Method Description

General Information

Type of Entry (*Academic, Practitioner, Researcher, Student*)	**Academic**
First Name	**Tiago**
Last Name	**Dantas**
Country	**Brazil**
Type of Affiliation (*University, Company-Organization, Individual*)	**University**
Affiliation	**PUC-RIO**

Team Members (*if applicable*):

1st Member	
First Name	Tiago
Last Name	Dantas
Country	Brazil
Affiliation	PUC-RIO
2nd Member	
First Name	Fernando
Last Name	Cyrino Oliveira
Country	Brazil
Affiliation	PUC-RIO

Information about the method utilized

Name of Method	BAGGED.CLUSTER.ETS
Type of Method (*Statistical, Machine Learning, Combination, Other*)	Combination
Short Description (up to 200 words)	The method combines Bagging, Exponential Smoothing and Clustering Methods. First, time series is transformed using Box-Cox transformation. The transformed time series are decomposed into additive terms (Seasonal, Trend and Remainder using STL decomposition if its seasonal or Trend and Remainder using Local polynomial regression if it is not seasonal). 1000 New versions of the remainder component are generated using bootstrap. The seasonal and trend (or just trend if it's not seasonal) are added back to the bootstrapped versions and then series are detransformed. The last observations of each new bootstrapped series are used as validation set. The ETS models are applied to each series, forecasts are made for the validation set and series are ranked according to their sMAPE. Afterwards, series are clusterized using PAM algorithm with euclidian distance. The number of clusters is determined using silhouette information. For each cluster a number of series (proportional to the size of each cluster) is picked. The series picked are the ones that performed better in each cluster (according to the rank in calculated in the validation set). In the 99 series are selected, forecasted (using ETS) and aggregated into one single output using the median.

Extended Description:

The codes were executed on aws cloud on c4.8xlarge machines.
In order to meet the deadline we were forced to make some choices. This are listed below:

- *Seasonality was identified using ETS() and only then the STL or Loess method was applied to decompose the time series.*
- *For weekly data: when time series were bigger than 1100 observations, only last 999 were considered (this is due to the amount of time necessary to run the procedure)*
- *For cases where the procedure couldn't run due to frequencies higher than ETS() function limit, the frequency was set to 1.*
- *Seed was set to 100*
- *The number o considered time series in the considered in the cluster phase was set to 300 but only 100 forecasts were aggregated using the median.*
- *A presentation about the method can be seen in the following link: https://forecasters.org/wp-content/uploads/gravity_forms/7-c6dd08fee7f0065037affb5b74fec20a/2017/07/isf2017_tiago_fernando_simple3.pdf*