

Inventory Forecasting

Description

At PubMatic, we receive billions of ad requests every day. Having a good sense of the number of ad requests we'd receive going ahead plays a vital role in infrastructure capacity planning, financial forecasting and sales planning (demand and supply)

Goal of this ML project is to accurately predict how many ad requests would we receive on a given day (up to 30 to 90 days in the future). A forecast would be reasonably accurate if it is within 10% of the actual number for 90% of the days.

Input

Input data set that can be used for training and validation can be downloaded from [here](#)

Data Description

The file has 3 columns – date, total requests and paid impressions where date represents the date in yyyy/mm/dd format for which the actual data is known. Total requests represent the number of ad requests that PubMatic received on the specific date. Paid impressions represent the number of ad requests that PubMatic monetized on that date.

Test data set that would be used to measure the accuracy of the ML model can be downloaded from [here](#). Please do not use the test data set for training nor tuning the model in any way. Please use this only to present the results around how accurate the trained model was while generating forecasts for days that were not seen during the training and validation phase.

Output / Deliverables

- ML model that meets the goals defined above. Please share the visualizations and results in a graphical manner (actual numbers versus forecasted / predicted numbers for the data in the test data set)
- Document describing the various algorithms and approaches that were evaluated (at least 3 different algorithms), their respective accuracy, pros and cons and the approach that was finalized and why. Briefly describe aspects like the loss functions used, hyper parameters that were finalized and why, data preparation / enrichment techniques that were applied if any and the impact on accuracy as you forecast further into the future for dates from the test set.

Stretch Goals / Enhancements

1. Besides generating a day wise forecast of ad requests, also predict the corresponding number of paid impressions / fill rate for the data in the test data set.
2. Granular forecasts – Forecast day wise ad requests at a geo, platform and publisher level for the next 7 to 30 days. A data set providing platform wise numbers can be found [here](#) (please split this into training and test data sets as part of the exercise)