



MerkleSokrati assignment1

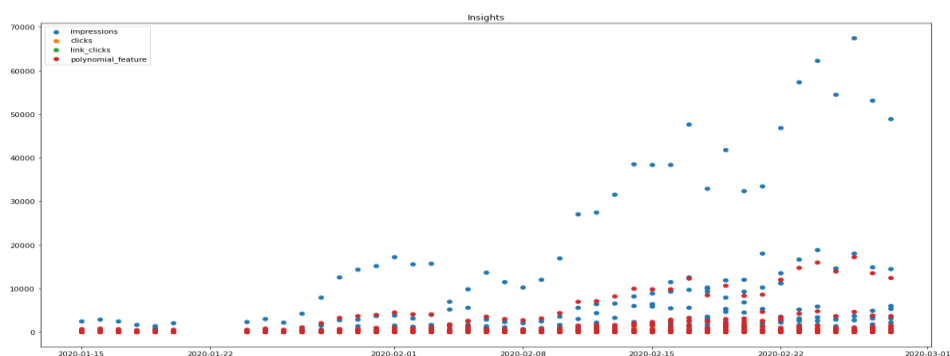
Machine learning model prediction of campaigns insights with Python and Scikit Learn

Summary:

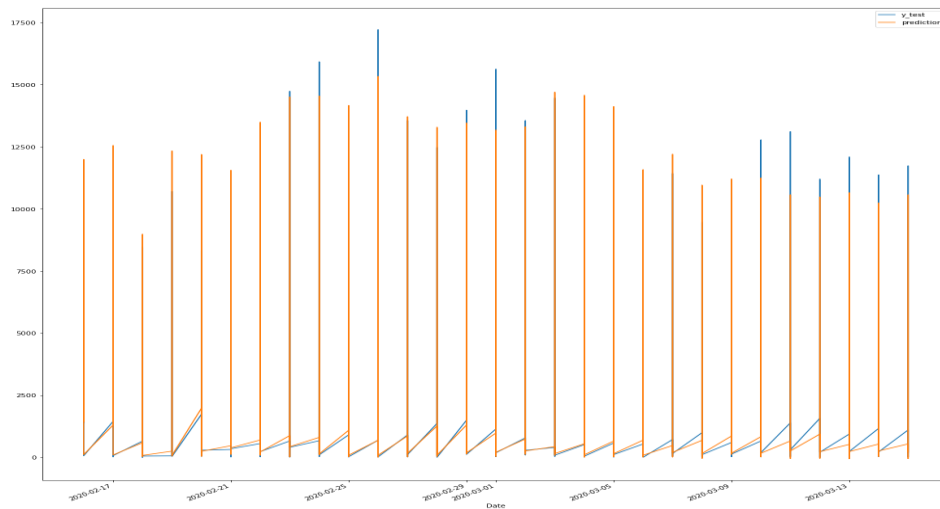
Methodology:

- Build a first model with standard hyper parameters values.
- Results measure.
- Build an optimized model with other values.
- Results measure and comparison with standard model.

My Machine Learning model is built in Python with Scikit Learn library. Problem to solve was a regression problem to predict insights of advertisements campaigns. After loading dataset provided, I have dropped missing values rows to start with clean data. I've preprocessing data in other different ways. Firstly, I have built a new column from three insights columns with Scikit Learn Polynomial Features. It's a feature engineering method and this column will be my target.



Secondly, I have made advanced preprocessing features pipelines to process differently numerical features and categorical features in order to standardize data. Then I have used SGDRegressor estimator to predict insight.



Score accuracy on test set are around 97%.

Finally, I've test some extra optimizations like features selections with SelectKBest and Anova statistical test to determine which feature could be dropped in order to increase accuracy. Unfortunately, results are close to standard results but these last are enough satisfying in my opinion.

Doc ref: MerkleSokrati_assignment1.ipynb

Author: Yoan Capron

capronyoan@gmail.com

+33610691623