Overview of DFE Run

Automated

December 21, 2022

Abstract

This is an automatically generated report. It is meant ot illustrate the kinds of analysis that can be performed in LightR.

1 Overview

- The report was generated on Jan 1, 2023.

- The number of data sets was 42.

- The number of formulas per model was 5.

2 Basic Feature Engineering

- The number of times this was invoked was 42

- The number of errors reported during the first round of feature engineering was 0.

- The average time taken was 1.234

Distribution of error codes is shown in Table 1

Error Code	Number
0	42
1	0

Table 1: Error Counts for basic feature engineering

Formula	Error Code	Number
no lag	0	42
no lag	1	0
lag 1	0	42
lag 1	1	0

Table 2: Error Counts, broken down by formula

3 Formula Specific Feature Engineering

- The number of times this was invoked was 42

- The number of data sets that had at least one error was 0.

- The average time taken was 1.234

Distribution of error codes is shown in Table 2

4 Model Building

- The number of data sets for which model building was attempted was X

- The number of models attempted was X

- The number of models that were built was X

- Summary statistics of time (in seconds) to build a model

  - Minimum X
  - Maximum X
  - Average X

- Distribution of build times is in Figure 1

5 Error Codes

5.1 Basic Feature Engineering

See Table 5.1

5.2 Formula Specific Feature Engineering

See Table 5.2

Step-by-Step Procedure

We use total pageviews as the example metric. Denote $X_{i,t}$ to be the pageviews from member i on day t.

Stage One

Input: per-member, per-day data (i.e. $X_{i,t}$)
Output: 6 numbers: (sum_treatment, sum_square_treatment, n_treatment) and (sum_control, sum_square_control, n_control)

For the Treatment variant:

1. Aggregate across days for each member in treatment. For member i in treatment, compute his total pageviews across all T days by summing his daily pageviews: $S_i = \sum_{t=1}^{T} X_{i,t}$
2. Aggregate across members. sum_treatment $= \sum_i S_i$, sum_square_treatment $= \sum_i S_i^2$, n_treatment = COUNT(DISTINCT members in treatment)

Repeat the same for the Control variant, and get sum_control, sum_square_control and n_control.

Figure 1: Distribution of model build times (seconds)

Error Code	Explanation
0	No Error

Table 3: Explanation of Error Codes

Error Code	Explanation
0	No Error

Table 4: Explanation of Error Codes