Financial Markets Customer Life Event Prediction

Introduction

The Customer Life Event Prediction accelerator includes a structured glossary of business terms and a set of sample data science assets. The glossary provides the information architecture that you need to predict major life events, such as buying a home or relocating. You can use the sample notebooks, predictive models, and dashboards to accelerate data preparation, machine learning modelling, and data reporting. Set your clients on the path to financial success by engaging with them at the right time with relevant investment options.

Tip: Download the PDF of these instructions from the 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 > New deployment space** and then complete these steps:
    1. Enter the name Customer Life Event Space
    2. Select your Watson Machine Learning service. If you don't have one, click the **Create** button and provision one.
    3. Click on **Create** to create the deployment space.
2. Navigate back to the accelerator project, select the **Assets** tab and scroll to the **Source Code** section.
3. Edit the **1-data-preprocessing** notebook by clicking the vertical ellipsis next to the notebook name and choose **Edit** option. This notebook loads the data, creates and saves the life_event_prep.py script to prepare and clean data for model training. It also selects life events to predict from the data set. Follow the instructions in the notebook to configure and step through running it.
4. Edit and run the **2-model-training** notebook. This notebook transforms the data, builds ML models and deploys the models.
5. Edit and run the **3-model-scoring**. This notebook deploys the data assets and a model scoring function.
6. Run the dashboard from RStudio console by completing these steps:

1. Download the customer-life-event-prediction-analytics-dashboard.zip file from the project's Data section of the **Assets** page.
2. Click **Launch IDE > RStudio** on the menu bar.
3. In the **Files** pane, select the **Upload** toolbar button and upload the customer-life-event-prediction-analytics-dashboard.zip file into RStudio.
4. Navigate to customer-life-event-prediction-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.

7. Optional step: edit and run the import-wkc-glossary notebook. This notebook imports the glossary of business terms into Watson Knowledge Catalog, to get started on data governance, and then maps the business terms to the data headers. You must have the [permission to create governance artifacts](). Alternatively, import the glossary of business terms using the steps below:
   1. Download the customer-life-event-prediction-glossary-categories.csv and customer-life-event-prediction-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 customer-life-event-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-life-event-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.

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. Additionally, once you start the **RStudio** Environment, you will be prompted to choose the RStudio environment definition. Some of the dashboards require larger memory allocation so would need a higher plans. Use the M plan in case your dashboard fails to start up in the XS plan.

- **RStudio XS - 2 vCPU and 8 GB RAM** - Offered as part of the Free Lite plan and consumes 1 capacity unit per hour.
- **RStudio M - 8 vCPU and 32 GB RAM** - Offered as part of the Paid plan and consumes 4 capacity units per hour.
- **RStudio L - 16 vCPU and 64 GB RAM** - Offered as part of the Paid plan and consumes 8 capacity units per hour.

Sample data assets

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

- event.csv: Event and type data, Temporal data, Life event category etc.
- customer.csv : Customer Data, Demographic data.
- census_probabilities.csv : Migration, birth, marriage and birth probabilities based on USA census. (This is an optional addition for the user to incorporate census data into the final dataset for modelling.)
- news.csv: Contains News information. This information is only used to depict relevant news and life event type information in the analytics dashboard.
- event_type.csv:It comprises of event type data, life event type data etc.
- training_user_inputs_and_prepped_column_names.json : Used for standardizing the numeric variables for modelling by scaling the values.
- customer-life-event-prediction-map-terms.csv : File containing the list of mappings between column headers and business terms.

Notebooks

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

- **1-data-preprocessing**: This notebook performs the following functions:
  - Load data
  - Create and save script life_event_prep.py to prepare and clean data for model training
  - Select life events to predict.
- **2-model-training**: This notebook performs the following functions:
  - Build ML models
  - Analyze and visualize the data
  - Select best performing ML model and save to Cloud Pak for Data
  - Select a Watson Machine Learning based deployment space, store the models in the space, and deploy the model.
- **3-model-scoring**: This notebook performs the following functions:
  - Get the deployment space and deployments
  - Deploy the data assets
  - Create and deploying a pipeline function for model scoring
  - Predict Life Events.
- **import-wkc-glossary**: This notebook performs the following functions:
  - Import the categories into Watson Knowledge Catalog using customer-life-event-prediction-glossary-categories.csv
  - Import the business terms and publish them into Watson Knowledge Catalog using customer-life-event-prediction-glossary-terms.csv
  - Publish event.csv,customer.csv and census_probabilities.csv files into an existing catalog and map business terms to column headers.

R Shiny dashboard

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

Dashboard View : Shows client activity, Top action Clients, Market Action, Recent Life events experienced by clients, and Revenue Opportunities.

Client View : Targets individual Client information (personal and financial), Depicts the Top Business Metrics, Visualizes Event distribution, opportunities and Client life event history. It also provides option to run the Model Scoring, Predicts Life Events and Visualizes the influential factors and data fields.

Sample business glossary for use with Watson Knowledge Catalog

Optionally, the glossary of business terms can be imported into Watson Knowledge Catalog, to get started on data governance.
The customer-life-event-prediction-glossary-categories.csv file defines the main and sub categories for the business terms.
The customer-life-event-prediction-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