

Utilities Payment Risk Prediction

Introduction

The Payment Risk Prediction accelerator helps identify the customers who are likely to miss paying their utility bill by the due date, allowing the business to proactively engage with these customers. 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, the predictive model and the dashboard to accelerate data preparation, machine learning modeling and data reporting.

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

Instructions

Follow these steps to implement the industry accelerator:

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.

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 Payment Risk Prediction 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 the RStudio console by completing these steps:
 1. 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.
 2. Click **Launch IDE > RStudio** on the menu bar.

3. In the **Files** pane, select the **Upload** toolbar button and upload the utilities-payment-risk-prediction-analytics-dashboard.zip file into RStudio.
4. Navigate to utilities-payment-risk-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.
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-payment-risk-prediction-glossary-categories.csv and utilities-payment-risk-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 utilities-payment-risk-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 utilities-payment-risk-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 demographic data.
- CST_PROFILES.csv: Customer profiling information.
- 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
- 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.

Notebooks

Follow the instructions in the first few cells of each 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
 - Build ML models, analyze and visualize the data
 - Select the 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:** 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

R Shiny dashboard

The R Shiny dashboard displays model insights and customer summaries as well as 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, whether or not 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.

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-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.

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.