Tunisian Republic
Ministry of Higher Education and Scientific Research

Carthage University
Higher Institute of Information Technologies and Communication

M E M O R Y Presented for the purpose of obtaining **National Master's degree**
Mention : Robotics, Computer Science and Communication Systems
Specialty : Data Science and Smart Services (D3S)

Artificial Intelligence for PV Output Power Forecasting

- **Presented by Mr: Walid MCHARA**
- **Under the direction of: Mr. Med-Abdellatif KHALFA**
- **Co-supervised By Mr: Lotfi KHEMISSI**

Academic year : 2023/2024

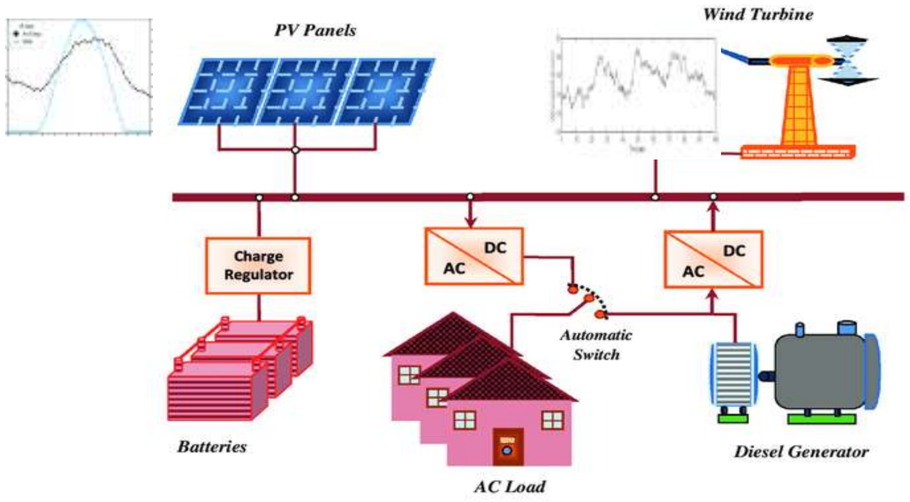
Plan

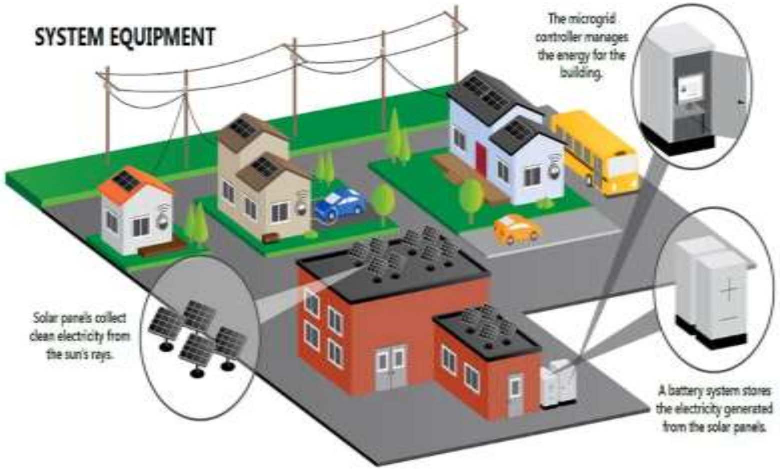
1. Introduction & problematic
2. Forecasting Platform KNIME
3. Data Preprocessing
4. P_{PV} Forecasting with Regression Models
5. Forecast Evaluation
6. P_{PV} Prediction using LSTM
7. Conclusion & Perspectives

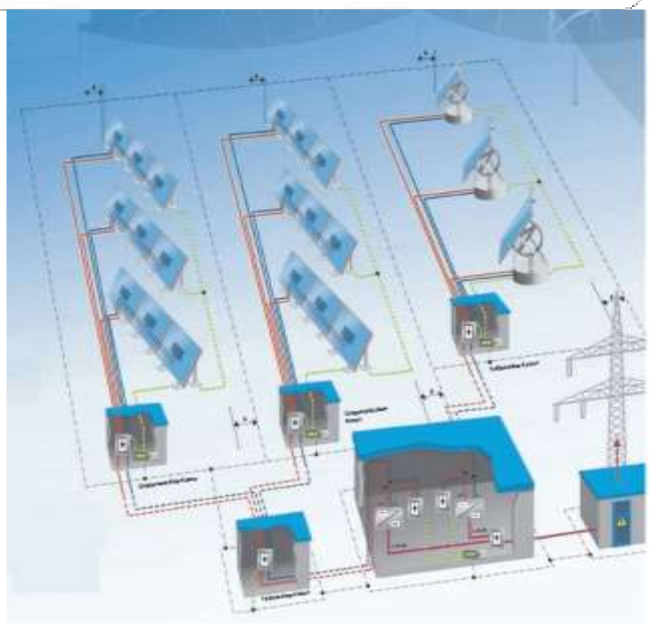
1

Introduction & problematic

Photovoltaic (PV) Power Generation Systems







Stand-alone PV system (SAPVS)

- **Without battery**
- **With battery**
- **Hybrid PV System**

Distributed PV system (DPVS)

- **Without battery**
- **With battery**

Centralized PV system (CPVS)

Types of PV Forecasting for DPVS & CPVS

Very Short-term
$[1\ min, 1\ hour]$

Short-term
$[1\ hour, 6\ hours]$

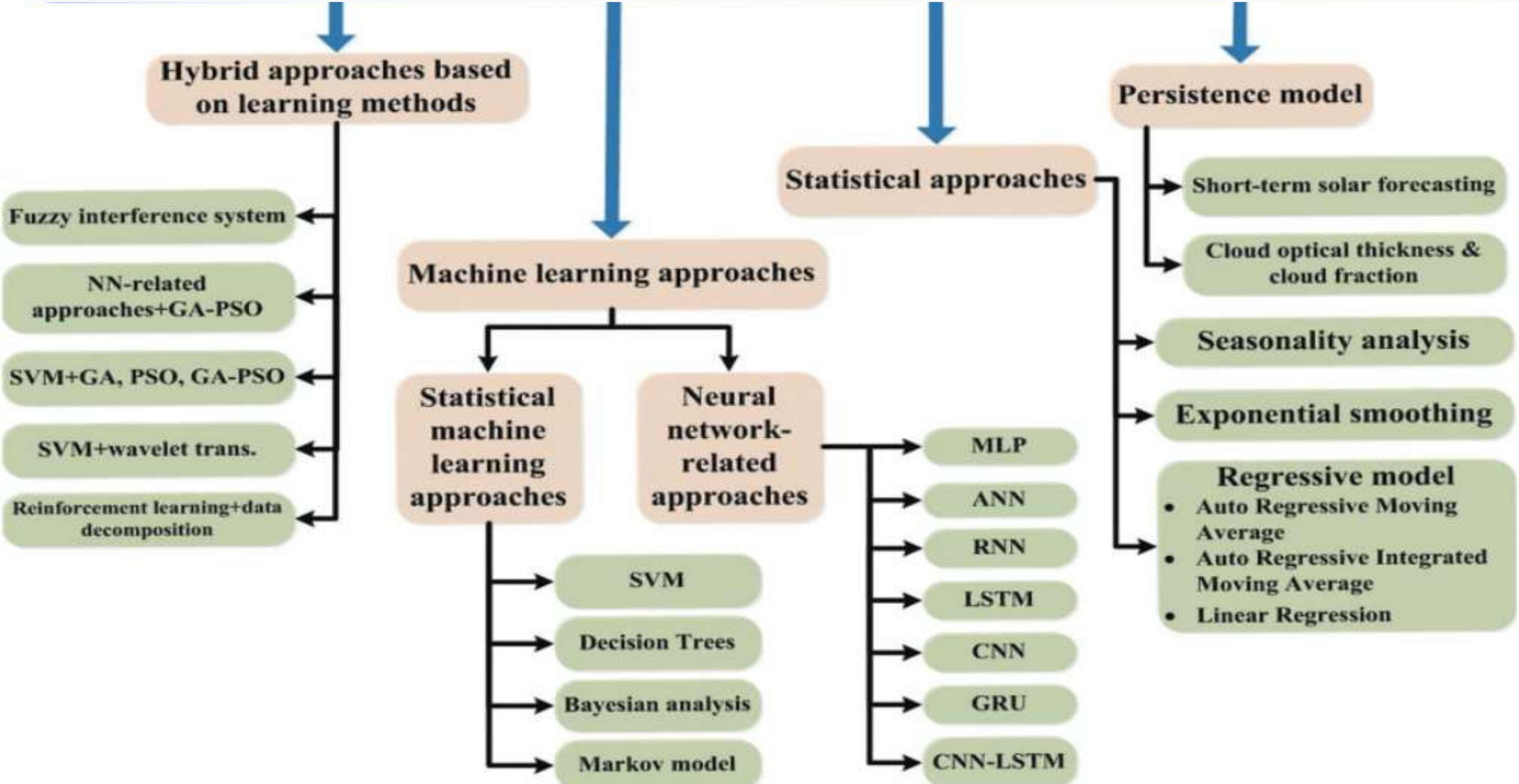
Long-term
$[6\ hours, 24\ hours]$

- Real-time dispatch
(Rapid variation of the load)

- **Load trading**
- **Reserves purchasing**
- **Unit commitment**
- **Power curtailment**
- **Power management**

- **Economic dispatch**
- **Unit commitment**
- **Maintenance**

Classification of solar photovoltaic forecasting based on historical data.



2

Forecasting Platform

Open for Innovation
KNIME

Platform Overview: Explore an open-source platform for data analytics.

Rich Toolkit: Offers a diverse range of tools and functionalities for data processing, analysis, and visualization.

User-Friendly GUI: Features a Graphical User Interface (GUI) that empowers users to build and execute data workflows effortlessly.

Versatile Tasks: Facilitates tasks such as data cleaning, feature engineering, machine learning modeling, and result visualization.

Language Compatibility: Provides seamless integration with popular programming languages such as R and Python.

3

Data Preprocessing

❖ Data Source:

The dataset is collected from the Research Center of Energy **(CRTEn)** located at **Borj Cedria** Science and Technology Park (Latitude: 36.717º, Longitude: 10.427º) from **january 1, 2005 to December 31, 2020**.

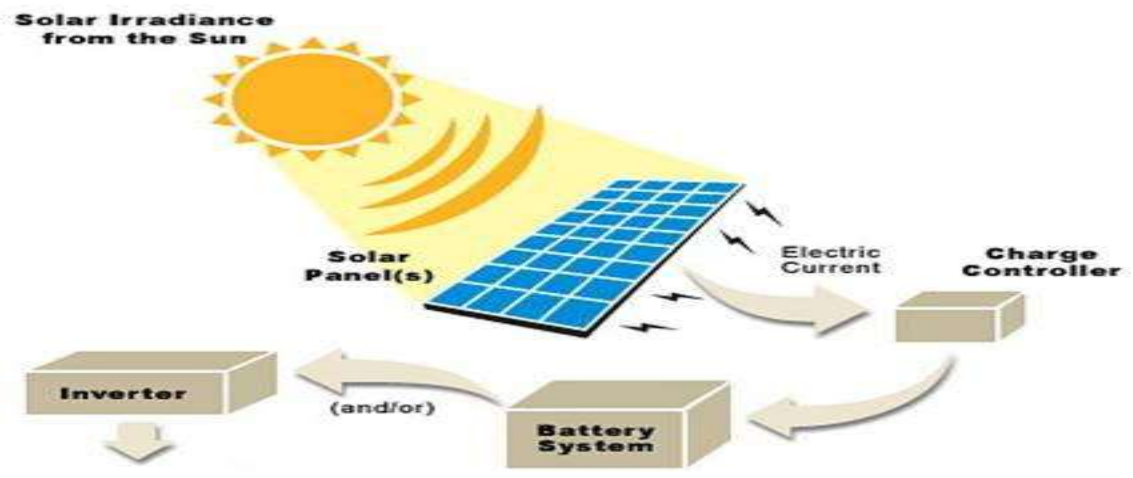
X_1	X_2	X_3	X_4	Y
Global Irradiance(t)	Ambient temperature (t)	Wind Speed (t)	Sun Heigh(t)	PV (t)
Global Irradiance$(t+1)$	Ambient temperature $(t+1)$	Wind Speed $(t+1)$	Sun Heigh$(t+1)$	PV $(t+1)$
…	…	…	…	…

3

Data Preprocessing

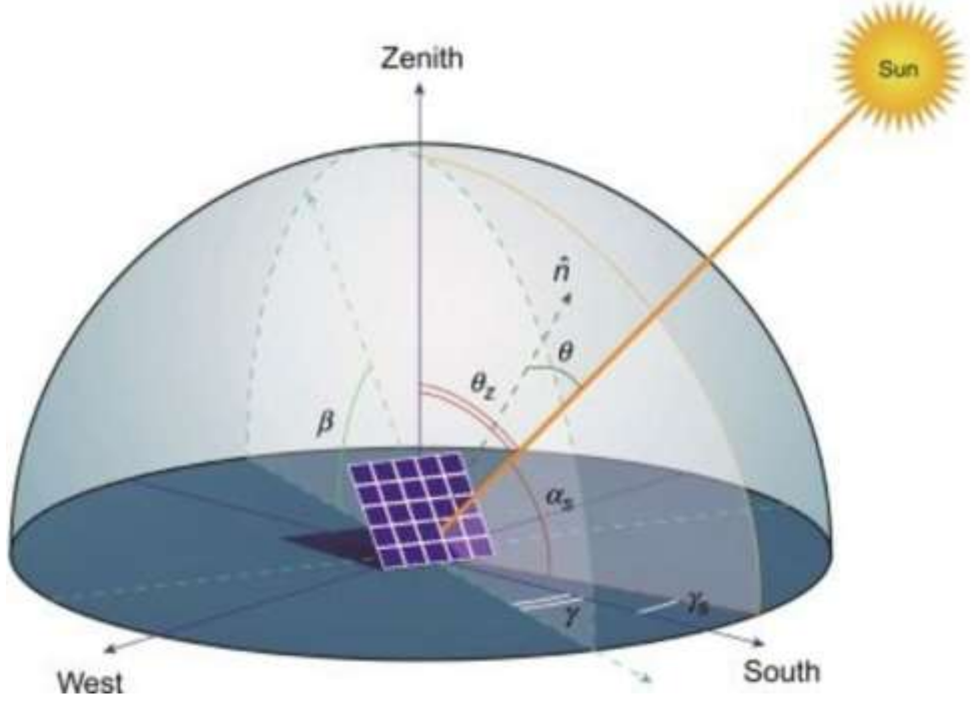
❖ Feature: Global Irradiance (GI)

- A value, denoted in watts per square meter (W/m2), signifies the complete solar radiation received by a flat horizontal surface.



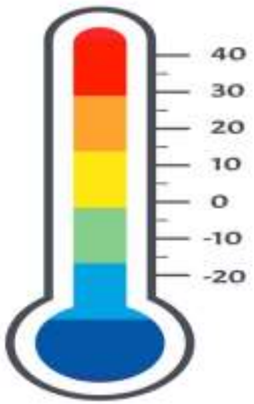
❖ Feature: Sun Heigh (SH)

- An angle, expressed in degrees, signifies the sun's elevation above the horizon.



❖ Feature: Ambient temperature (Ta) :

- Denotes the air temperature, measured in degrees Celsius (°C), at a height of 2 meters above the ground.



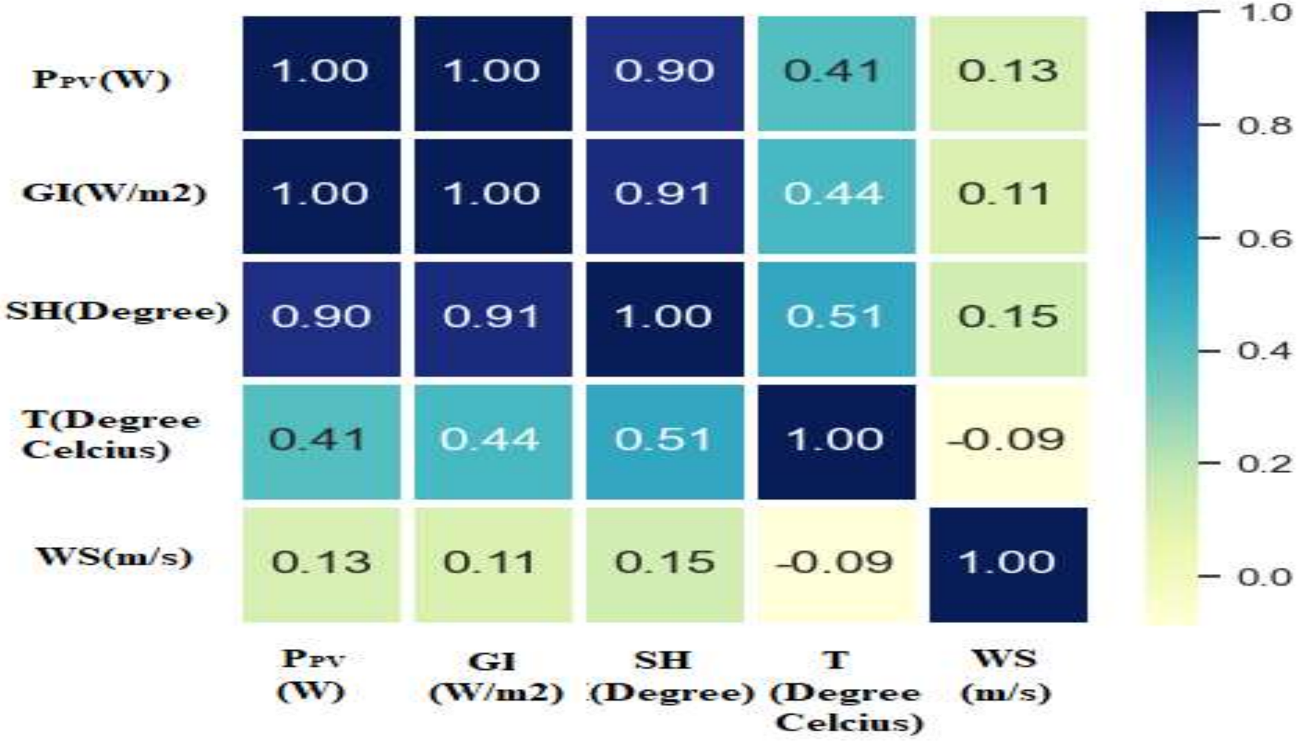
❖ Feature: Wind Speed (Ws)

- Indicates the wind speed, measured in meters per second (m/s), at an elevation of 10 meters above the surface.

3

Data Preprocessing

❖ **Correlation Matrix**



❖ **Data Normalization: Min-Max normalization**

$$x_{scaled} = \frac{x - Min(x)}{Max(x) - Min(x)}$$

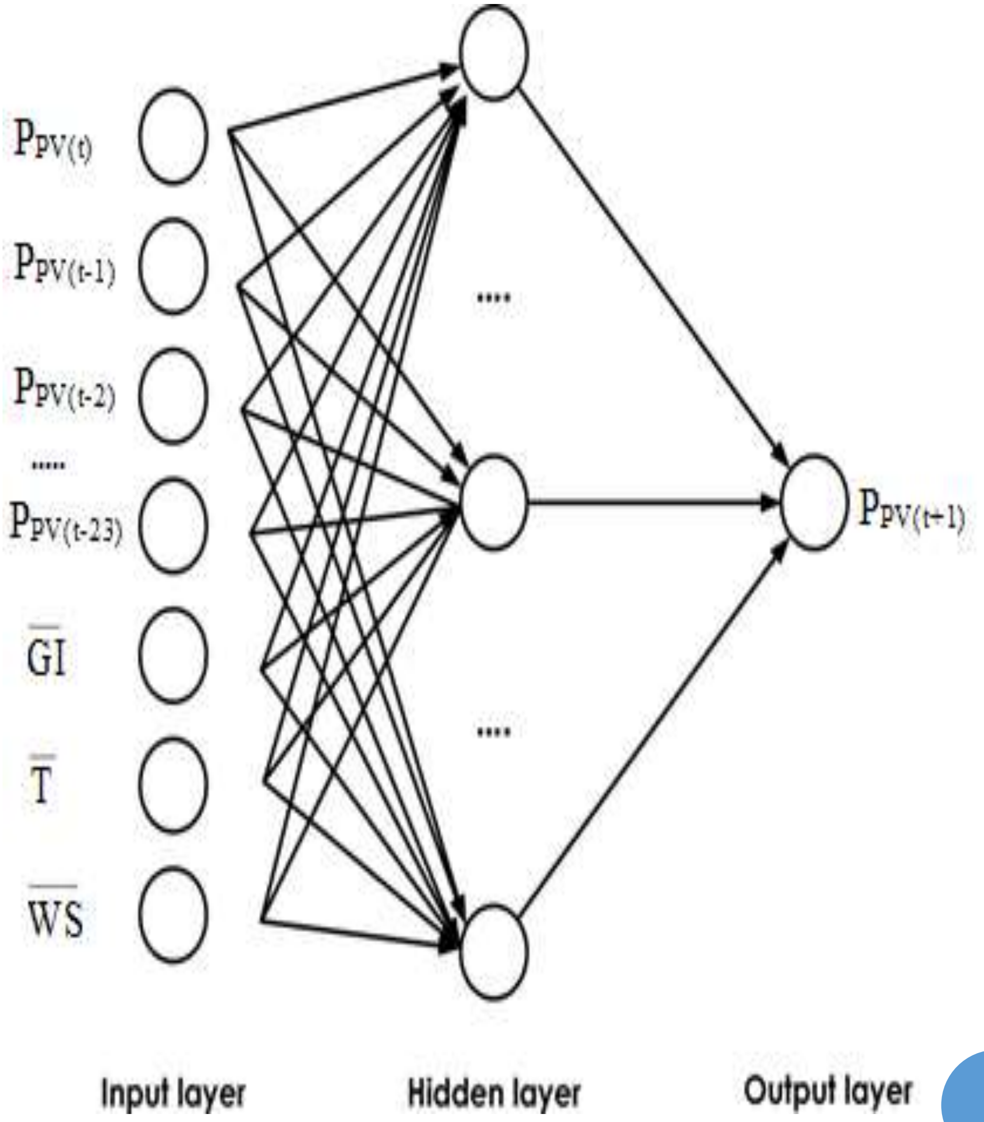
3 Data Preprocessing

❖ Contribution: Restructuring Input Data

Notation	Input variables
$X_{1\text{-}24}$	Hourly output PV power for the last 24 hours in (W).
X_{25}	Average GI of the last 24 hours in (W/m²).
X_{26}	Average temperature of the last 24 hours in (°C).
X_{27}	Average wind speed of the last 24 hours in (m/s).
Notation	**Output variables**
y	PV output power for the next hour in Watts.

X_1	…	$X24$	X_{25}=AverageGI	X_{26}=AverageT	X_{27}=AverageWS	Y
PV (t)	…	PV (t−23)	GI(last24hours)	T (last24hours)	WS(last24hours)	PV (t+1)
PV (t+1)	…	PV (t−22)	GI(last24hours)	T (last24hours)	WS(last24hours)	PV (t+2)
…	…	…	…	…	…	…

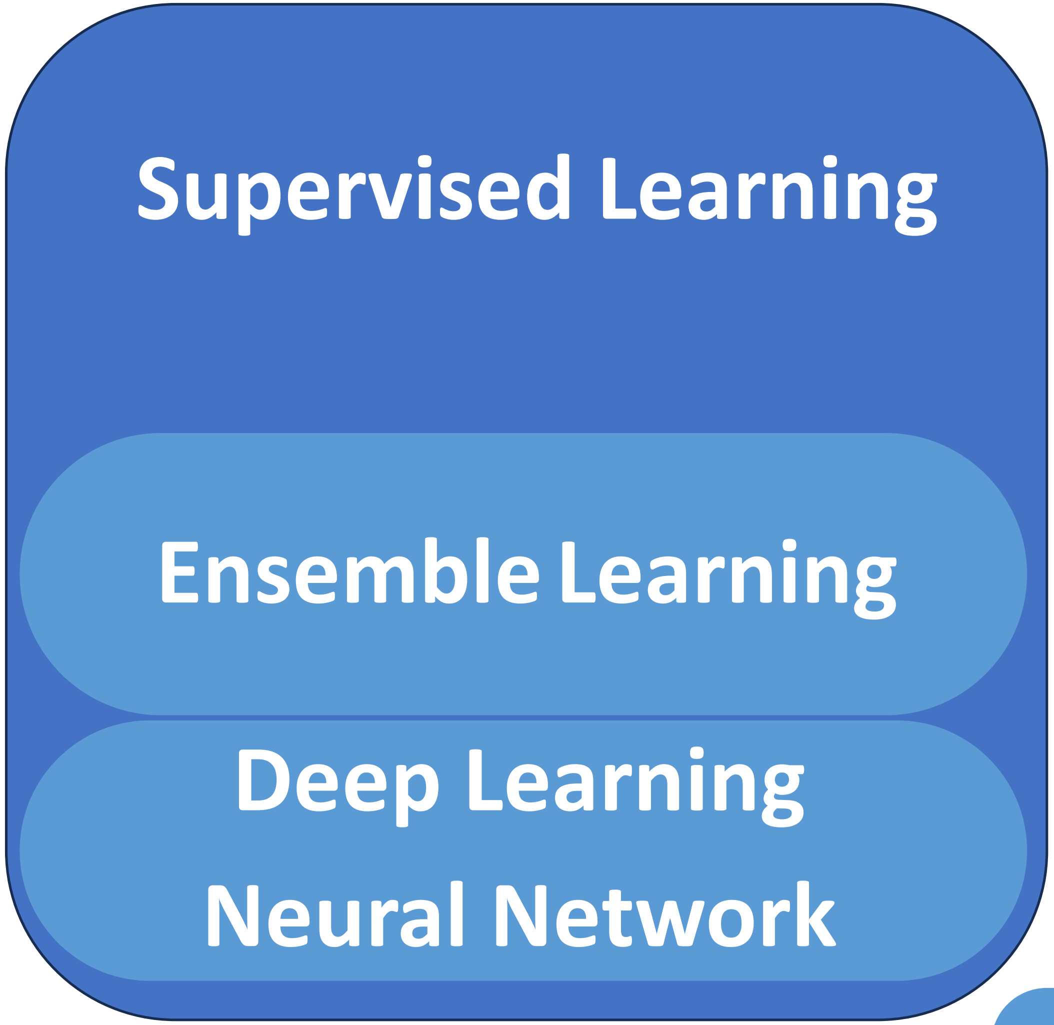
❖ Neural Network architecture:



4 P_{PV} Forecasting with Regression Models

- Simple Regression Tree
- Polynomial Regression
- Linear Regression
- Gradient Boosted Trees
- Random Forest
- Multilayer Perceptron(MLP) (ANN)

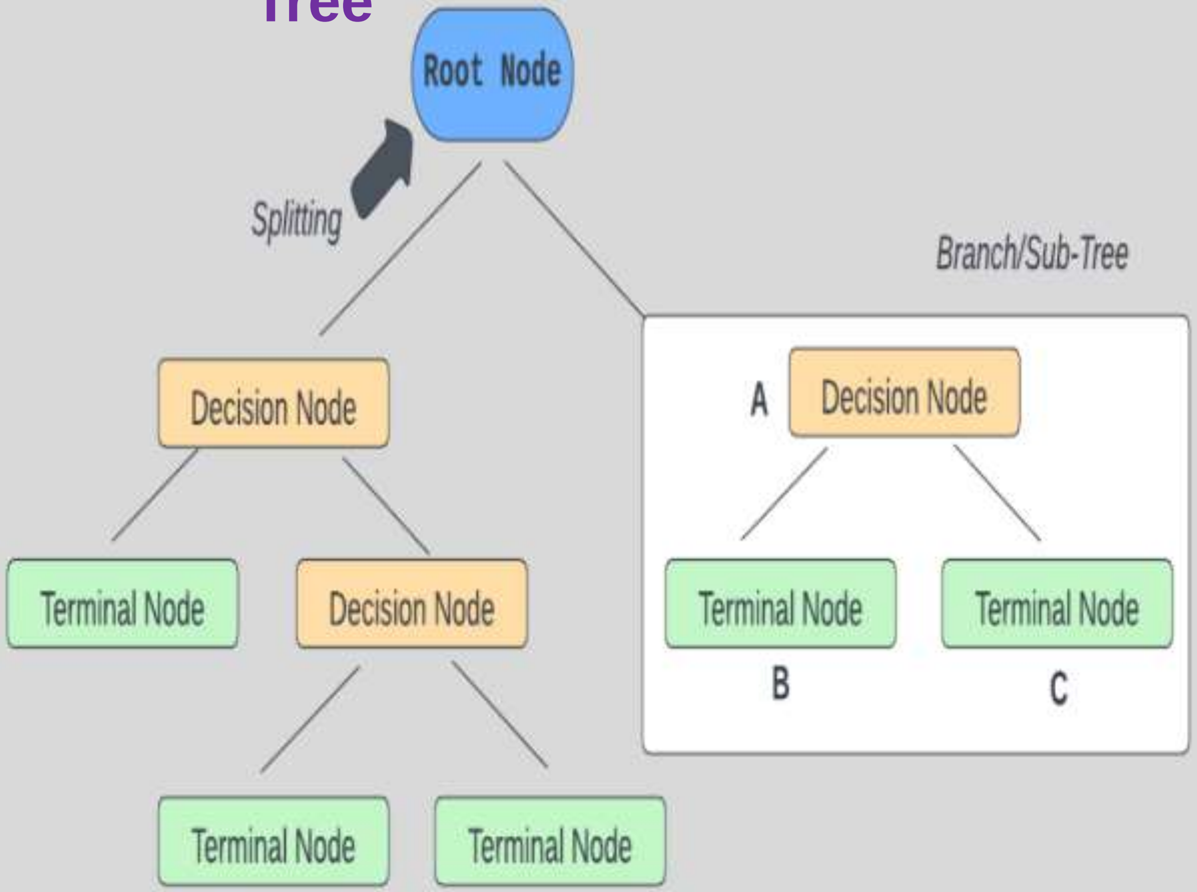
DL4J RPROP



4

P_{PV} Forecasting with Regression Models

❖ **Simple Regression Tree**



Advantages

- **Simplicity**: Easy to understand and interpret.
- **Non-linearity:** Can capture non-linear relationships in the data.
- **Versatility:** Suitable for various types of data and problems.

Challenges

- **Overfitting:** Prone to overfitting when the tree becomes too complex.
- **Instability:** Small changes in the data can result in significantly different trees.
- May struggle with capturing complex relationships compared to more advanced models.

4

P_{PV} Forecasting with Regression Models

❖ Polynomial Regression

Degree of the Polynomial

Determines the degree of the polynomial equation, n.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n$$

Higher-degree polynomials can fit more complex patterns but may be prone to overfitting.

Advantages

Flexibility

Interpretability

Versatility

Challenges

Overfitting

Interpretation

Requires a sufficient amount of data for reliable results.

4

P_{PV} Forecasting with Regression Models

❖ Linear Regression

Types

- **Simple Linear Regression** ($y = \beta_0 + \beta_1 x_1$).
- **Multiple Linear Regression**($y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$).

Advantages

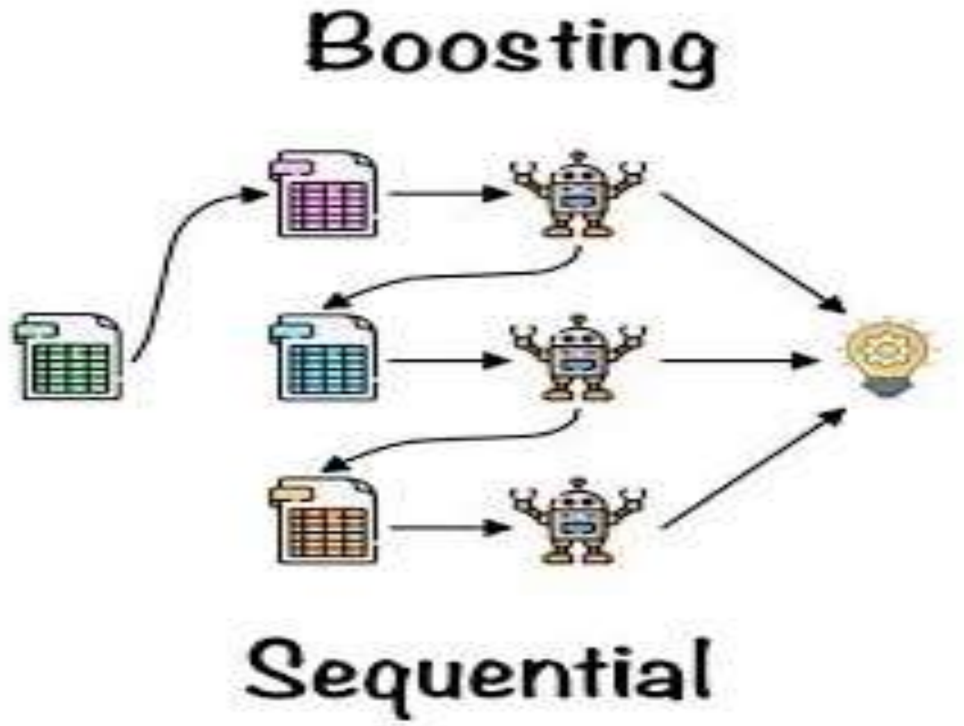
- **Simplicity**
- **Transparency**
- **Versatility**

Challenges

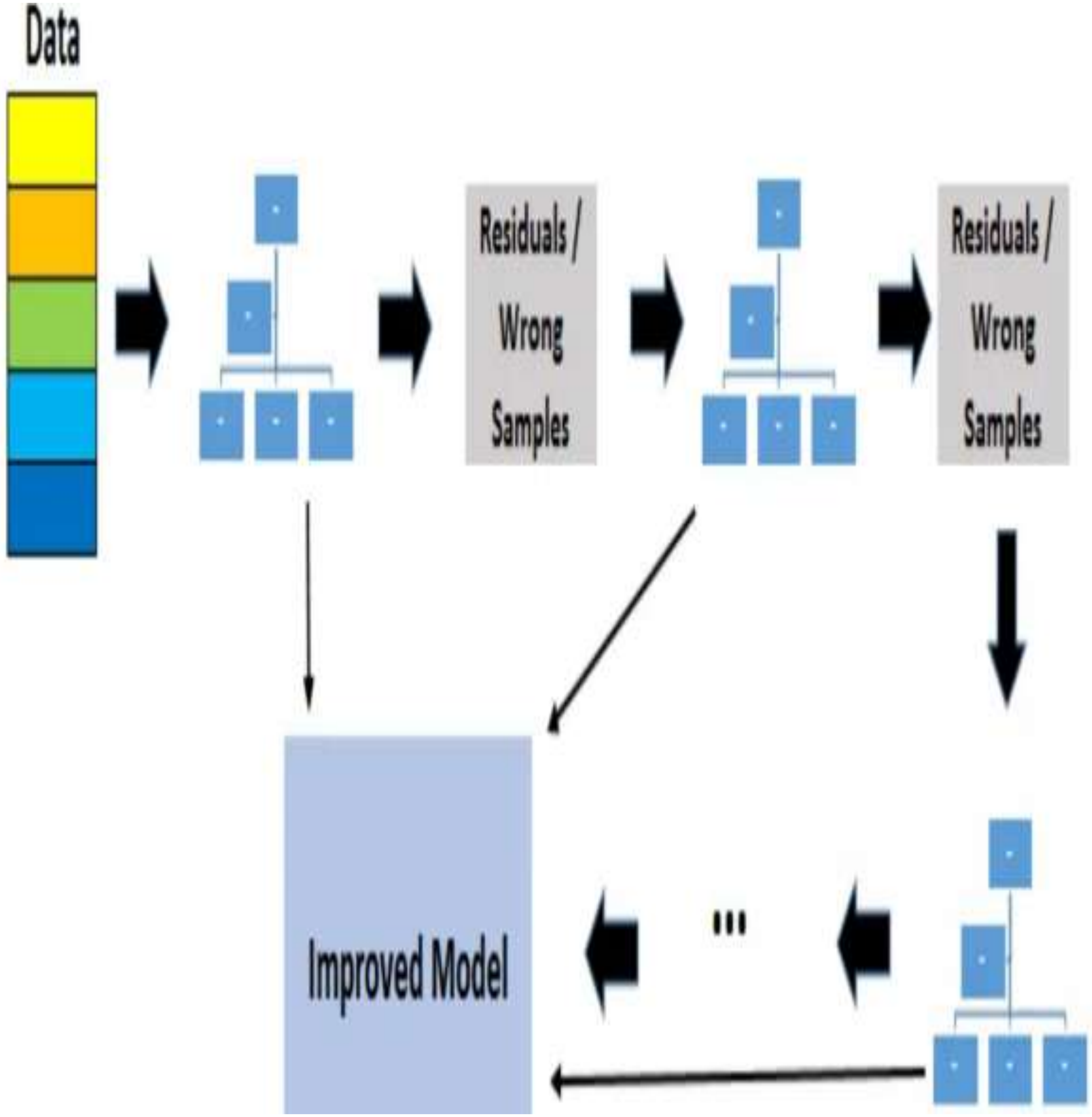
- **Assumptions**: Relies on the assumption of a linear relationship, which may not always hold.
- **Complex Relationships**: Limited in capturing complex, non-linear patterns in data.
- **Outliers**: Sensitive to outliers that can influence the model's accuracy.

4 P_{PV} Forecasting with Regression Models

Boosting: Combing weak learners into strong learners by creating sequential models such that the final model has the highest accuracy.



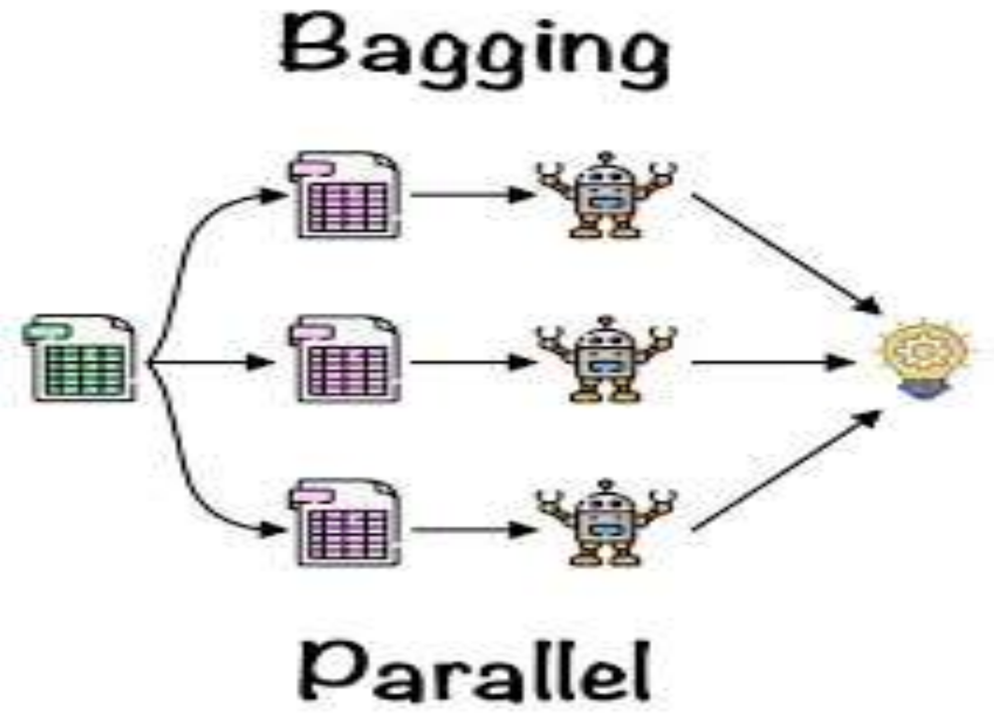
❖ Gradient Boosted



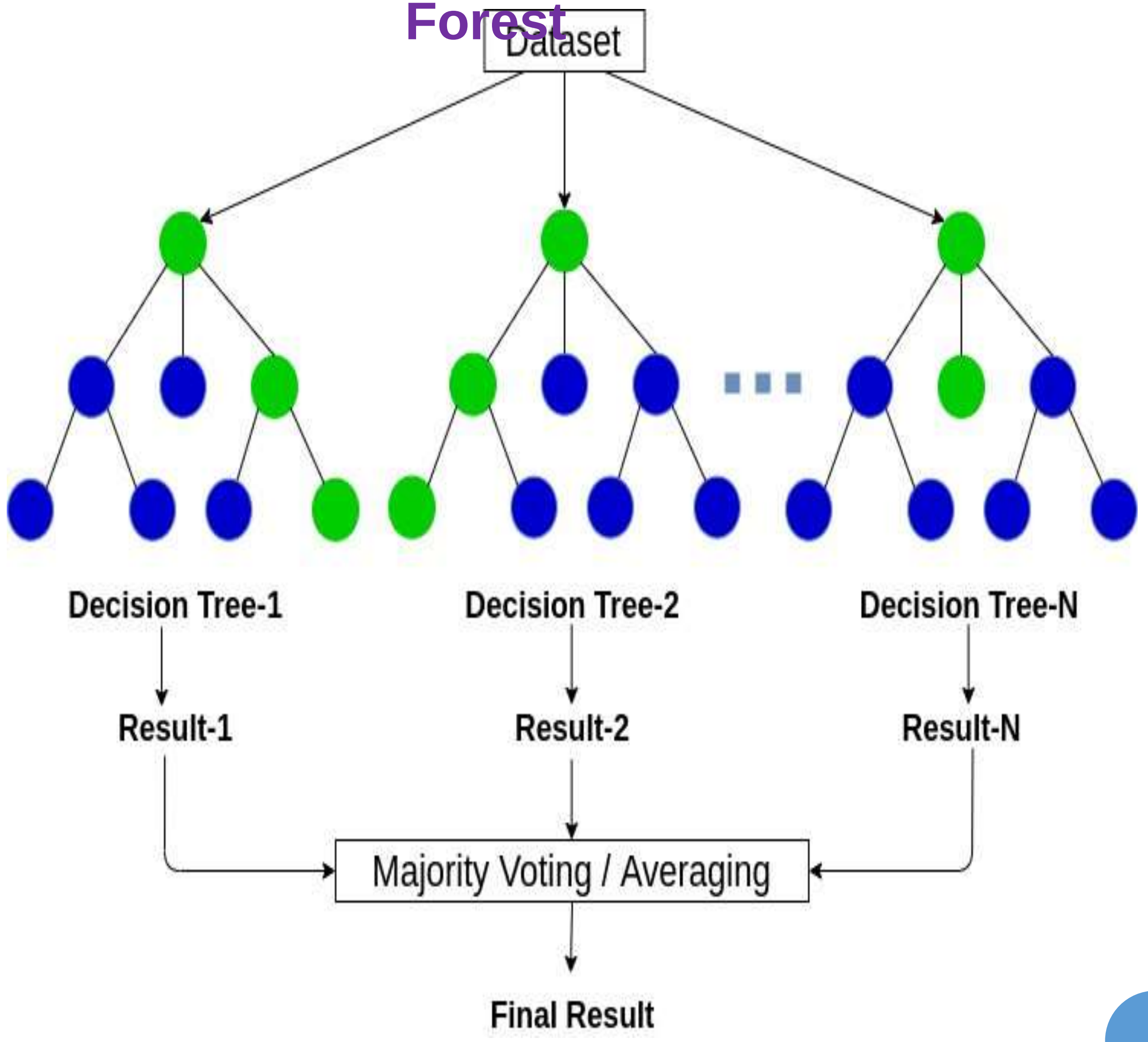
4

P_{PV} Forecasting with Regression Models

Bagging: Creating a different training subset from sample training data with replacement. The final output is based on majority voting,averaging,....



❖ **Random Forest**



4 P_{PV} Forecasting with Regression Models

❖ Multi Layer Perceptrons (MLP)

Neural Network Structure

- **Input Layer** receives the input data.
- **Hidden Layers**: Intermediate layers
- **Output Layer:** The final layer provides the model's prediction.

Advantages

- **Complex Patterns**: MLPs can model highly non-linear and complex relationships in data.
- **Feature Learning**
- **Versatility**

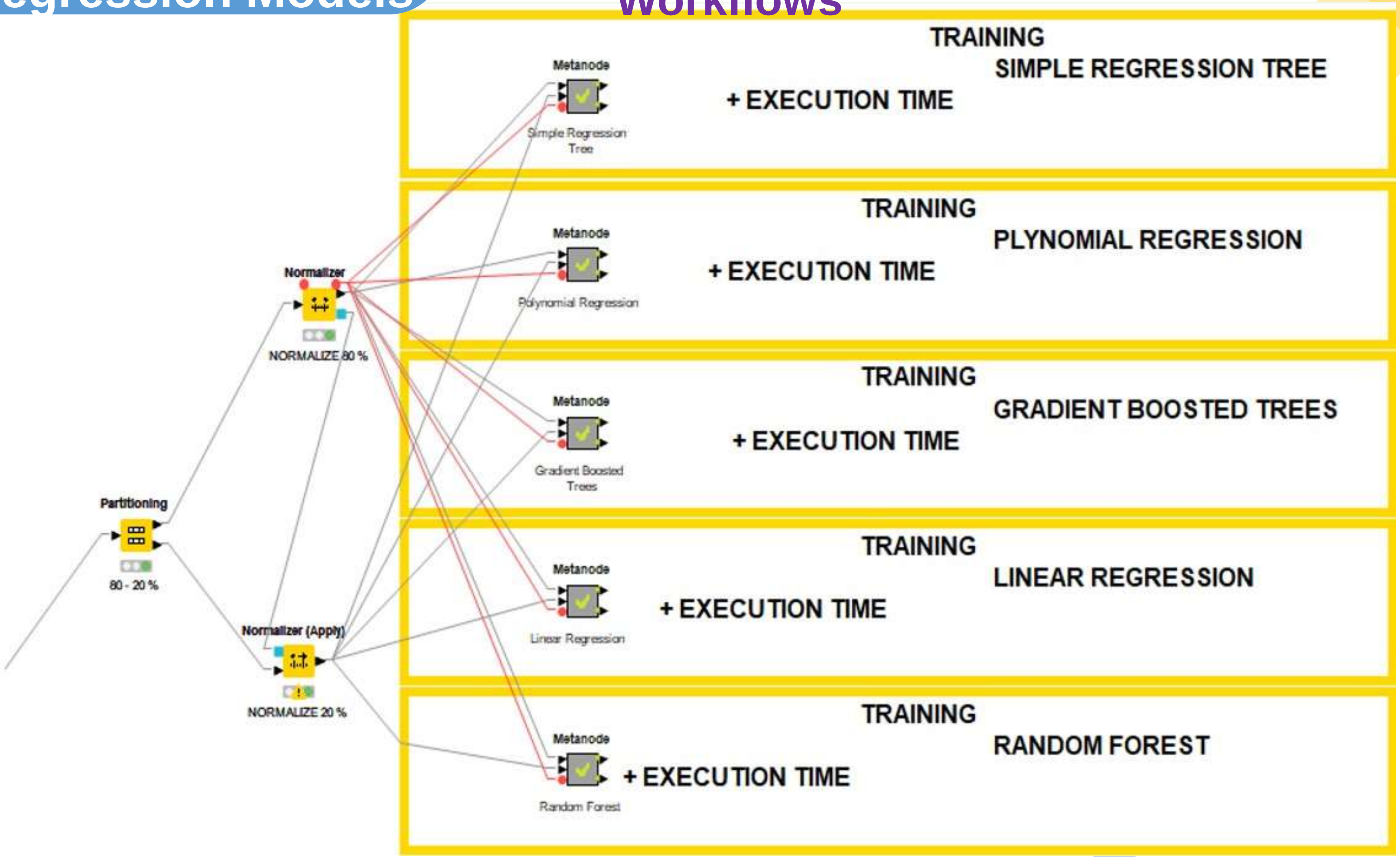
Challenges

- **Overfitting**
- **Hyperparameter Tuning**
- **Training Data**: Requires a substantial amount of data for optimal performance.

4

P_{PV} Forecasting with Regression Models

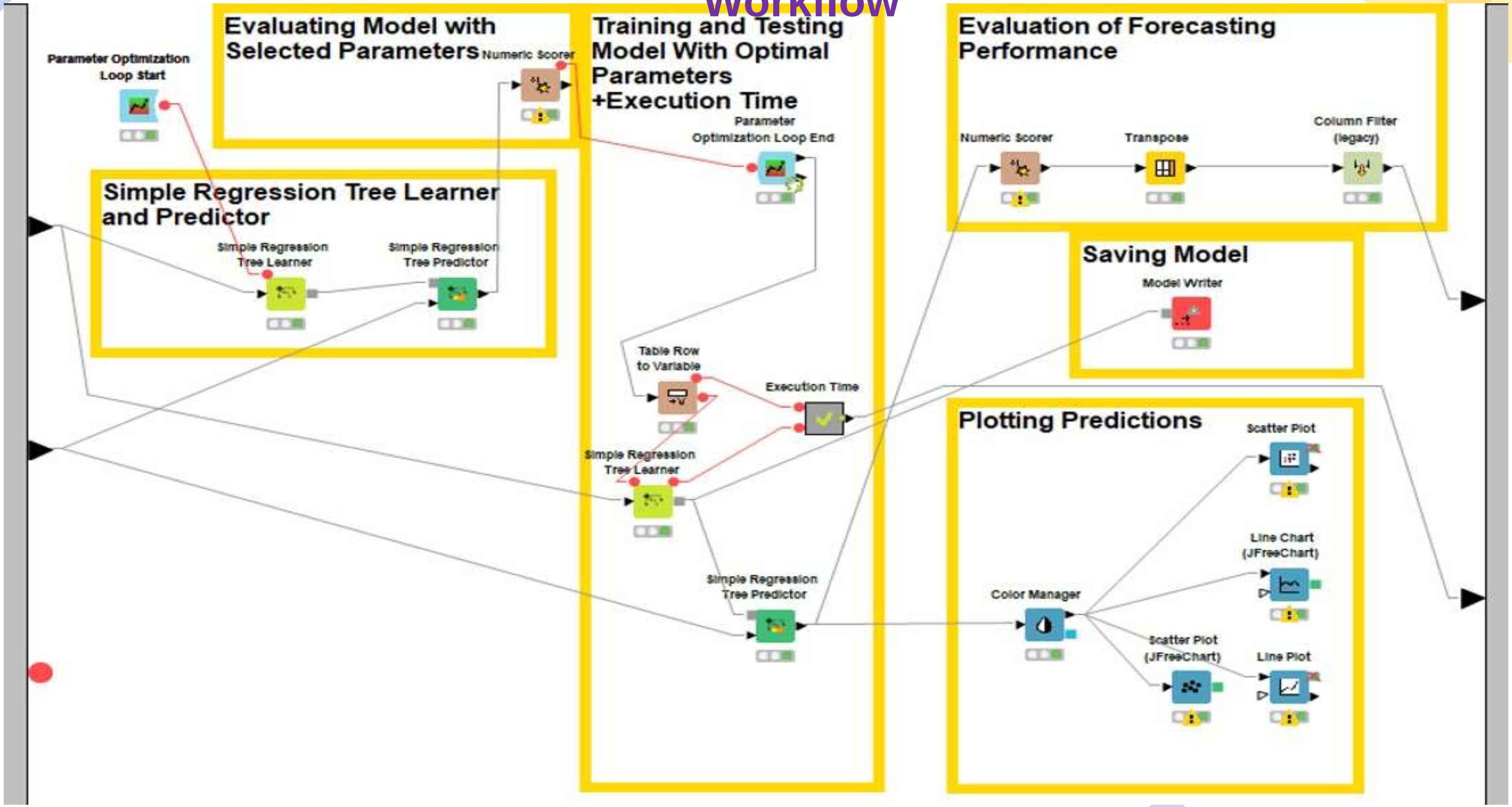
❖ The five Machine Learning Workflows



4

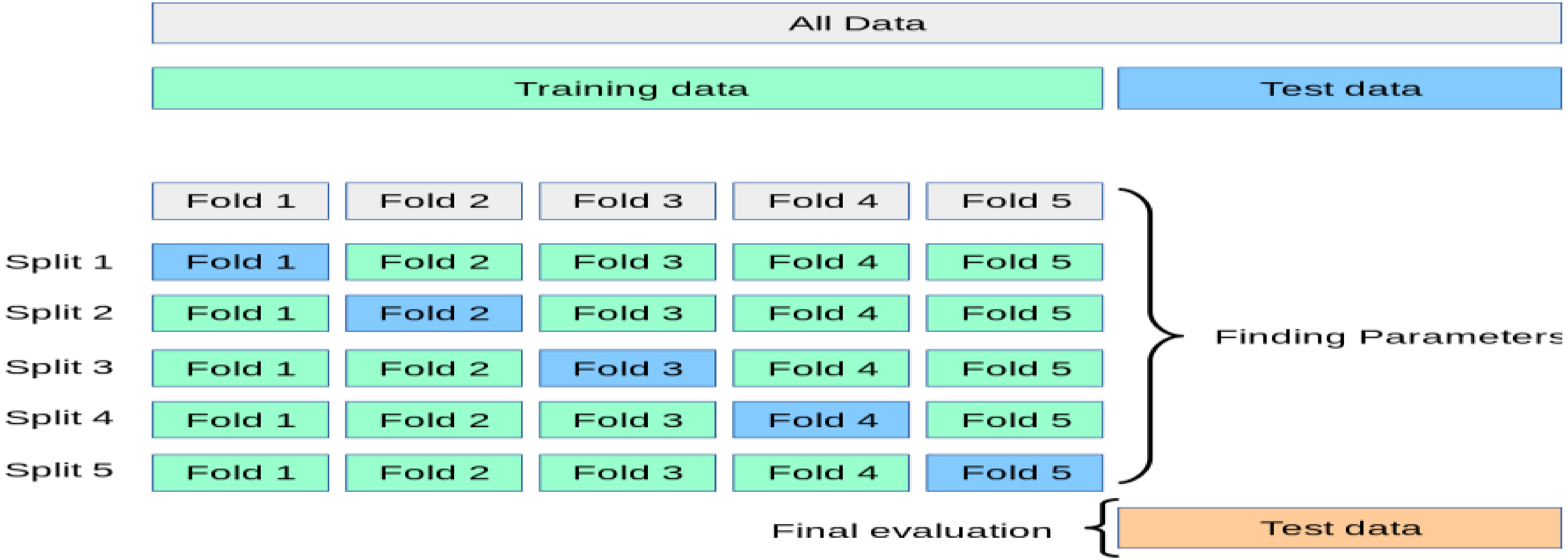
P_{PV} Forecasting with Regression Models

❖ Regression Tree Workflow



4 P_{PV} Forecasting with Regression Models

❖ **Hyperparameter tuning: Grid-Search cross-validation**



❖ **Performances indexes for regression models**

$$R^2 = 1 - \frac{\sum \left(y_i - \bar{y}\right)^2}{\sum \left(y_i - \bar{\bar{y}}\right)^2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$$

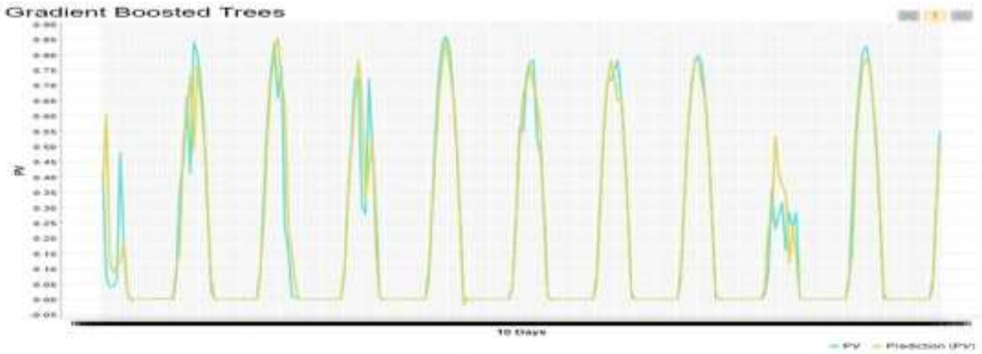
$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \bar{y}\right|$$

Comparison between train and test scores for forecasting models of PV energy

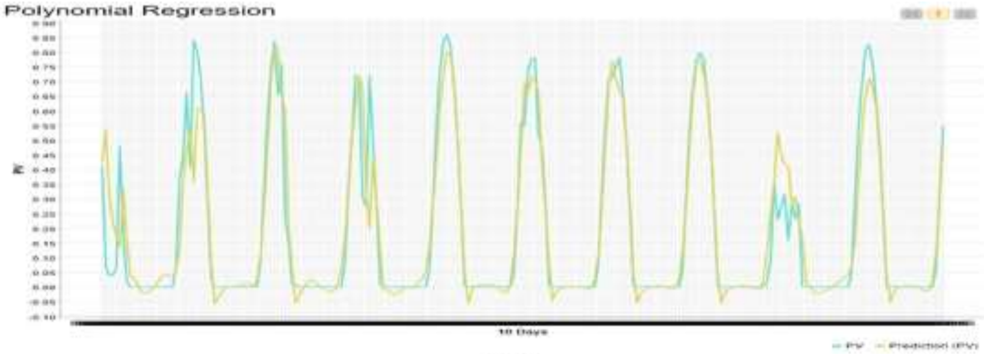
Model	Execution time		R_2	MAE	MSE
Simple Regression Tree	16s	Train	0.960	0.009	0.003
		Test	0.845	0.045	0.013
Random Forest	1m24s	Train	0.970	0.017	0.002
		Test	**0.922**	**0.034**	**0.007**
Polynomial Regression	4s	Train	0.900	0.045	0.008
		Test	0.904	0.046	0.008
Linear Regression	2s	Train	0.900	0.050	0.008
		Test	0.898	0.051	0.008
Gradient Boosted Tree	24s	Train	0.920	0.030	0.007
		Test	**0.912**	**0.032**	**0.007**

5 Forecast Evaluation

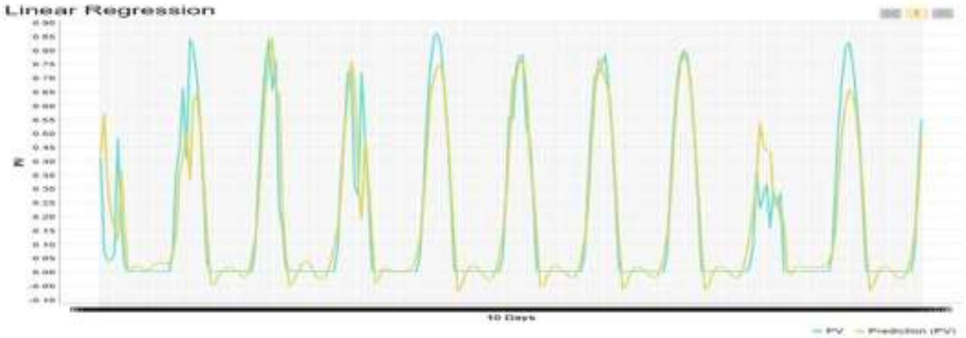
❖ **Comparison between train and test scores for forecasting models of PV energy**



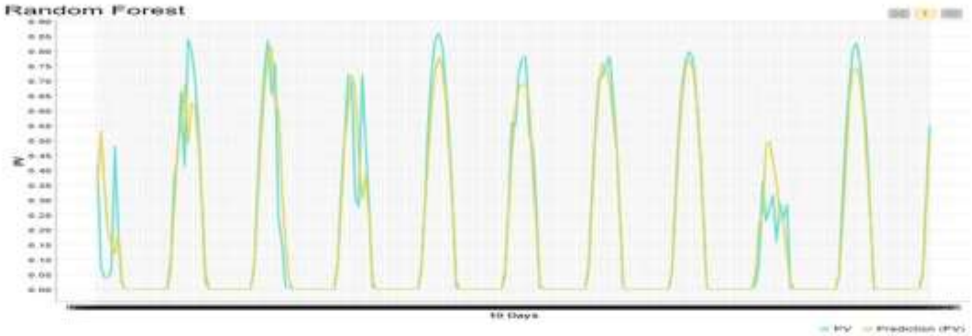
(a)



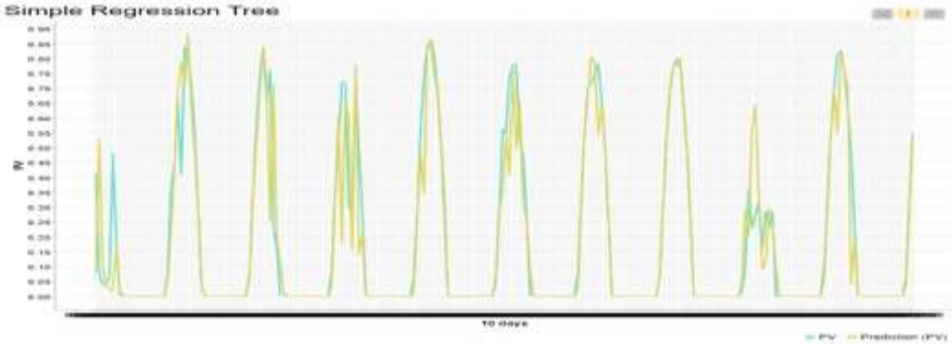
(b)



(c)



(d)

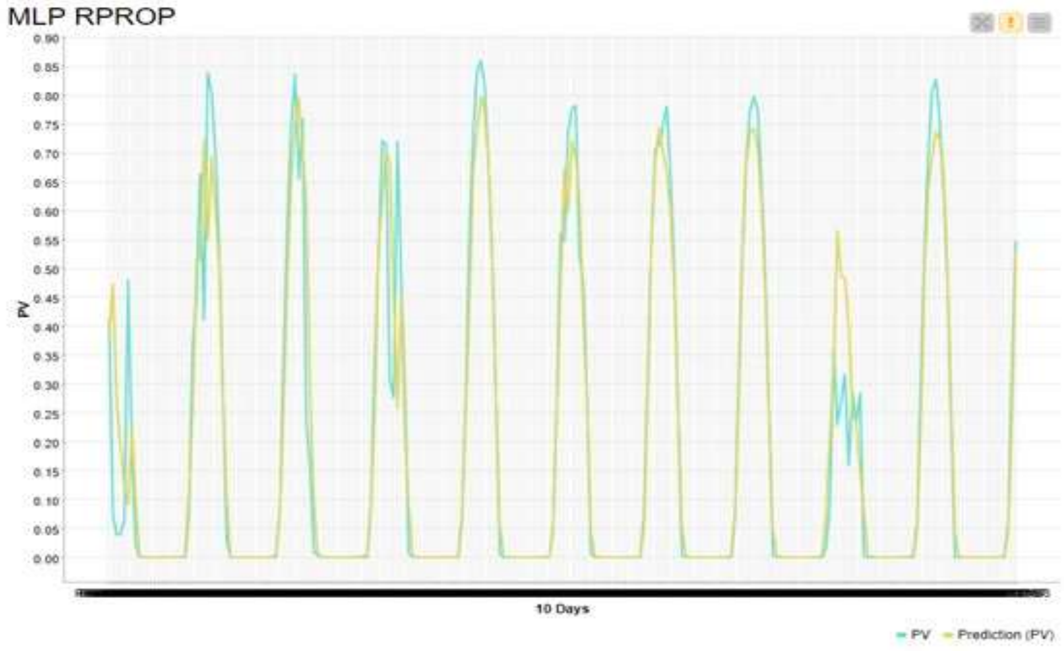


(e)

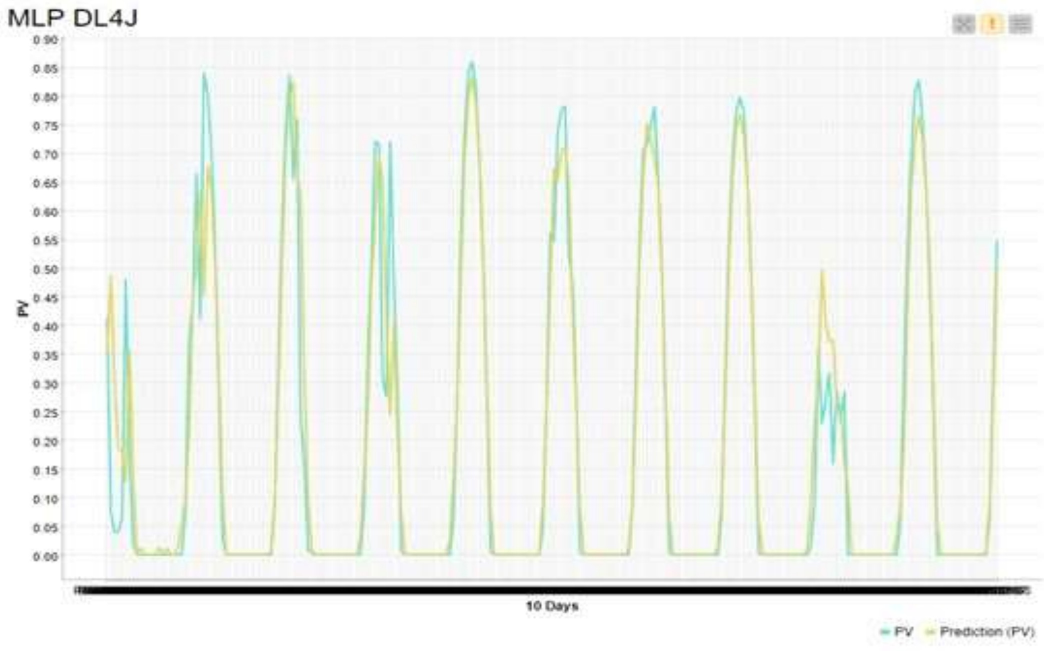
5 Forecast Evaluation

❖ **Comparison between train and test scores for forecasting MLP models of PV energy :**

	R_2	MAE	MSE	Execution time
DL4J	0.915	0.036	0.007	19m 52s
RPROP	0.917	0.035	0.006	3m 47s



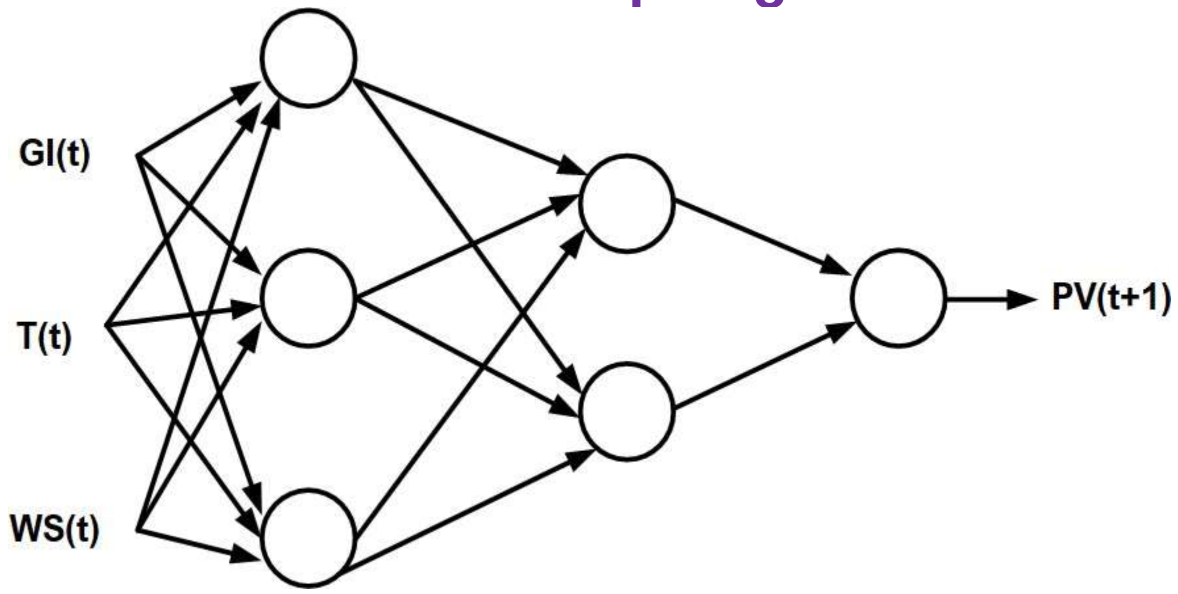
(a)



(b)

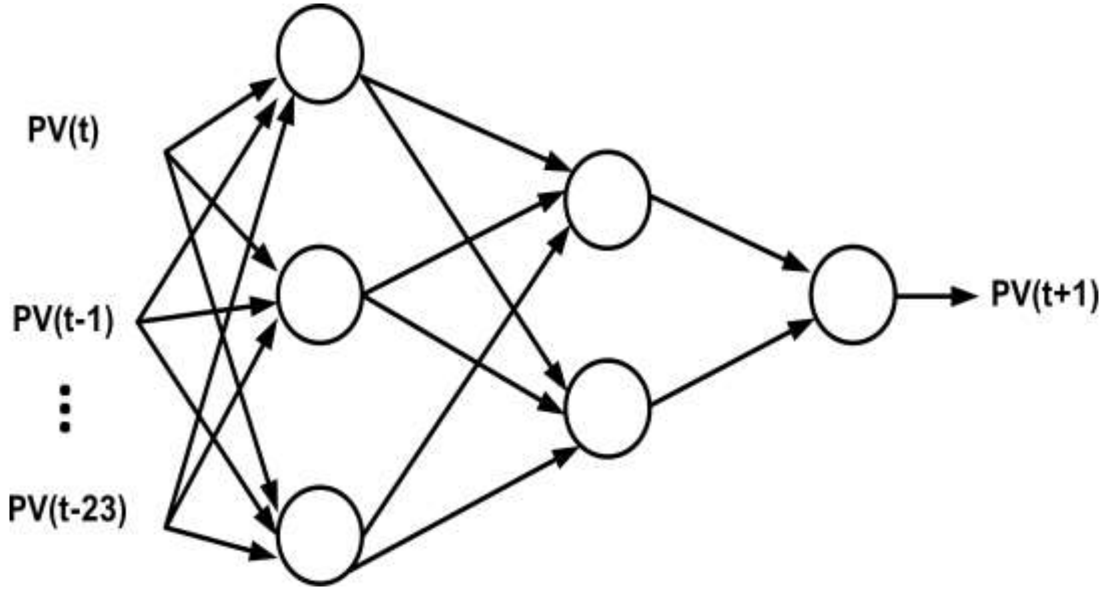
5 Forecast Evaluation

❖ MLP TP1 and TP2 Topologies:



TP1

X_1	X_2	X_3	Y
$GI(t)$	$T(t)$	$WS(t)$	$PV\,(t+1)$
$GI(t+1)$	$T(t+1)$	$WS(t+1)$	$PV\,(t+2)$
…	…	…	…

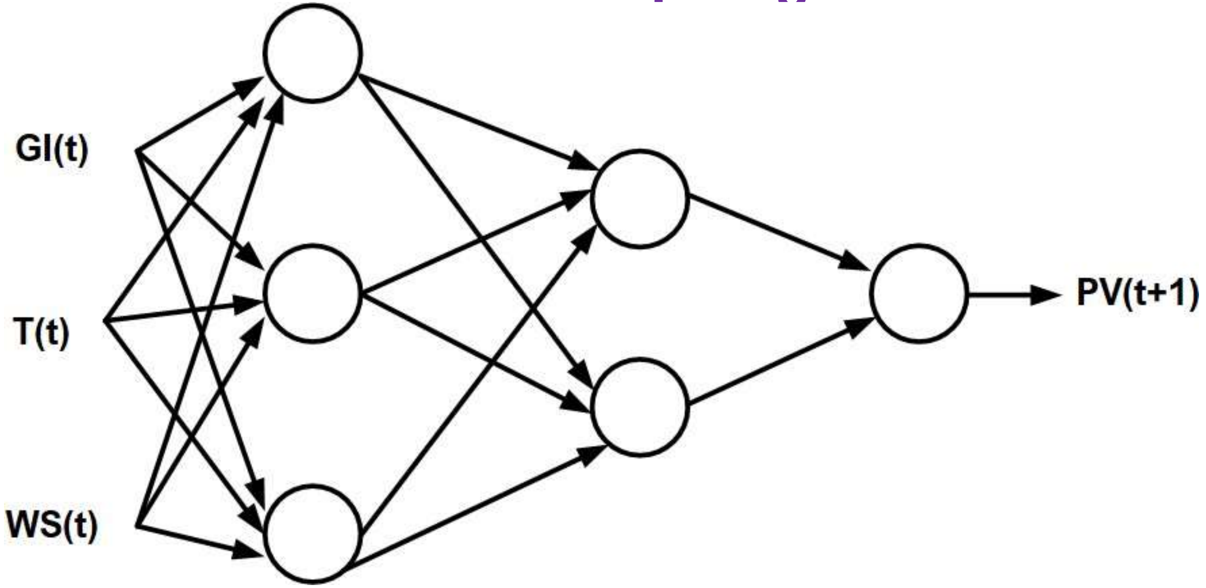


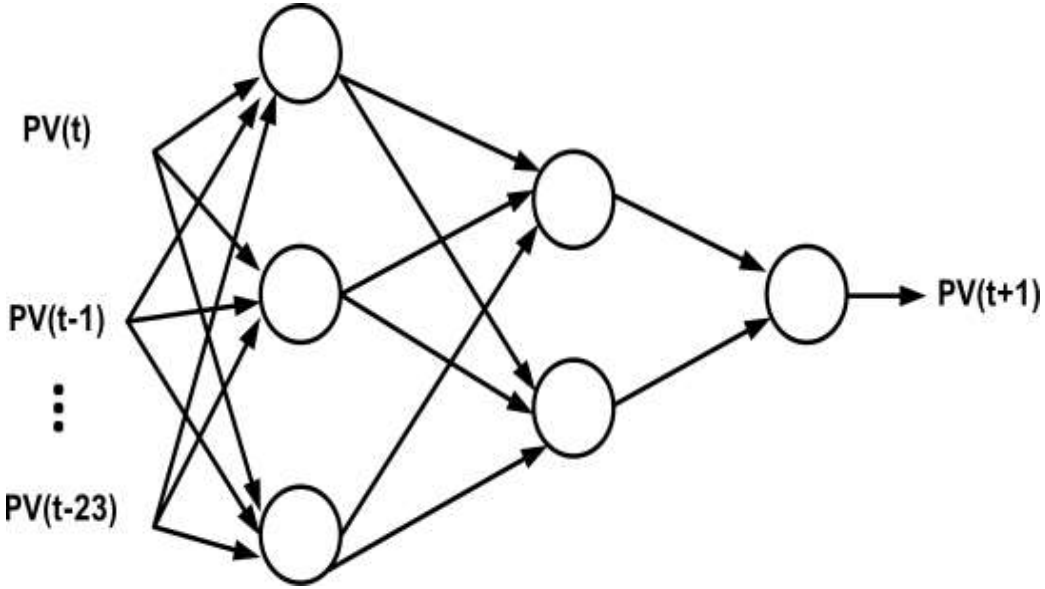
TP2

X_1	X_2	X_3	…	$X27$	Y
$PV\,(t)$	$PV\,(t-1)$	$PV\,(t-2)$	…	$PV\,(t-23)$	$PV\,(t+1)$
$PV\,(t+1)$	$PV\,(t)$	$PV\,(t-1)$	…	$PV\,(t-22)$	$PV\,(t+2)$
…	…	…	…	…	…

5 Forecast Evaluation

❖ **MLP TP1 and TP2 Topologies:**





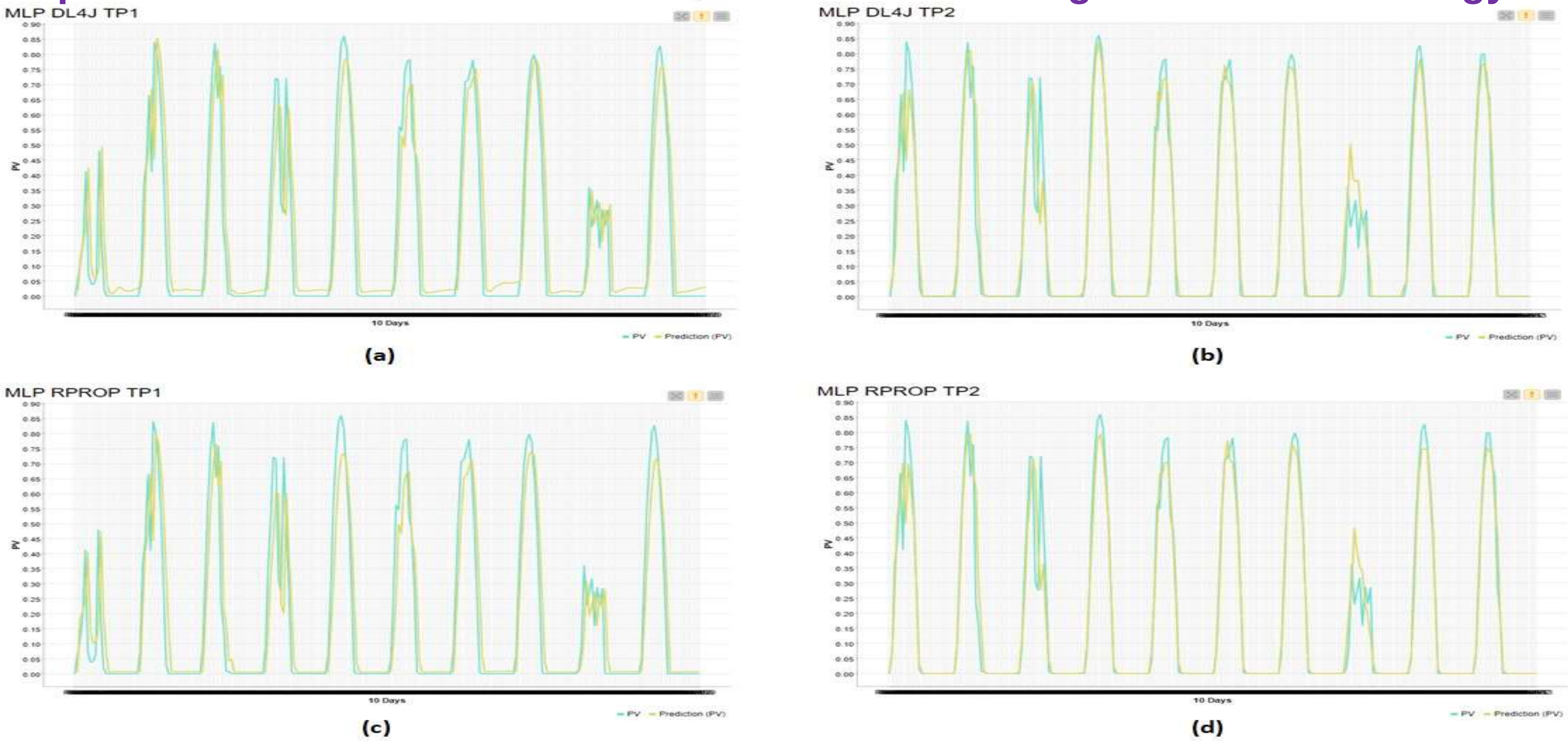
TP1

TP2

Model	**Execution time**		R_2	**MAE**	**MSE**
DL4J TP1	**24m 35s**	Train	0.816	0.084	0.016
		Test	0.814	0.037	0.007
DL4J TP2	**24m 35s**	Train	0.915	0.037	0.007
		Test	0.914	0.037	0.007

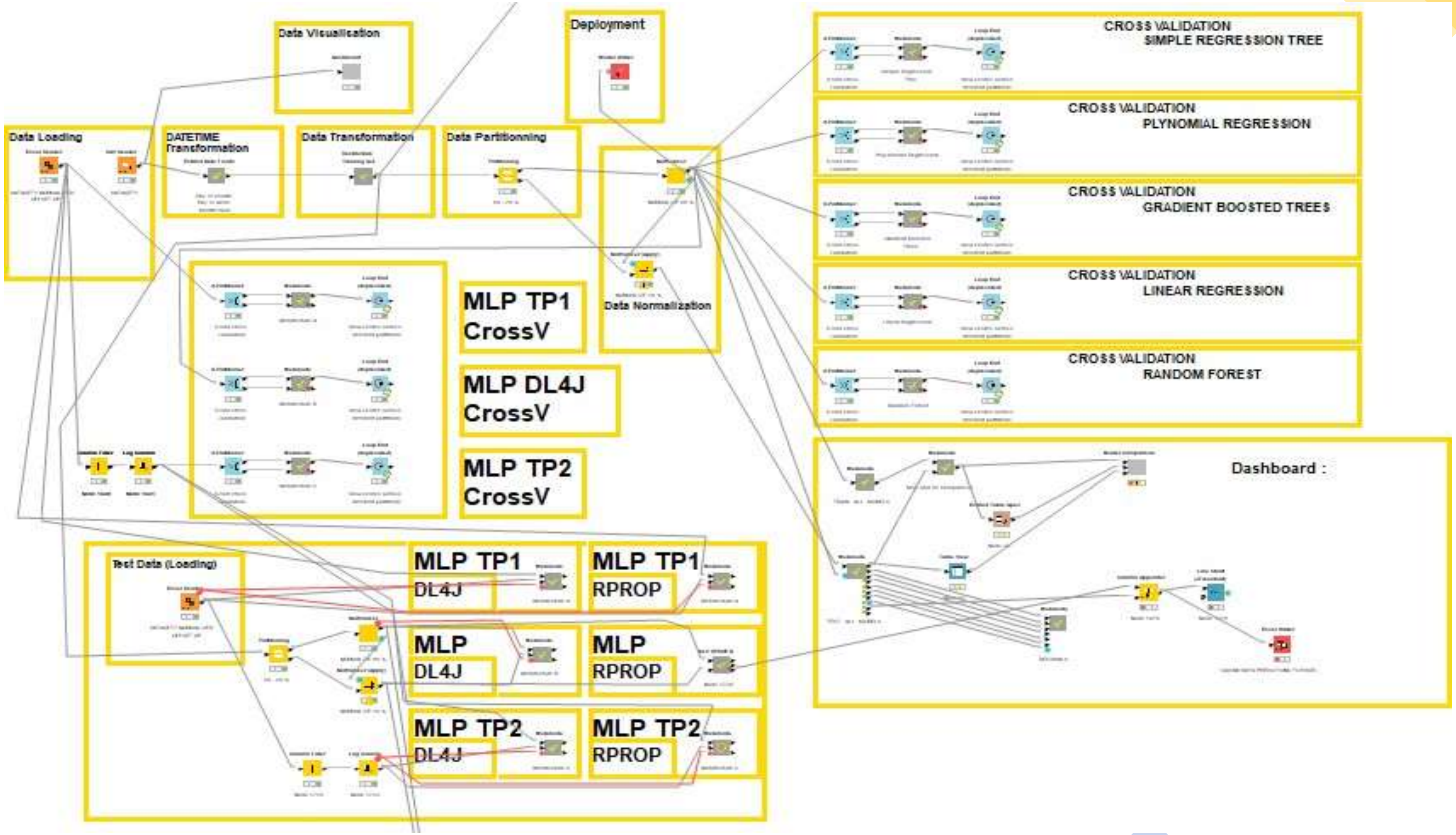
Model	**Execution time**		R_2	**MAE**	**MSE**
RPROP TP1	**10m 23s**	Train	0.828	0.071	0.014
		Test	0.828	0.070	0.014
RPROP TP2	**10m 55s**	Train	0.922	0.033	0.006
		Test	0.921	0.034	0.006

❖ **Comparison between train and test scores for forecasting models of PV energy**



5 Forecast Evaluation

❖ Global Workflow System:



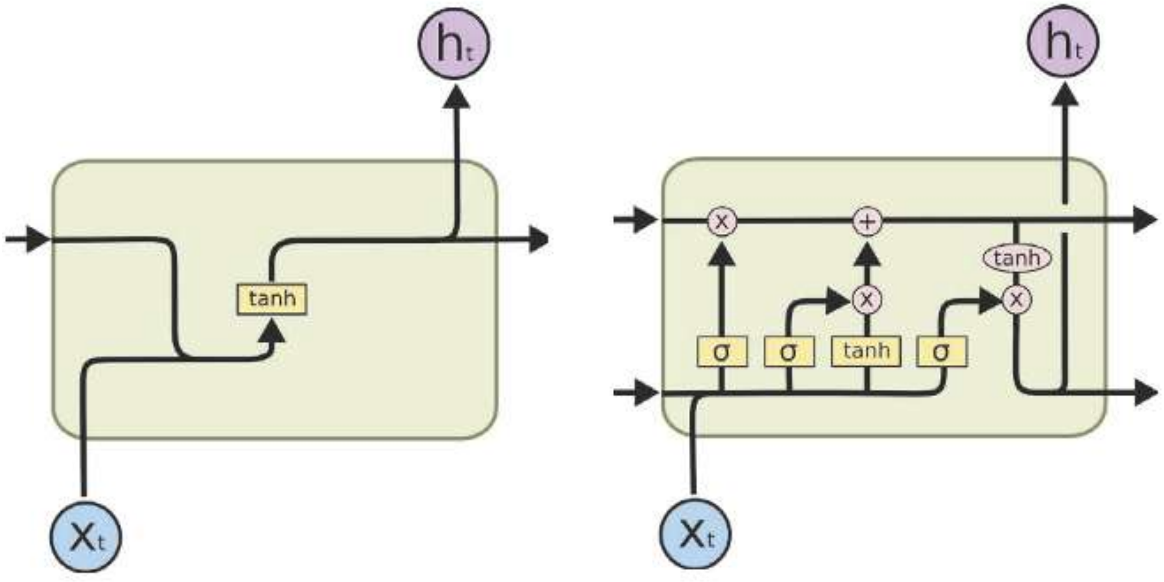
RNN vs LSTM

- Vanishing gradient problem.

When dealing with a time series, it tends to forget old information. When there is a distant relationship of unknown length, we wish to have a “memory” to it.

P_{PV} Prediction using LSTM

RNN vs LSTM

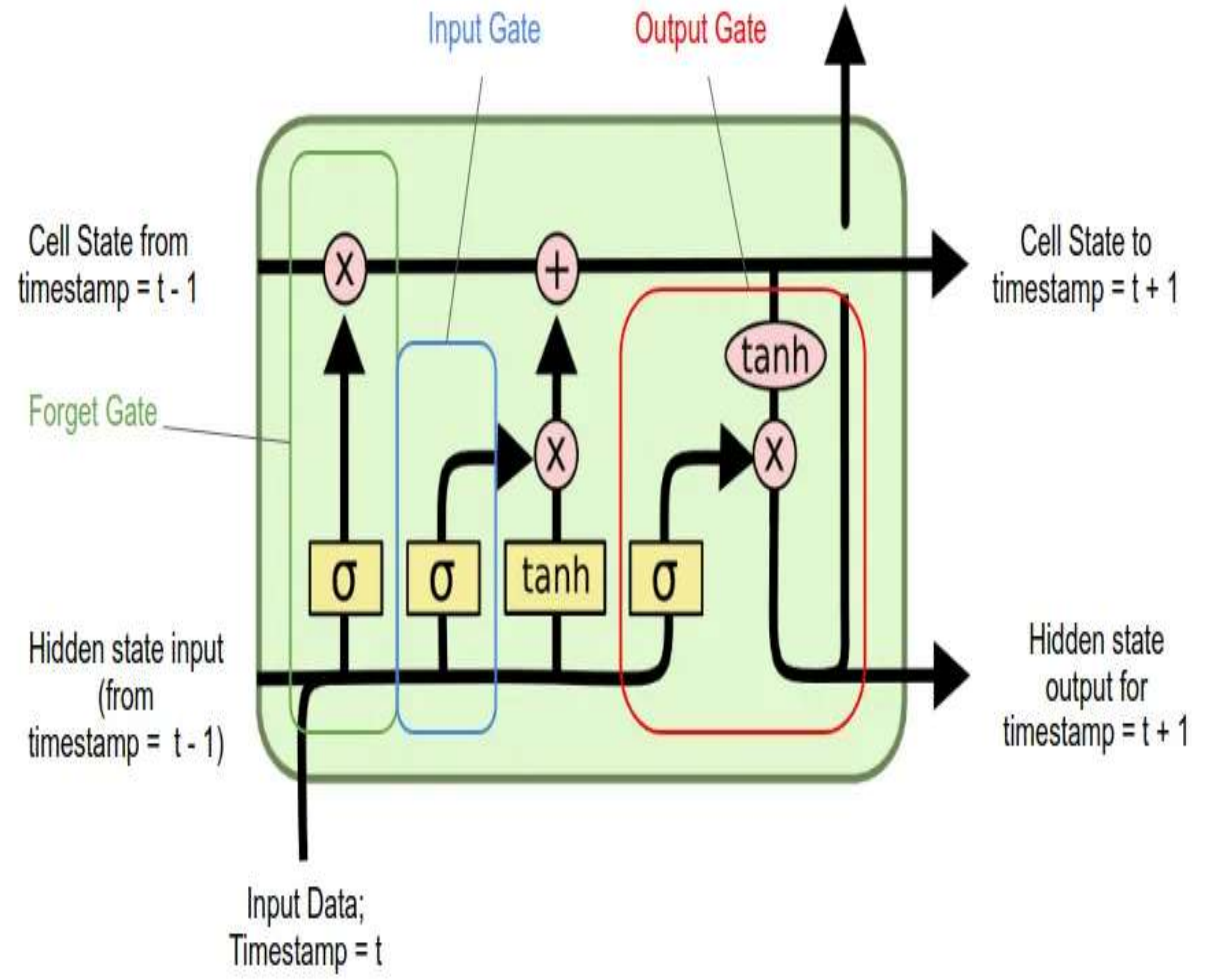


(a) RNN (b) LSTM

6

P_{PV} Prediction using LSTM

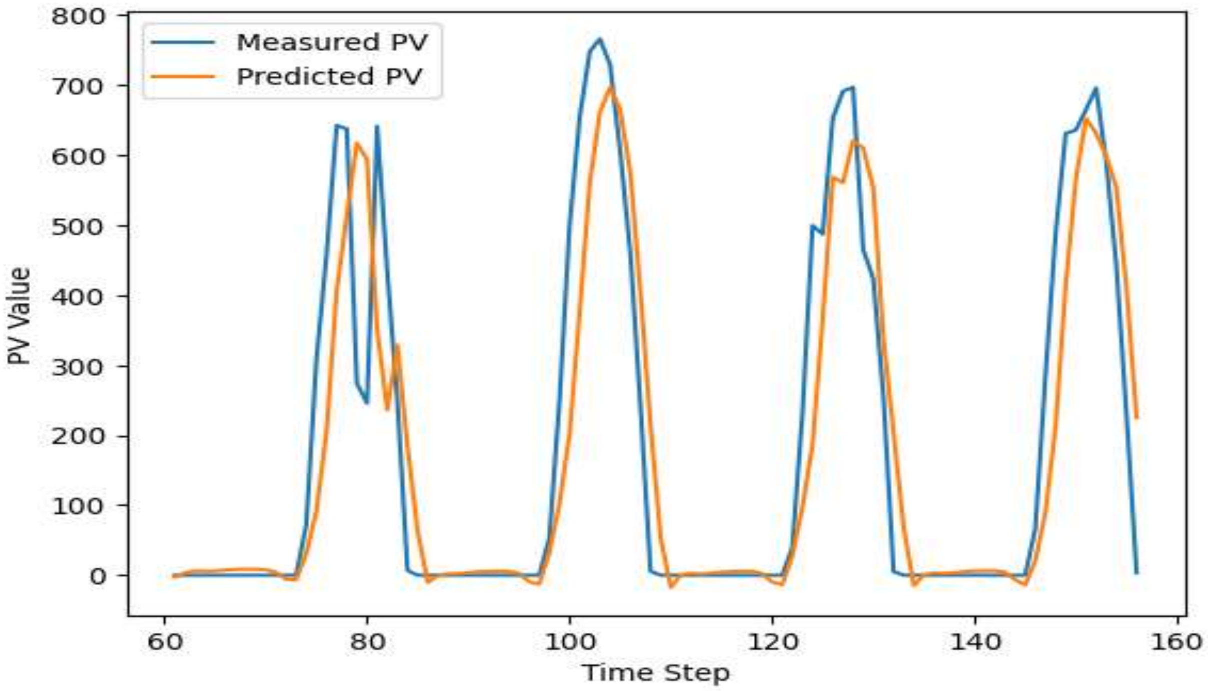
LSTM Network Architecture



6

P_{PV} Prediction using LSTM

LSTM Performance

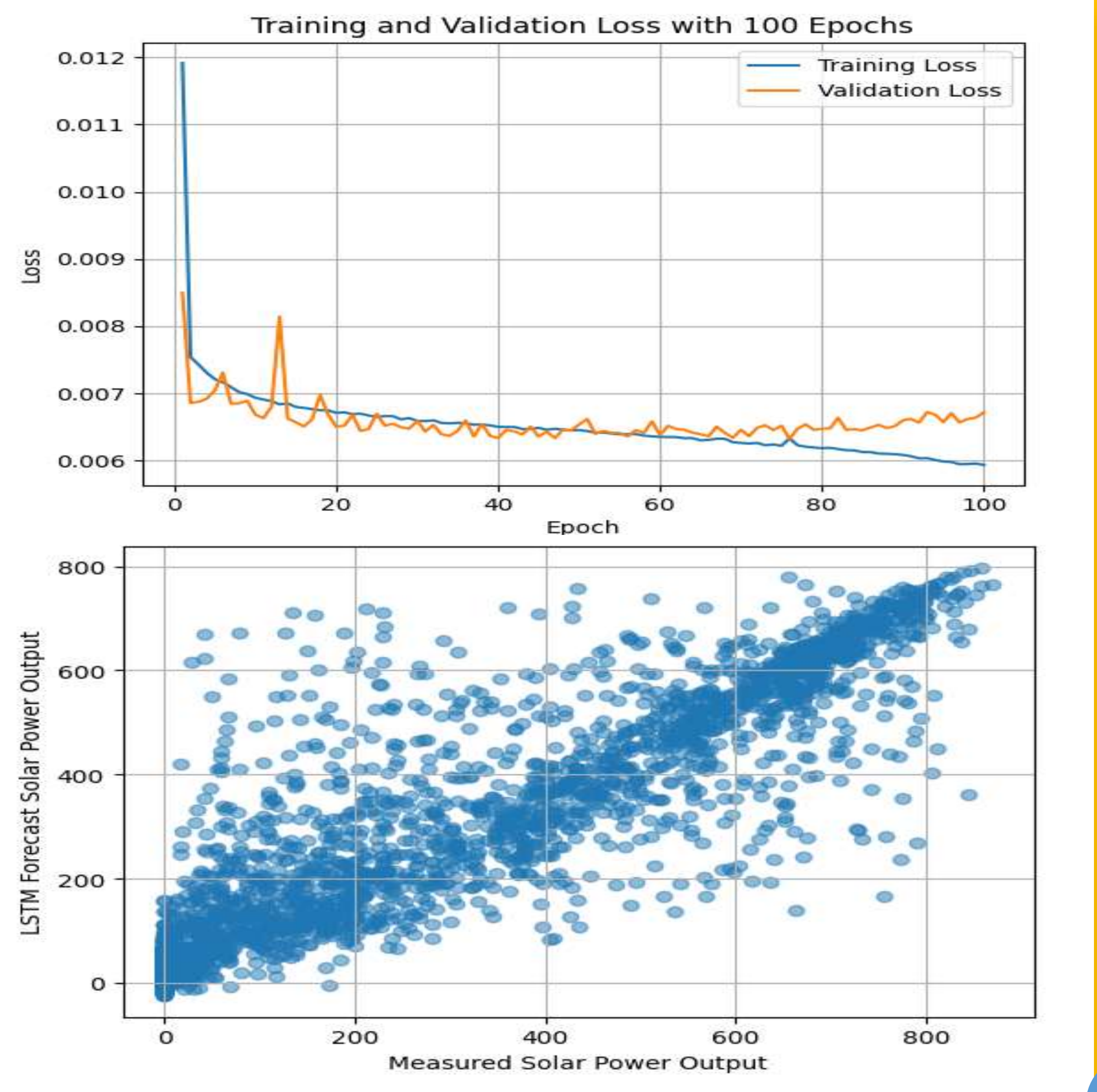


Algorithm	Valor
MSE	0.00653384468361185
RMSE	0.08083220078416677
MAE	0.036050230863963
R^2	0.9208829381609711

6

P_{PV} Prediction using LSTM

LSTM Performance



7

Conclusion & Perspectives

- The generation of **photovoltaic** energy is closely **dependent** on prevailing **weather conditions**.

- The construction of forecasters involves the utilization of 27 inputs, among which the inclusion of **PV output power** for the previous 24 hours and **mean daily climatic conditions** has been confirmed as the **optimal choice** for achieving accurate forecasts.

- Results from simulations using the KNIME platform indicate that **Neural Networks** (MLP) and **Ensemble Learning** outperform other methods, achieving an accuracy of over 0.91 R-squared and an MSE of less than 0.007 for one-hour ahead PV output power forecasting.

- Long-term forecasting (24 hours ahead) of PV energy is explored using an RNN algorithm called **LSTM**, offering both **short and long-term** predictions with an **MSE** of less than **0.0065** and an **R2** of more than **0.92**.

7

Conclusion & Perspectives

- There are **plans** to create commercial software based on the Python script code, featuring a graphical user interface within a Visual Studio environment.

- **Future work** may involve implementing and evaluating additional RNN models, such as the GRU) model, and incorporating additional data, such as meteorological information or images , to enhance forecast quality.

Thank you for your attention