two and three on the Jupyter Notebook.

2. Run Cell 2/3 **Alternative – Import Workshop Twitter Data.** This will import the data asset in your project folder.

Part 3: Setting up Watson Discovery.

1. Open https://cloud.ibm.com/catalog/services/discovery
2. Choose the **Lite** plan and click **Create**.
3. You will be redirected to the **Resource list page**. Click on the **Discovery** service when it is done provisioning.
4. Click on **Launch Watson Discovery**.

Watson Discovery will open in a new tab, but do not close the previous tab as it contains an **API Key** that you will need later in the workshop.

5. Select **Upload your own Data**.
6. Name your collection and click **Create**.
7. Click on the **API** button in the top right corner.

Here, you will see a collection ID, configuration ID, and Environment ID. We will be adding these IDs along with the API's key in the code.

Part 4: Data analysis using Watson Discovery.

1. We will go back to the notebook and run the cells starting from 4, the batch cell allows us to access Watson discovery in the scope of the notebook.
2. The next cell 4.1 converts each tweet to json. The discovery service will then extract sentiment from these json tweets.
3. Here we import our credentials from Watson discovery to establish the connection in cell 4.2. This is in reference to **Part 3 Step 7** of this workshop.
4. The next cells in 4.2 pushes the tweets into the discovery collection we created before.
5. Running cell 5 will query the discovery collect to get back all the enriched and analyzed tweets from Discovery.
6. Next, we will apply our first filter to **extract non-positive tweets** by running cell 5.1.
7. Similarly, apply the second filter to **extract relevant tweets by category and content** by running cell 5.2
8. The last filter we apply is to **extract tweets with category confidence score above 70%** by running cell 5.3.
9. Lastly, we will **remove signs such as "@" and "#"** to clean the content of the tweets by running cell 5.4.
10. Now we will move onto Knowledge Studio to create our own custom model.

Part 5: Knowledge Studio

Step 1 - Setup

1. Navigate to https://cloud.ibm.com/catalog/services/knowledge-studio
2. Select the **Lite** plan and click **Create**.
3. You will be directed to the Knowledge studio Getting Started tab. **In the side bar** click on **Manage.**
4. Click **Launch Watson Knowledge Studio**.

Step 2: Uploading assets

1. Click **Create entities and relations workspace.**
2. Name the workspace and click **Create**.
3. Click on **Upload** and select **wks_types.json.**
4. In the **left-hand side sidebar** under **assets**, click on **Documents**.
5. Click on **Upload Document Sets**.
6. Select **wks_documents.zip**; click **Upload**.

Note: Make sure the upload corpus documents option is checked.

7. In the left-hand side sidebar under assets, click on **Dictionaries**.
8. click the 3 dots (image) symbol beside the Create Dictionaries.
9. Click on **Upload Dictionaries**.
10. Select **wks_dictionaries.zip** and click **Upload**.

Step 3: Train and evaluate

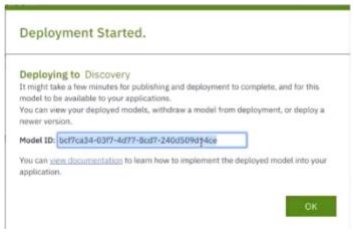
1. In the sidebar click on **Machine Learning Model**.
2. Click on **Performance**, then click on **Train and evaluate**.
3. Ensure that all the files are selected and click on **Train and evaluate**.
4. Wait for the model to finish training. This can take a few minutes.

After Training & Evaluate is done you can go into View Ground Truth to see in depth what the model did.

Step 4: Export the model

1. In the left sidebar**, under Machine Learning Model**, click on **Versions**.
2. Click on **Create version**.
3. A new version will be created below. Click on **Deploy.**
4. Select **Discovery** as the service to deploy and click **Next.**
5. Under **Space or resource group**, select **resource groups.**
6. Select **Default** and your discovery service name.

7. Click on **Deploy.**
8. After we select **Deploy**, we will get the following pop up (see image). Make sure to save Model ID as we will be using it in the future and select OK.



Note: Make sure you copy the Model ID and save it to your notebook.

Part 6 - Adding custom model to the discovery collection.

1. Navigate back to Discovery by going to https://cloud.ibm.com/resources and selecting the **Discovery service you created in part 3**.
2. Click on **Upload your own data**.
3. Name your collection and click **Create**.
4. Click on the api button in the top right corner and keep those values for reference.
5. Copy and paste the **Collection ID**, **Configuration ID**, and **Environment ID** into cell 7.0

Note: The api key is the same one you have entered in cell 4.2

6. We will then run the next cell to **create a JSON for each tweet**.
7. The following cells 7.2 will **send the tweet data to the default model**.
8. Send a **Custom Model to Discovery** by running the cells under section 7.3.

Running the cell below will ensure that there we are not working with more documents than allowed in the Lite plan.

Now that the custom model has been added, follow the instructions to add our trained Watson Knowledge Studio model to the custom configuration. ¶

In [ ]:
```
# Since the documents we just uploaded were enriched using the default model, we will delete them and upload again.
query = discovery_custom.query(environment_id=env_id_custom, collection_id=col_id_custom ,query='*.*', count=50)

for doc in query.result['results']:
    delete_doc = discovery_custom.delete_document(env_id_custom, col_id_custom, doc['id']).get_result()
```

9. Navigate back to Discovery.

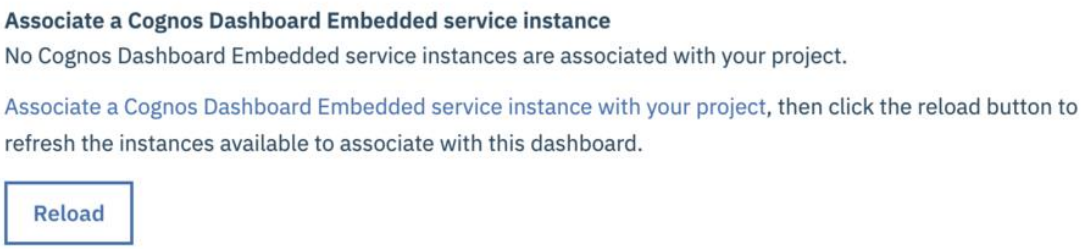
10. Select the new collection and scroll down to the bottom to select **Add enrichments**.
11. Click on **Add enrichments.**
12. Locate **Entity Extraction** and add the **model ID** that we got **from part 5 step 4.8**.
13. Close the popup window and select **Apply changes to collection** in the top right corner.
14. Back in the notebook, run the cells under section 7.4 to **send the Data to the Custom Model**.
15. Run the cells under section 7.5 to **analyze the Enriched Data from Custom Model**.
16. Run the cells under section 7.6 to **merge the datasets and Deriving the Correct Location Country**.

Part 7: Creating the dashboard visualization

Step 1: Add a dashboard service

1. Navigate back to the project folder and click on **add to project**.
2. Select the **dashboard** card.
3. Fill in the name: "**Map**"
4. Select the **cognos embedded service** (for e.g Cognos-dashboard-embedded-jp)

If you do not have a cognos embedded service, follow the steps 5-9



5. Click on **Associate a Cognos Dashboard Embedded service instance with your project**.
6. A new window will open; click on the **new** tab.
7. Select the **lite** plan and click on **create**.
8. In the **confirm creation** popup, double-check the following settings:
    - Region: Dallas

- Plan: Lite
- Resource group: Default.
- Service Name: Auto-Generated

9. Click the **Confirm** button.
10. Back in the **dashboard creation** page, click on **reload**.
11. Select your service from the list; click on **Save**.

Step 2 - Add the data source.

1. On the main dashboard, select the following options:
    a. Dashboard: tabbed
    b. tabbed layout: freeform
2. Click **Ok.**
3. On the new page, we will **add our data source** by click the **plus symbol** next to selected sources.
4. Click on **Data assets** to expand the list.
5. Select the data asset **new_merged_data.csv**.
6. Click on **Select**.

Step 3: Add the map visualization

1. Click on the **visualization** button from the left-hand tab.
2. Drag **the map visualization** into the canvas.
3. Click on the **data asset** to expand the list.
4. Drag and the drop the following columns into the fields.
    a. Locations: wks_location
    b. Location color: wks_location
5. Click on the **collapse** button on the top right-hand corner on the visualization.

Step 4: Change the map visualization theme

1. Select the **visualization** and click on the **properties** button on the **toolbar**.
2. From the style list, select the **dark theme**.
3. Uncheck the **display legend option**.

Step 5: Change the canvas theme.

1. Click on the **properties** button on the **toolbar**.
2. From the **visual themes**, select the **dark theme**.

Step 6: Add the companies word cloud visualization

1. Click on the **visualization** button from the left-hand tab.
2. Drag the **word cloud** visualization into the canvas.
3. Click on the data asset to expand the list.
4. Drag and the drop the following columns into the fields.
    a. Words: wks_Company
    b. Color: wks_Company
5. Click on the **collapse** button on the top right-hand corner on the visualization.

Step 7: Add the categories word cloud visualization

1. Click on the **visualization** button from the left-hand tab.
2. Drag the **word cloud** visualization into the canvas.
3. Click on the data asset to expand the list.
4. Drag and the drop the following columns into the fields.
    a. Words: wks_Categories
    b. Color: wks_Categories
5. Click on the **collapse** button on the top right-hand corner on the visualization.

Do Step 8 for both word cloud visualizations.

Step 8 - Change the word cloud theme

1. Select the **visualization** and click on the **properties** button on the **toolbar**.
2. From the word orientation, select the **horizontal** option.

Step 9 - Add the summary visualization

1. Click on the **visualization** button from the left-hand tab.
2. Drag the **summary** visualization into the canvas.
3. Click on the data asset to expand the list.
4. Drag and the drop the following columns into the fields.

   a. #Value: tid

5. Click on the **collapse** button on the top right-hand corner on the visualization.

Step 10: Add the list visualization

1. Click on the **visualization** button from the left-hand tab.
2. Drag the **List** visualization into the canvas.
3. Click on the data asset to expand the list.
4. Drag and the drop the following columns into the fields.

   a. Level One: Text

5. Click on the **collapse** button on the top right-hand corner on the visualization.