

Experiment: Temporal Features

 Creator **Olusanmi Hundogan**

 Created **Jan 10, 2024, 00:16**

 Last updated **Apr 30, 2024, 11:50**

Files

- notebooks/analyze_temporal_features.py
- experiments/experiment_temporal_features.py
- temporal-feature-analysis.pdf / analyze_temporal_features.html

Motivation

Currently the model treats every data point without integrating any temporal information. Every row is treated independently from every other row. For this experiment, we investigated the effects of temporal features on the model's performance. This write up is just a summary of results coming from an [earlier](#) analysis. Refer to this study to see the details.

Design

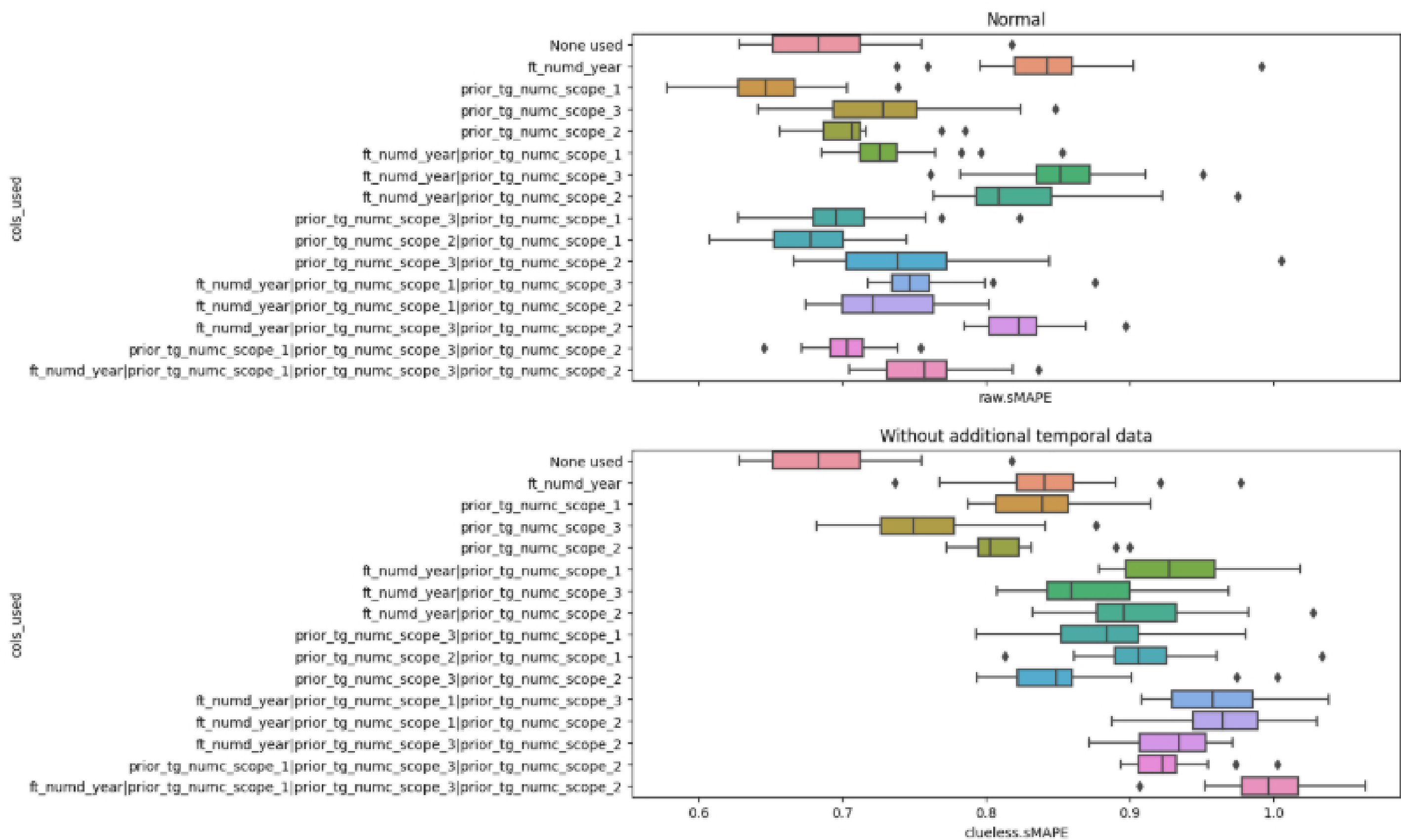
For this experiment we trained a model-pipeline for scope 1 and evaluated its performance. Each pipeline configuration uses an element taken from the power set of feature combinations. These features were:

- ft_numd_year: The current year of the data point
- ft_prior_tg_numc_scope_1: The company's previous scope 1 value
- ft_prior_tg_numc_scope_2: The company's previous scope 2 value
- ft_prior_tg_numc_scope_3: The company's previous scope 3 value

Each configuration was trained and evaluated 10 times on 25% of the companies with at least one previous year. We evaluate the model with the feature set present and without the feature set present during the validation phase to discern the impact of these features.

Results and Insight

This figure shows the effects of each feature set configuration on the sMAPE value with and without the features present during validation. Scope 1 sMAPE is lowest if the previous year was used as a feature during training. Introducing the year as a feature always leads to worse sMAPE results. Introducing previous scope 2 or 3 to predict scope 1 seems detrimental to the model. We achieve the best results, by only including the corresponding previous scope to the scope currently predicted.



Decision

Update 10.01.2024

It is reasonable to use only the scope of the previous year of the scope category that we attempt to predict. Hence, if scope one is predicted, only include previous years scope 1 value and so on.

Update 29.03.2024

After running the same experiment with the new data set the results remain largely the same. Only `ft_numd_year` does not appear to keep its strong detrimental effect. However, the effect is not advantageous as well.

