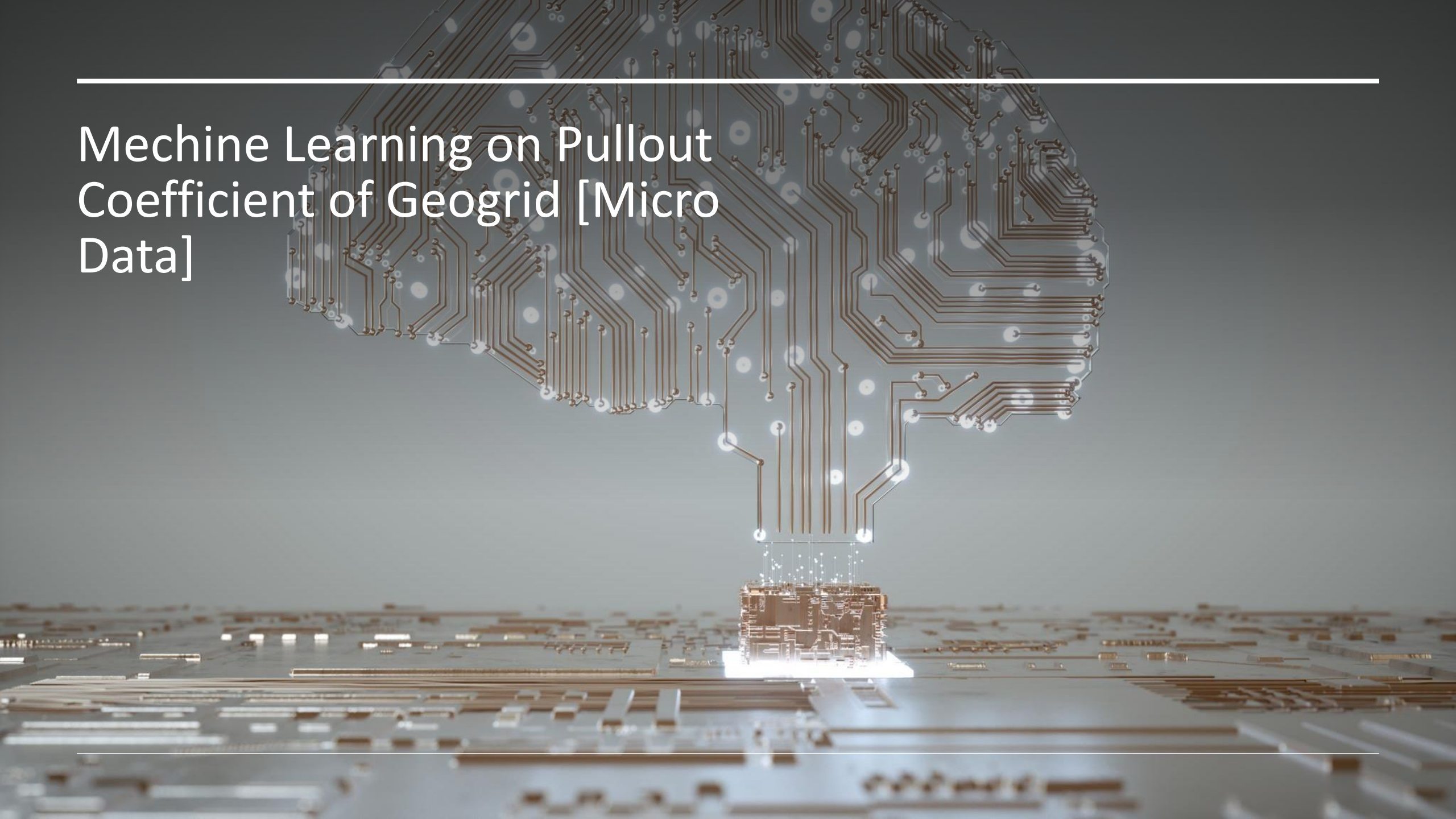

Machine Learning on Pullout Coefficient of Geogrid [Micro Data]



Dataset and Info

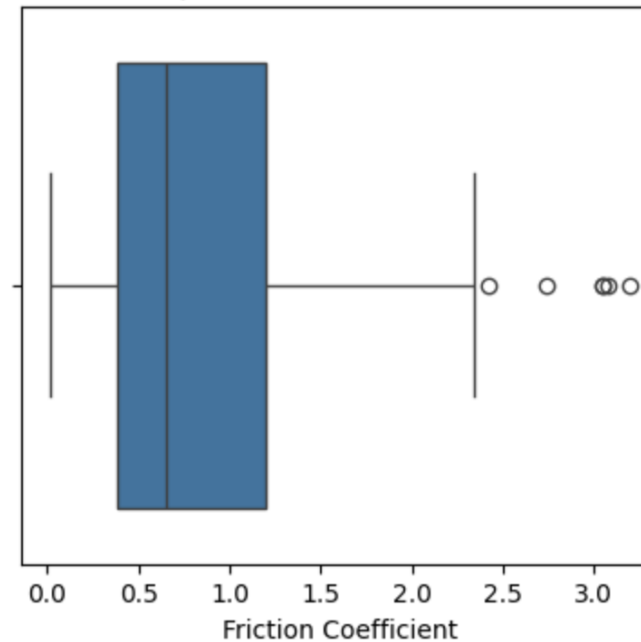
	L-filter: None	L-filter: 0.2 mm	L-filter: 0.25 mm	L-filter: 0.5 mm	L-filter: 0.8 mm	Normal Stress	Hardness	Friction Coefficient
count	296.000000	296.000000	296.000000	296.000000	288.000000	296.000000	296.000000	275.000000
mean	265.287703	2.040676	3.082432	9.727703	22.474306	18.750000	0.555419	0.847342
std	116.676692	1.243903	1.680672	5.050105	8.582137	13.427457	0.891389	0.598380
min	57.860000	0.250000	0.850000	1.750000	10.530000	5.000000	0.046000	0.014000
25%	165.800000	1.290000	2.000000	6.930000	16.825000	8.750000	0.046000	0.380750
50%	274.345000	1.770000	2.700000	8.570000	20.065000	15.000000	0.057000	0.652000
75%	351.780000	2.600000	3.870000	12.660000	27.087500	25.000000	0.057000	1.196000
max	532.410000	6.390000	9.200000	28.910000	58.470000	40.000000	2.125000	3.198000

The Dataset contains many null values which were removed.

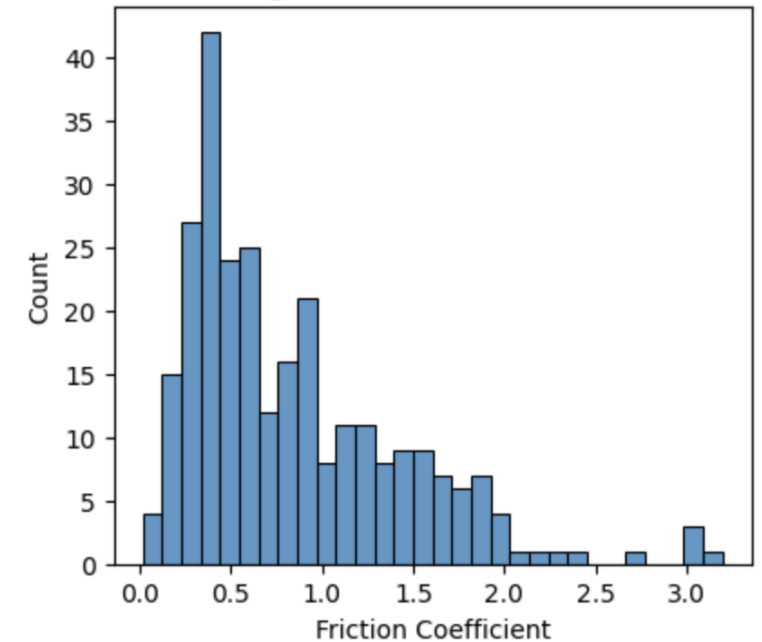
Outliers

```
Lower bound: -0.8421249999999999
Upper bound: 2.418875
Number of outliers below lower bound: 0
Number of outliers above upper bound: 6
Total number of outliers: 6
```

Box plot of Friction Coefficient



Histogram of Friction Coefficient



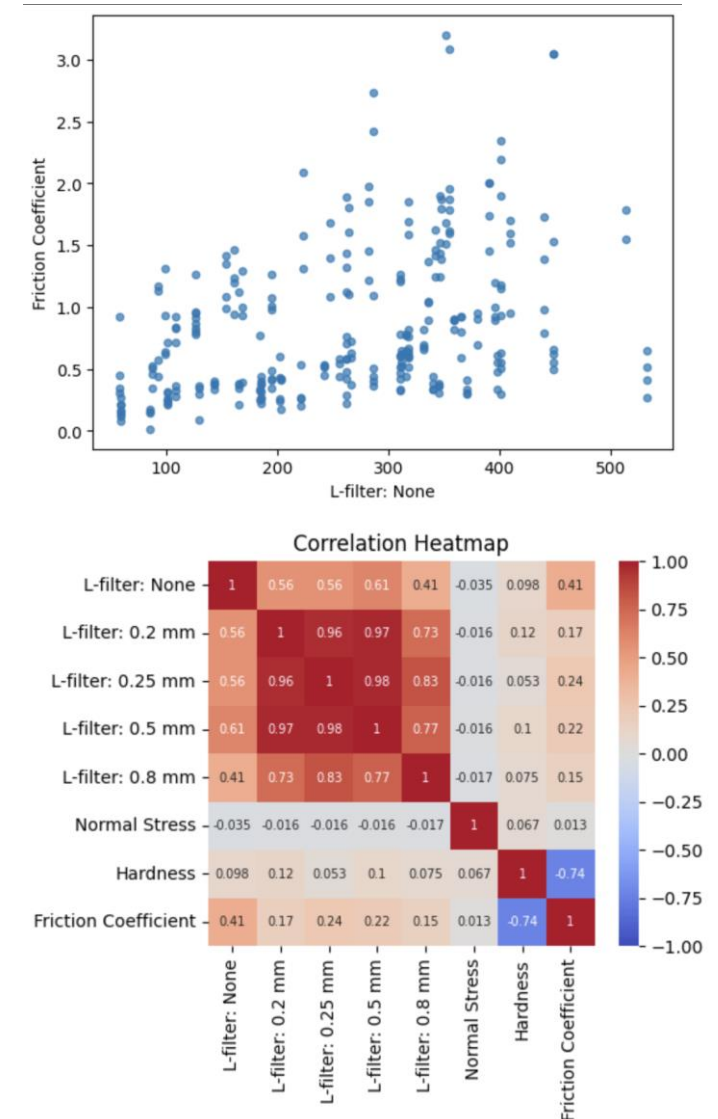
Using the IQR method, 6
outlying values were removed.

Final Cleaned Dataset

	L-filter: None	L-filter: 0.2 mm	L-filter: 0.25 mm	L-filter: 0.5 mm	L-filter: 0.8 mm	Normal Stress	Hardness	Friction Coefficient
count	269.000000	269.000000	269.000000	269.000000	261.000000	269.000000	269.000000	269.000000
mean	259.365428	2.034684	3.073420	9.694944	22.343908	17.620818	0.606387	0.801087
std	116.270793	1.267206	1.717098	5.140313	8.799966	12.990715	0.919803	0.515832
min	57.860000	0.250000	0.850000	1.750000	10.530000	5.000000	0.046000	0.014000
25%	165.800000	1.290000	2.000000	6.800000	16.600000	5.000000	0.046000	0.379000
50%	263.940000	1.770000	2.630000	8.570000	20.050000	10.000000	0.057000	0.638000
75%	346.900000	2.600000	3.870000	12.600000	27.060000	20.000000	2.125000	1.128500
max	532.410000	6.390000	9.200000	28.910000	58.470000	40.000000	2.125000	2.340000

Correlations Values

- Among the Roughness Values, no filter values had the highest correlations, so it was used as parameter in the model.
- Normal Stress showing low correlation value due to the fact that it is non monotonically related to Friction Coefficient.
- The figures depict correlation heatmap and Fric. Coeff vs Roughness



Splitting and Pipeline

The data was split into 3 sets, Train set, Dev set and Test set.

75% of data was kept in for training of models, 15% for development of models and validation, 10% on test set for performance evaluation.

Pipeline was used from scikit learn, in which standard scaler and log transformer was used to scale the data. Log transformer for Hardness, while SC for other two parameters.

Log Transformer was later removed due to no significant performance increase.

The data was then fit and transformed using the pipeline.

Models Used



Random Forests

Implemented using **ensemble module** of scikit learn.

Hyperparameter tuning was done using Dev set.

It is the best performing model.



Decision Trees

Implemented using **trees module** of scikit learn

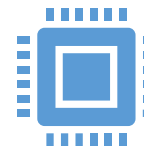
Performance was on par.



Multi-Layer Perceptron (MLP)

Implemented using **TensorFlow**.

Performance was below par, probable reason being no. of parameters only 3.



Generalized Additive Model (GAM)

Implemented using pyGAM module

Hyperparameter tuning done using Dev set.

Second Best Performing Model

Performance of models

Training Set				
Model Used	Results in training set			Rank
	RMSE	R^2	MAPD	
Random Forests	0.12	0.94	16.88	2
Decision Trees	0.005	0.99	0.11	1
MLP	0.35	0.50	48.53	4
GAM	0.19	0.85	27.43	3
Testing Set				
Model Used	Results in testing set			Rank
	RMSE	R^2	MAPD	
Random Forests	0.18	0.89	15.20	1
Decision Trees	0.21	0.83	21.26	3
MLP	0.44	0.31	41.21	4
GAM	0.22	0.83	19.74	2

Reliability Conditions

Reliability of models											
Model	Criteria						Cond. 1	Cond. 2	Cond. 3	Cond. 4	Cond.s met
	R^2	k	k'	R_0^2	$R_0'^2$	R_S					
RF	0.89	1.07	0.90	0.97	0.96	0.63	Yes	Yes	Yes	Yes	4
DTR	0.83	1.03	0.92	0.99	0.97	0.49	Yes	Yes	No	No	2
MLP	0.31	1.08	0.74	0.95	0.79	0.06	Yes	No	No	No	1
GAM	0.83	1.06	0.89	0.98	0.96	0.50	Yes	Yes	No	Yes	3

Random Forests is easily the best performing model, and is fully acceptable, since it satisfies all 4 conditions.

Resultant Graphs plots on testing set

