Product Tracing

Description

This repository contains the codes for performing product tracing. The program takes HDM, Inovice and Declaration data, takes a timeframe for each data as an input and retruns the product paths by whcih the product reached the final seller starting from importer. The prgoram is also responsible for reproting the statistics of the the paths found. The products in this case are identified as a category brand pair. In detail documentation of thiss scripts can be found here.

Data Sources

This program uses the the raw hdm, invoice and declaratiosn data recevived from SRC along with preprocessed data receiced from NLP Pipeline. The aggregation method that we are using for tracing data can be found in `data_aggregation.ipynb` jupyter notebook.

Before inputing the data into our program you have to aggregate the raw and processed data with the logic in the jupyter notebook.

The samples for raw and processed data can be found in folder `Data/data_aggregation_examples`. The final prepared sample datas that can be inputed into program can be found in folder `Data/final_tracing_data_samples`

Configuration

The `config.json` file contains the paths to the input data and configurable parameters for the pipeline. Key sections include:

- **Data**: Specifies paths to input data files.
- **Params**: Specifies input parameters for tracing category, brand(optional), HDM tin and optional path penght configuraation parameters.
- **Tracing**: Enables or disables specific Tracing method.
- **DataRequrirments**: Input data requiremnts for all three datassets.s

Worklflow

1. Data Processing: Recieving raw and preprocessed data, aggregating and merging data.
2. Preparing Declarations data: Extracting and normalizing imported product quantities.
3. Price Processing: Extracting and Normalizing all prices from all three datasets.
4. Tracing the brand category pair product.

Usage Guide

Follow these steps to set up the project on your local machine:

1. **Clone the repository**:

```bash git clone https://github.com/MindwiseLLC/Product-tracing.git

1. **Ensure Python 3.11 is installed**:

This project requires Python 3.11. You can download and install it from the official Python website: https://www.python.org/downloads/.

1. **Create a virtual environment**

After installing Python 3.11, create a virtual environment for the project:

**For Windows:**

```
python -m venv tracing-env
tracing-env/Scripts/activate
```

**For macOS/Linux:**

```
python3.11 -m venv tracing-env
source tracing-env/bin/activate
```

4. **Install dependencies**

After activating the virtual environment, install the required dependencies:

```
pip install -r requirements.txt
```

5. **Run the main script**

First, navigate to the ***'code'*** directory using the command:

```
cd code
```

To run the main script, execute the following command:

```
python main.py
```

This will execute the pipeline according to the settings defined in config.json.
```