Day 53

DIY

Q1. Problem Statement: Model Evaluation Metrics for Regression

Write a Python program that reads the *winequality-white.csv* (provided on LMS) file into a DataFrame, the following are the tasks that are to be taken into consideration while constructing diffract model and in the end evaluate them based on RMSE, MAPE, RMSLE.

1. Load the given dataset into a data frame
2. Find missing values and drop them if you find any
3. Check data types for all features
4. Extract dependent and independent variables into the y & x data frame ("alcohol" is our dependent feature)
5. Split your data into train and test, by 20% as test size
6. Create a new data frame for comparison of all models containing column as model name, RMSE, MAPE, RMSLE
7. Build linear regression, SVM, ridge, lasso, Decision Tree and measure their RMSE, MAPE, RMSLE and make the final data frame

Dataset:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.00100	3.00	0.45	8.8	6
1	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.99400	3.30	0.49	9.5	6
2	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.99510	3.26	0.44	10.1	6
3	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.99560	3.19	0.40	9.9	6
4	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.99560	3.19	0.40	9.9	6
...	...	...	...	...	...	...	...	...	...	...	...	...
4893	6.2	0.21	0.29	1.6	0.039	24.0	92.0	0.99114	3.27	0.50	11.2	6
4894	6.6	0.32	0.36	8.0	0.047	57.0	168.0	0.99490	3.15	0.46	9.6	5
4895	6.5	0.24	0.19	1.2	0.041	30.0	111.0	0.99254	2.99	0.46	9.4	6
4896	5.5	0.29	0.30	1.1	0.022	20.0	110.0	0.98869	3.34	0.38	12.8	7
4897	6.0	0.21	0.38	0.8	0.020	22.0	98.0	0.98941	3.26	0.32	11.8	6

4898 rows × 12 columns

Sample Output:

7. Build linear regression, SVM, ridge, lasso, Decision Tree and measure their RMSE, MAPE, RMSLE and make the final data frame

	Model Name	MAPE	RMSE	RMSLE
0	Linear regression	2.947415	0.398478	0.035545
1	SVM regression	7.310045	1.010017	0.085515
2	Ridge regression	2.952137	0.398849	0.035566
3	Lasso regression	3.381967	0.444636	0.039187
4	Decision Tree regression	4.154571	0.572211	0.049829