Project Vision

This product will give our industry partner a tool at hand, that can effectively increase conversion of their leads to customers, primarily by providing the sales team with valuable information. The modular architecture makes our product future-proof, by making it easy to add further data sources, employ improved prediction models or to adjust the output format if desired.

Project Mission

The mission of this project is to enrich historical data about customers and recent data about leads (with information from external sources) and to leverage the enriched data in machine learning, so that the estimated life-time value of leads can be predicted.

Usage

To use the application a user must adjust the input and output files of the data pipeline the program employs. For this the user must adjust the corresponding variables in the `src/demos.py` file. To achieve results with actual client data ensure access to the correct S3 buckets.

Demo

To run the current demo of the program make sure to have the environment installed and run the `src/main.py` either locally or via the build process. The user will be prompted with the following options:

```
(1) BDC
(2) EVP
(3) DB
(4) Pipeline
(5) Exit
```

(1) BDC

This will give two options `(1) Read CSV` and `(2) Dummy API`. The first option will read the base data CSV file and confirm that it is processed correctly. The second option shows the communication with a dummy API as a proof of concept.

(2) EVP

This will ask the following series of questions: `Load model from file? (y/N)` : When `y` the program will ask for a file location of a previously saved model to use for predictions. `Split dataset (y/N)` : Should the enriched lead data be split into train/validation/test sets. When running the demo for the first time chose `y`. `Add dummy labels (the lead value) (y/N)` : If `y` is chosen the data will be enhanced with random labels that can be used as a proof of concept of the EVP predictions. Choose `y` if the data stored locally is unlabeled.

Now the user is prompted with the following options

```
(1) Train
(2) Test
(3) Predict on single lead
(4) Save model
(5) Exit
```

- (1): Train the current model on the current trainig dataset.
- (2): Test the current model on the current test dataset and print the mean squared error.
- (3): Choose a single lead from the test dataset and show the prediction and true label of the data point.
- (4): Save the current model to a specified file location.
- (5): Exit the EVP submenu.

(3) DB

The user can display the currently stored data of a chosen lead.

(4) Pipeline

The user can run a predefined pipeline including all data enhancement steps. This is good for demo purposes. Otherwise, the user is prompted to choose from all available steps to create a custom pipeline. It is possible to store the results of the pipeline either locally or in the defined S3 location. To store the data remotely do not set a limit for processing and make sure access to S3 storage is established.

(5) Exit

Gracefully exit the program.