

Brief Code Walk-Through

In order to create a system that can forecast consumers' next purchases based on information from prior transactions as well as from customer meta data, we specifically targeted the dataset of the H&M Group (in-person shop and online store). For our use, we have used the Kaggle json credential file and script to get and import this dataset from Kaggle.

To confirm that the distribution of data points matched the values available in the dataset's features, we first did an exploratory data analysis (EDA). We next took the necessary steps for any missing values and unimportant features.

We established a calendar of dates spanning from 22nd March 2020 to 22nd June 2020 in order to achieve our goal of forecasting the purchases made by clients in the following 90 days from their individual prior transactions.

The feature engineering technique was used to shape our data points into the necessary data kinds, as well as to merge different data sets together to produce better dimensionality and information in a single final data set to be used for predictive model construction.

We also built our desired (needed) dependent variable using the cumulative output of different features in the data set, as well as some independent variables. In our data set, we took into account customers' historical purchasing behaviour, and the dependent variable 'next 90 days purchase' was serviced in accordance with the models' specifications.

Following all of these stages, we used three distinct models: Decision Tree Classifier, Random Forest Classifier, and Gradient Boosting Model to predict whether or not a consumer will make a purchase within 90 days of their previous transaction. We also looked for crucial features and trained our models in each scenario after applying hyper-parameter tuning to each one to acquire the optimal parameters set.