

# Prometheus: machine learning toolbox for marketing applications

Natalia Efremova, Naoto Minakawa, Felipe Thomaz  
Saïd Business School, University of Oxford, Oxford, UK

## 1 Overview

### 1.1 Repository location

Prometheus is a collection of educational notebooks for development of machine learning (ML) models in Python. They closely follow course lectures (MBA courses: Marketing, AI for Marketing, Marketing Analytics). You will find the list of lectures in Sec.2: each lecture is accompanied with a link to a piece of code, called a notebook.

You can run these notebooks in [Google CoLab](#) from your browser. Just click on the link in the document and it will automatically open the notebook in your google drive. After opening the link in CoLab, press the button "Open in Google Colaboratory" in the top of the page.

Alternatively, you can download the notebooks from this [GitHub page](#) and run them [Conda environment](#) on your machine. To be able to download the content, you need to have GitHub account (or create one, it's free). If you do not have any experience in Python programming, we highly recommend using CoLab notebooks (and opening them in a Google Chrome browser).

To get started, you will need to access the notebook [here](#).

This is a beginner tutorial, which will help you with navigating in CoLab. Google CoLab notebooks are shared and saved just like Google Docs. Upon visiting the notebook, go to "File → Save a copy in Drive". You will then be able to see a copy of the notebook in a folder named "Colab Notebooks" inside your Google Drive. This is the file you should work with.

- The notebook is divided into individual cells. When running code in the notebook, one can run one cell at a time, and the output of the code will populate under the cell. You can run a cell with the *Run* button on the left or pressing *Shift + Enter* keys on your keyboard.
- The values of any variables used in this cell will be stored and can be used when running future cells. The variable values will be written over if the cell is run again.
- Some cells load data from online repositories. This takes a good bit of time (up to 10 mins). We highly recommend that you run this cell once and then use the stored data in your code without reloading the dataset. The notebooks are set up in a fashion conducive to this use.

### 1.2 Project overview

This toolkit was produced as part of a "Prometheus" research project, developed under the [Future of Marketing Initiative](#). The Prometheus initiative is driving the future of marketing with a unique partnership between academia (FOMI) and industry (Teradata). The main goal of the Prometheus project is to facilitate the adoption of artificial intelligence and ML tools for marketing applications.

The core of this project is a collection of open-source ML tools for common marketing problems. We aim to deliver easily accessible code for both marketing academics and industry professionals. The project includes three main topics: 1) traditional ML tools for marketing applications, 2) natural language processing (NLP) and 3) computer vision (CV) methods.

Traditional ML tools include such topics as clustering, attribution models and recommender systems. NLP section covers such topics as word embeddings, language models and sentiment analysis. CV section discusses image classification and creative content development. This toolbox will allow researchers to access marketing tools, to understand their basic functionality and even design simple analytics tools from scratch.

## 2 Code description

This code follows the lectures on Executive MBA class on artificial intelligence (AI). The suggested order in following the code blocks is the following (although it is not necessarily needs to be the same, most notebooks are self-explanatory).

1. **Introduction: Definitions, AI in Marketing Landscape, and AI as Competition.**  
Code: [CoLab and Python tutorial](#).
2. **Outputs, Governance, and Controls.**  
Code: [Customer segmentation with unsupervised learning\(k-nns\)](#).
3. **Inputs: Data, Algorithms, Compute, People Beyond Digital Transformation.**  
Code: [Exploring ML models in Python](#).
4. **Natural Language Processing, Computer Vision, Graph Embeddings.** Code:
  - Computer vision: [Convolutional Neural Networks](#);
  - NLP: [Word embeddings](#);
  - NLP: [RNNs and LSTMs](#);
  - NLP: [Transformers \(BERT\)](#).
5. **Decision Support vs Decision Automation.**
6. **MMMs, MTAs, Customer Paths, Pricing.** Code:
  - [Multi-touch attribution with logistic regression](#);
  - [Cojoint analysis](#).
7. **AI in Execution: Service, Support, and Recommendations.**  
Code: [Recommendation systems](#).
8. **AI in creativity, Branding, and Advertising.**  
Code: [Generative Adversarial Networks \(GANs\)](#).

## 3 Additional sources

Lecture slides for this course are accompanied by the AI glossary, which contains basic AL and ML definition and description of most popular algorithms.

## 4 Data upload

In order to run some of the notebooks, you will need to load external (e.g. a \*.csv file). In this case, there will be a link to this file. Please load the data file by pressing the "Files" link on the left (Fig.1). The file should appear in the list of files on the left.

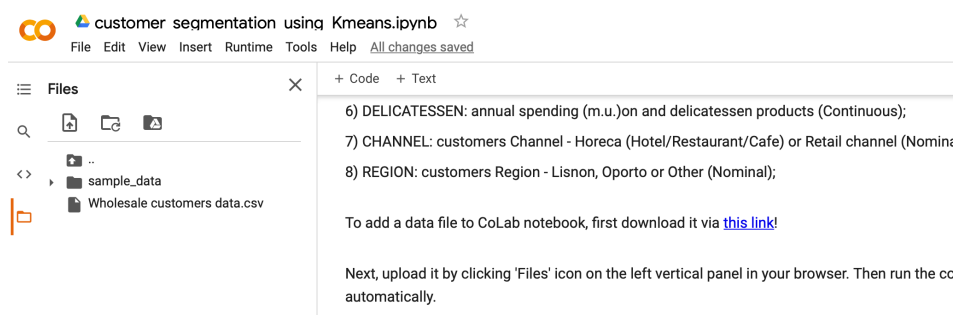


Figure 1: How to upload a data file.

Remember that uploaded files will be deleted when this you close the notebook.