

Utilities Customer Attrition Prediction

Introduction

The Utilities Customer Attrition Prediction accelerator includes a structured glossary of business terms, a set of sample data science assets, and a sample dashboard to visualize the results. The glossary provides the information architecture that you need to understand why customers leave. Your data scientists can use the sample notebooks, predictive model, and dashboard to accelerate data preparation, machine learning modeling, and data reporting. Understand the likelihood of Customer Attrition occurring & analyse the business metrics influencing the Attrition.

Tip: Download the PDF of these instructions from the Data assets section on the **Assets** page so you can keep these instructions open while you work.

- [Instructions](#)
- [Sample data assets](#)
- [Notebooks](#)
- [R Shiny dashboard](#)
- [Sample business glossary](#)

Instructions

Follow these steps to implement the industry accelerator:

1. Create a deployment space, which you'll need when you run the notebooks. From the navigation menu, select **Deployments > View All Spaces > New deployment space** and then complete these steps:
 1. Enter the name **Utilities Customer Attrition Space**.
 2. Select your Watson Machine Learning service. If you don't have one, from the **Select machine learning service** pull-down menu, select **Create a new machine learning service**, provision one, and then click the **Create** button.
 3. Click **Create** to create the deployment space.
2. Navigate back to the accelerator project, select the **Assets** tab and scroll to the **Notebooks** section.
3. Edit the **1-model-training** notebook. This notebook prepares the data, builds ML models, and deploys the model. Follow the instructions in the notebook to configure and step through the execution.
4. Edit and run the **2-model-scoring** notebook. This notebook deploys data assets and a model scoring function.
5. Run the dashboard from RStudio console by completing these steps:
 1. Download the **utilities-customer-attrition-analytics-dashboard.zip** file from the Data assets section of the **Assets** page. If you don't see the file, click **View All** to display the full list of assets.
 2. Click **Launch IDE > RStudio** on the menu bar.
 3. In the **Files** pane, select the **Upload** toolbar button and upload the **utilities-customer-attrition-analytics-dashboard.zip** file into RStudio.

4. Navigate to `utilities-customer-attribution-analytics-dashboard`, select the app.R file, and click the **Run App** toolbar button to launch the dashboard. If you see a warning message that certain packages are not installed, you can ignore it because the packages will be installed first time you run the app.
5. Once the app has launched, you can perform model scoring in real time by entering your API key and selecting your region on the **Client View** tab.
6. Optional. You can import the glossary of business terms into Watson Knowledge Catalog to get started on data governance. With the **Lite plan** only 5 Business terms can be imported. You must have [the permission to create governance artifacts](#).
 1. Download the `utilities-customer-attribution-glossary-categories.csv` and `utilities-customer-attribution-glossary-terms.csv` files from the Data assets section of the **Assets** page.
 2. Navigate to **Governance > Categories**.
 3. Click **Add Category > Import From File**.
 4. Import the `utilities-customer-attribution-glossary-categories.csv` file. Select **Replace all values** as your merge option.
 5. Navigate to **Governance > Business Terms**.
 6. Click **Add Business Term > Import From File**.
 7. Import the `utilities-customer-attribution-glossary-terms.csv` file. Select **Replace all values** as your merge option.
 8. Once the Import completes successfully, click on **Go to task** and then click **Publish** in the next page.
 9. Navigate to **Governance > Categories > Industry Accelerators** to explore the business terms.

*Note: Importing the accelerator and running the notebooks and dashboard consumes approximately 1.5 - 2 Watson Studio CUH and 3 - 3.5 Watson Machine Learning CUH.

Sample data assets

These sample data files that act as dimensional and fact tables are included in the project on the **Assets** page:

- `CUSTOMER.csv`: Customer demographic data.
- `STANDARD_YEARLY_USAGE.csv`: Historical annual energy usage for each customer for previous 7 years.
- `CST_PROFILES.csv`: Customer profiling information.
- `ISSUE.csv`: Dimension table with Issue category.
- `EMPLOYMENT.csv`: Dimension table with different Employment categories.
- `LOCATION.csv`: Dimension table with location data such as addresses and coordinates.
- `EDUCATION.csv`: Dimension table with different Education categories.
- `MARITAL_STATUS.csv`: Dimension table with marital status categories. M - Married, S - Single, U - Unknown
- `OFFER.csv`: Dimension table with different offers which were available to customers.
- `CONTRACT.csv`: Dimension table with contracts which were available to customers.
- `CST_SEGMENT.csv`: Dimension table with segment categories for customers.
- `Attrition View.csv`: Joining the above datasets, we created a csv file that is used as raw data input for the data preparation in 1-model_training notebook. Refer to `Attrition View Creation Query.sql` for the SQL query used to merge the tables.

Notebooks

Follow the instructions in the first few cells of the notebook to configure the project token and API key. Then execute the notebooks step-by-step.

- **1-model_training:** This notebook performs the following functions:
 - Load data
 - Prepare and clean data for model training
 - Analyze correlations
 - Build ML models
 - Analyze and visualize the data
 - Select best performing ML model, create the final pipeline and save to Cloud Pak for Data
 - Store the pipeline in the space and deploy the model.
- **2-model_scoring:** This notebook performs the following functions:
 - Get the deployment space and deployments
 - Deploy the data assets
 - Create and deploy a function for model scoring
 - Predict customer attrition.

R Shiny dashboard

The R Shiny dashboard displays model insights, customer summaries and scores new data. The dashboard has the following tabs:

- **Model Insights :** This tab contains results from 1-model_training notebook. The model was applied to the training data and the results on this data is displayed. The user can see how many customers were in the training data, how many were predicted to attrit and the attrition rate. The cumulative gains and lift charts for this data are also plotted. This data can be filtered by customer segment, city, historical complaints, warranty and tenure. The tab also contains a table of data which was the validation and test data in 1-model_training notebook. This data is used to simulate new data which has just been scored by the model and for which we don't yet know the true target value, whether they actually attrited or not. By clicking on a point on the cumulative gains chart this table is filtered and can be exported. An analyst or marketer can use this to target a specific percentage of customers instead of having to contact all customers.
- **Client View :** Targets individual client information, depicts the top account and Utilities retail details for the customer. It provides the option to run the model scoring webservice, predicting Customer Attrition for the selected customer.
- **Simulation Tool :** This tab contains a form with all model inputs. The user can change any of these inputs and see the impact that the change has on the attrition prediction probability.

Sample business glossary for use with Watson Knowledge Catalog

Optionally, you can import the glossary of business terms into Watson Knowledge Catalog to get started on data governance.

The `utilities-customer-attrition-glossary-categories.csv` file defines the main and sub categories for the business terms.

The `utilities-customer-attribution-glossary-terms.csv` file defines the business terms, category of the business terms and their Related Terms/Part of Terms, if applicable.

Note: This optional step requires the Watson Knowledge Catalog service instance.

Terms and Conditions

This project contains Sample Materials, provided under this [license](#).

Licensed Materials - Property of IBM.

© Copyright IBM Corp. 2020, 2021. All Rights Reserved.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.