

Hospital Readmission Prediction for diabetic Patients

Introduction

According to a [study](#), every one in ten patient is suffering with Diabetes and diabetes has higher chances of hospitalisations. Because of which CMS introduced many programs to reduce hospitalizations and readmissions. e.g. [Hospital Readmission Reduction Program](#)

Hospital Readmission Prediction model focusses on assessing the risk of hospital readmissions within 30 days of discharge and factors affecting readmissions in diabetic patients. The model analyses various features such as patients medical records, admission details, discharge details, lab test results, types of diagnoses, time in hospital, payer type etc. to predict probability of readmission in diabetic patients.

Tip: Download the PDF of these instructions from the Data 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. 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 diabetes data and patients data and merges them and then cleans and prepares training data, and saves it for model training. 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-and-deployment** 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-and-deployment-notebook-job** .
4. Edit and run the **3-shiny-app-deployment** notebook. This notebook deploys 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-shiny-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-shiny-app-deployment** notebook.
 - Alternatively, navigate to **Deployments -> Spaces -> Hospital Readmission Prediction Space -> Deployments -> hospital-readmission-prediction-Shiny-App** to find the URL of deployed r-shiny dashboard and open it in a new tab.
 - To, run the dashboard from RStudio console by completing these steps:
 - i. Download the `hospital-readmission-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 `hospital-readmission-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 use button on Patient View tab of the dashboard to predict readmission probability for the patient.
 7. Edit and run the **4-monitor-wml-model-with-Watson-OpenScale** notebook. This notebook sets up and configures OpenScale, creates a data mart for the model with Watson OpenScale and configures OpenScale to monitor that deployment. It also injects records and measurements for viewing in the OpenScale insights dashboard.
 8. 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:

- `diabetic_data` : The data from [UCI repo](#). The data contains individual patient admission and readmission records, patients medical records, admission details, discharge details, lab test results, types of diagnoses, time in hospital, payer type etc.
- `patient-data.csv` : Patient number with mocked up first name and last name of patients.

Additionally, there are other datasets created via the analytics project :

- `training-data.csv` : The above two datasets are merged, cleaned and mapped against original values to produce a dataset with one record per patient indicating whether he is readmitted or not in the notebook `1-data-preparation` . The prepared dataset is called `training-data.csv` .
- `explanation-results.csv` : Data explaining the openscale results.

Notebooks

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

- **1-data-preprocessing**: This notebook performs the following functions:
 - Loads datasets
 - Merges the datasets
 - Analyse the columns, create mappings to column values
 - Filter the data
 - Save the result
- **2-model-training-and-deployment**: This notebook performs the following functions:
 - Load training data
 - Analyze and visualize the data
 - Handle correlation and categorical features
 - Build ML models
 - Select the best performing ML model and save it
 - Analyse feature importance and explanation of the model
 - Store the model in the deployment space and deploy the model
- **3-shiny-app-deployment** : This notebook performs the following functions:
 - Get the deployment space
 - Store and deploy R Shiny app
 - Generate URL to view the app.

- **4-monitor-wml-model-with-Watson-OpenScale:** This optional notebook performs the following functions:
 - Get the model and deployments
 - Configure OpenScale, Create/Setup datamart
 - Score the model to configure monitors - Quality monitoring and feedback logging - Configure Fairness monitoring for detection of model bias
 - Configure Explainability features
- **0-map-business-terms-to-data-headers:** This optional notebook performs the following functions:
 - Publish the files `diabetic_data.csv` and `patient_data.csv` into a specified catalog.
 - Read mappings from **hospital-readmission-prediction-map-terms.csv** and applies business terms to the published dataset 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-and-deployment-notebook-job: Runs the `2-model-training-and-deployment` notebook end to end.

3-shiny-app-deployment-notebook-job: Runs the `3-shiny-app-deployment` notebook end to end.

R Shiny dashboard

Hospital Readmission Prediction dashboard includes various set of visualizations to understand risk of readmissions in diabetic patients within 30 days of discharge. The dashboard has the following tabs:

Home: Home tab contains charts and tables to display low, medium and high risk risk patients. It also consists of visualizations to understand admission and readmission risks in different age groups, payer types and patients discharge disposition. When clicked on a patient in a data table, the dashboard navigates to Patient View.

Patient View: Patient View tab provides details for individual patients clicked in the home tab. Patient view tab displays patients personal and health records, various test results, diagnoses types, time in hospital etc. This tab includes a button to predict readmission probability for the patient. Additionally, this probability will be explained with a chart using `Explanation` button on the screen.

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

The `hospital-readmission-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 > Hospital Re-Admission Prediction** to explore the glossary contents.

Terms and Conditions

This project contains Sample Materials, provided under license.

Licensed Materials - Property of IBM.

© Copyright IBM Corp. 2021, 2022. All Rights Reserved.

US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Citation

The "Diabetes 130-US hospitals for years 1999-2008" dataset used in this accelerator is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license. <https://creativecommons.org/licenses/by/4.0/legalcode> Clore, John, Cios, Krzysztof, DeShazo, Jon & Strack, Beata. (2014). Diabetes 130-US hospitals for years 1999-2008. UCI Machine Learning Repository.