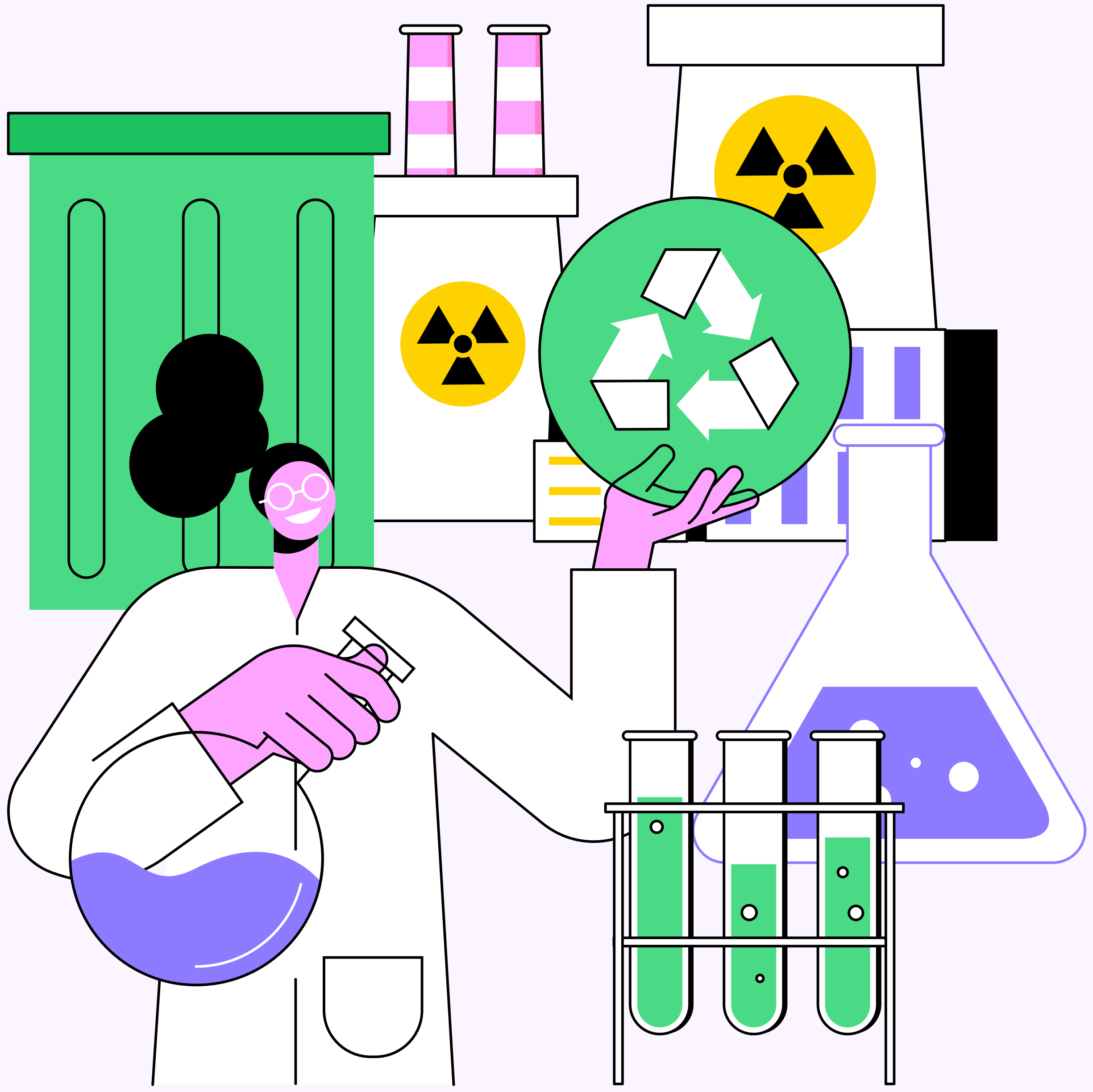
FUEL IGNITION PROPERTY PREDICTION USING REGRESSION MODELS

‘Igniting the Future of Fuels’

Presented by Group 1



PROBLEM AT HAND

Experimental determination of ignition quality indicators such as the **Derived Cetane Number (DCN), Research Octane Number (RON), and Motor Octane Number (MON)** is time-consuming, costly, and limited by data availability.

Hence, there is a need for accurate, data-driven models that can predict these properties directly from molecular structure.

In this project, we have worked on different models, and have tried to predict the needed quality indicators.

MODELS WE HAVE WORKED ON

1. USING RANDOM FOREST

2. GRAPH NEURAL NETWORK

MAP ON USING **RANDOM FOREST**

A complete journey from raw data to deployment

STEP 01

Define the Problem

We identified what type of model is needed for the problem. We have decided to use **Regression models**.

STEP 02

Data Collection

We have went through research papers and have collected datasets for model training.

STEP 03

Data Filling

Handled missing (NaN) values using Regression Impute to ensure data completeness.

STEP 04

Feature Engineering

Generated **molecular descriptors** from SMILES for representing properties of fuels.

STEP 05

Model Development

Performed train validation split and trained regression model to predict ignition quality.

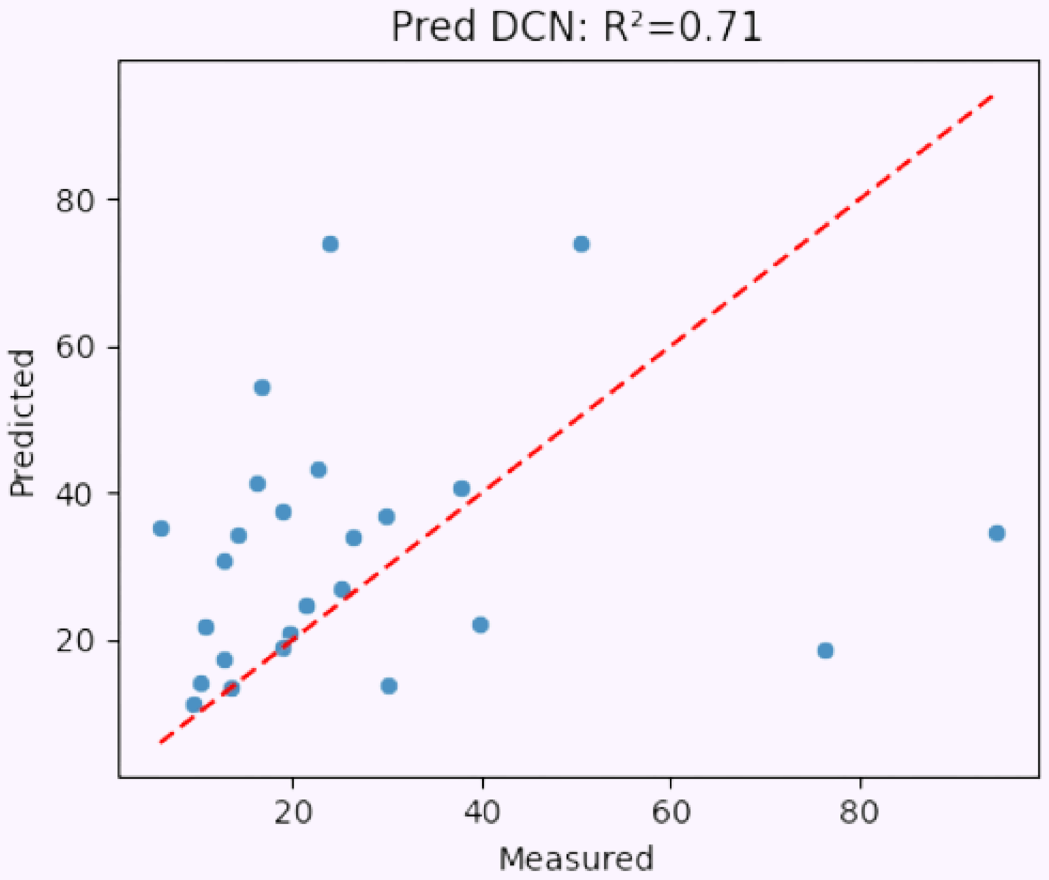
STEP 06

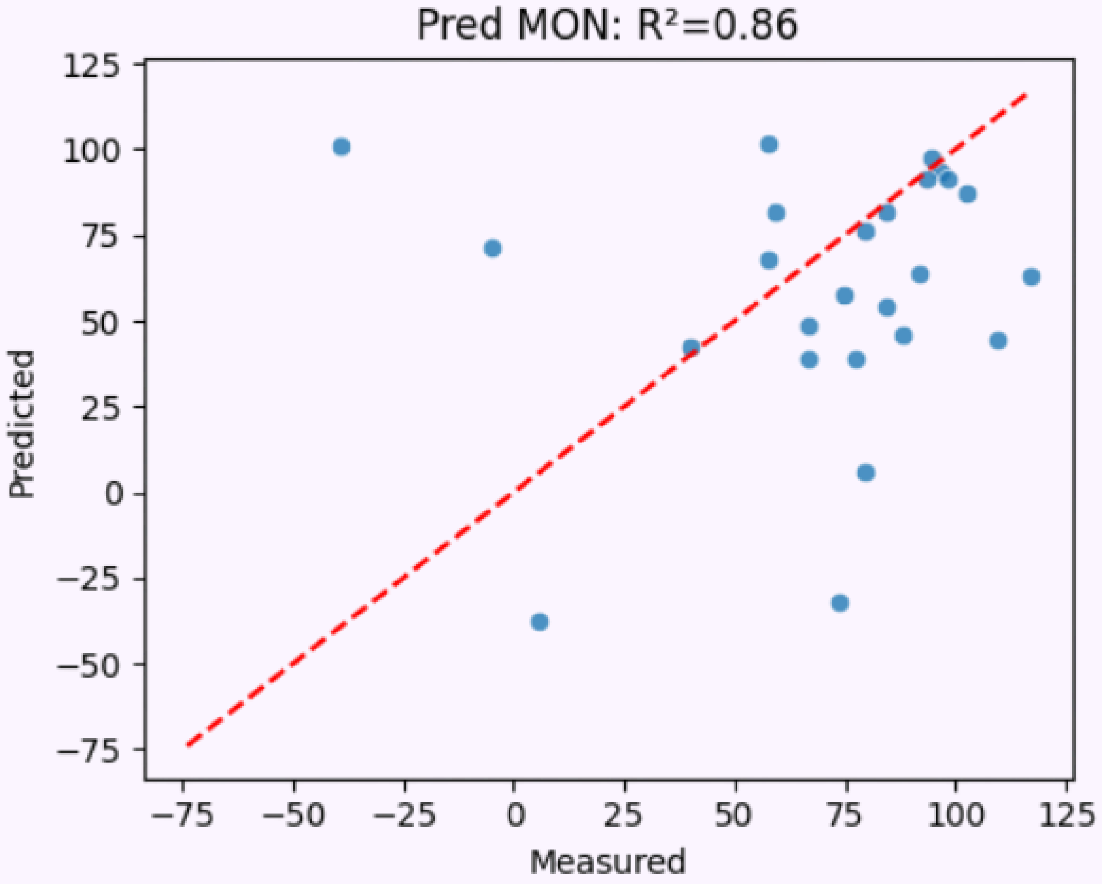
Model Evaluation

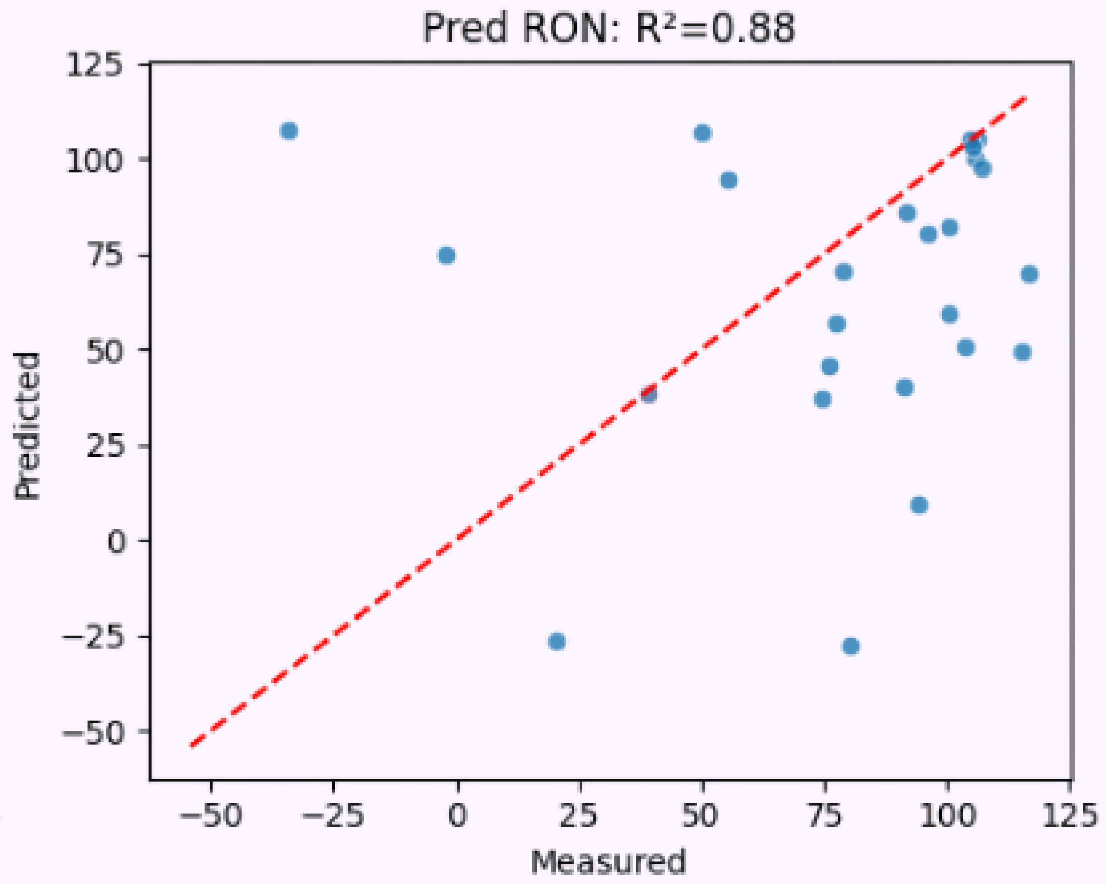
Evaluated performance using **R^2 scores** to assess model accuracy.

RANDOM FOREST

Model Training







Validation for DCN

1. MAE = 6.181
2. RMSE = 11.543
3. R2 Score = 0.708

Validation for MON

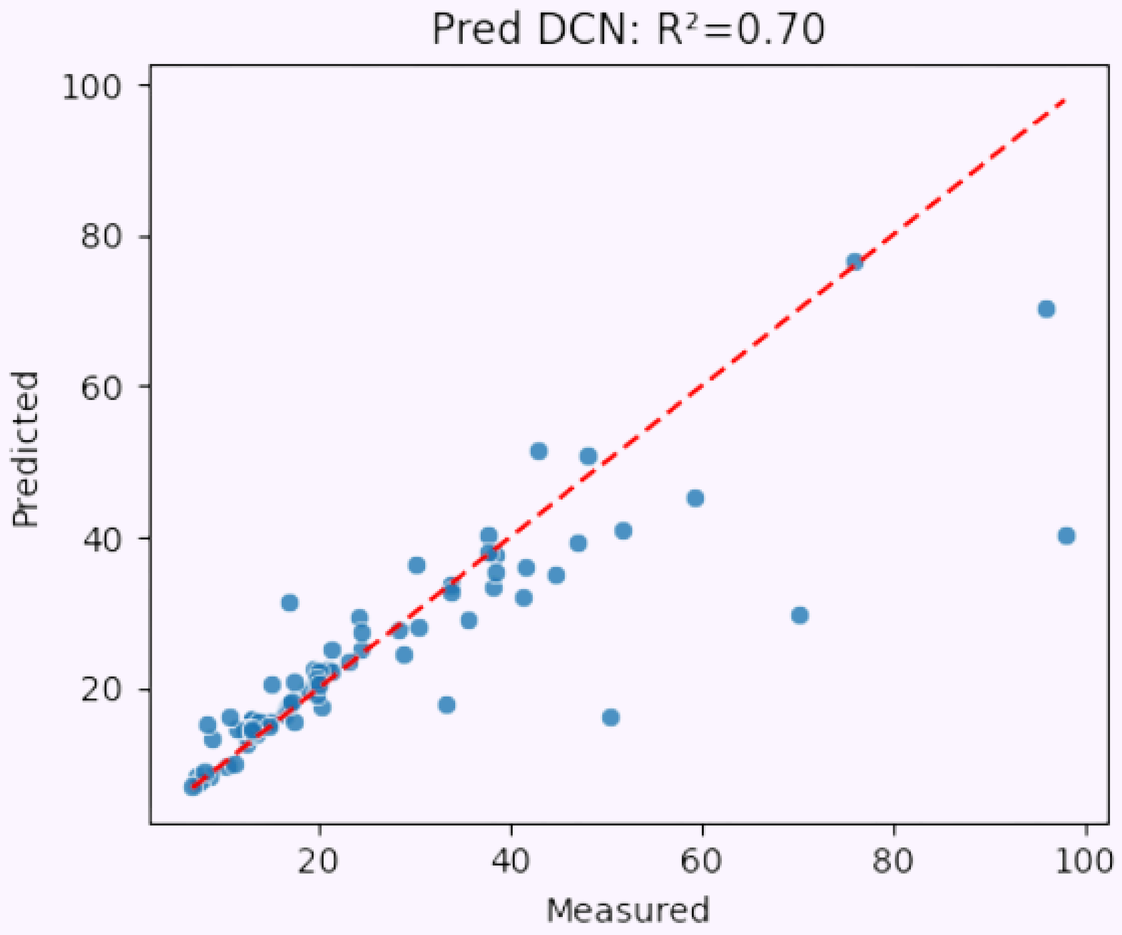
1. MAE = 9.709
2. RMSE =14.566
3. R^2 = 0.859

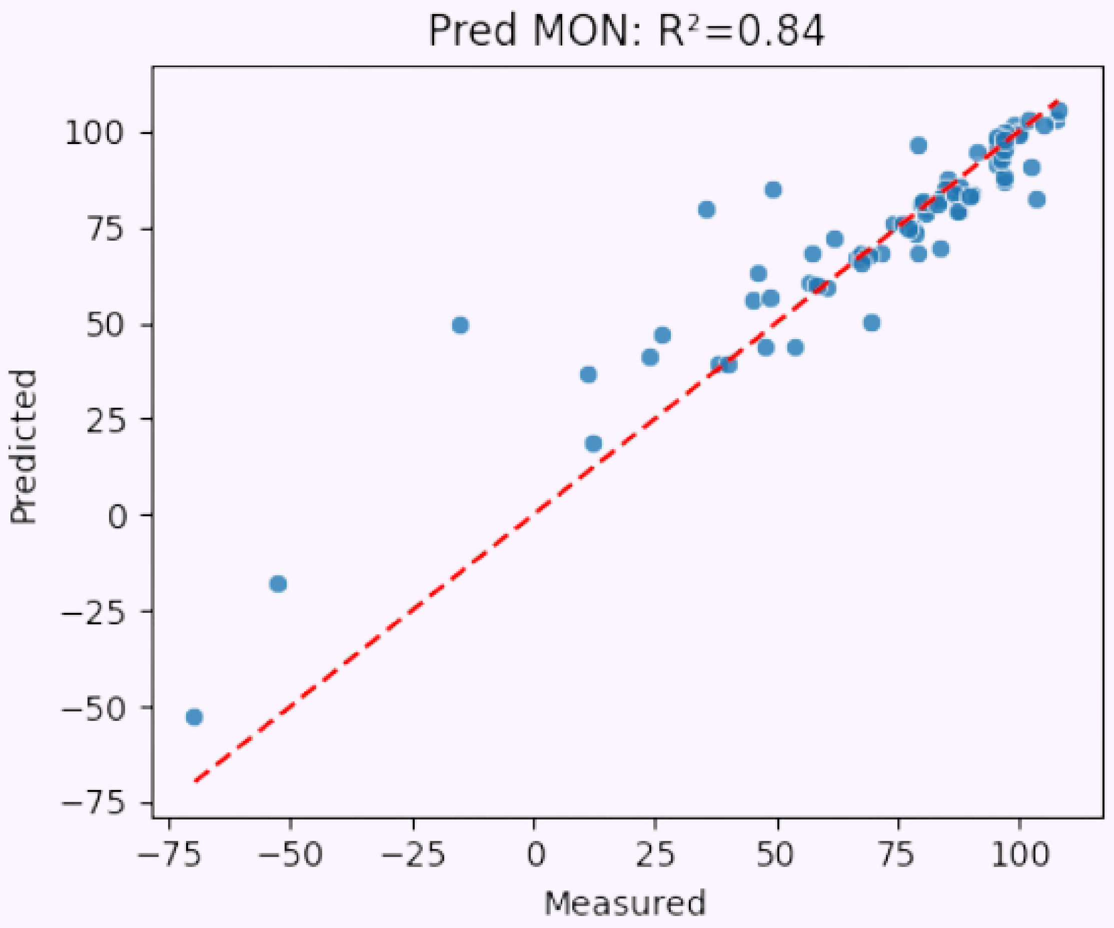
Validation for RON

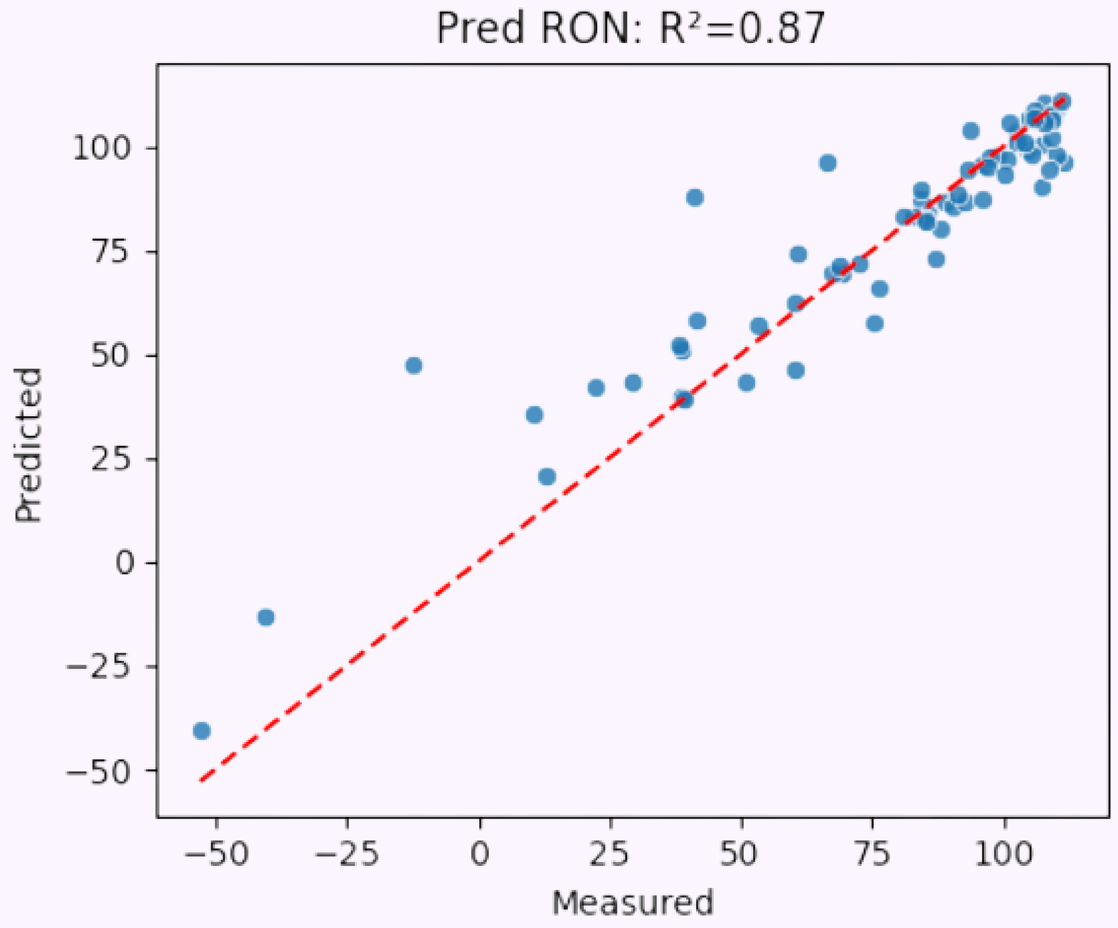
1. MAE = 8.709
2. RMSE = 13.497
3. R^2 = 0.885

RANDOM FOREST

Model Testing







Validation for DCN

1. MAE= 4.992
2. RMSE=10.504
3. R^2 =0.699

Validation for MON

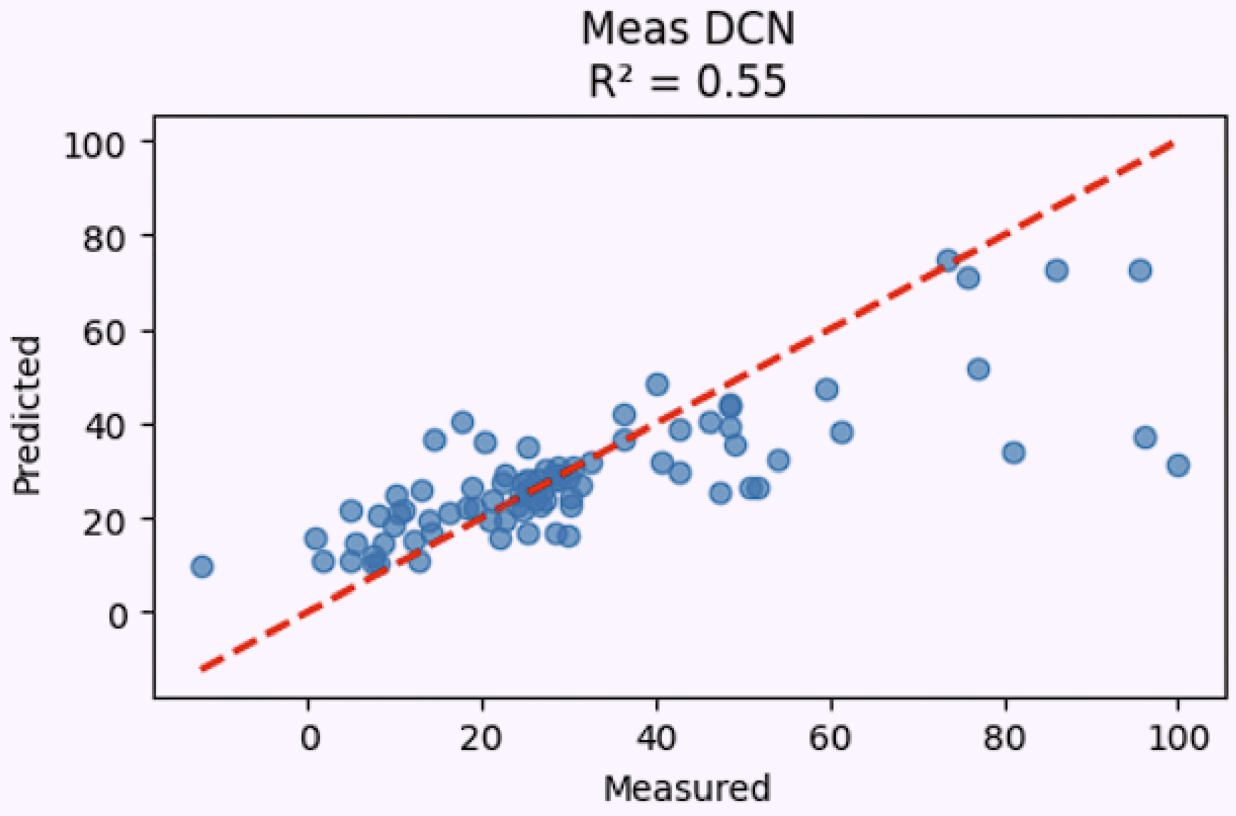
1. MAE= 7.621
2. RMSE=13.424
3. R^2 = 0.837

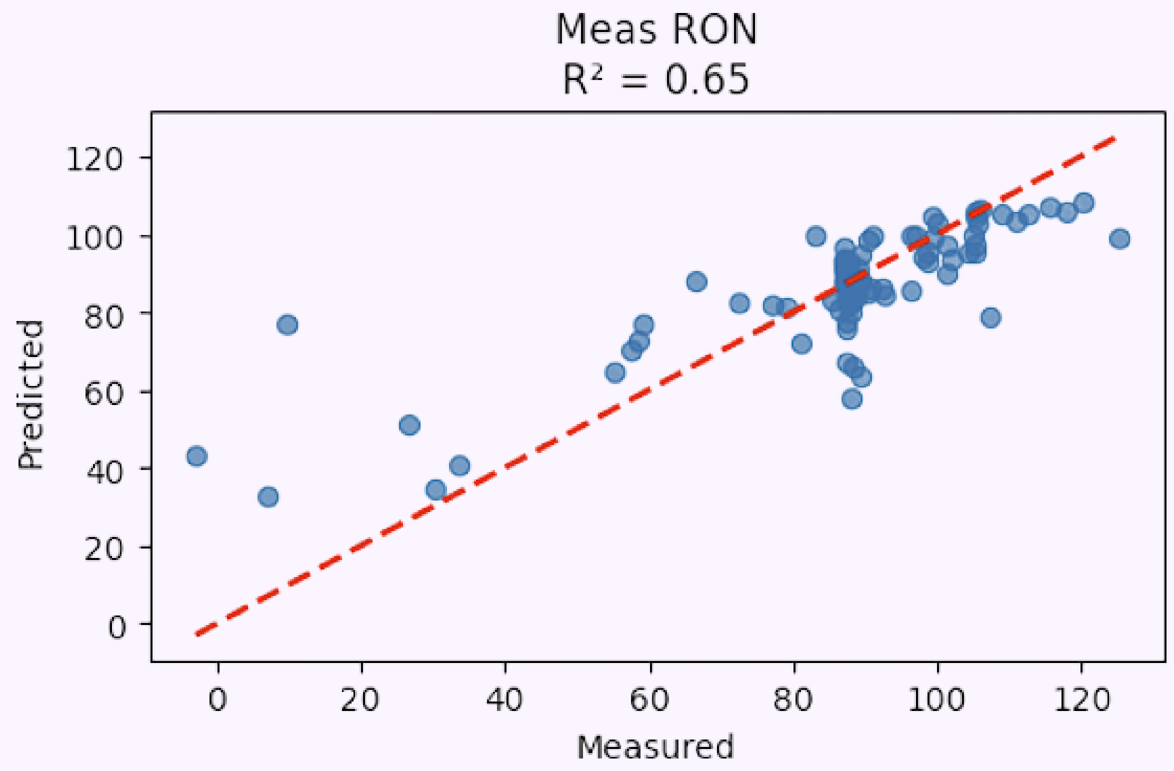
Validation for RON

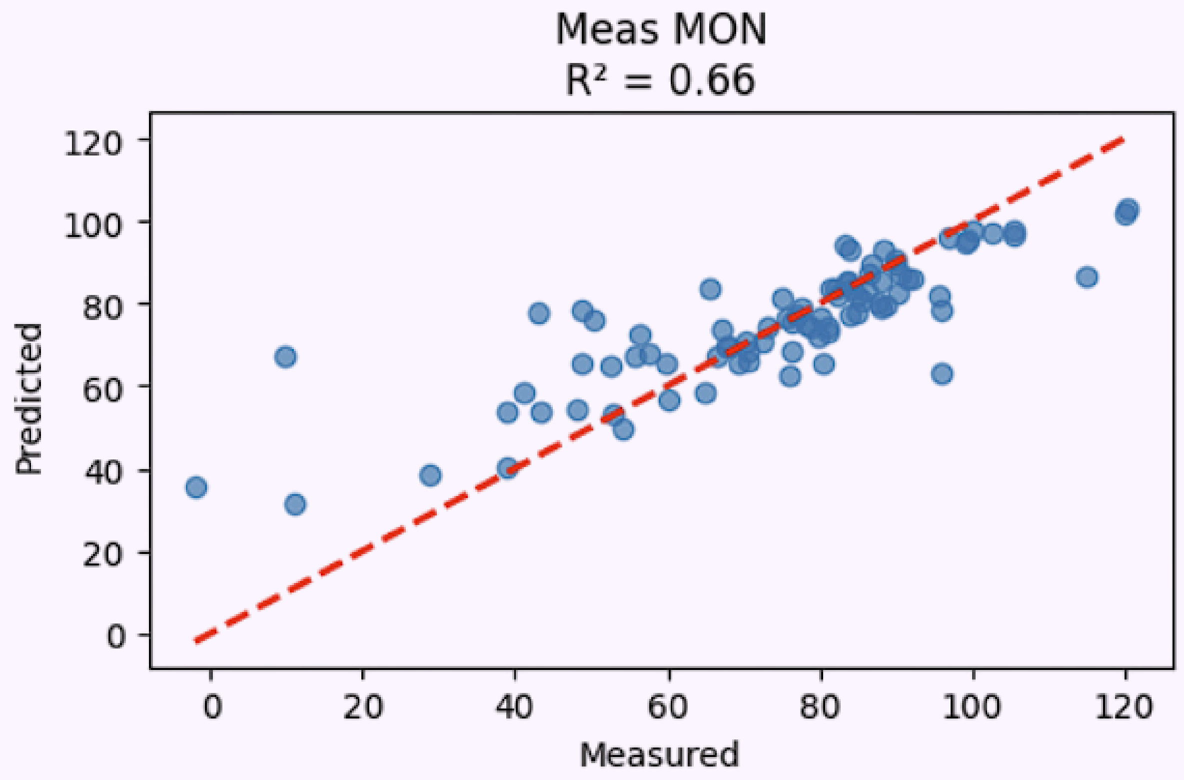
1. MAE= 7.633
2. RMSE=12.754
3. R^2 = 0.869

RANDOM FOREST

Regression Impute for estimating na values

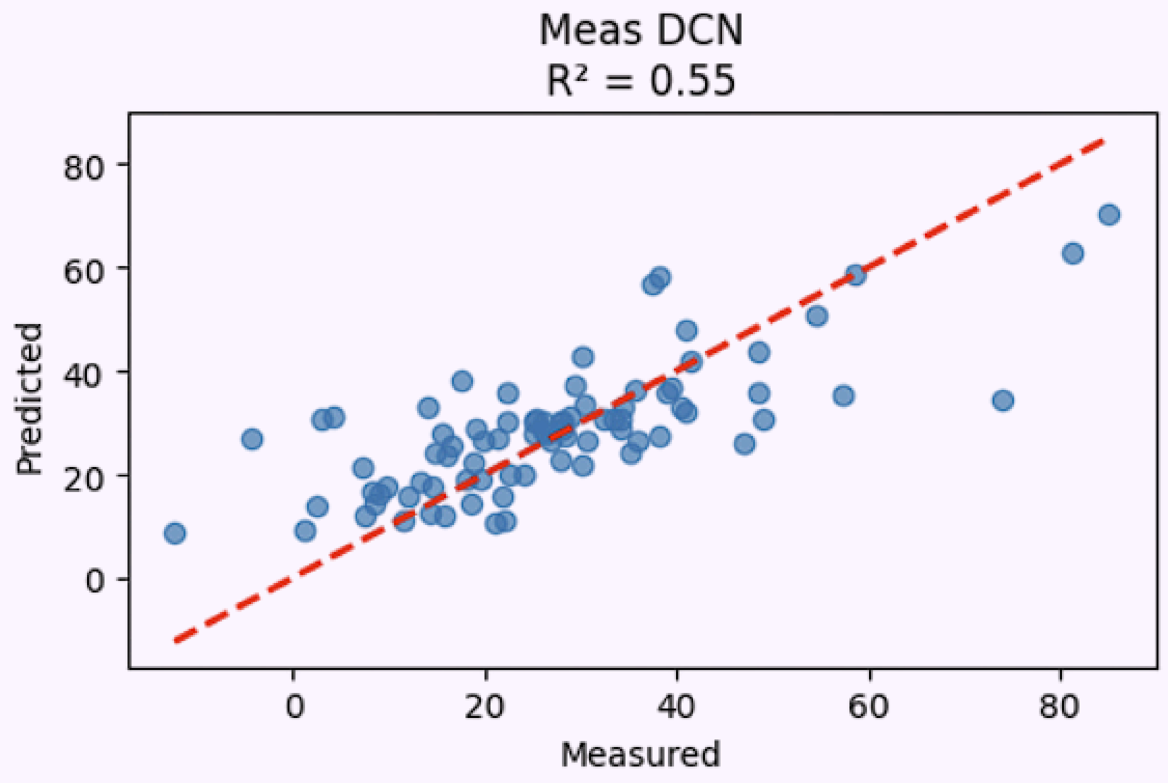


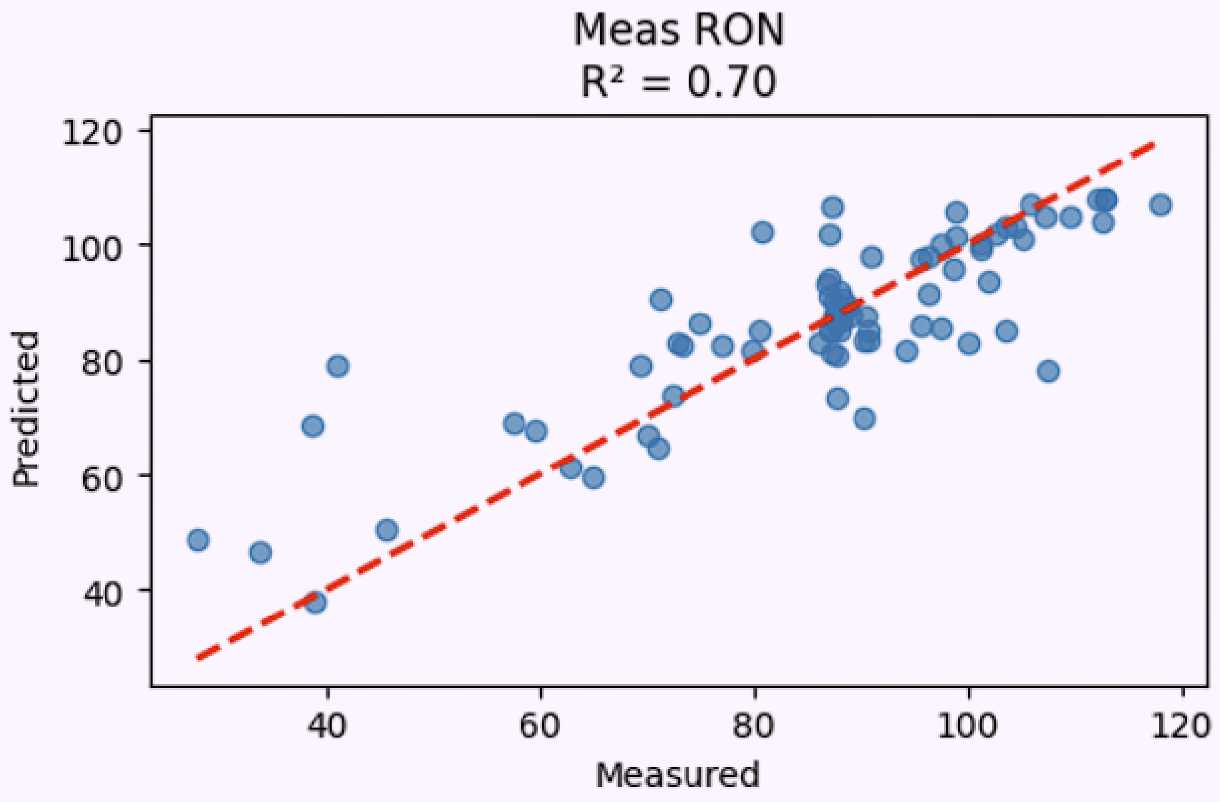


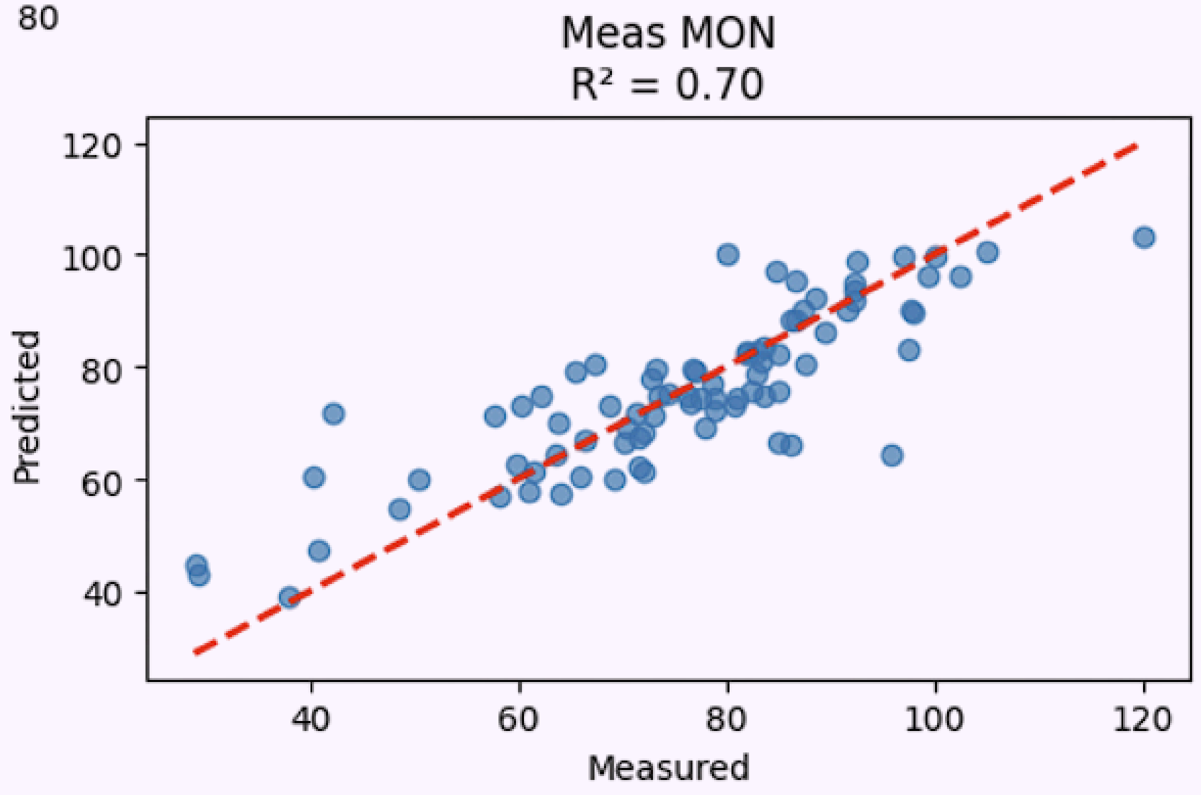


RANDOM FOREST

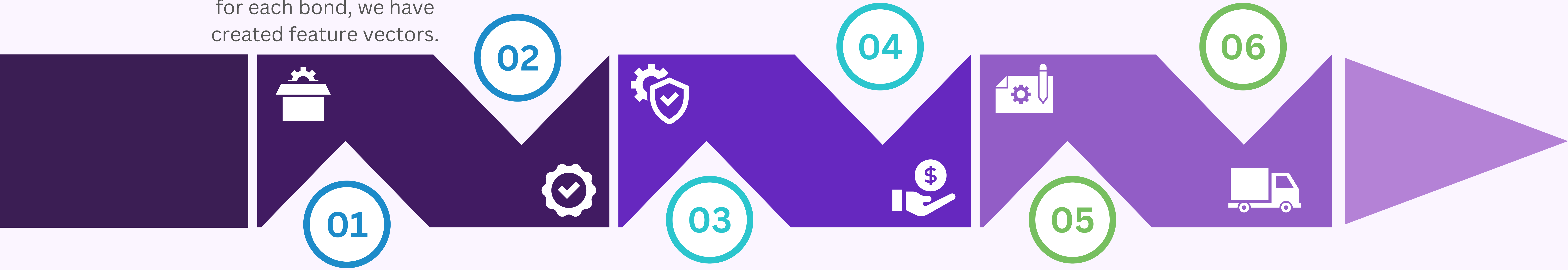
Regression impute for filling na and removing outliers







Map on using **GNN**



Data Acquisition

Remove missing or invalid **SMILES** and converted all values to numerical form so that we can normalize our target values

Feature Engineering

For each atom in molecule, we have used atomic features at node level and for each bond, we have created feature vectors.

Standardization

We have used standard scalars (mean = 0, standard deviation = 1).

GNN Architecture

In this we have divided into message passing phase and readout phase.

Model Training

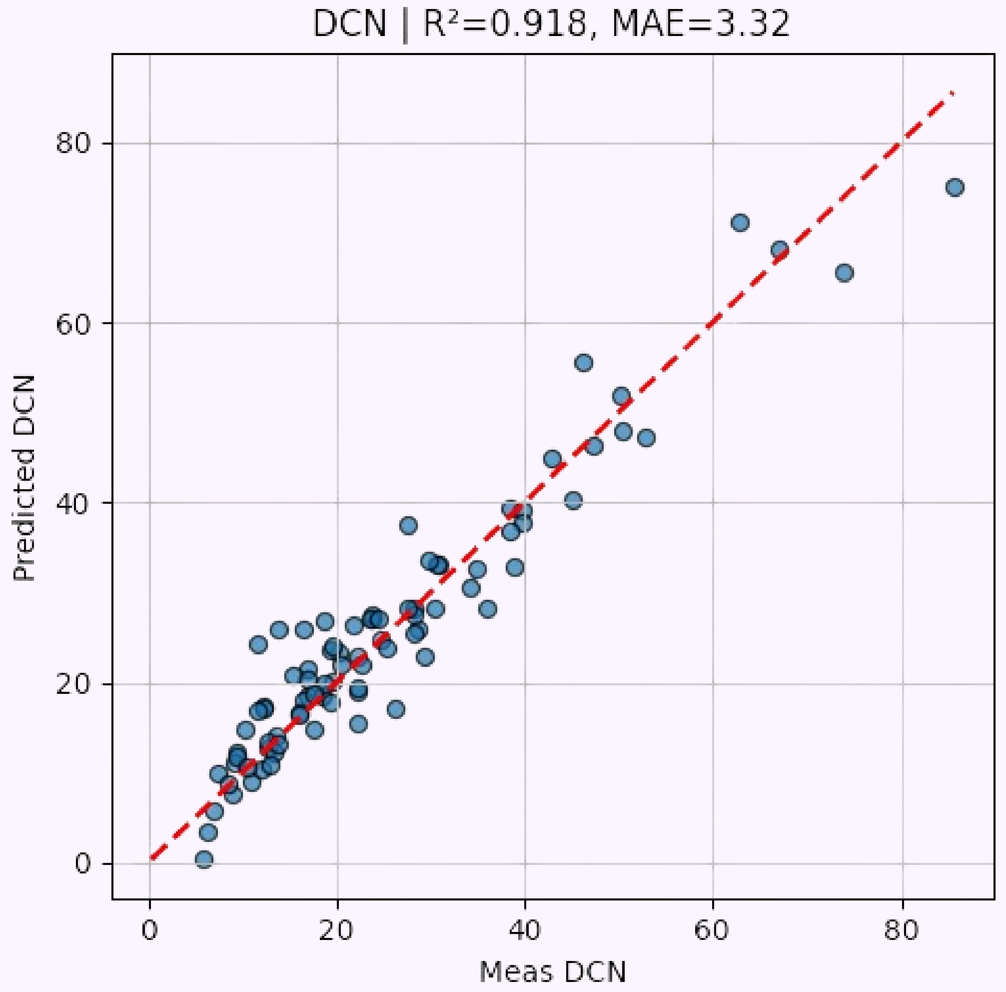
We have used early stopping technique for regularization and evaluated every epoch.

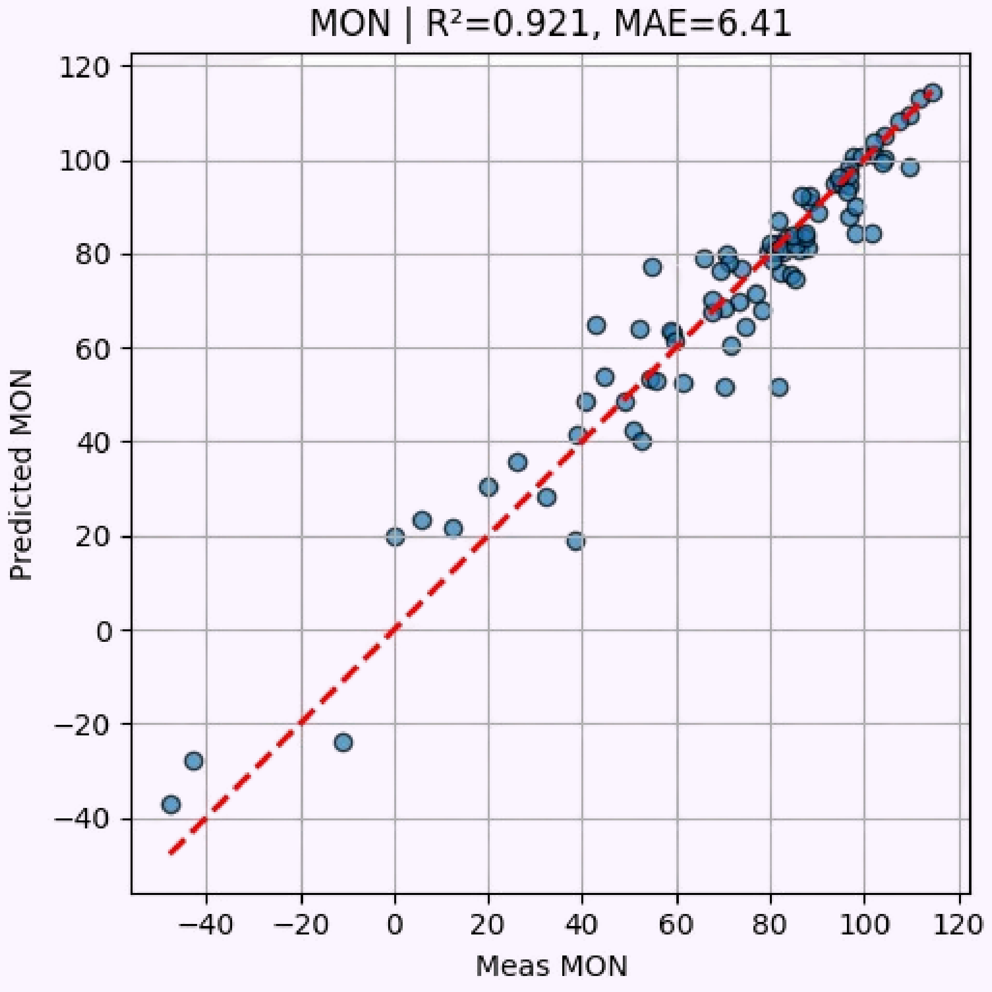
Model Evaluation

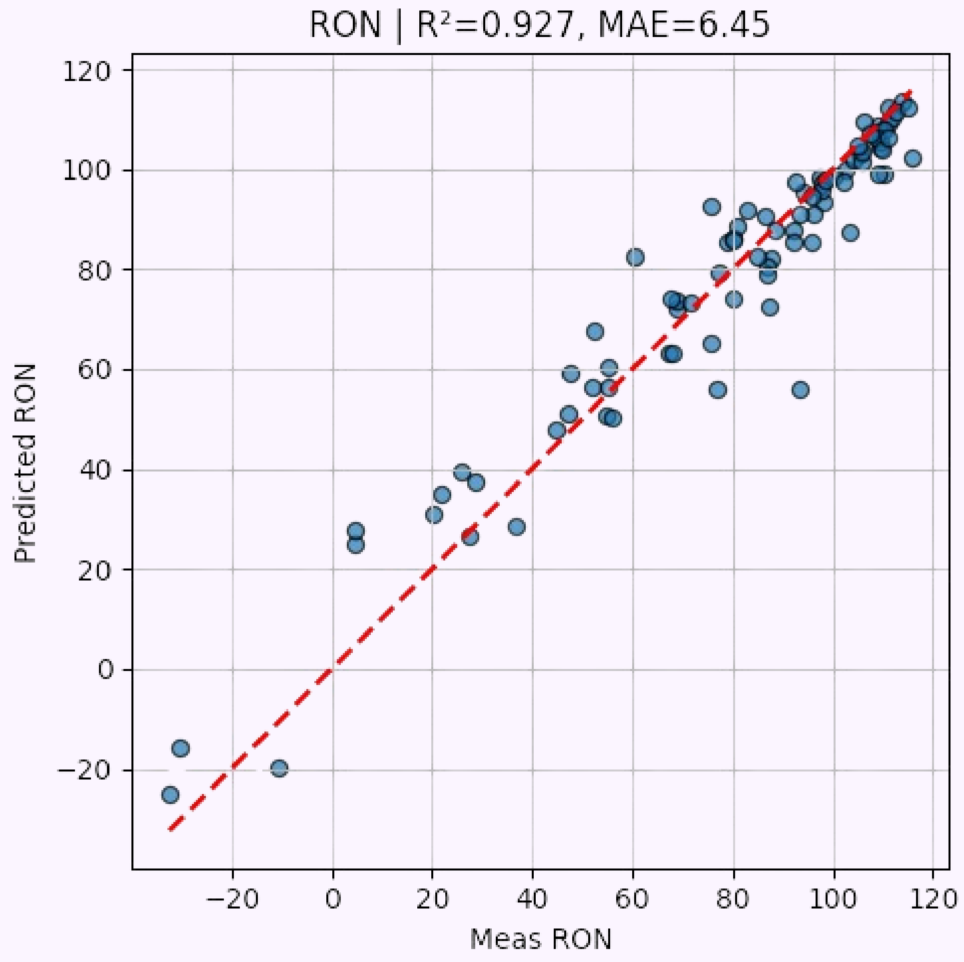
Used R2 score and MAE for analysis of our trained model.

GRAPH NEURAL NETWORK

Model Training on validation loss







Validation for DCN

1. MAE = 6.181
2. RMSE = 11.543
3. R2 Score = 0.708

Validation for MON

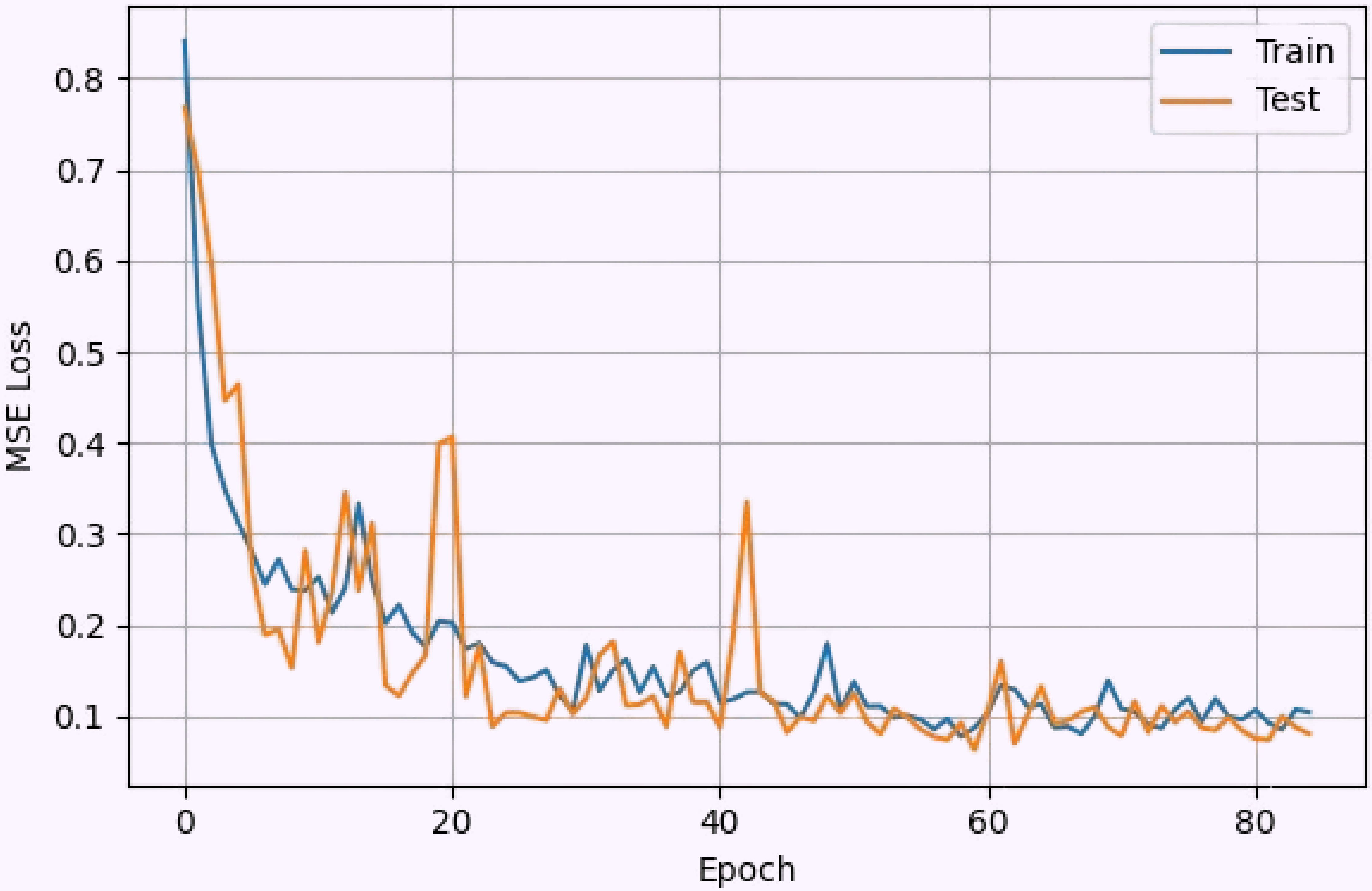
1. MAE = 9.709
2. RMSE =14.566
3. R^2 = 0.859

Validation for RON

1. MAE = 8.709
2. RMSE = 13.497
3. R^2 = 0.885

GRAPH NEURAL NETWORK

Using Early Stopping Regularization Technique



GITHUB REPOSITORY

Scan the QR code to get acces to code files and dataset

THANK YOU