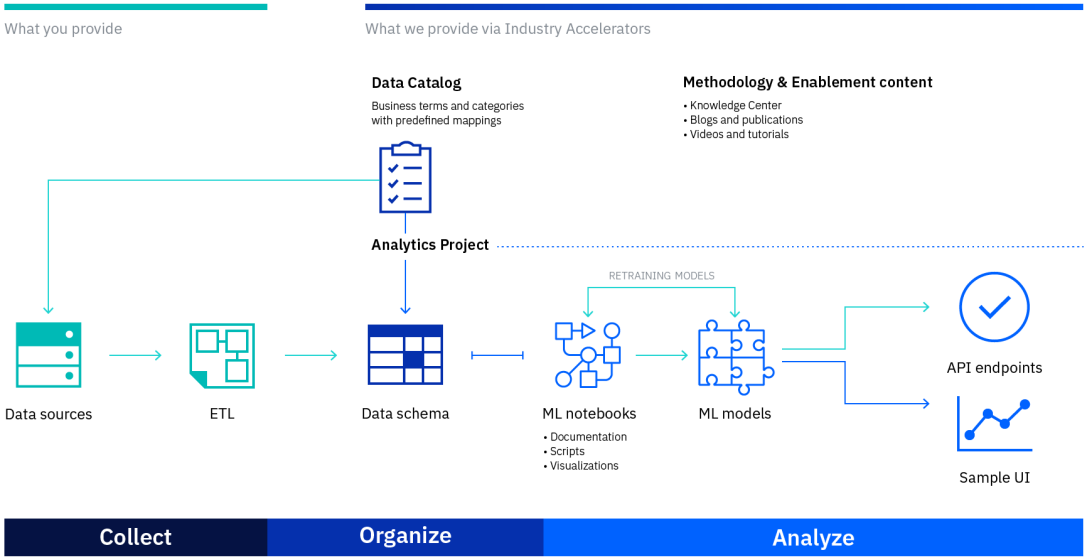
Utilities Customer Attrition Prediction

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.



- Instructions
- Sample data assets
- Notebooks
- R Shiny dashboard
- Sample business glossary

Instructions

Follow these steps to implement the industry accelerator:

1. Navigate to the **Assets** tab and scroll to the **Notebooks** section.
2. Edit the **1-model-training** notebook by clicking the edit icon that looks like a tiny pencil next to the notebook name. This notebook prepares the data, builds ML models, and deploys the model. Follow the instructions in the notebook to step through the execution.

Alternatively, you can run the notebook from the `Jobs` tab by executing **1-model-training-notebook-job** .

3. Edit and run the **2-model-scoring-and-app-deployment** notebook. This notebook deploys data assets, a model scoring function and the r-shiny dashboard. It also generates a URL to launch the r-shiny dashboard. Alternatively, you can run the notebook from the `Jobs` tab by executing **2-model-scoring-notebook-job** .
4. Launch the r-shiny dashboard from one of the following ways.

    - Open the URL generated by deploying r-shiny dashboard in **2-model-scoring-and-app-deployment** notebook.
    - Navigate to **Deployments -> Spaces -> Utilities Customer Attrition Space -> Deployments -> Utilities-Customer-Attrition-Shiny-App** to find the URL of deployed r-shiny dashboard and open it in a new tab.
    - Run the dashboard from RStudio console by completing these steps:
      i. Download the `utilities-customer-attrition-prediction-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.
      ii. Click **Launch IDE > RStudio** on the menu bar.
      iii. In the **Files** pane, select the **Upload** toolbar button and upload the `utilities-customer-attrition-prediction-analytics-dashboard.zip` file into RStudio.
      iv. 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 username and password on the **Client View** tab.
6. Optional. To connect the data assets used in this accelerator to the business terms in Watson Knowledge Catalog, you can edit and run the **0-map-business-terms-to-data-headers** notebook. Enter the authentication details required in the first few cells.

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.

Additionally, there is another dataset created via the analytics project :

- `model output summary.csv` : Consolidated prepped data after attrition prediction for Exploratory data analysis and data visualization in the R shiny dashboard.

Notebooks

Follow the instructions in the notebooks to step through the execution.

- **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-and-app-deployment**: 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.
  - Store and deploy R Shiny app
  - Generate URL to view the app.
- **0-map-business-terms-to-data-headers**: This optional notebook performs the following functions:
  - Publish the **Attrition View.csv** file into a specified catalog.
  - Read mappings from **utilities customer attrition prediction map terms.csv** and applies business terms to the published dataset headers.

Jobs

Navigate to **Jobs** tab to execute following jobs.
1-model-training-notebook-job: Runs the `1-model-training` notebook end to end.
2-model-scoring-notebook-job: Runs the `2-model-scoring-and-app-deployment` notebook end to end.

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.

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 using the below files available in the project tar file.
The `utilities-customer-attrition-glossary-categories.csv` file defines the main and sub categories for the business terms.
The `utilities-customer-attrition-glossary-terms.csv` file defines the business terms, category of the business terms and their Related Terms/Part of Terms, if applicable.
Once the glossary is imported into Watson Knowledge Catalog, Publish the Business Terms, Navigate to **Governance > Categories > Industry Accelerator > Utilities Customer Attrition Prediction** to explore the glossary contents.

Terms and Conditions