CBIC 2019
Belém . Pará
14th BRAZILIAN COMPUTATIONAL INTELLIGENCE MEETING
3-6 NOVEMBER 2019 . BELÉM . BRAZIL

Very short-term wind energy forecasting based on stacking ensemble

Sinvaldo Rodrigues Moreno
Ramon Gomes da Silva
Matheus Henrique Dal Molin Ribeiro
Naylene Fraccanabbia

PUCPR
GRUPO MARISTA

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

Viviana Cocco Mariani
Leandro dos Santos Coelho

Presentation agenda

Introduction

Objectives

Dataset

Methodology

Results

Conclusion

Acknowledgments

Introduction

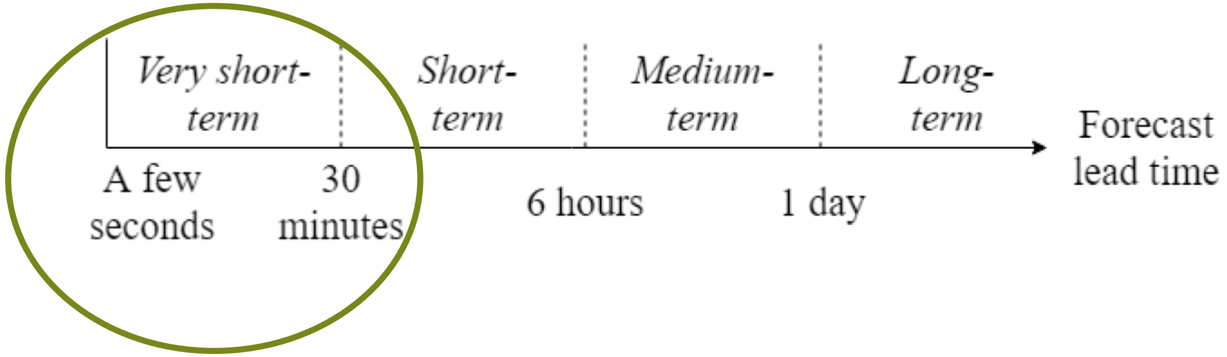
- **Wind power generation** is one of the technologies of electric production which **still in development in Brazil**;
- **In 2018**, wind energy represented **13.98% of the national energy consumption** and it supplied 38% of the population of Brazil, according to 2018 annual report of the Brazilian Wind Energy Association (ABEEolica);
- Due to the high level of uncertainty and the **chaotic fluctuations in wind speed**, the wind energy is classified as **intermittent source**;
- Because of the chaotic and uncontrollable behavior, **predicting wind energy accurately** as possible is a challenge;

Mega Watts	NEWS/2017	Total 2017	NEWS/2018	Total 2018
Total Global	48996	521774	46820	568409
Total Americas	10572	123091	11940	135041
EUA	7071	89047	7588	96635
Canada	341	12240	566	12816
Brazil	**2027**	**12769**	**1939**	**14707**
Mexico	478	4006	929	4935
Argentina	24	228	494	722
Chile	269	1418	204	1621

Introduction

Objective

- This paper proposes an application of a **stacking ensemble learning forecasting model** that combines **different algorithms (heterogeneous ensemble) to forecasting the wind power generation one-step ahead** (10 minutes ahead).



Dataset

Observation of wind turbine power generation **every 10 minutes;**

Turbine in a wind farm **located at Parazinho, RN, Brazil**;

The dataset period starts from August 01 2017 00:00h to August 31 2017 23:50h

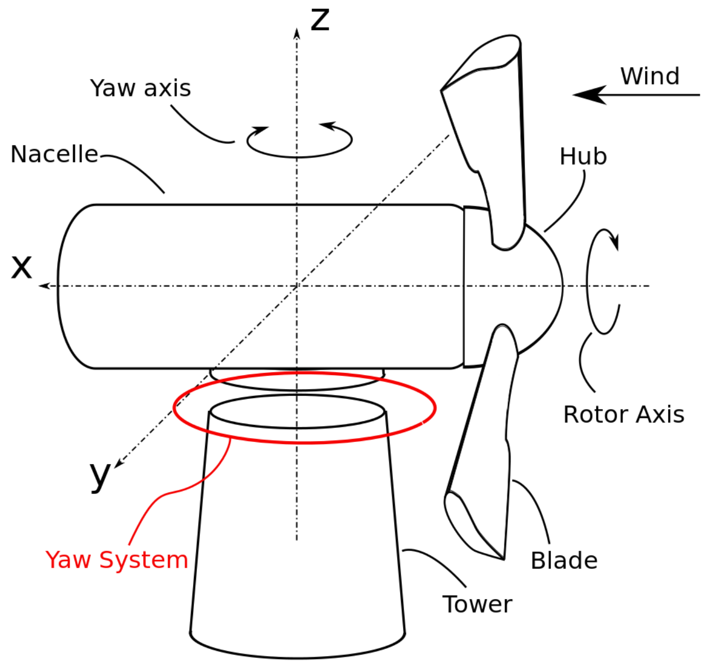
Observations number: 4439

Dataset

Eight variables, **Power is the system output**, others are inputs.

TABLE I
INPUTS AND OUTPUT OF THE SYSTEM

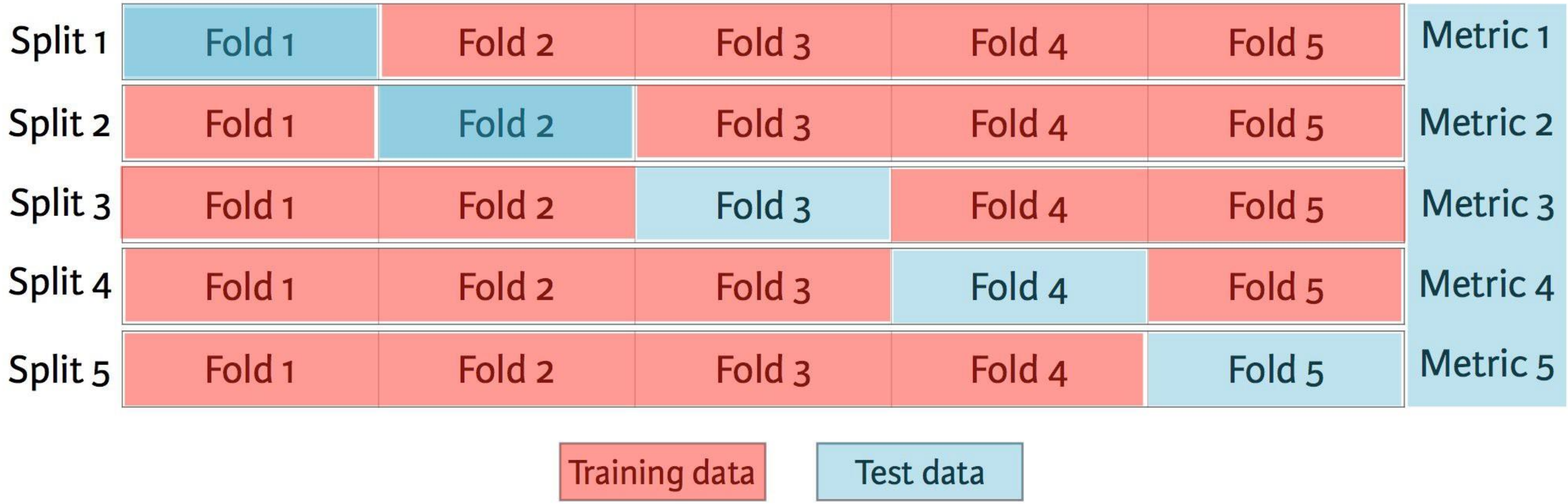
Type	Description	Unit Measure
Output	Power	KW
Input	Generator Bearing Temperature	Celsius
Input	Generator Bearing 2 Temperature	Celsius
Input	Generator Speed	RPM
Input	Wind Speed	m/s
Input	Wind Direction Absolute	Degrees
Input	Nacelee Direction	Degrees
Input	Ambient Temperature	Celsius



Dataset

TABLE II
NUMBER OF OBSERVATIONS USED FOR WIND POWER DATA ANALYSIS

Dataset	Percentage	Number of observations
Observed	100%	4438
Training	70%	3106
Test	30%	1332



Methodology

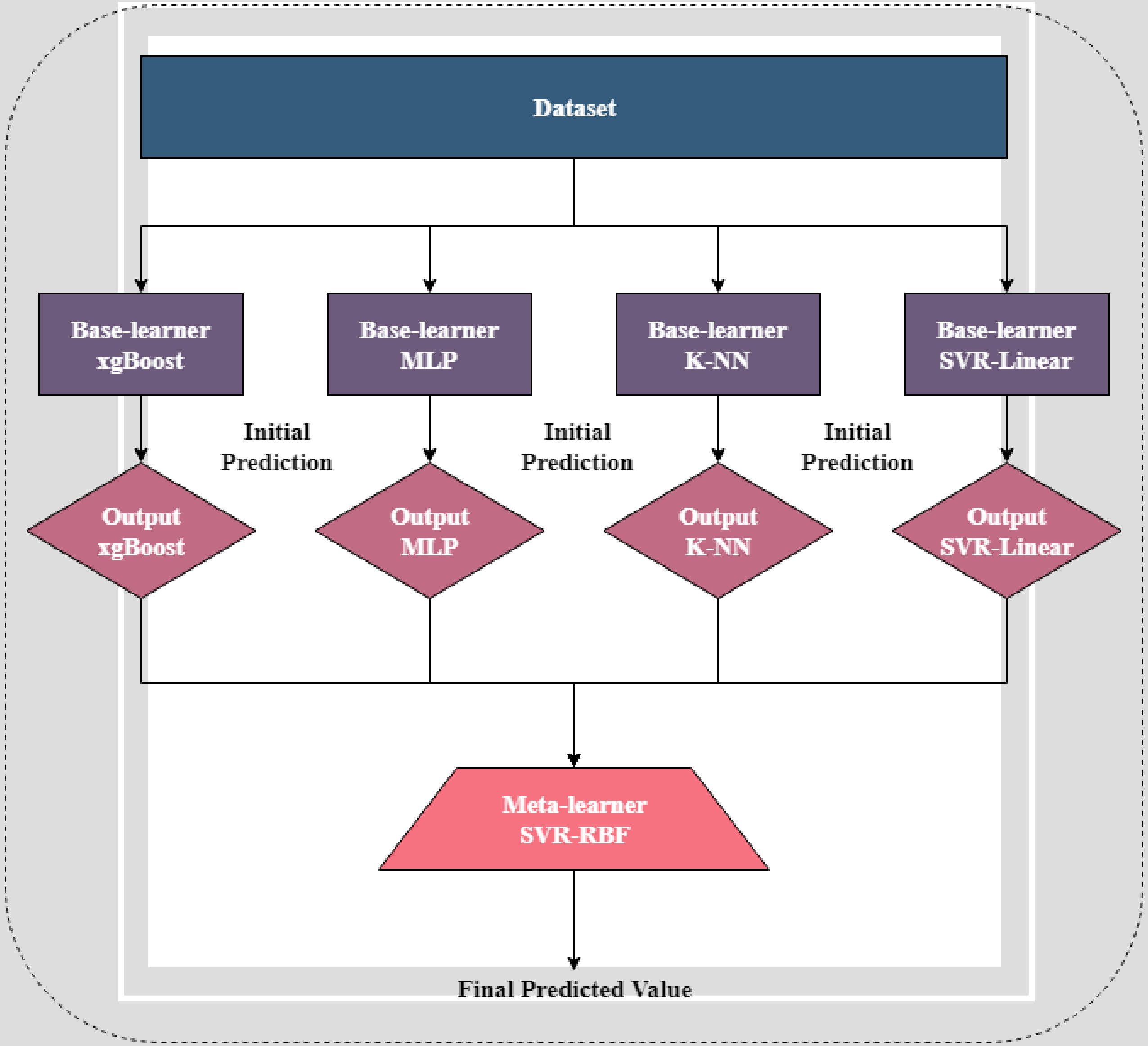
A Box-Cox transformation as preprocess;

Four different algorithms (models) as base learners:

- **XGBoost** (eXtreme Gradient Boosting)
- **MLP** (multilayer perceptron) neural network
- **K-NN** (K-nearest neighbors)
- **SVR-Linear** (Support Vector Regression - Linear)

Stacking

Meta learner: **SVR-RBF** (SVR radial basis function)



Methodology

- Performance measures: $\mathrm{RRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i},$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left[y_i(t) - \hat{y}_i(t)\right]^2}{\sum_{i=1}^{n}\left[y_i(t) - \overline{y}_i(t)\right]^2},$$

$$\mathrm{SSE} = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2,$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

- Statistical tests: $\mathrm{FD} = \frac{12n}{k(k+1)}\left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right],$

Friedman test

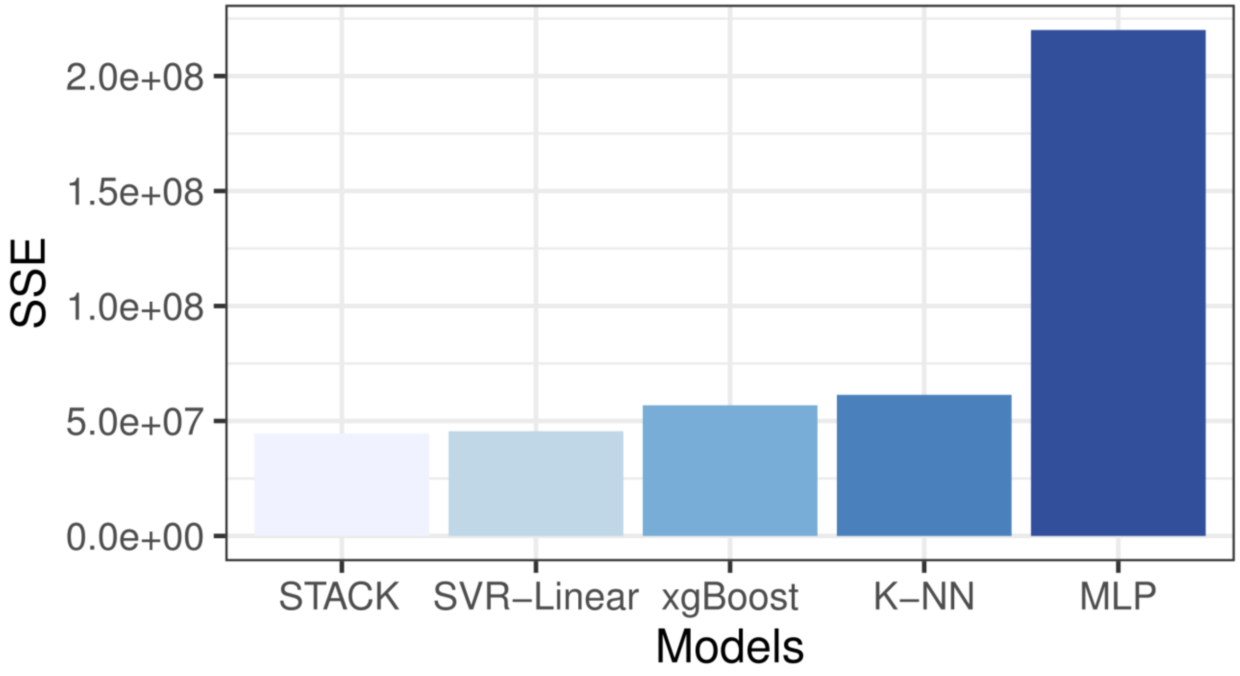
critical difference (CD)

$$\mathrm{CD} = \frac{q_{\infty,k,\alpha}}{\sqrt{2}}\sqrt{\frac{k(k+1)}{6}},$$

Nemenyi test

TABLE VI
CONTROL HYPER-PARAMETERS FOR META AND BASE-MODELS

Model	Control Hyperparameters	
SVR-RBF (STACK)	Kernel	Radial
	Sigma	150
	Cost	0.1
xgBoost	Boosting Iterations	50
	L2 Regularization (λ)	0.1
	L1 Regularization (α)	0.001
	Learning Rate	0.3
MLP	Hidden Units layer1	3
K-NN	Neighbors	13
SVR-Linear	Kernel	Linear
	Cost	4



Results

Results

TABLE VII
PERFORMANCE MEASURES OF THE BASE-MODELS AND STACKING ON TRAINING AND TEST SETS

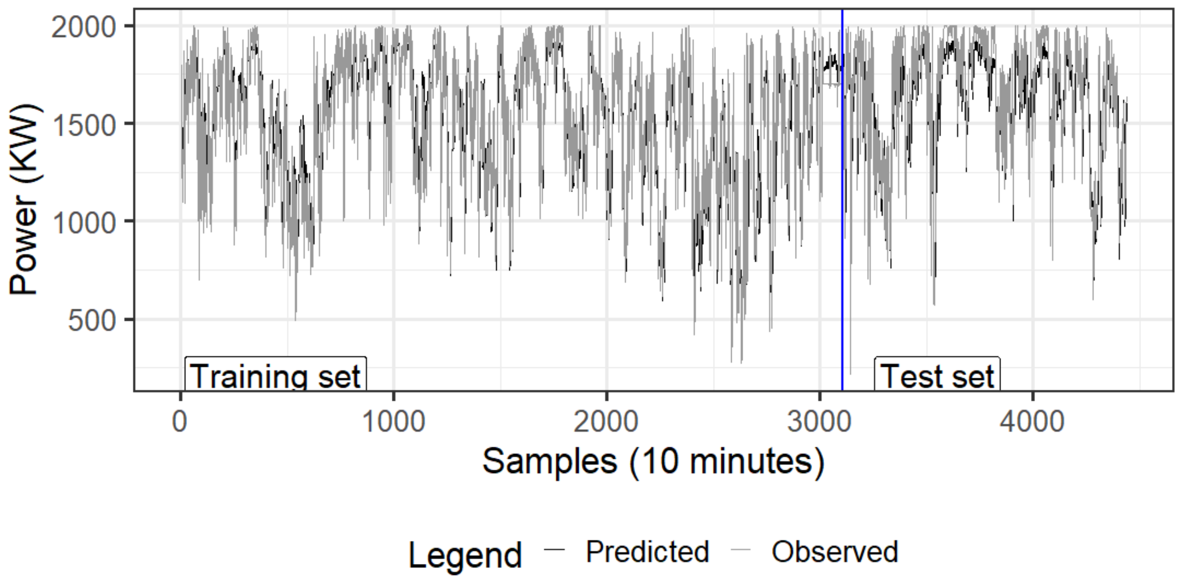
Training set			
Models	RRMSE	MAPE	R^2
STACK	0.0980	0.0818	0.8152
xgBoost	**0.0648**	**0.0527**	**0.9219**
MLP	0.2389	0.2209	-
K-NN	0.1283	0.1062	0.6809
SVR-Linear	0.1310	0.1093	0.6728
Test set			
Models	RRMSE	MAPE	R^2
STACK	**0.1101**	**0.0934**	**0.6690**
xgBoost	0.1243	0.1072	0.5938
MLP	0.2445	2.4529e+35	0.0008
K-NN	0.1291	0.1160	0.6001
SVR-Linear	0.1112	0.0971	0.6588

TABLE IV
RRMSE CRITERIA

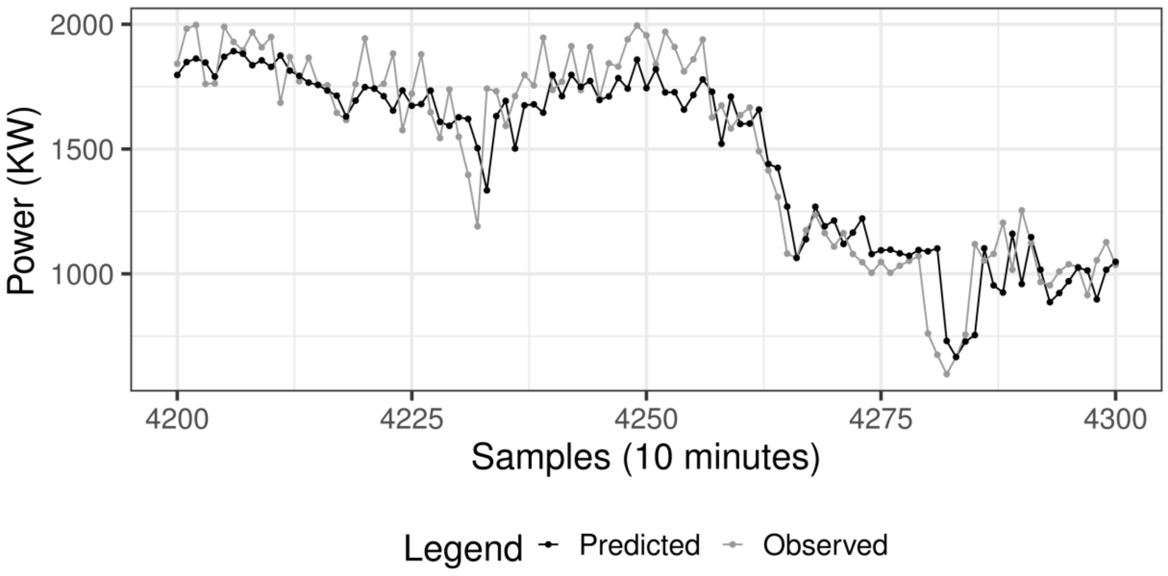
RRMSE (%)	Forecasting power
< 10	Excellent
10 - 20	Good
20 - 30	Reasonable
> 30	Incorrect

TABLE V
MAPE CRITERIA

MAPE (%)	Forecasting power
< 10	Excellent
10 - 20	Good
20 - 50	Reasonable
> 50	Incorrect



Prediction for whole dataset



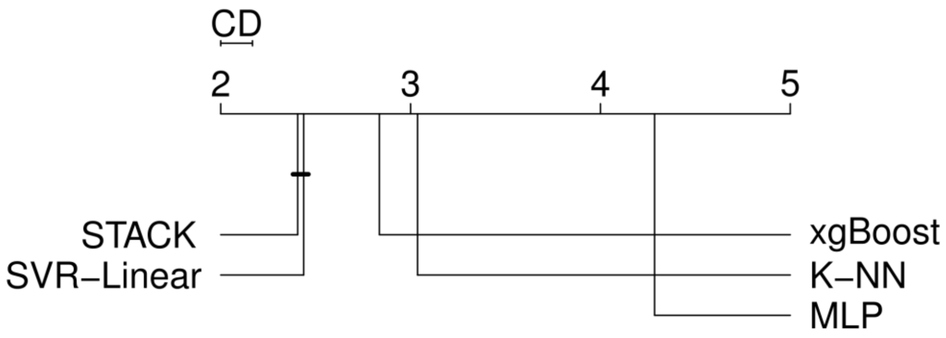
Samples from 4200 to 4300

Results

Results

- **Friedman test:**
- $\chi_4^2 = 1252.6, p\text{−value} < 0.05$

- There is **no** statistical difference between STACK and SVR-Linear
- However, STACK presents statistically lower error than others models.



Those algorithms that are **not joined** by a line can be regarded as **different**.

Nemenyi test:

- $\text{CD} = 0.16717, \text{degrees of freedom} = 6655,$ $p\text{−value} < 0.05$

critical difference (CD)

Conclusion

- A stacking ensemble of 4 heterogeneous base-learner models and 1 meta-learner was proposed;
- Wind power generation forecasting one observation ahead was made;
- The stacking ensemble was composed by XGBoost, MLP, SVR-Linear and k-NN, as base-learners in the first layer, and using SVR-RBF as meta-learner in the second layer.
- Stacking ensemble had a **better** performance than other approaches with individually analysis.
- **For future works**
  - adopt different combinations of models in both layers of the stacking ensemble;
  - increasing the number of base-learners.;
  - increasing the number of steps ahead to forecasting.

Acknowledgments

CAPES

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

FUNDAÇÃO ARAUCÁRIA
Apoio ao Desenvolvimento Científico e Tecnológico do Paraná

SCIENTIA, VITA ET FIDES
PUCPR

CBIC 2019
Belém . Pará
14th BRAZILIAN COMPUTATIONAL INTELLIGENCE MEETING
3-6 NOVEMBER 2019 . BELÉM . BRAZIL


WHY?
WHAT?
WHERE?
WHEN?
HOW?
WHO?
HOW?
WHERE?
WHAT?
WHO?
WHERE?
HOW? WHO?
WHAT? HOW? WHY?
HOW?


ザ・コーヒー
the coffee.

THANK YOU