Financial Markets Customer Attrition Prediction

Introduction

The Financial Markets Customer Attrition Prediction accelerator provides a set of sample data science assets, a structured glossary of business terms, that help you to predict which customers will leave.

- 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-data-preprocessing** notebook by clicking the edit icon that looks like a vertical ellipsis next to the notebook name. This notebook loads the data and creates and saves the `attrition_prep.py` script to prepare and clean data for model training. It also analyses correlations in the data set. Alternatively, you can run the notebook from the Jobs tab by executing **1-data-preprocessing-notebook-job** .
3. Edit and run the **2-model-training** notebook. This notebook transforms the data, builds machine learning models, and deploys a model. Alternatively, you can run the notebook from the Jobs tab by executing **2-model-training-notebook-job** .
4. Edit and run the **3-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 **3-model-scoring-and-app-deployment-notebook-job** .
5. Launch the r-shiny dashboard from one of the following ways.

   - Open the URL generated by deploying r-shiny dashboard in **3-model-scoring-and-app-deployment** notebook.
   - Navigate to **Deployments -> Spaces -> Customer Attrition Space -> Deployments -> Customer-Attrition-Prediction-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 `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 `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.

6. Once the app has launched, you can perform model scoring in real time by entering your username and password on the **Client View** tab.
7. Optional. Edit and run the `import-wkc-glossary` notebook. This notebook imports glossary of business terms into Watson Knowledge Catalog to get started on data governance and then maps the business terms to the data headers. Alternatively import the glossary of business terms using below steps,

1. Download the `customer-attrition-prediction-glossary-categories.csv` and `customer-attrition-prediction-glossary-terms.csv` files from the Data section of the **Assets** page.
2. Navigate to **Governance > Categories**.
3. Click **Add Category > Import From File**.
4. Import the `customer-attrition-prediction-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 `customer-attrition-prediction-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.

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 Data, Demographic data, Temporal data.
- `account.csv`: Account type and Account Information Data, Investment Information, Temporal data.
- `customer_summary.csv`: Detailed Customer Transaction Data, Business Metrics, Investment and Income Stats.
- `customer_history.csv`: A join of the previous three data sets based on `customer.customer_id`, `customer.effective_date`, `customer_summary.customer_id`, `customer_summary.end_date`, `account.primary_customer_id`, `account.open_date` and `account.close_date` columns. This data set is used as raw data input for the data preparation in the **1-data-preprocessing** notebook. See the `CUSTOMER_HISTORY_VIEW.sql` file for the SQL query used to merge the tables. The resulting data is filtered, transformed, and aggregated to contain one record per customer, so you can use it for modelling.
- `account_summary.csv`: Related to the `account.csv` data set. This data set also contains account related data and investment data. This information is only used to depict financial statistics of the customers in the analytics dashboard.
- `training_user_inputs_and_prepped_column_names.json`: Column headers of the prepped data which is passed as the user inputs for training the model.
- `customer-attrition-map-terms.csv`: Holds details used for mapping business terms to the published dataset column headers.
- `training_data.csv` : Prepared dataset to save the prepped data for modelling in **2-model-training** notebook

Notebooks

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

- **1-data-preprocessing**: This notebook performs the following functions:
  - Load data
  - Create and save script `attrition_prep.py` to prepare and clean data for model training
  - Analyze correlations
- **2-model-training**: This notebook performs the following functions:

  - Build ML models
  - Analyze and visualize the data
  - Select the best performing ML model and save it
  - Store the model in the deployment space and deploy the model
- **3-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 pipeline function for model scoring
  - Predict customer attrition
  - Store and deploy R Shiny app
  - Generate URL to view the app.
- **import-wkc-glossary**: This notebook performs the following functions:
  - Imports the categories into Watson Knowledge Catalog using `customer-attrition-prediction-glossary-categories.csv`
  - Imports the business terms and publishes them into Watson Knowledge Catalog using `customer-attrition-prediction-glossary-terms.csv`
  - Publishes `customer_history.csv`, `customer_summary.csv`, `customer.csv` and `account.csv` files into an existing catalog and maps business terms to column headers.

Jobs

Navigate to **Jobs** tab to execute following jobs in sequence.

1-data-preprocessing-notebook-job: Runs the `1-data-preprocessing` notebook end to end.

2-model-training-notebook-job: Runs the `2-model-training` notebook end to end.

3-model-scoring-and-app-deployment-notebook-job: Runs the `3-model-scoring-and-app-deployment` notebook end to end.

import-wkc-glossary-notebook-job: Runs the `import-wkc-glossary` notebook end to end.

R Shiny dashboard

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

- Dashboard View: Shows top action clients, monthly customer attrition, and customer attrition risk level.
- Client View: Targets individual client information, depicts the top business metrics and account details, provides an option to run the model scoring web service, predicts customer attrition, and visualizes the influential factors and data fields.

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

The `customer-attrition-prediction-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, Navigate to **Governance > Categories > Industry Accelerator > Customer Attrition Prediction** to explore the glossary contents.

Terms and Conditions