

ITMO UNIVERSITY

Data Analysis Task: Jaipur Weather Forecasting

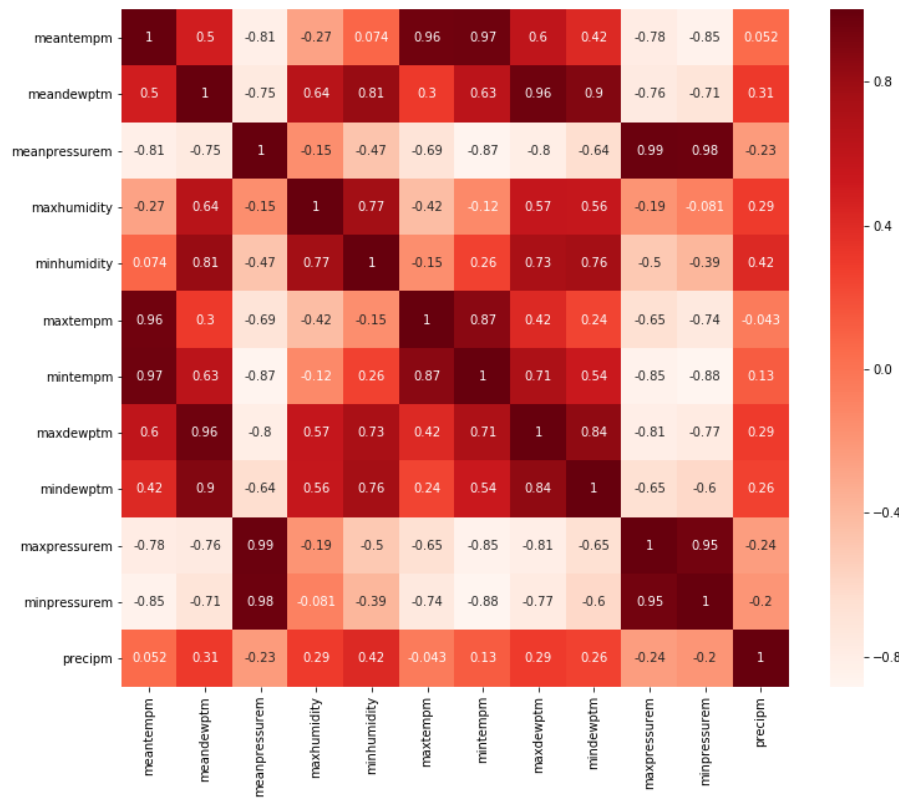
Shivam Mehta, M41331

Why This Dataset?

- ✓ Jaipur is one of the most Dry Regions in India, and Rain Prediction can be helpful for farmer of that area
- ✓ I aimed to develop a model which helped the farmer for saving from huge loss during cultivation time by knowing about the amount of rainfall
- ✓ Thus the farmer can plant his crops accordingly

EDA – Exploratory Data Analysis

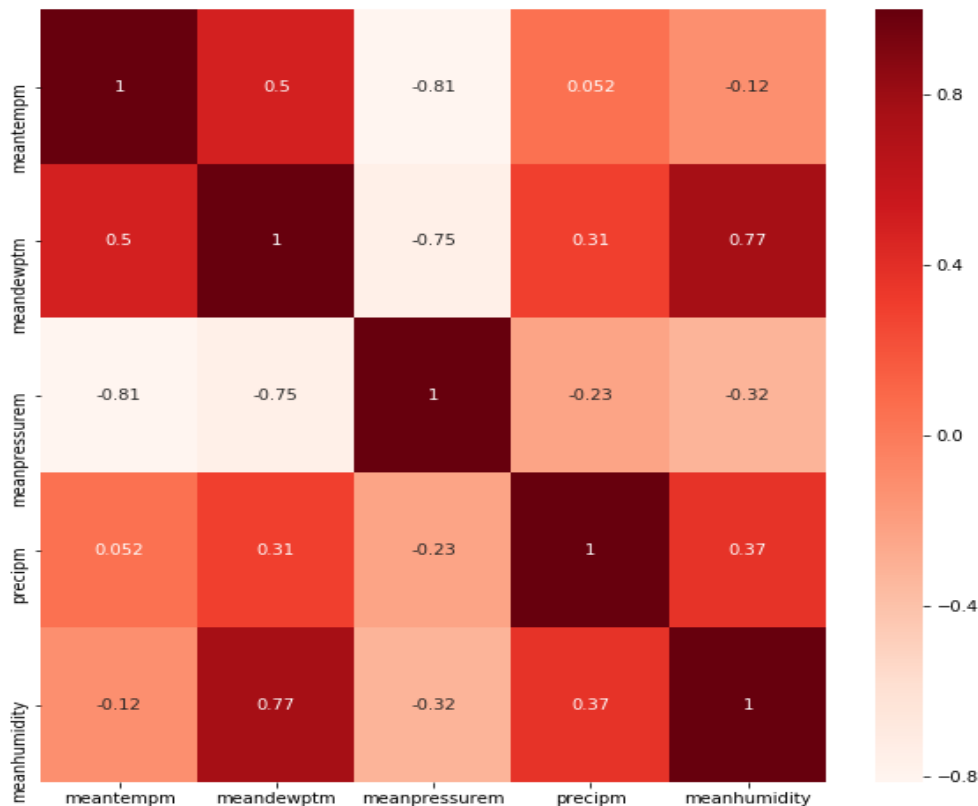
✓ Correlation of Features



EDA – Feature Engineering

- ✓ Feature Engineering : Based on EDA
 - Merging Column
 - Cleaning Data
 - Normalizing
 - Remove values that are constant

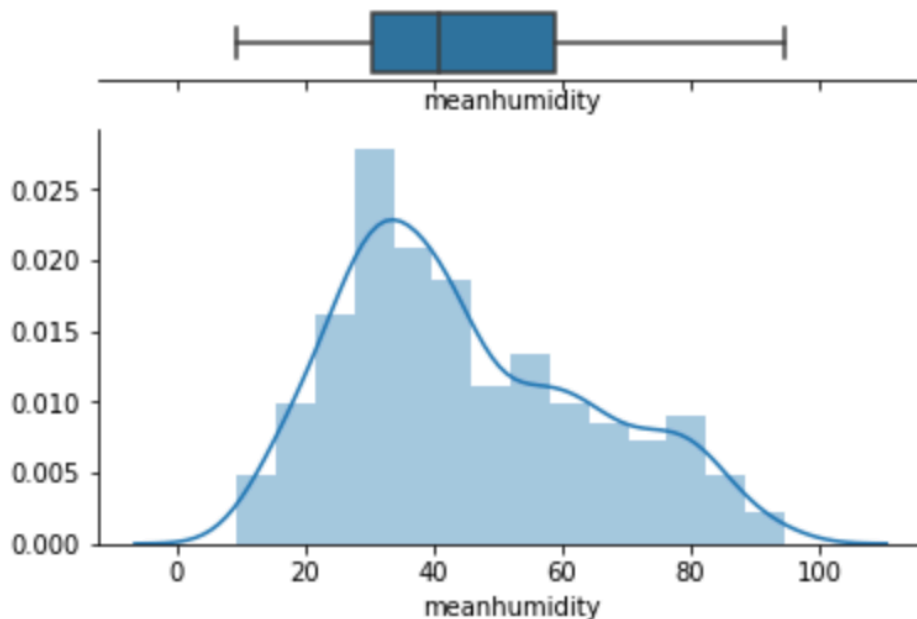
EDA – Feature Correlation



EDA – Feature Distribution (1)

- ✓ Features with Maximum Correlations had no outlier and their distributions were:

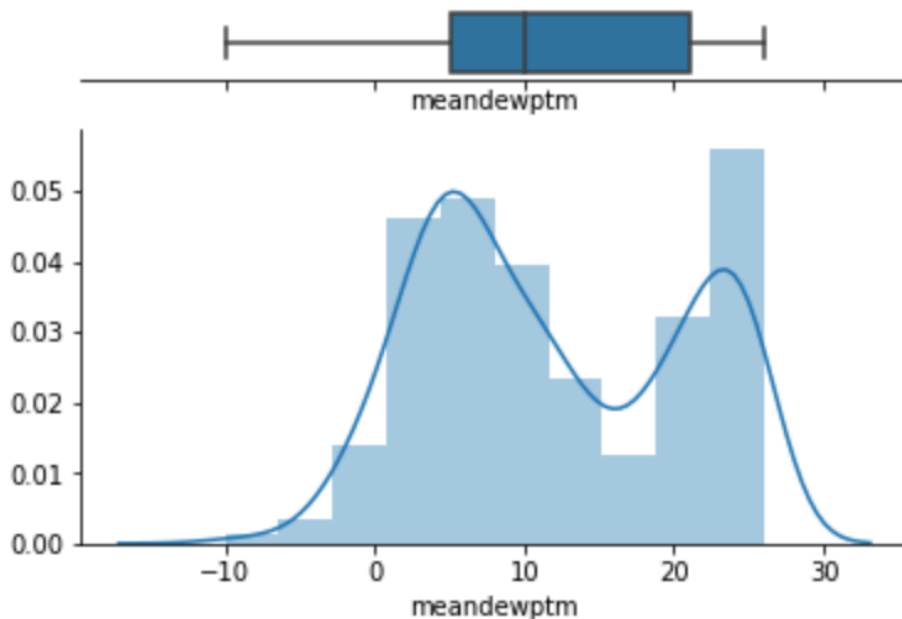
1. Mean Humidity



EDA – Feature Distribution (2)

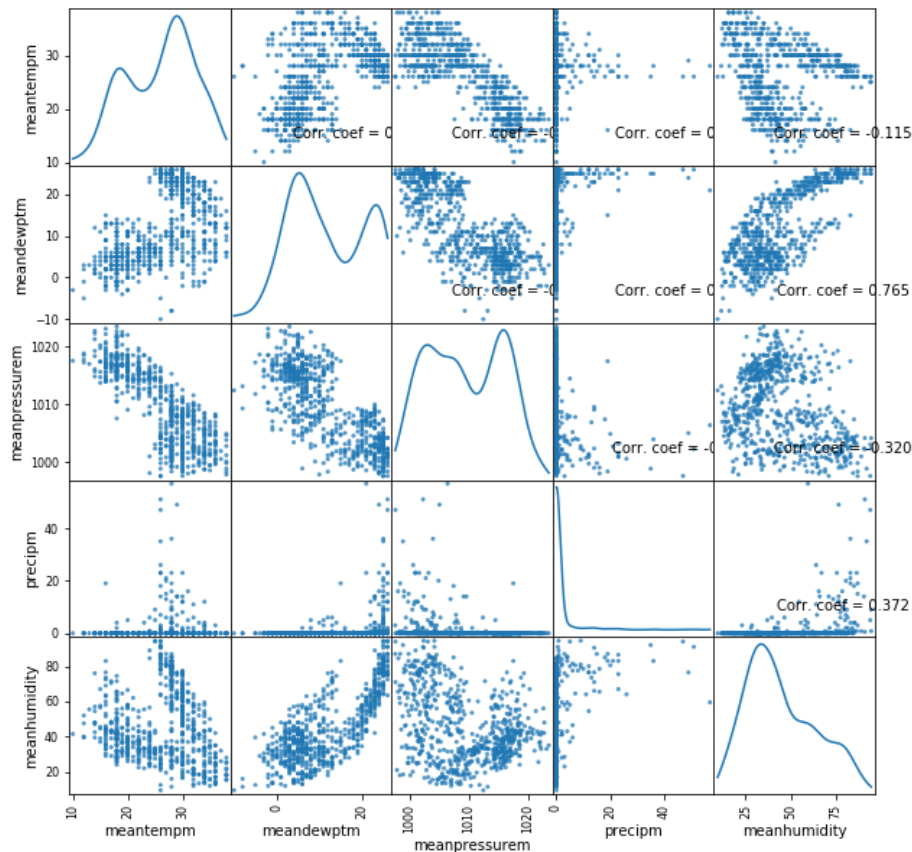
- ✓ Features with Maximum Correlations had no outlier and their distributions were:

2. Mean Dew Point



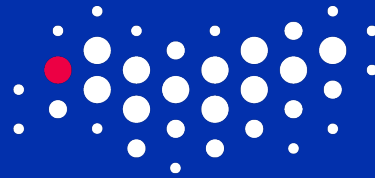
EDA – Feature Scatter Plots

Scatter and Density Plot



Improvement to Dataset

- ✓ Padding Past 2 day's information
- ✓ Total Features changed from Initial 10 to merged 5 to padded 14.



ITMO UNIVERSITY

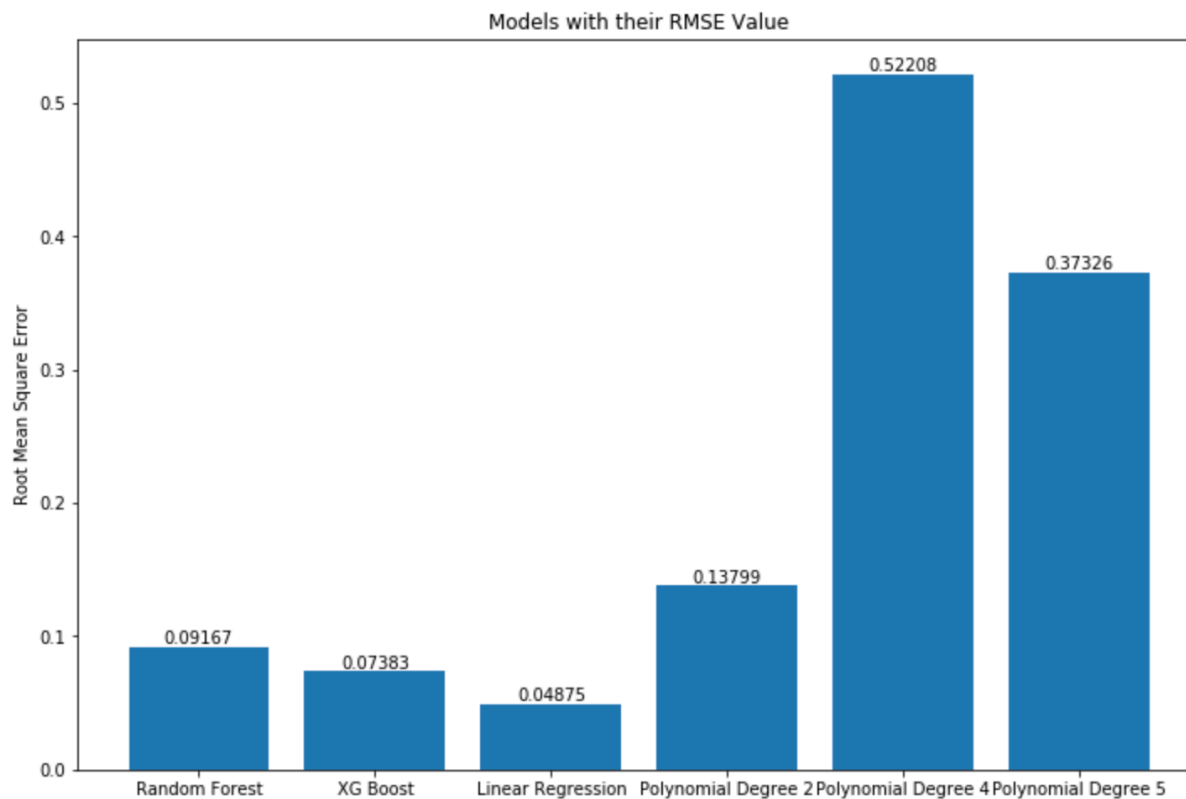
Applying Machine Learning Models

Accuracy Measure : RMSE (Root Mean Square Error)

Applying ML Models (1)

- ✓ Applied Random Forest Regressor : [0.09167]
- ✓ XGBoost Regressor: [0.07383]
- ✓ Linear Regression : [0.04875]
- ✓ Polynomial Regression of Degree – [2,4,5] :
[0.13799, 0.52208, 0.37326]

Applying ML Models (2)

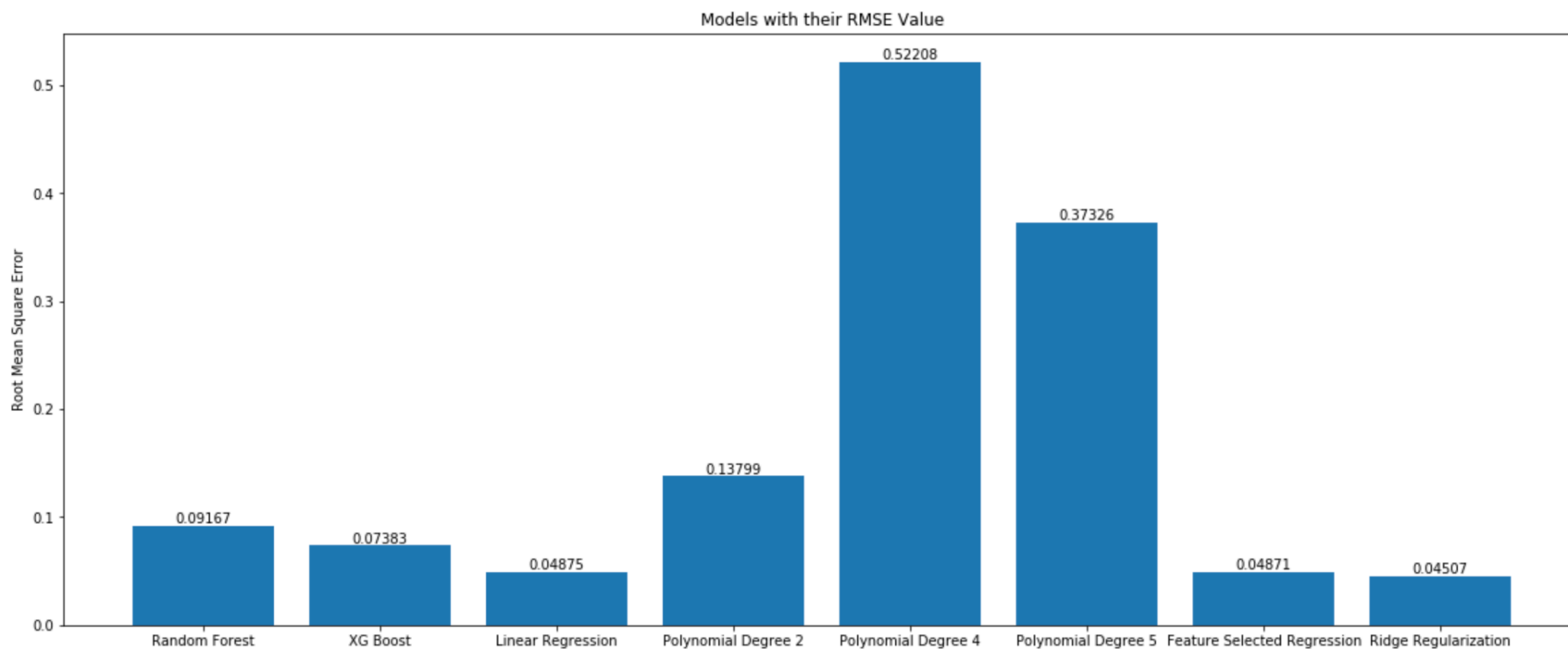


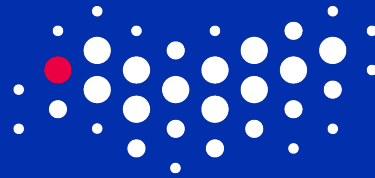
In-between Analysis :
Linear Regression Fit the
Data with the lowest RMSE

Applying ML Models (3)

- ✓ Recursive Feature Selection – $N=10$:
[0.04871]
- ✓ Ridge Regression with Feature Selection :
[0.04507]

Applying ML Models (4)





Applying Deep Learning Models

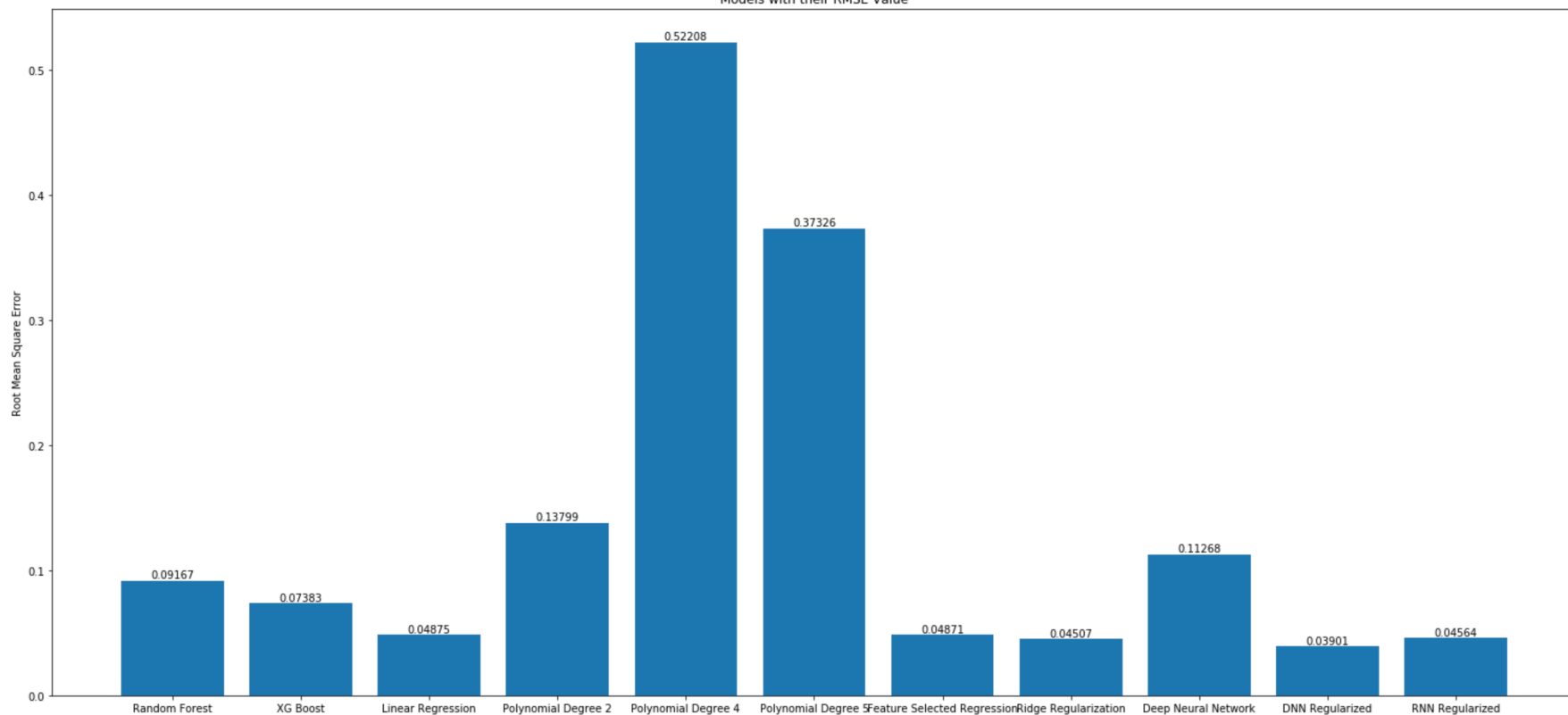
Accuracy Measure : RMSE (Root Mean Square Error)

Applying DL Models (1)

- ✓ Neural Network – { Layers : (6, Dense) , Activation : Relu } : [0.10314]
- ✓ Regularized Neural Network – { Layer : (6, Dense with Dropouts) , Activation : Relu, Regularizer : (L2 Ridge) : [0.03917]
- ✓ Recursive Neural Network – { Layer: (1, Embedding with Dropout), (2, LSTM with Dropouts), (3 Dense with Dropouts), Activation : Relu, Regularizer : (L2 Ridge) : [0.04564]

Applying DL Models (2)

Models with their RMSE Value



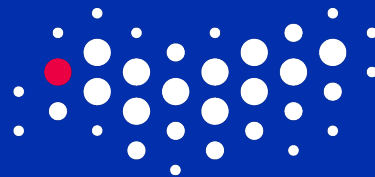
Final Result

For this dataset and This hyperparameters Models can be ranked liked :

Rank	Model	RMSE
1	DNN Regularized	0.03901
2	Ridge Regularization	0.04507
3	RNN Regularized	0.04564
4	Feature Selected Regression	0.04871
5	Linear Regression	0.04875
6	XG Boost Regressor	0.07383
7	Random Forest Regressor	0.09167
8	Deep Neural Network	0.11268
9	Polynomial Degree 2, 5, 4	0.13799, 0.37326, 0.52208

Conclusion

- ✓ For the Given Hyperparameters:
 1. Deep Neural Network With Regularization predicted the best results
 2. Some Other usable methods were Linear Regression with Ridge Regularization and Feature Selection and Recursive Neural Network, XGBoost Regressor.
 3. Random Forest, Neural Network and Polynomial Regression ranked low with this Dataset and Hyperparameters.



ITMO UNIVERSITY

Thank you!