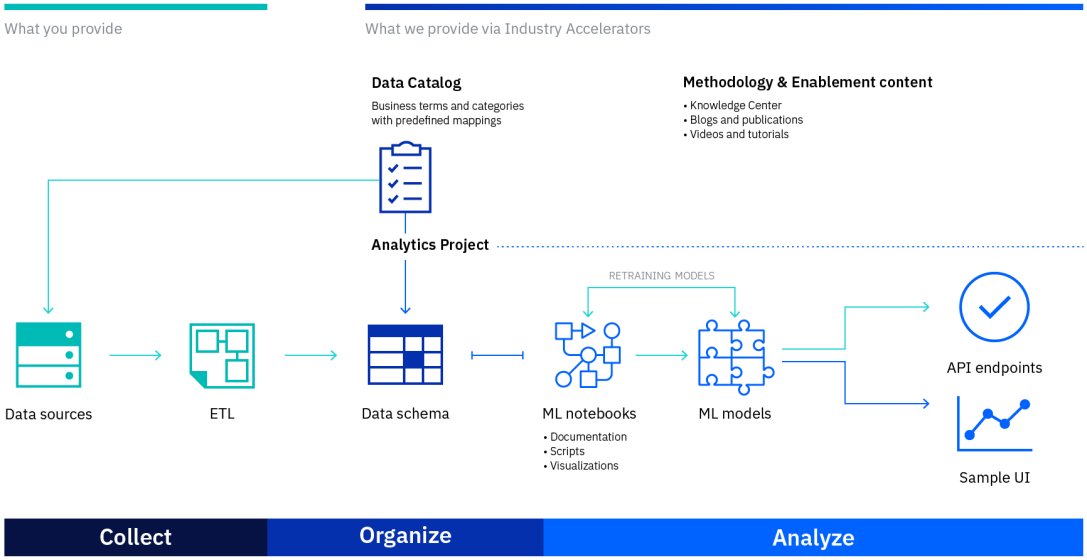
Utilities Payment Risk Prediction

Introduction

The Payment Risk 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 understand why customers miss their utility bill payment. Your data scientists can use the sample notebooks, predictive model and dashboard to accelerate data preparation, machine learning modeling and data reporting. Identify those customers who are likely to miss paying their utility bill by the due date, allowing the business to engage proactive with these customers.



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 Payment Risk Prediction Space -> Deployments -> Utilities-Payment-Risk-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 `utilities-payment-risk-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-payment-risk-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.
- `CST_BUILDING_PROFILES.csv` : Table with building details for each customer.
- `BUILDING_TYPE.csv` : Dimensional table with categories for building type.
- `GENDER.csv'`: Dimensional table with gender status'.
- `INVOICE.csv` : Historical utility bill payment details. Includes monthly billing amounts, energy usage, billing and due dates for each customer.

- `ACCOUNT.csv`: Account details, such as rate start and end dates and credit history for each customer account.
- `Bill Payment 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 `Bill Payment View Creation Query.sql` for the SQL query used to merge the tables.
- `model output summary.csv`: Data file generated in `1-model-training` notebook, contains records for all customers and all billing cycles, including the latest cycle for each customer, where the actual target value is not known. Additionally, there is another dataset created via the analytics project :

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
  - Build ML models, Analyze and visualize the data
  - Select best performing ML model and save to Cloud Pak for Data
  - Create a Watson Machine Learning based deployment space
  - Store the model 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 the probability that a customer will miss their payment
  - 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 **Bill Payment View.csv** file into a specified catalog.
  - Read mappings from **utilities-payment-risk-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:

- Billing Cycle View : This tab gives the user a view of the important customer information for the current billing cycle, such as the number of customers billed, total amount owed and overdue amount owed in the period. In our example, the billing cycle is the last month of data that we have, June 2019. The data can be filtered by city, customer segment, if they

missed the payment last month, total bill amount range and history of missing payments. The tab also contains charts showing the historical billing cycle number of customers who missed their payment, the overdue amount owed each billing cycle and the frequency of missed payments. Each customer was scored in 1-model-training and assigned a probability of missing this month's bill payment. Summary data for each customer is contained in the table. Clicking on any customer brings the user to the Client View tab for that customer.

- Client View : Provides client information and current billing information for the selected individual. This tab also contains historical billing amounts, usage and predictions for the selected customer. The option to run the model scoring webservice, predicting the risk of the selected customer missing their current bill payment, can also be found in this tab.

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-payment-risk-prediction-glossary-categories.csv` file defines the main and sub categories for the business terms.
The `utilities-payment-risk-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, **Publish** the Business Terms and Navigate to **Governance > Categories > Industry Accelerator > Utilities Payment Risk Prediction** to explore the glossary contents.

Terms and Conditions