

LB AdEx

Overview

This page is designated to scope, plan and deliver the Proof of Concept the AdEx application that was running on Dataiku and then moved to GCP in June 2023.

Product

Vision

We want Researchers and smart tools to help each other at finding quickly the best products for the right applications.

Goal

AdEx is an application that allows Lab Researchers to upload a dataset so that it can predict best results based on recommended inputs.

The value: allow the researchers to reduce the number of trials with non-successful outputs.

Documentation

AdEx deck for "From Dataiku to GCP" migration [here](#)

Quick User Manual Documentation [here](#).

Full User Manual [here](#).

LLD [here](#).


Projects

BM Aubervilliers & ARO Lyon

Product Features

This Product composed of a few modules represented in 5 steps for the user going from the upload to the recommendation:

issuekey summary issuetype created updated duedate assignee reporter priority status resolution

 Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Upload and customize Dataset

The upload

- A Unique file can be uploaded from computer only
- Type of files that can be uploaded are: ".csv" or ".pk" (export from a previous AdEx analysis-export), else:
 - If a wrong file format is uploaded: nothing happens, no error message, nothing stops the user to reupload a file

Dataset customization and selection

- The user can visualize the uploaded dataset on the right side/panel. It exactly shows the same number of columns and cells, and values that are in the uploaded file in the data table on the original radio button/tab
 - There is no column/row limits

- Format cleanup, list of accepted rules
- Clear session button allows the clear the dataset of the session to reupload a new one.
- 2 type of uploads with different format - the file must already be in that format:
 - Trials in rows
 - Trials in columns
- Select the "Variable Selection" tab
- Select Trial ID column if identifier (Primary key) is available
- Select inputs and outputs from available list in dropdown menus
 - Minimum of 2 inputs and 1 output
- Click on the Verify "Only ID, Inputs & Targets" radio button/tab to see selected columns from original dataset
- If those steps has been fully completed, the button in the next tab will be orange

Set Design space - Select Dataset Range

- Select "Design Space" tab
- Click "USE DEFAULT DESIGN SPACE AND MAXIMIZE TARGETS"
 - Grey if previous step was incomplete
 - Orange shows if the previous step to select variables was completed
 - Green if the design space is set
- Or Click on the dropdown, select to change the range of the value of the columns. (cannot change back and forth with the USE DEFAULT DESIGN SPACE...)
- Reset to default value can still make you use the default range
- Proceed to "Model Optimize" story and perform tasks.
- Select each column and change ranges.

Model Optimize

- Select "FIT MODEL & SEARCH NEXT TRIAL" to compute the model.
 - Grey if previous step was incomplete
 - Orange shows if the previous step to select variables was completed
 - Green if the design space is set
- To reset the design space, click the green button to change it back to orange
- If double clicking before computation is complete and button turns green, button turns red with message "TOO MANY CLICKS - REFRESH PAGE" - at this point user needs to start from a fresh page
- After computation is complete and button turns green, the user can visualize the results of the fit:
 - "Model Info": the user selects one output from the drop down menu and a graph displays predicted vs measured output along with error bars (5-Model Info Graph)
 - SHAP graphs (See attached 11-SHAP Graphs):
 - A bar chart with horizontal bars
 - A graph displaying features, feature values vs SHAP value
 - One graph per input showing SHAP values vs input values
 - "Recommended Trials":
 - A table sorted by trial ranking is displayed (See attached 6-Table Sorted by Trial Ranking)
 - X Output graphs (See attached 7-X Output Graphs):
 - In red, it shows output value from the design table (historical trials) sorted by the identifier (primary key)
 - In green, it shows the top 10 recommended trails
 - Below the graphs, a contour plot can be visualized for each output by selecting X and Y in drop down menus below these graphs (See attached 8-Contour Plots)
 - "One-Dimensional Profile of the Model": the user selects the trial ID if available (selected in "Select Variables" tab), the input name and the target to display a banded graph. A red cross is for each historical trail. Reference values for the profile are also displayed below graph (See attached 9-Banded Graph)
 - "Prediction Visualizer": only available for multiple targets. Select two targets at the time and display scatter plot with historical trails (red points and error bars) and suggested trials (green points). Clicking on points displays trial experimental values below scatter plot (See attached 10-Prediction Plot)

Update Scores

Timeline

First Phase

The data Scientists, who are part of Materials R&I, [TAMBURRO, Alessio](#) and [Ongari, Daniele](#) have been developing a small application running on Dataiku where researchers can upload their dataset and get recommendation.

The current example taken in developing the app shows how the yield can be optimized.

This proof of concept also shows that the application running on Dataiku is not scalable to our users due too poor performance.

Current documentation can be found here: https://docs.google.com/presentation/d/1VPZLjZ05u780Y9Unwead3OEk_TOrSmL2Tpzr4teGaGg/edit#slide=id.g1683c30a2ec_0_1

Second Phase

In order to have a few users testing and using the application, we need to move it GCP.

How?

We need the DataLab squad, UI/UX designer and full stack developers to re-design the application coded in python on DataLab.

The application would behave exactly with the same feature set we currently have.

What business requirements we can improve:

- DataLab UI/UX experience
- Users would still upload their dataset in the DataEx module
- Solvay users could use the Google SSO to access DataEx (No need to sign an NDA if the user will only have strict access to DataEx)

Steps

1. We need to review the scope, create epics, stories and design with the current DataLab Team
 - a. Meetings with the Architect, Product Manager, Product Owner, Business Analyst and UI/UX Designer
2. Breakdown and estimate non-functional re-design with the current team
 - a. Meetings with squad
3. Prioritize with an additional statement of work (growing the DataLab team with Lab Booster current budget)
 - a. Meeting with Cloud/Vanenburg team to pitch the additional workload with a solution