Model development

- Models applied initially : Linear regression, polynomial regression, KNN, SVR, DT, For., Lasso regression, Ensemble models (voting and bagging regressor) and Neural network (keras)

- Scaled on RobustScaler for optimal coefficients and have better performance

- checked: performance metrics of R2, RMSE, mse, mae

Here are initial results of ML models:

Out [2]:

	Model	Mae	Mse	RMSE	R2
0	Linear Reg	0.576316	0.572359	0.756544	0.294815
1	SVR	0.696816	1.156140	1.075240	-0.424444
2	Descion Tree	0.622455	0.643793	0.802367	0.206802
3	DT bag	0.589913	0.610591	0.781403	0.247710
4	DT vot	0.622455	0.643793	0.781403	0.206802
5	RF	0.574472	0.570548	0.755347	0.297045
6	KNN	0.717206	0.869494	0.932467	-0.071276
7	Lasso	0.637596	0.686831	0.828753	0.153777
8	Polynomial	0.580622	0.590209	0.768250	0.295846

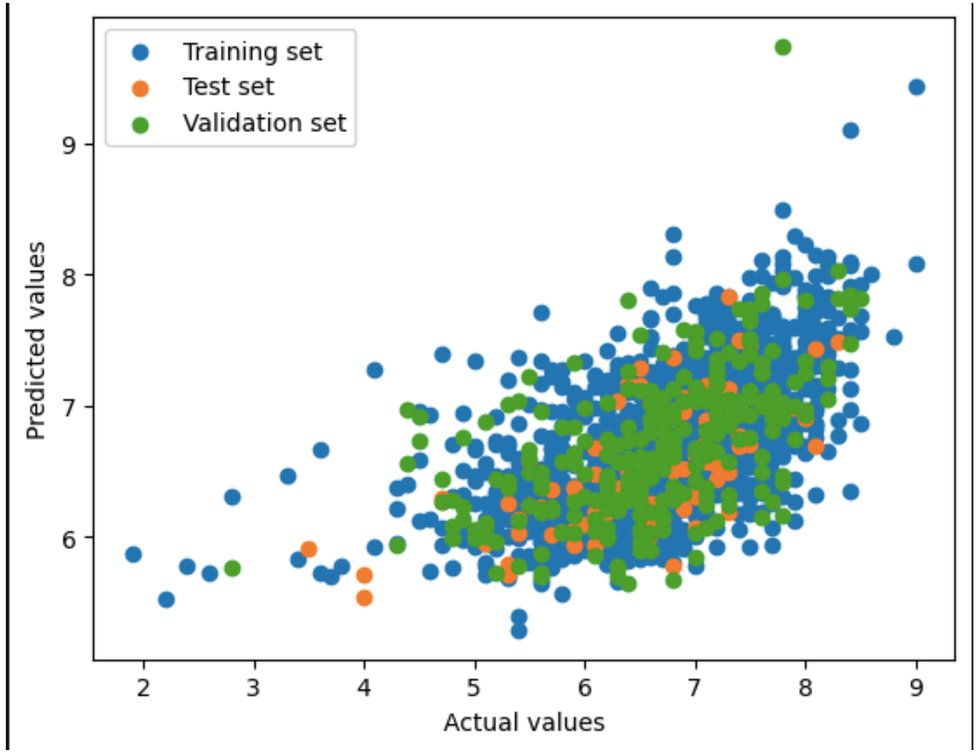
Model evaluation:

Models hypertuned after Grid search & performance metrics

	Model	Val_Mae	Val_Mse	Val_Rmse	Val_R2	Test_Mae	Test_Mse	Test_Rmse	Test_R2	Best_Params
0	RF	0.284007	0.374281	0.140087	0.775429	0.561224	0.734286	0.539176	0.335698	{'max_depth': 10, 'min_samples_leaf': 2, 'min_...
1	Lin	0.490564	0.420064	0.648123	0.326602	0.573687	0.567870	0.753572	0.300345	{'fit_intercept': True, 'normalize': False}
2	Pol	0.546667	0.499145	0.706502	0.199828	0.682752	1.010921	1.005445	-0.245524	{'linear__fit_intercept': False, 'linear__norm...
3	DT	0.521600	0.464177	0.681305	0.255885	0.607513	0.630077	0.793774	0.223702	{'criterion': 'mse', 'max_depth': 4, 'max_feat...
4	DT Vot	0.408421	0.288454	0.537079	0.537584	0.600668	0.623011	0.789310	0.232408	{'n_jobs': -1, 'verbose': 1, 'weights': [1, 3,...
5	DT Bag	0.503116	0.450275	0.671026	0.278170	0.627766	0.707227	0.840968	0.128648	{'max_depth': 5, 'min_samples_split': 2}

plotting train, test val splits

Polynomial regressor:



Polynomial regressor visuals: