Machine Learning Documentation

Artificial Intelligence – Sales and Operations Planning for Dynamics 365 Business Central

Last updated: November 17, 2023

TABLE OF CONTENTS

1. About

The Artificial Intelligence feature of the Sales and Operations Planning for Dynamics 365 Business Central application allows users to consume sales forecasts generated using state of the art Machine Learning and Big Data techniques.

For more information, please go to https://www.websan.com/sales-and-operations-planning-for-business-central .

2. INSTALLATION & EXECUTION

The following steps cover the installation process of the Artificial Intelligence (AI) feature of the Sales and Operations Planning for Dynamics 365 Business Central module, including all the necessary Azure resources, as well as the generation of Sales forecasts from Business Central.

The next steps must be covered to install the AI feature of the Sales and Operations Planning and run a forecast:

1. Security. Azure admin access is required to install the AI model and the resources needed.
2. Installation. Scripts are available via the Sales and Operations Planning for Dynamics 365 Business Central application page. The following Azure resources will be installed.
    a. Data Lake Storage Gen. 2
    b. Data Factory Version 2
    c. Machine Learning Workspace
    d. Functions
3. Azure Pipeline Resources Connection Setup. Users will need to confirm that the Azure resources installed in the pipeline are connected to one another. This will be completed in the user's Azure portal.
4. Load 3 years of data to the Historical Sales Data window in Business Central.
5. Sync the data from Business Central to Azure. Synching the data triggers the Machine Learning process and deploys the validated model to Azure functions.
6. Setup API for Model Consumption in Business Central. Users will copy a link generated by the Azure Functions resource in their Azure portal and paste it in Business Central. This will trigger the generation of the Sales predictions and populate the Sales Budget Overview Matrix in Business Central.

3. MACHINE LEARNING MODEL

The Machine Learning models built to predict Sales over time, by Item, Customer, and Location use a panel dataset (a combination of time series with cross sectional data). Summary:

- It is expected that the model will encounter datasets containing a high volume of entries where the Sales variable is equal to zero.
- As a response to an expected high cardinality of zeroes, the Croston smoothing method was implemented. Click here to learn more about Croston.
- The variable '*croston'* was created to replace the Sales. '*croston*' has fewer instances of zeroes.
- *'month', 'quarter'*, and *'year'* are variables derived from the *'date'* variable. These variables were created to introduce additional variance to the model and capture potential seasonality effects.
- Data points prior to the latest year in the dataset were assigned to the training partition. The remaining were assigned to validation.
- Hyper-parameter tuning found that the best estimator is an XG Boost model where maximum depth=3 and n_estimators=100.
- RMSE and R^2 were chosen as the main accuracy metrics. R^2=83.8% and RMSE=1581 on the validation dataset. No signs of overfitting were found.

Data

It is recommended that a minimum of three years of data are loaded for the purposes of training and validation. Data imported from Business Central to Azure Data Lake has a daily cardinality (by-day basis). The following columns (variables) are imported to Azure Data Lake:

- Posting date
- Sell-to Customer No.
- No. (item)
- Quantity
- Amount
- Location Code

Data Cleaning and Variable Engineering

It is assumed that users will load datasets with a high cardinality of entries where the s*ales* variable contains zeros (sales may not be registered for a product for days or even months). To address this situation, the Croston methodology has been implemented as an alternative to Exponential Smoothing and simple value imputation. To learn more about Croston, click here.

The features *'month', 'quarter'*, and *'year'* were derived from *'date'* to introduce additional variance to the model, which is expected to improve model accuracy, and capture potential seasonality effects.

Data Partition

Data splitting is performed on a date basis. Entries belonging to the last year of data are assigned to the validation partition (i.e., the latest datapoint in the dataset is July 31, 2023, therefore, validation data will run from August 1, 2022, to July 31, 2023). The remaining data is assigned to training. Pipelines were implemented in the code to prevent statistical data leaking.

Model Training & Validation

A XG Boost model with hyper-parameter optimization (CV=5, n_jobs=-1) was found to be the best predictor. Best parameters are max_depth=3 and n_estimators=100.

Accuracy Metrics

Given that target variable for this machine learning project is *sales,* which can easily reach high numbers, the accuracy metrics selected for are Root Mean Squared Error (RMSE) and R^2. The accuracy results on the validation dataset are found below. No signs of over/under fitting were found. The differential between train and validation datasets R^2 was 0.02.

R^2=83.8%

RMSE=1581.