Lab Guide

Hands-on-Lab:
Model building and deploying with Watson Machine Learning with notebook

Shivam R Solanki
Data Scientist
Shivam.raj.solanki@ibm.com

IBM

Summary

In this lab tutorial, we'll use IBM Cloud Pak® for Data to go through the whole data science pipeline to solve a business problem and predict customer churn using a Telco customer churn dataset. IBM Cloud Pak for Data is an interactive, collaborative, cloud-based environment. It can help data scientists, developers, and others interested in data science use tools to collaborate, share, and gather insights from their data — as well as build and deploy machine learning, and deep learning models.

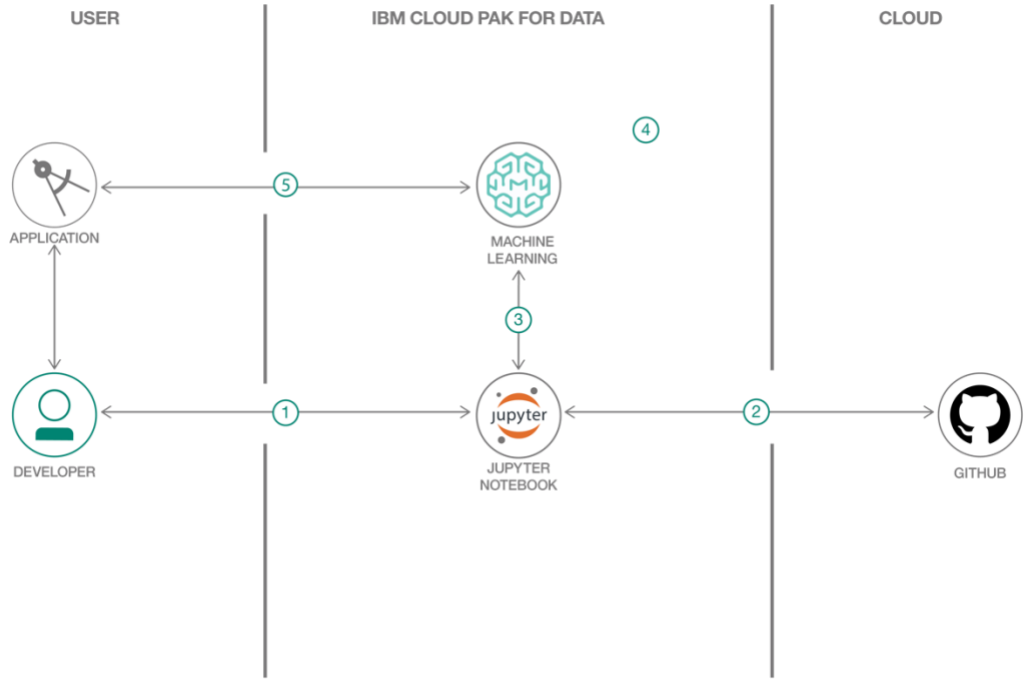
Description

Customer churn (when a customer ends their relationship with a business) is one of the most basic factors in determining the revenue of a business. You need to know which of your customers are loyal and which are at risk of churning — and you need to know the factors that affect these decisions from a customer perspective. This lab tutorial explains how to build a machine learning model and use it to predict whether a customer is at risk of churning. This is a full data science project, and you can use your model findings for prescriptive analysis later or for targeted marketing.

After you've completed this code pattern, you'll understand how to:

- Use Jupyter Notebooks to load, visualize, and analyze data.
- Run Notebooks in IBM Cloud Pak for Data.
- Build, test, and deploy a machine learning model using Scikit-learn on IBM Cloud Pak for Data.
- Deploy a selected machine learning model to production using IBM Cloud Pak for Data.
- Test the deployed model using the exposed endpoint

Flow



1. User loads the Jupyter Notebook into the IBM Cloud Pak for Data platform.
2. Telco Churn dataset is loaded into the Jupyter Notebook as virtualized data after following the Data virtualization and Data refinery tutorial.
3. Preprocess the data, build machine learning models, and save to Watson® Machine Learning on IBM Cloud Pak for Data.
4. Deploy a selected machine learning model into production on the IBM Cloud Pak for Data platform and obtain a scoring endpoint.
5. Test the deployed model using the scoring endpoint

Instructions

Complete details on how to get started running and using this application are in the README, including how to:

1. Import Jupyter Notebook to IBM Cloud Pak for Data.
2. Run the notebook.
3. Create a Space for Machine Learning Deployments
4. Deploy the model
5. Test the model.

Conclusion

This code pattern showed how to use IBM Cloud Pak for Data and go through the whole data science pipeline to solve a business problem and predict customer churn using a Telco customer churn dataset.