Università degli Studi di Modena e Reggio Emilia
Dipartimento di Scienze e Metodi dell’Ingegneria

Corso di Laurea Magistrale

in Ingegneria Gestionale

Temperature Forecasting

Insegnamento: Data Science and Management

Studente: Rivi Riccardo

A.A.: 2022/2023

Jena Dataset

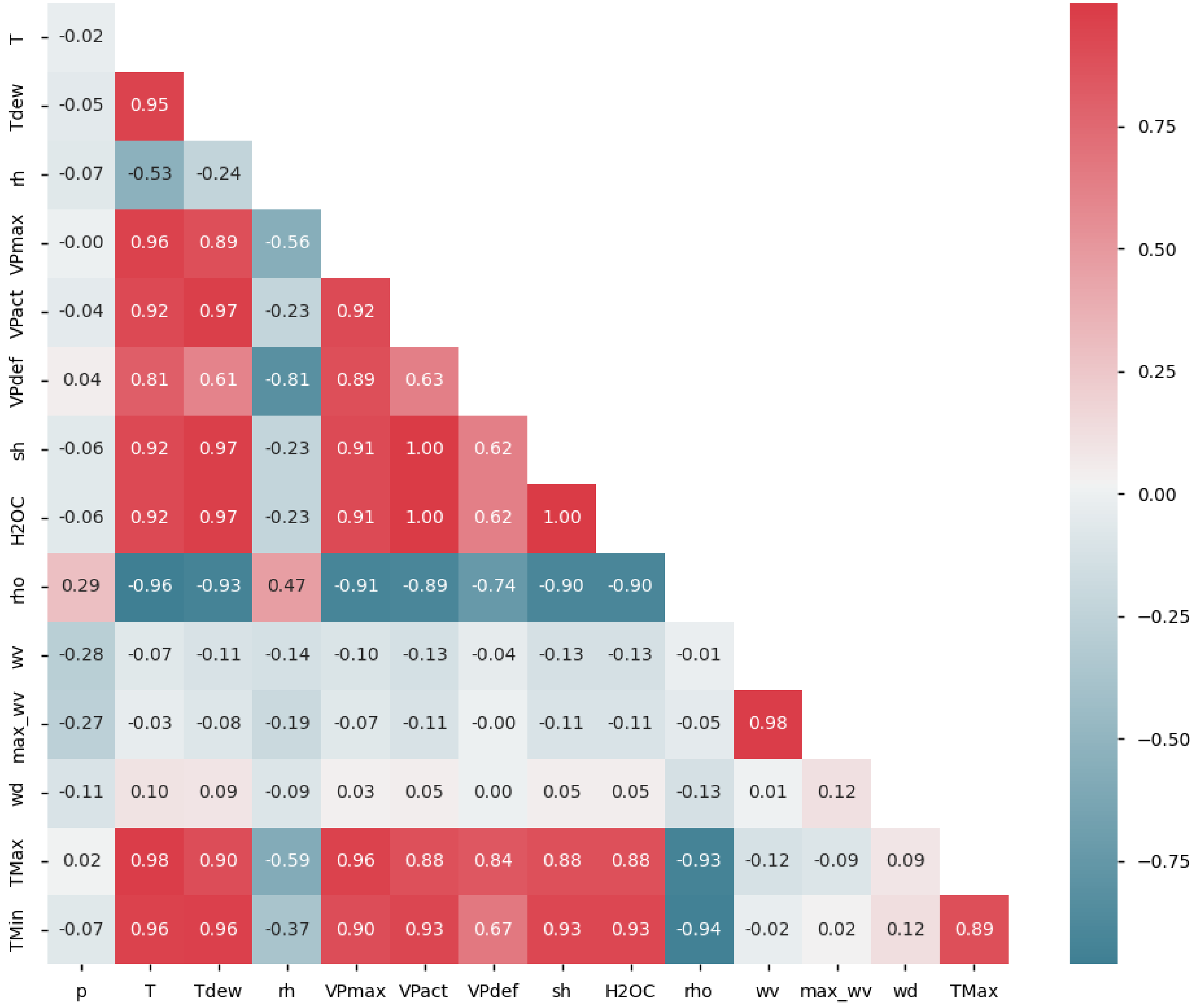
1/1/2009 - 1/1/2017

- *Date-time reference*: riferimento temporale di una certa misurazione
- *P (mbar)*: pressione atmosferica
- *T (degC)*: temperatura in Celsius
- *T (K)*: temperatura in Kelvin
- T_{dew} *(degC)*: temperatura di rugiada
- VP_{max} *(mbar)*: pressione di saturazione
- VP_{act} *(mbar)*: pressione di vapore nella miscela
- VP_{def} *(mbar)*: deficit di pressione di vapore: $VP_{max} - VP_{act}$
- *rh (%)*: umidità relativa
- *sh (g/kg)*: umidità specifica
- H_2*OC (mmol/mol)*: concentrazione o frazione molare di vapore
- *rho (g/*m^3*)*: densità dell’aria
- *wv (m/s)*: velocità del vento
- *max wv (m/s)*: massima velocità del vento
- *wd (deg)*: direzione del vento

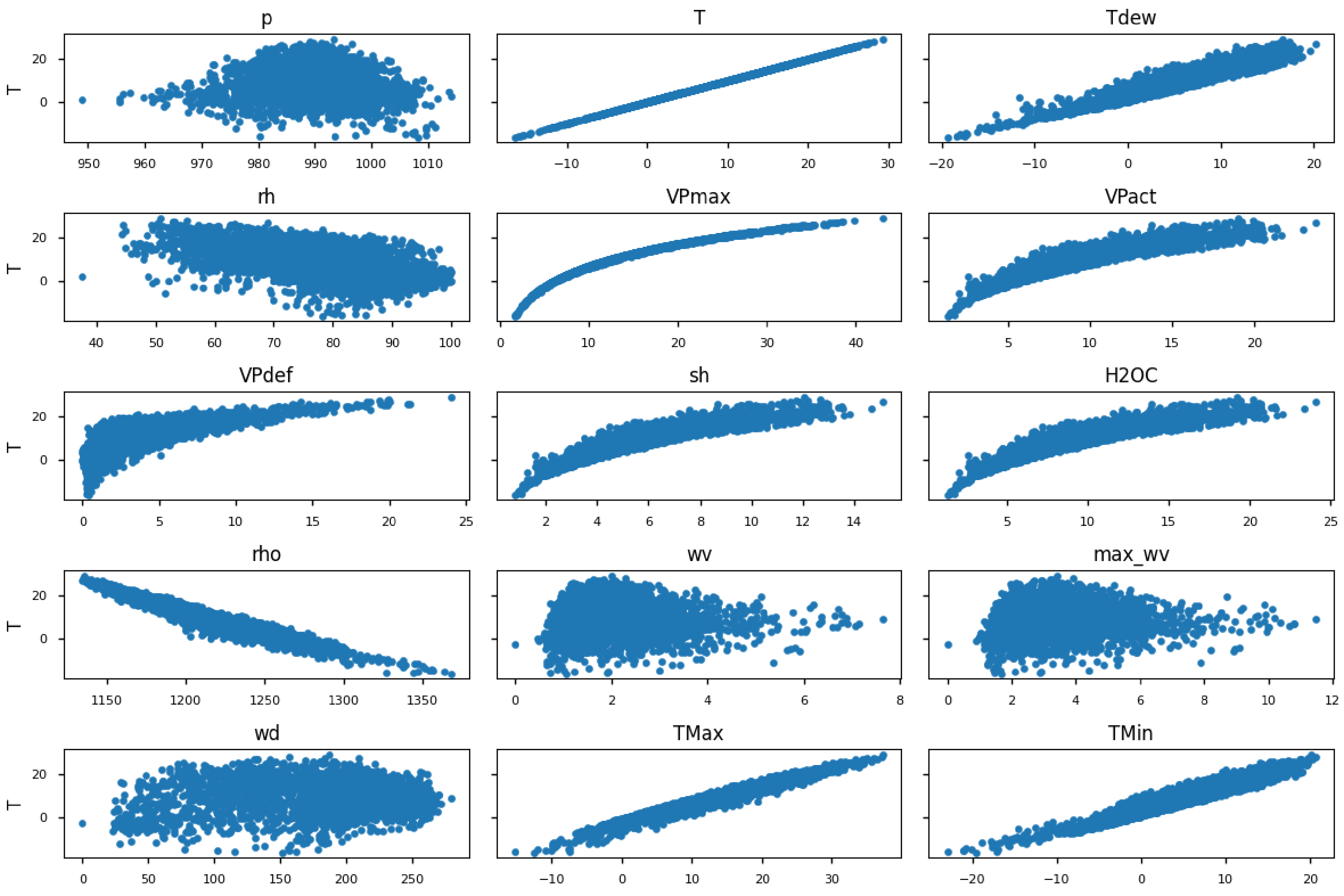
Settings

- **Granularità**: 10 min ⇨ valori medi giornalieri
- **Target**: *T (degC)*
- **Feature eliminata**: *T (K)*
- **Features aggiunte**: TMax e TMin, temperature massime e minime giornaliere
- **Lag temporale**: 3 giorni
- **Orizzonte di previsione**: 1 giorno
- **Repository**: https://github.com/rrivi17/TemperatureForecasting

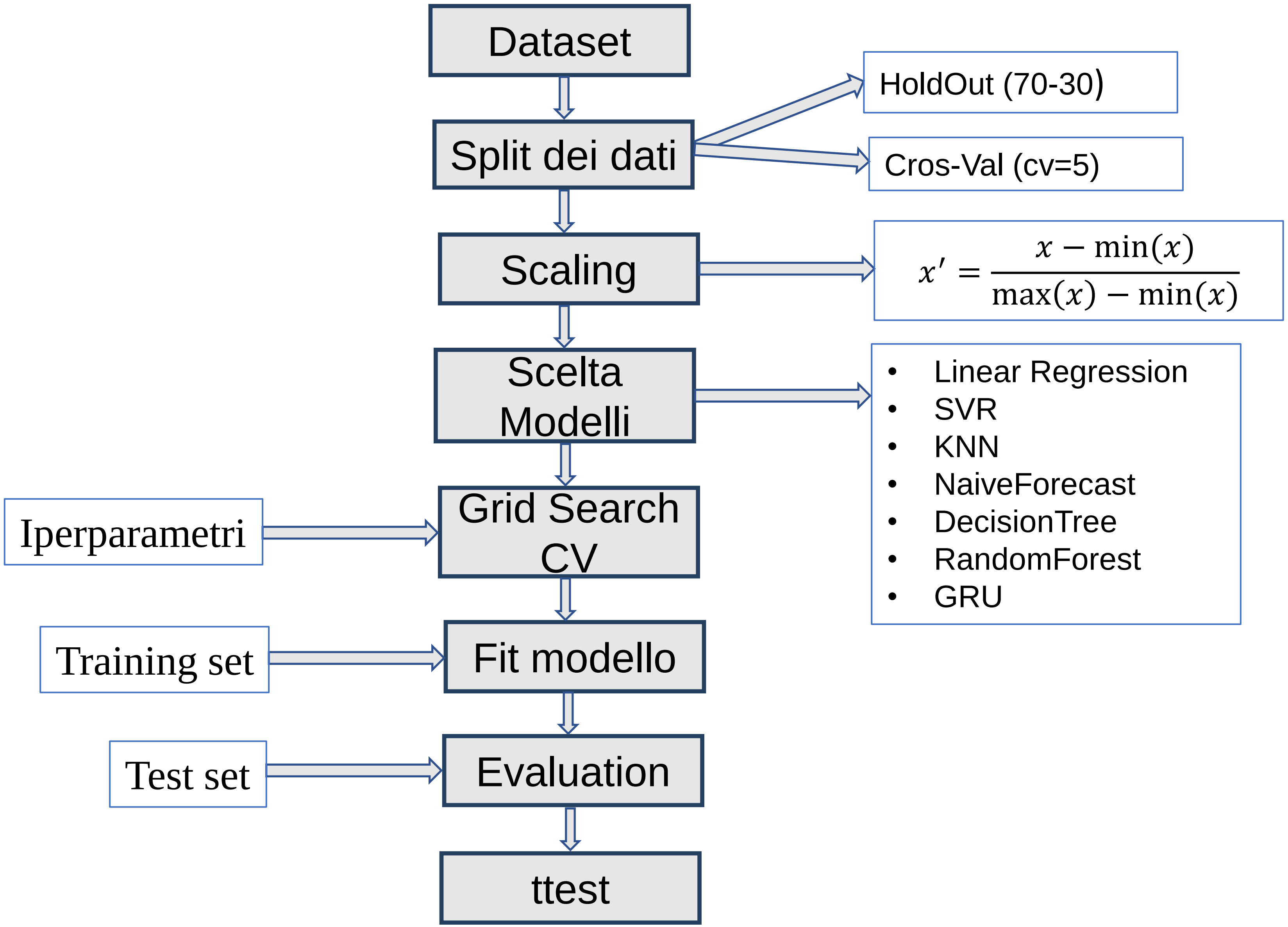
Matrice di Correlazione



Matrice Scatter plot



Procedimento



Metriche per Evaluation

$$MSE = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}$$

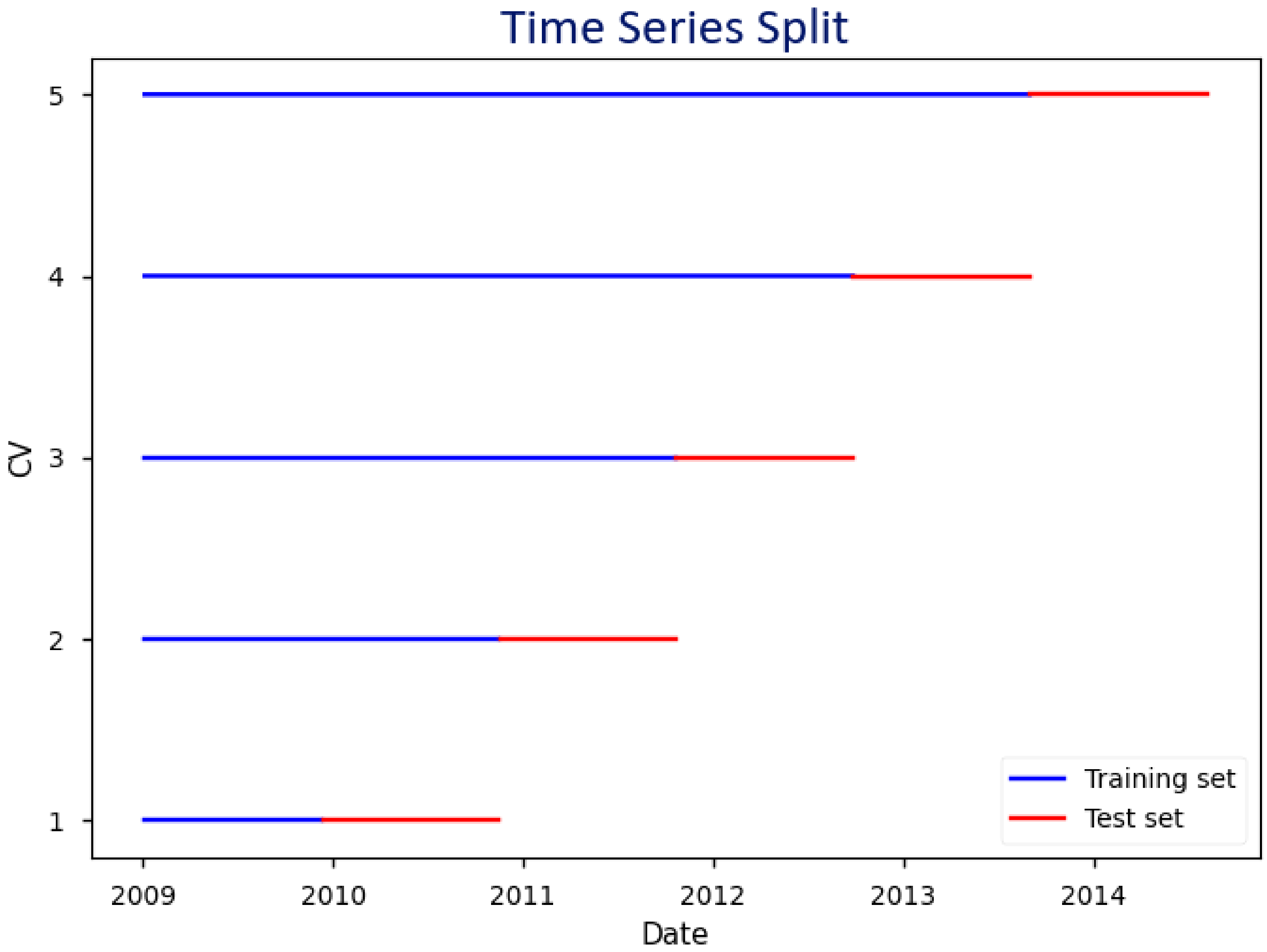
$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

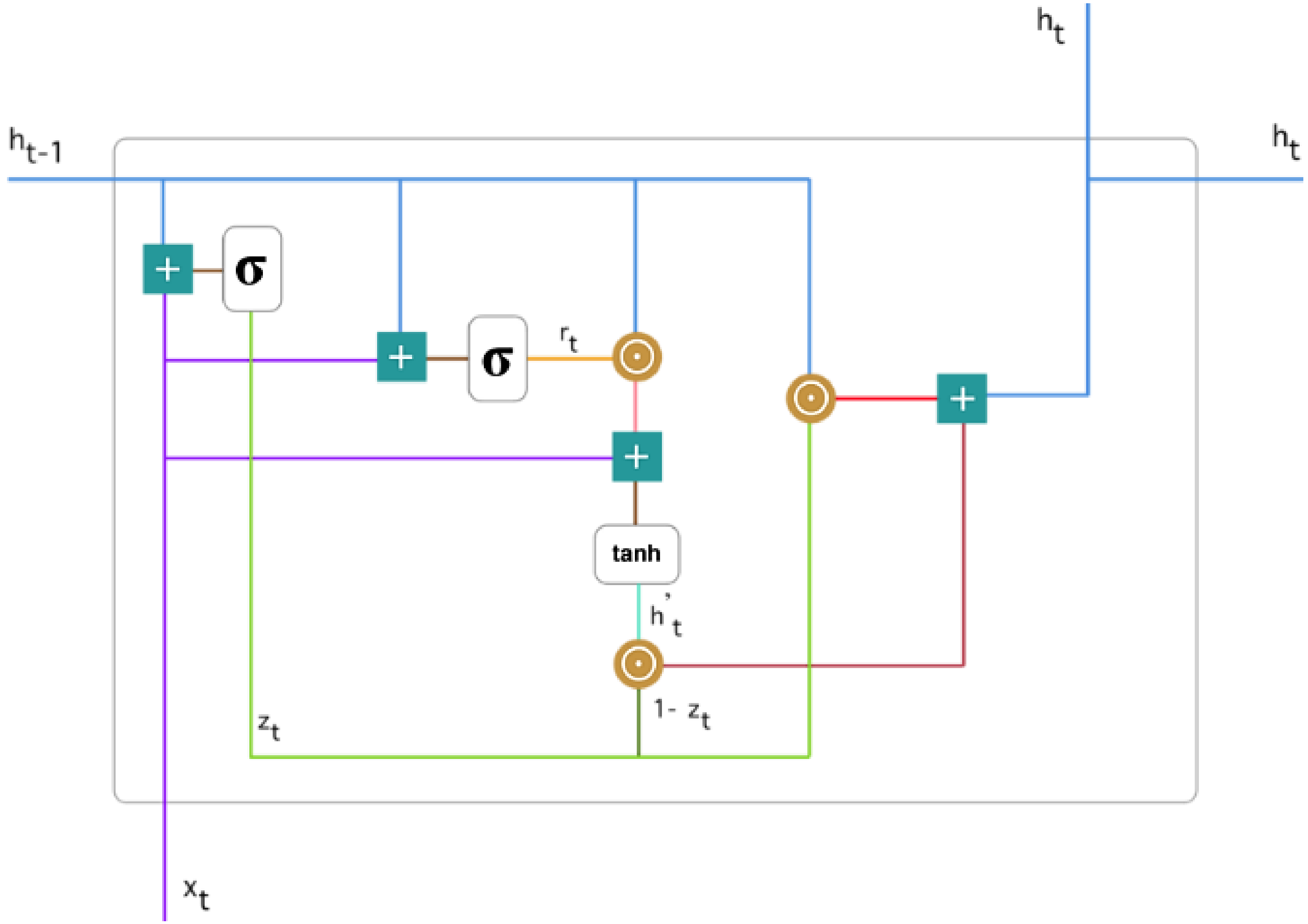
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{|y_i|} \times 100$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Cross Validation



Cella GRU

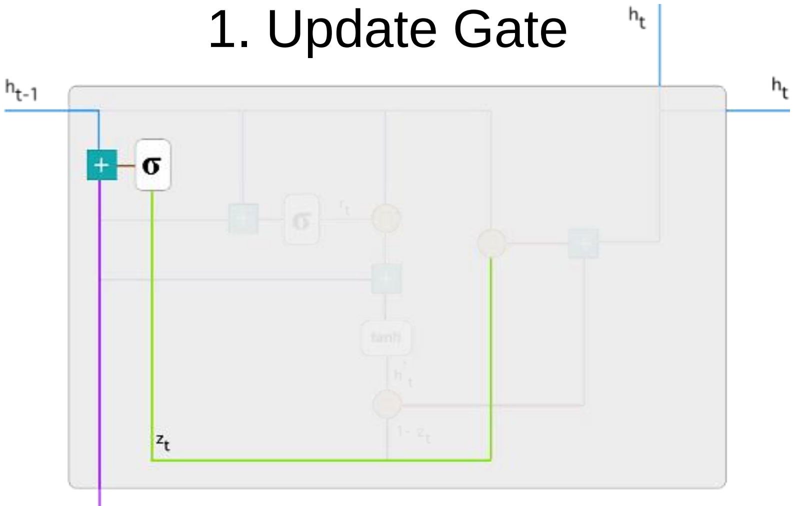


h_t
h_{t-1}
h_t
σ
σ
r_t
tanh
h'_t
z_t
1- z_t
x_t

σ
tanh
“plus” operation
“sigmoid” function
“Hadamard product” operation
“tanh” function

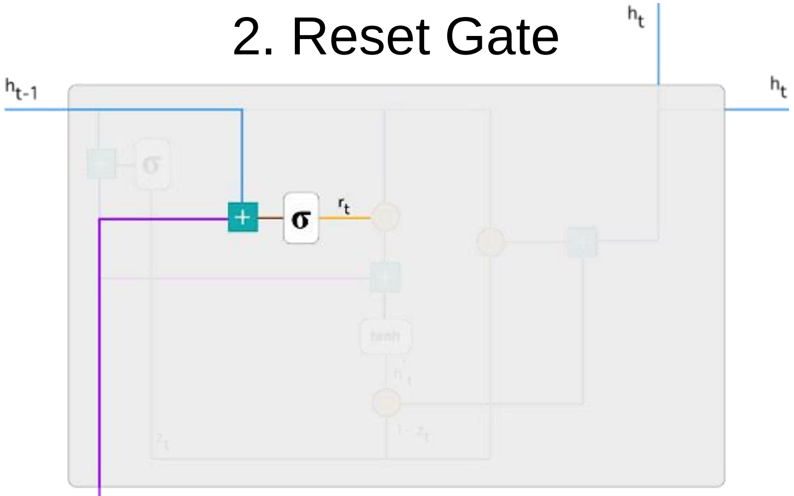
Cella GRU

1. Update Gate



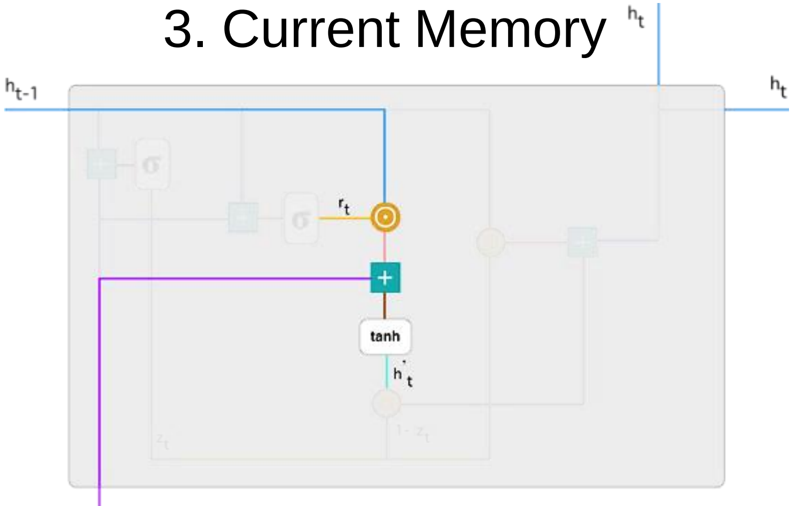
x_t $z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1})$

2. Reset Gate



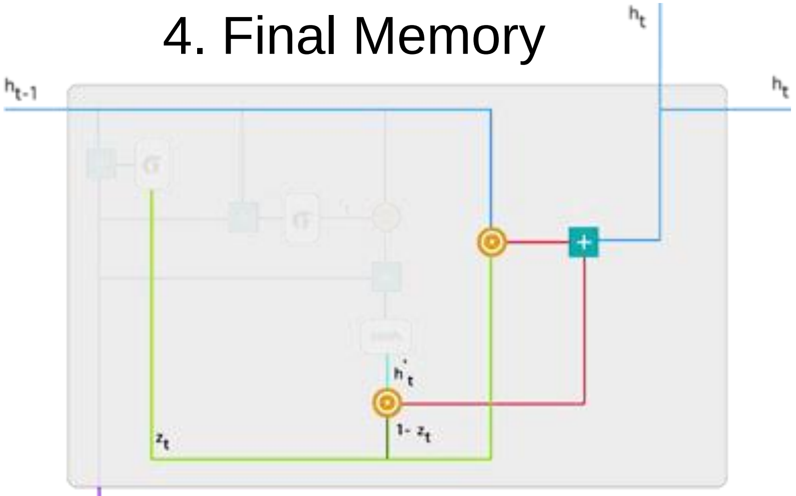
x_t $r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1})$

3. Current Memory



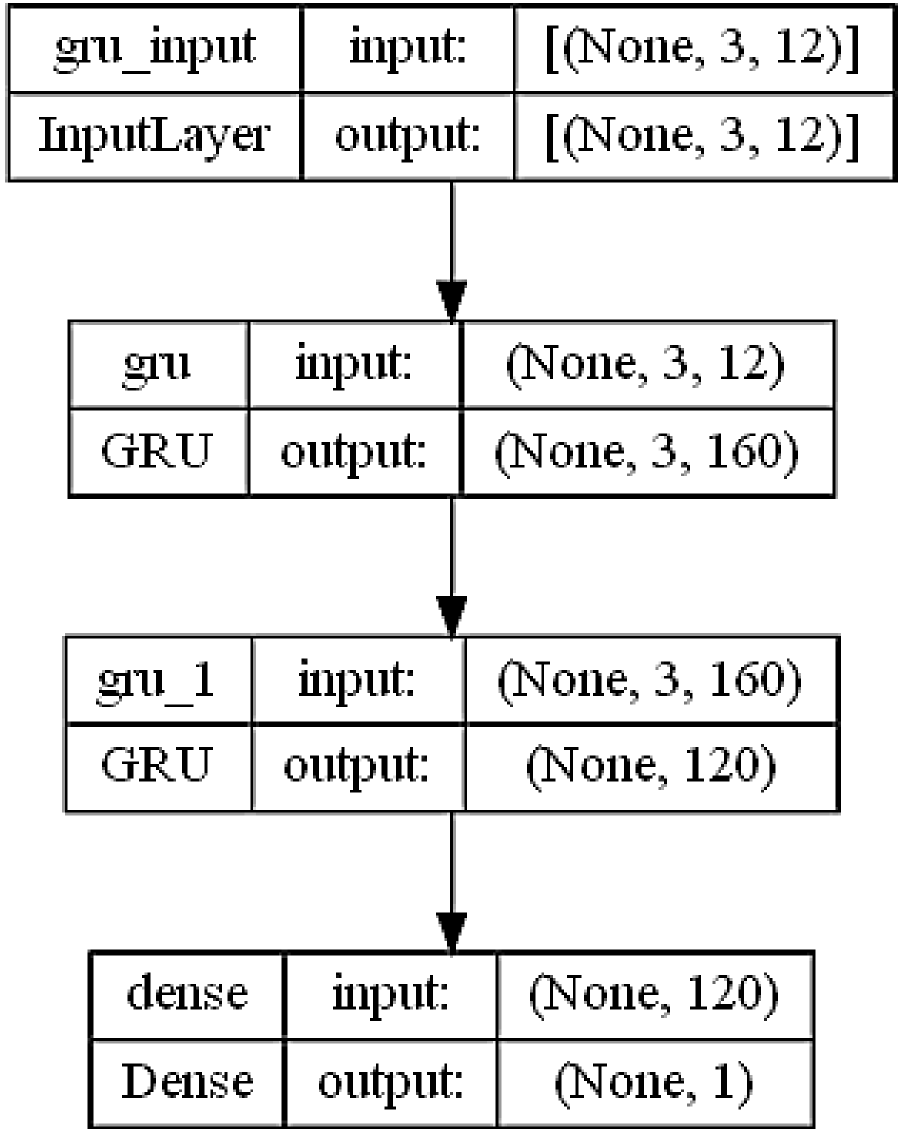
x_t $h'_t = tanh(Wx_t + r_t \odot Uh_{t-1})$

4. Final Memory

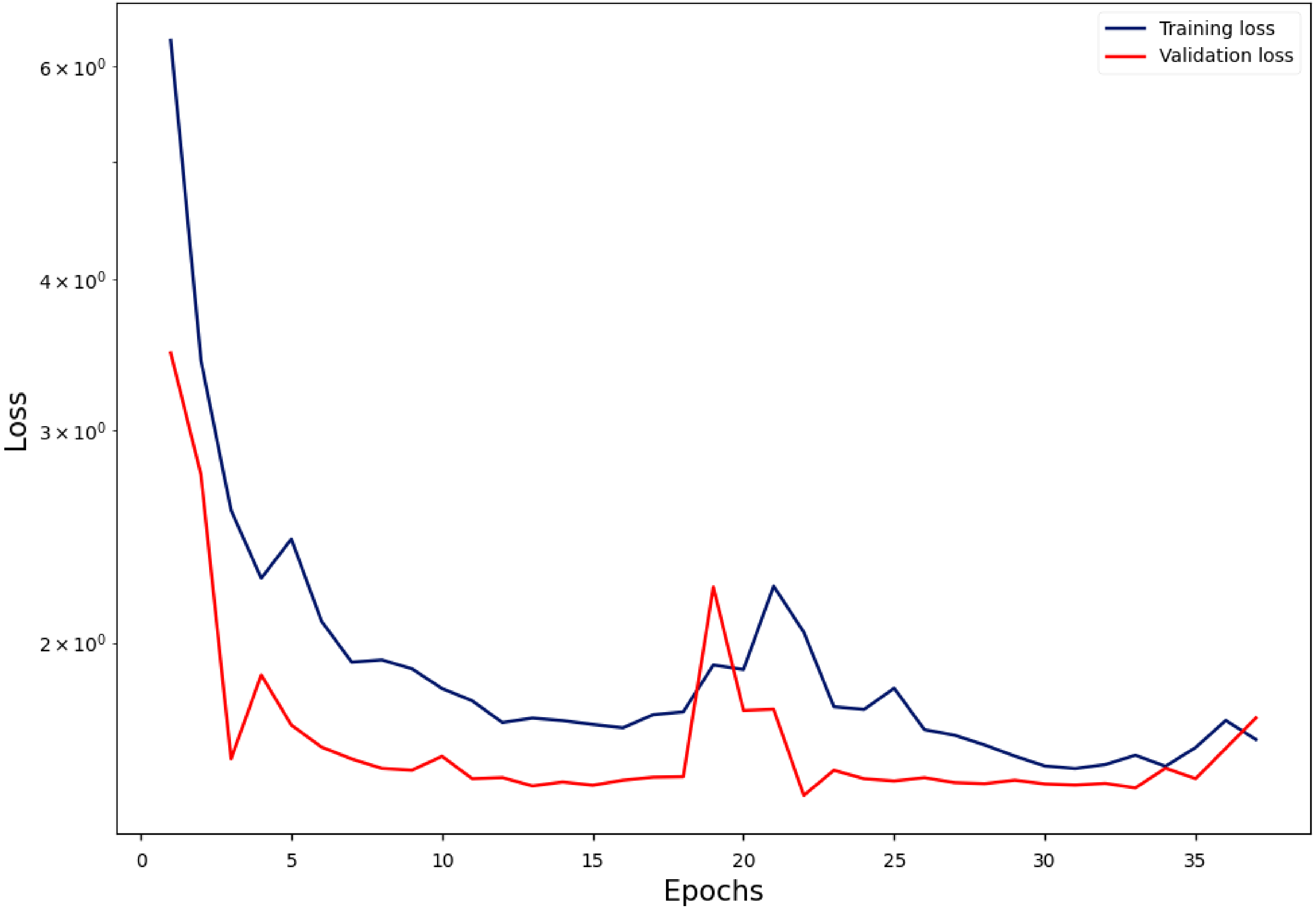


x_t $h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h'_t$

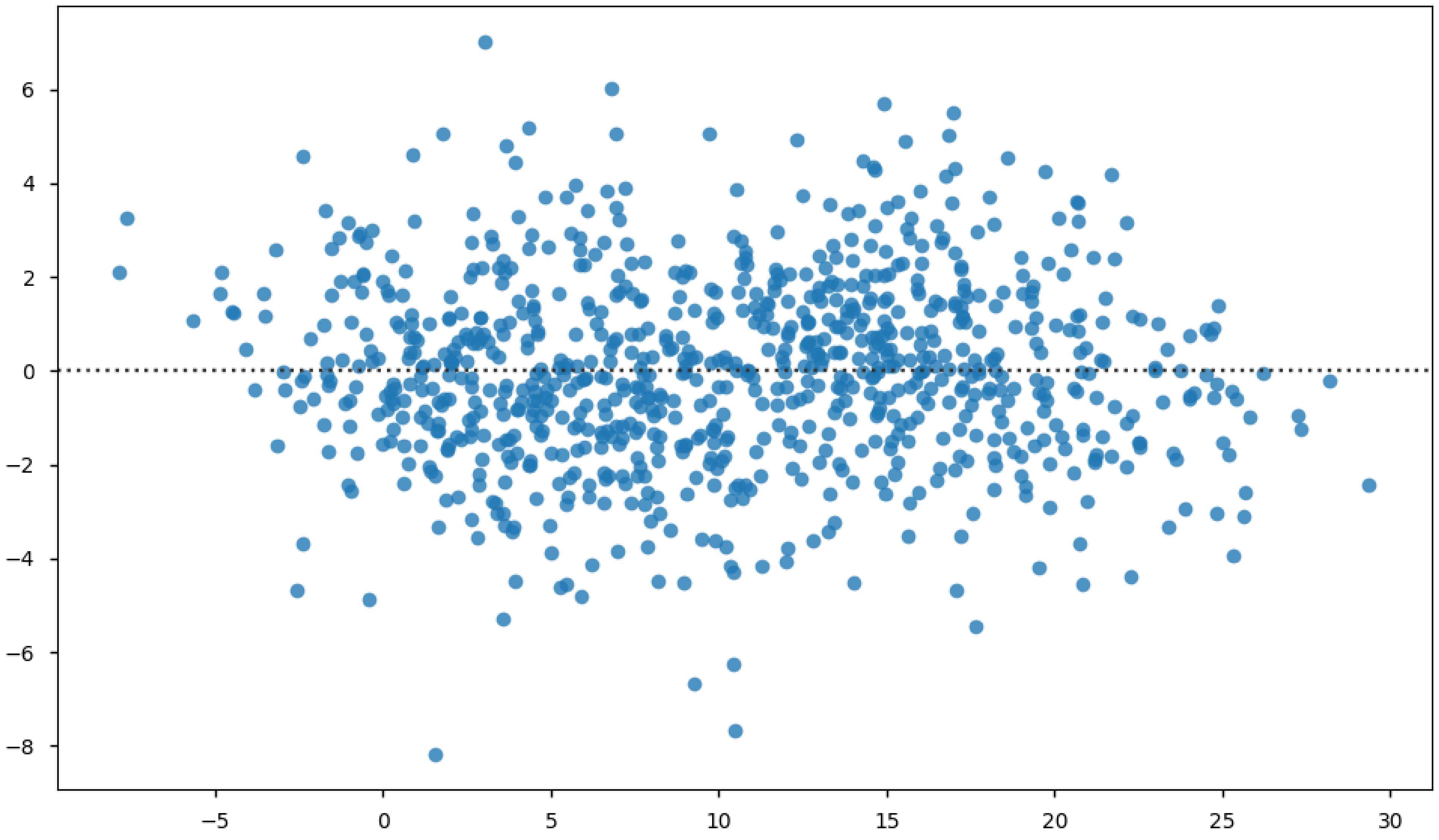
Struttura GRU



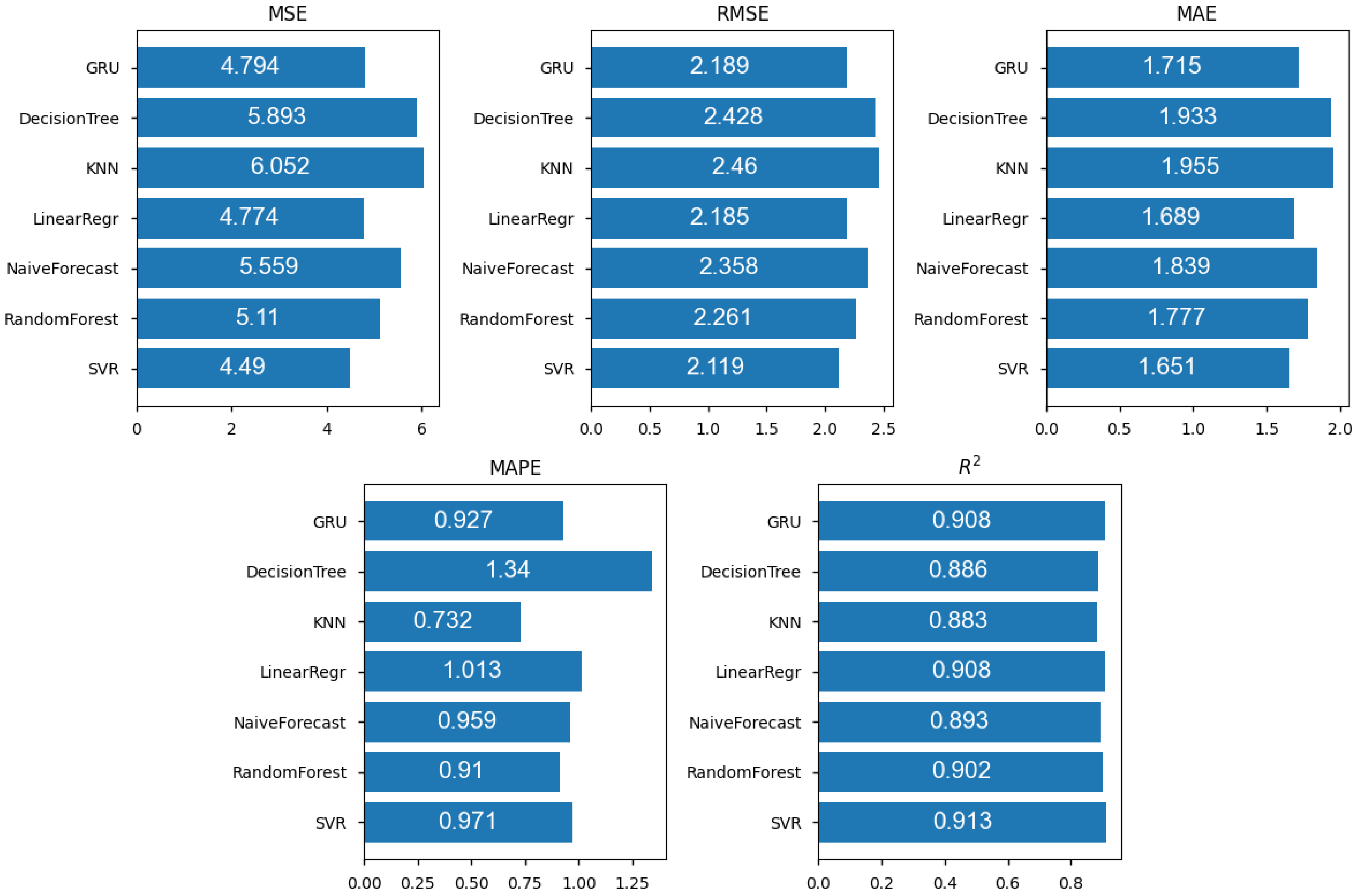
Learning Curve

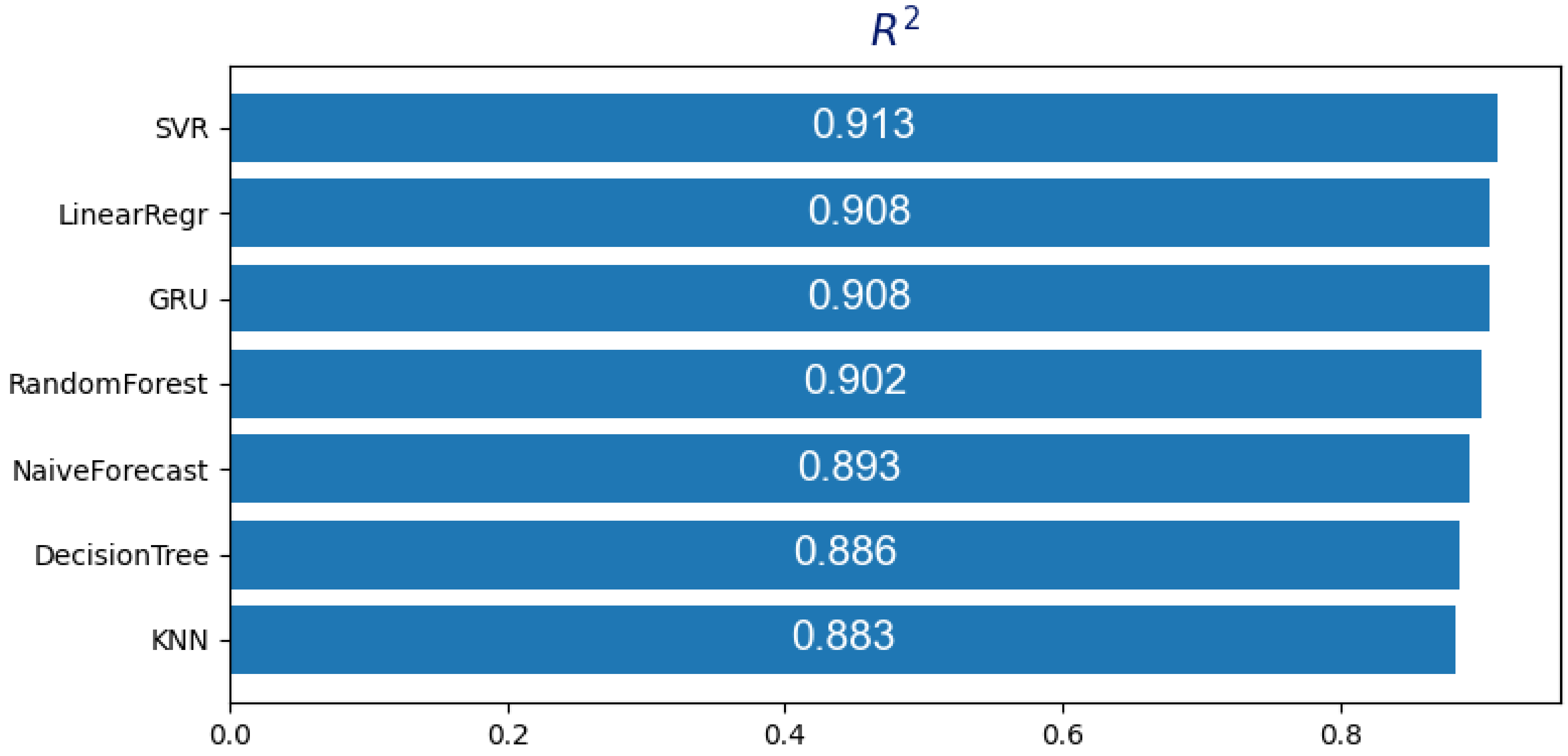


Residual Plot



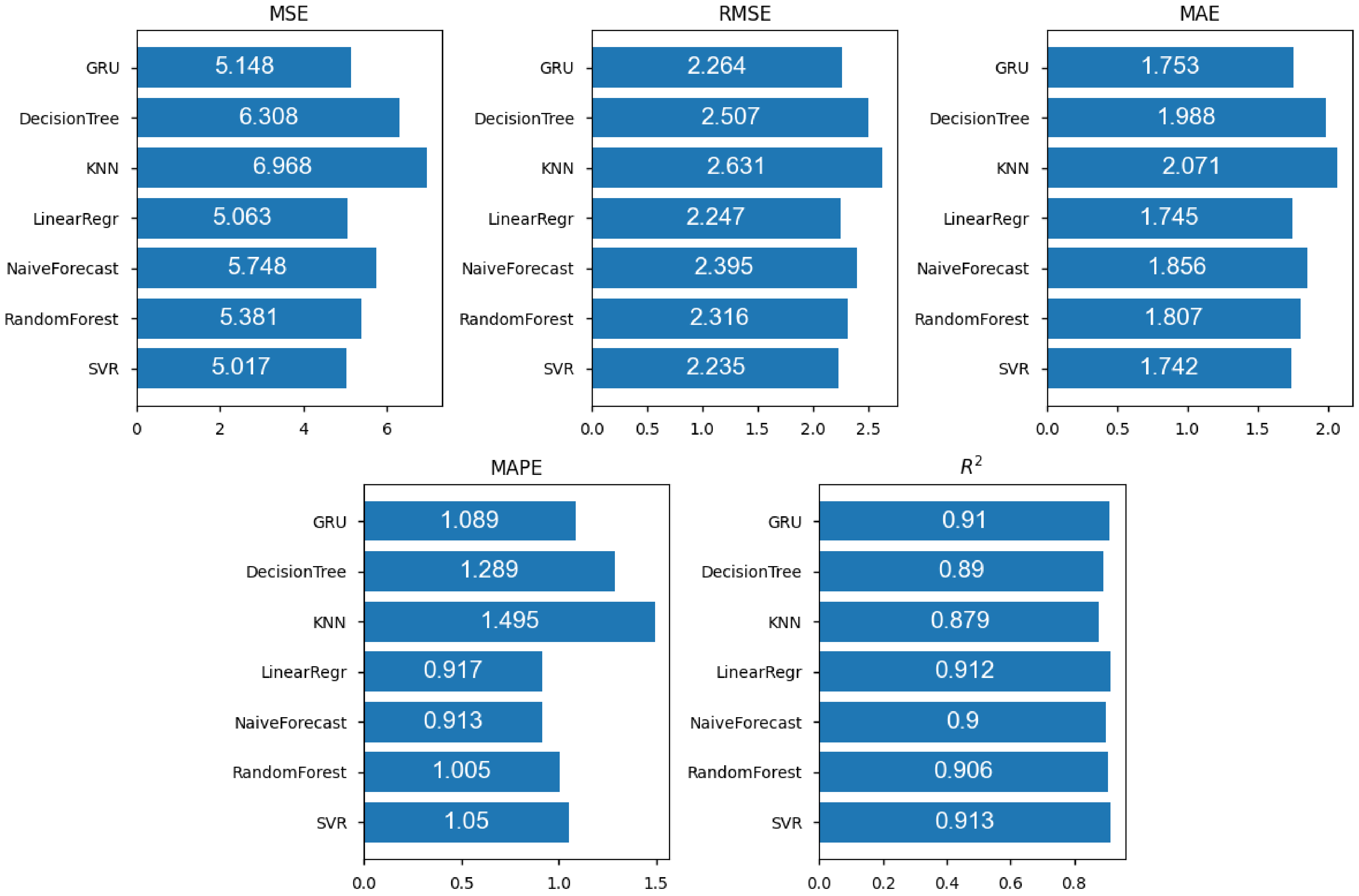
Evaluation-HoldOut

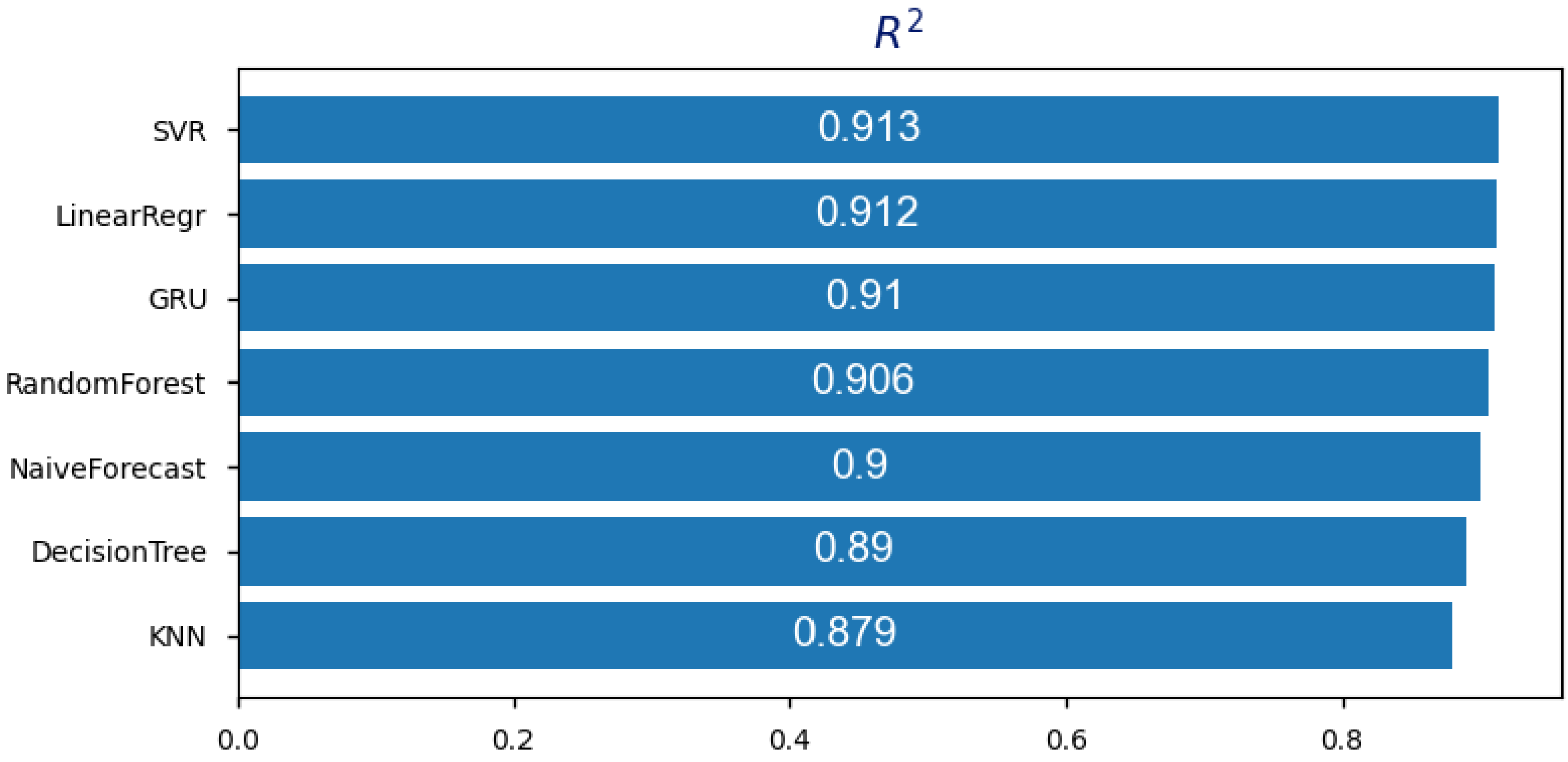




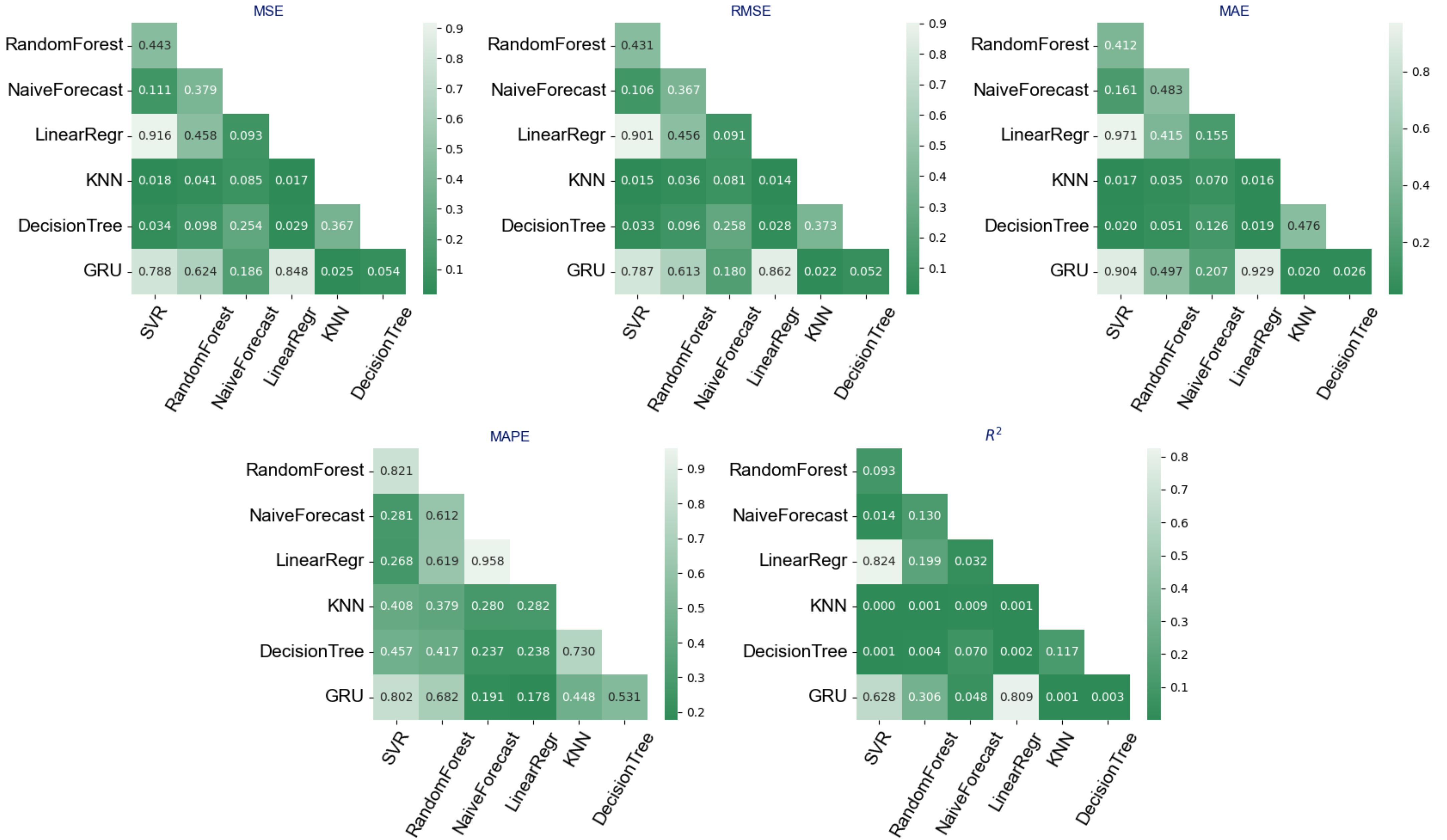
R^2
SVR
LinearRegr
GRU
RandomForest
NaiveForecast
DecisionTree
KNN
0.913
0.908
0.908
0.902
0.893
0.886
0.883
0.0
0.2
0.4
0.6
0.8

Evaluation-CV

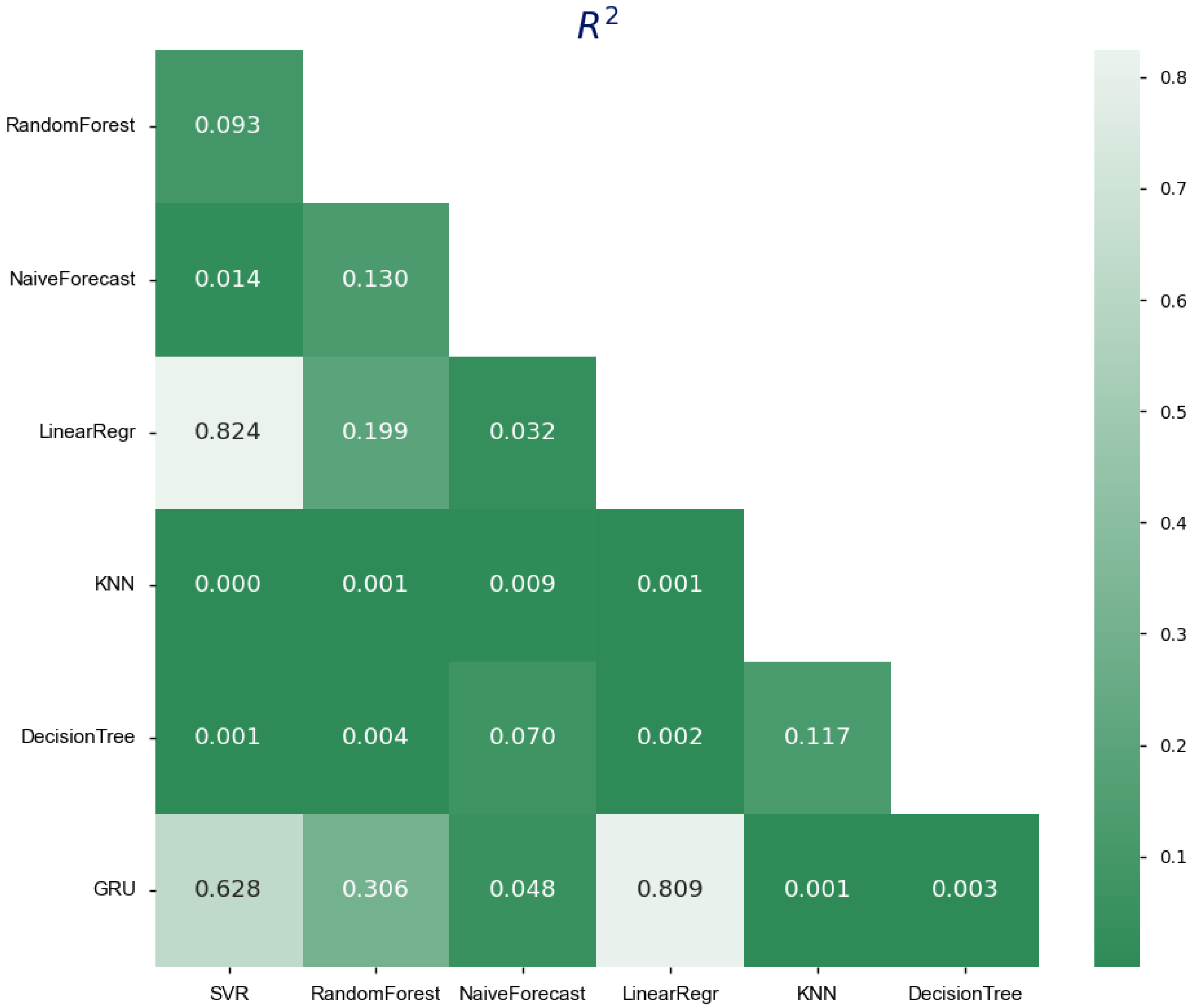


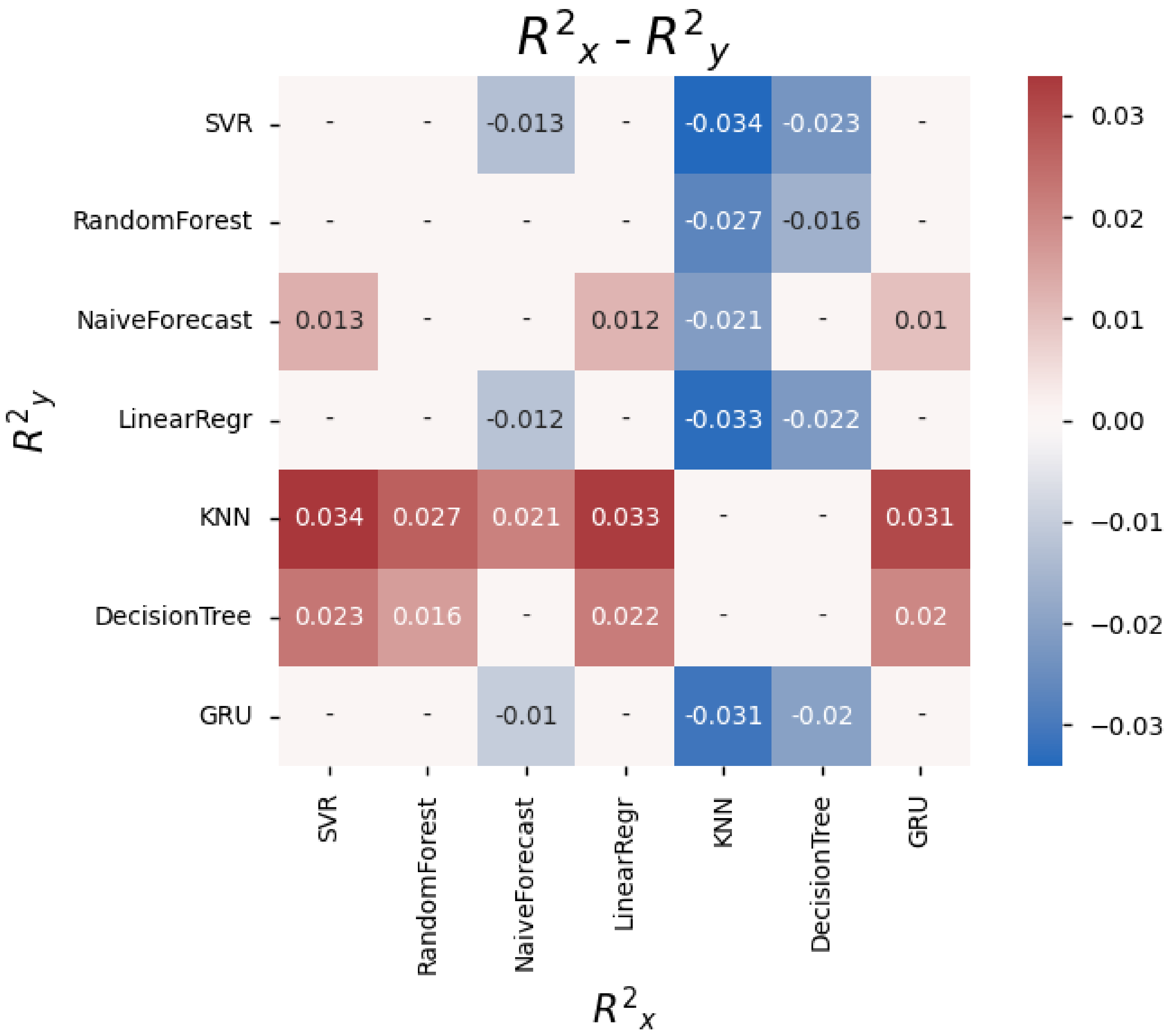

R^2
SVR
LinearRegr
GRU
RandomForest
NaiveForecast
DecisionTree
KNN
0.913
0.912
0.91
0.906
0.9
0.89
0.879
0.0
0.2
0.4
0.6
0.8

Ttest


MSE
RandomForest 0.443
NaiveForecast 0.111 0.379
LinearRegr 0.916 0.458 0.093
KNN 0.018 0.041 0.085 0.017
DecisionTree 0.034 0.098 0.254 0.029 0.367
GRU 0.788 0.624 0.186 0.848 0.025 0.054
SVR RandomForest NaiveForecast LinearRegr KNN DecisionTree
0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1
RMSE
RandomForest 0.431
NaiveForecast 0.106 0.367
LinearRegr 0.901 0.456 0.091
KNN 0.015 0.036 0.081 0.014
DecisionTree 0.033 0.096 0.258 0.028 0.373
GRU 0.787 0.613 0.180 0.862 0.022 0.052
SVR RandomForest NaiveForecast LinearRegr KNN DecisionTree
0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1
MAE
RandomForest 0.412
NaiveForecast 0.161 0.483
LinearRegr 0.971 0.415 0.155
KNN 0.017 0.035 0.070 0.016
DecisionTree 0.020 0.051 0.126 0.019 0.476
GRU 0.904 0.497 0.207 0.929 0.020 0.026
SVR RandomForest NaiveForecast LinearRegr KNN DecisionTree
0.8 0.6 0.4 0.2
MAPE
RandomForest 0.821
NaiveForecast 0.281 0.612
LinearRegr 0.268 0.619 0.958
KNN 0.408 0.379 0.280 0.282
DecisionTree 0.457 0.417 0.237 0.238 0.730
GRU 0.802 0.682 0.191 0.178 0.448 0.531
SVR RandomForest NaiveForecast LinearRegr KNN DecisionTree
0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2
R^2
RandomForest 0.093
NaiveForecast 0.014 0.130
LinearRegr 0.824 0.199 0.032
KNN 0.000 0.001 0.009 0.001
DecisionTree 0.001 0.004 0.070 0.002 0.117
GRU 0.628 0.306 0.048 0.809 0.001 0.003
SVR RandomForest NaiveForecast LinearRegr KNN DecisionTree
0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1

Ttest


R^2
RandomForest
NaiveForecast
LinearRegr
KNN
DecisionTree
GRU
SVR
RandomForest
NaiveForecast
LinearRegr
KNN
DecisionTree
0.093
0.014
0.130
0.824
0.199
0.032
0.000
0.001
0.009
0.001
0.001
0.004
0.070
0.002
0.117
0.628
0.306
0.048
0.809
0.001
0.003
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1



$R^2_x - R^2_y$
SVR
RandomForest
NaiveForecast
LinearRegr
KNN
DecisionTree
GRU
R^2_y
R^2_x
-0.013
-0.034
-0.023
-0.027
-0.016
0.013
0.012
-0.021
0.01
-0.012
-0.033
-0.022
0.034
0.027
0.021
0.033
0.031
0.023
0.016
0.022
0.02
-0.01
-0.031
-0.02
0.03
0.02
0.01
0.00
−0.01
−0.02
−0.03

Conclusioni

- SVR, Linear Regression e GRU i più prestazionali
- KNN e DecisionTree con performance ridotte
- KNN peggiore del NaiveForecast

Grazie per l’attenzione