Analysis of Different Datasets on Model Performance: Without Feature Selection vs. With Feature Selection

Tarun Chintada

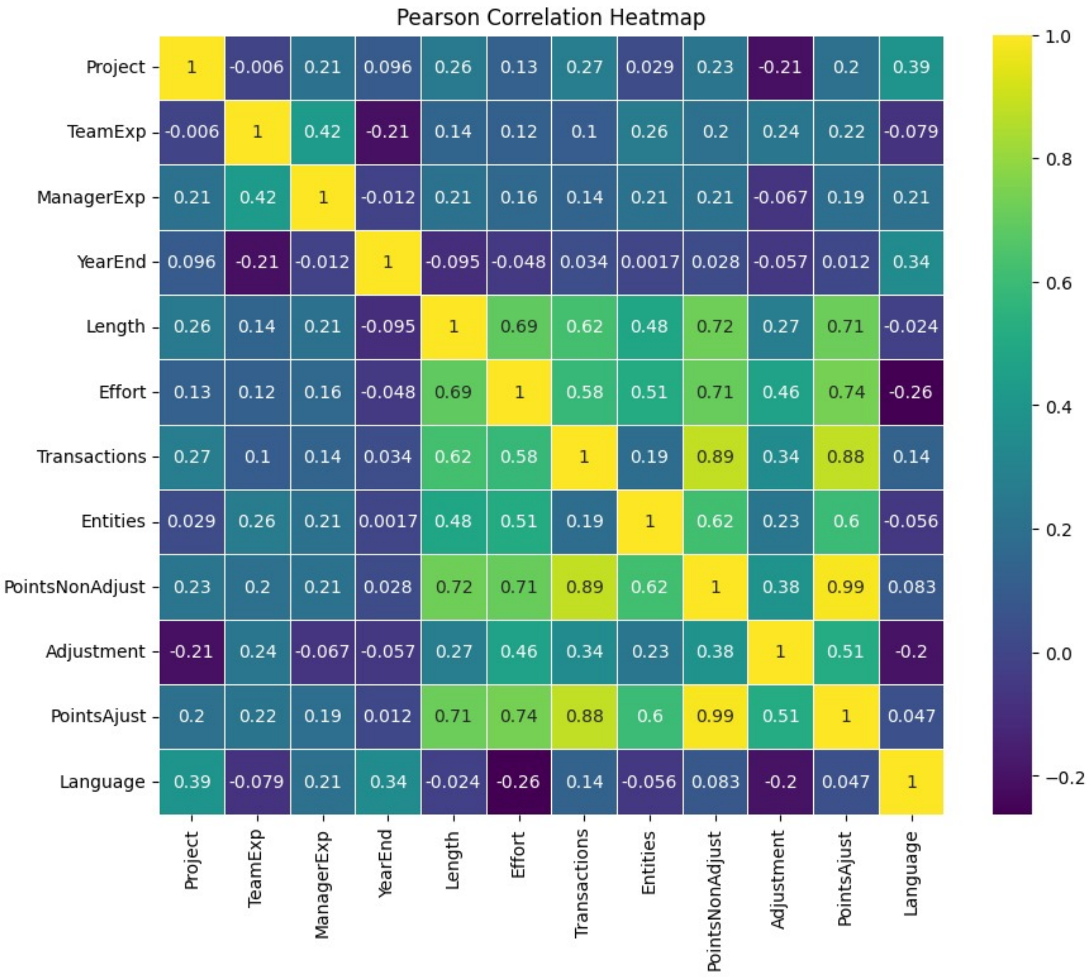
June 29, 2024

Neural Networks(30-30) Regression

1. Desharnais Dataset

The Desharnais dataset analyzed consists of information from 81 software projects from a Canadian company. Each project has 12 attributes

feature selection



The selected features are Length ,Transactions,Entities,PointsNonAdjust ,PointsAjust which are highly correlated features whose correlation coefficient is greater than 0.5

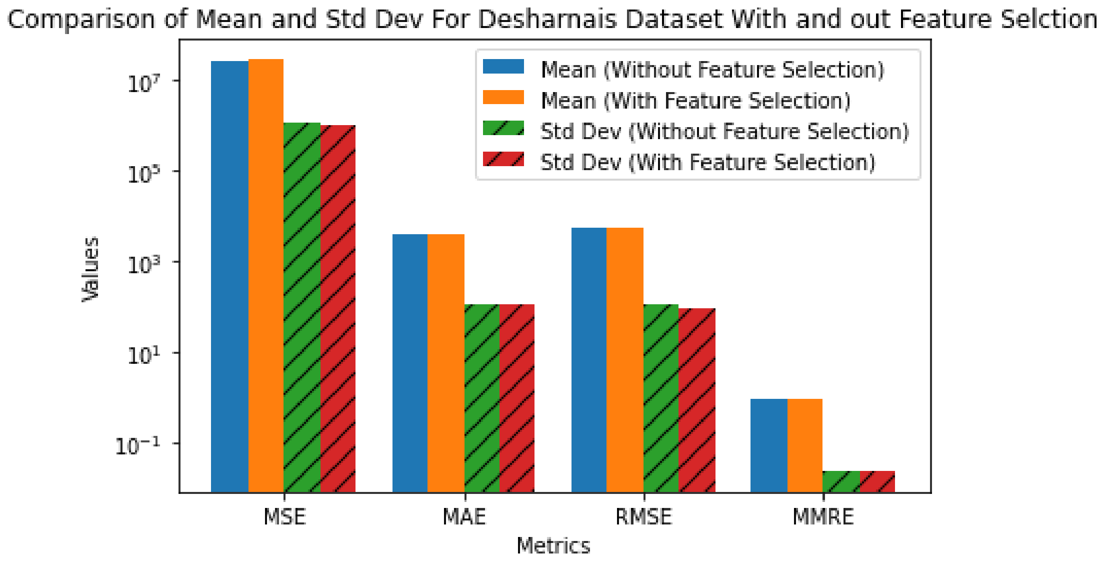


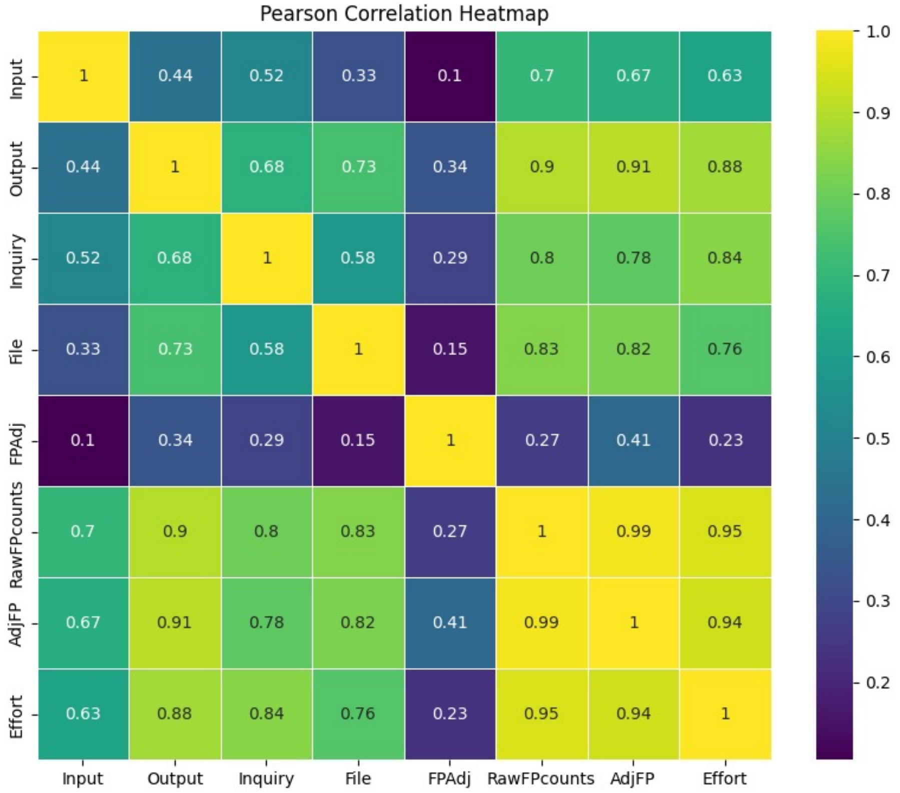
TABLE 1 .Comparison of the results Means and Standard Deviation (SD) of Deshrnais dataset

Measures	**Without Feature Selection**		**With Feature Selection**	
	Mean	**SD**	**Mean**	**SD**
MSE	27843741.3613	1199167.6066	28757662.3042	1021246.4407
MAE	4031.8336	111.1507	4013.2488	109.9385
RMSE	5275.4759	114.4352	5361.7675	95.4550
MMRE	0.8889	0.0245	0.8848	0.0242

2. albrecht Dataset

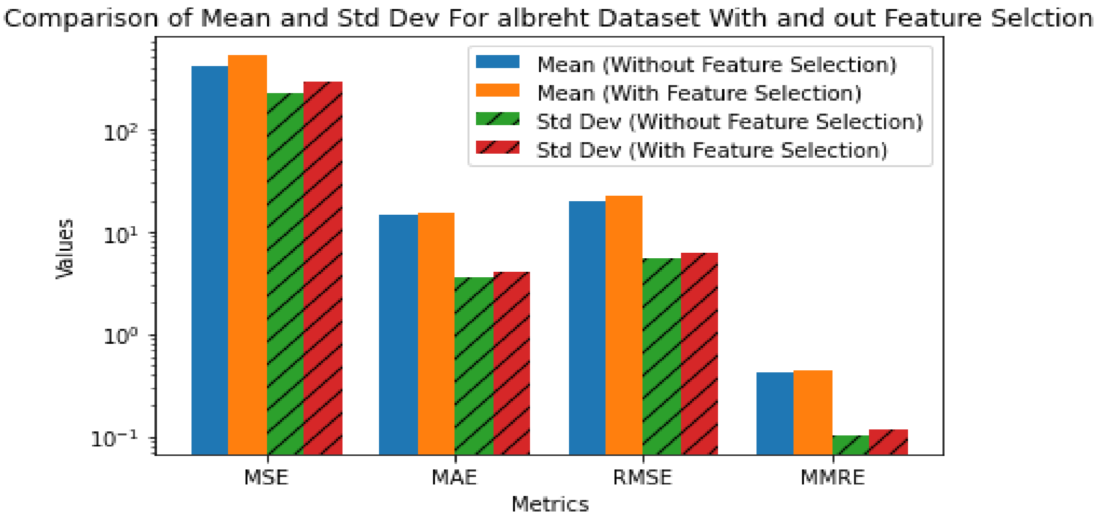
albrecht consist of 21 samples each sample has 9 attributes.

feature selection



	Input	Output	Inquiry	File	FPAdj	RawFPcounts	AdjFP	Effort
Input	1	0.44	0.52	0.33	0.1	0.7	0.67	0.63
Output	0.44	1	0.68	0.73	0.34	0.9	0.91	0.88
Inquiry	0.52	0.68	1	0.58	0.29	0.8	0.78	0.84
File	0.33	0.73	0.58	1	0.15	0.83	0.82	0.76
FPAdj	0.1	0.34	0.29	0.15	1	0.27	0.41	0.23
RawFPcounts	0.7	0.9	0.8	0.83	0.27	1	0.99	0.95
AdjFP	0.67	0.91	0.78	0.82	0.41	0.99	1	0.94
Effort	0.63	0.88	0.84	0.76	0.23	0.95	0.94	1

The selected features are Input,Output,Inquiry,File,RawFPcounts,AdjFP which are highly correlated features whose correlation coefficient is greater than 0.5

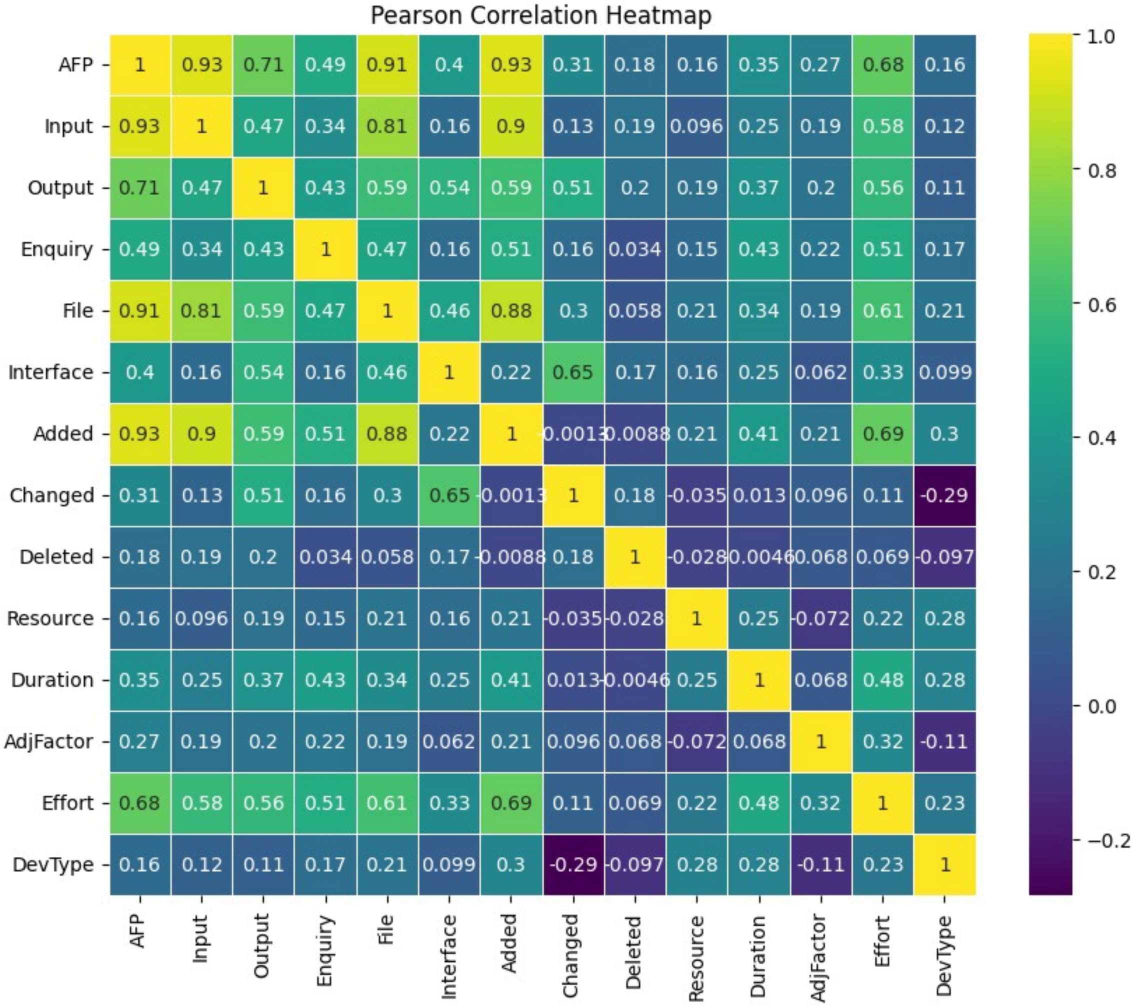


Measures	Without Feature Selection		With Feature Selection	
	Mean	SD	Mean	SD
MSE	415.8218	227.5610	520.1206	291.2830
MAE	14.5934	3.6008	15.2890	4.0386
RMSE	19.6571	5.4241	21.9569	6.1656
MMRE	0.4179	0.1031	0.4378	0.1157

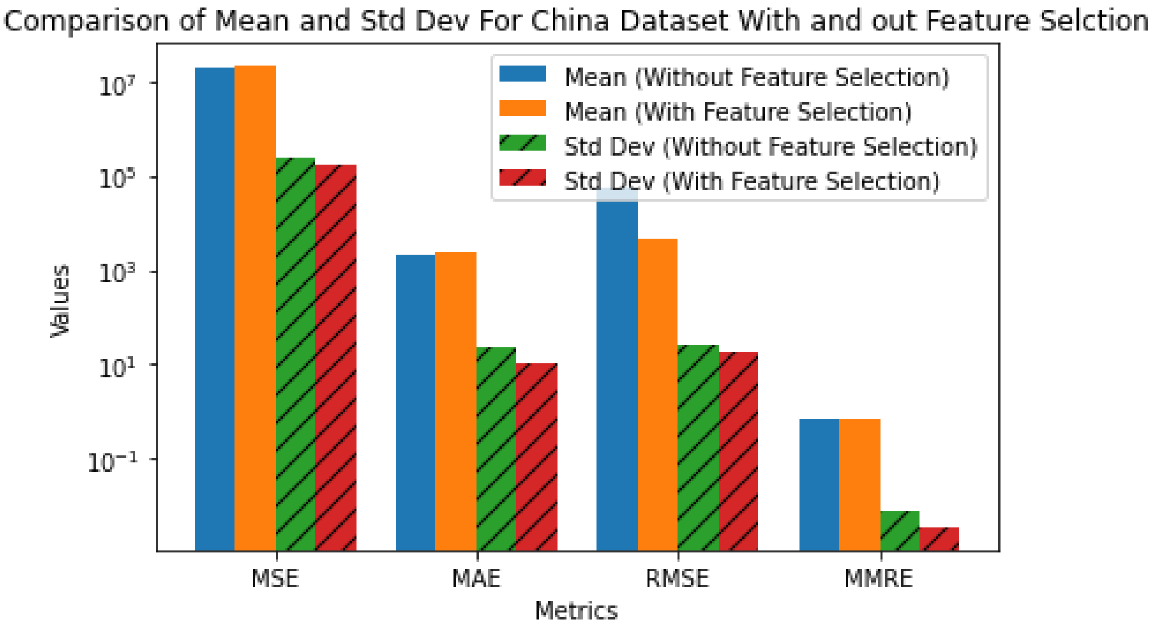
3. china Dataset

china consist of 499 samples each sample has 15 attributes.

feature selection



The selected features are AFP,Input,Output,Enquiry,File,Added which are highly correlated features whose correlation coefficient is greater than 0.5

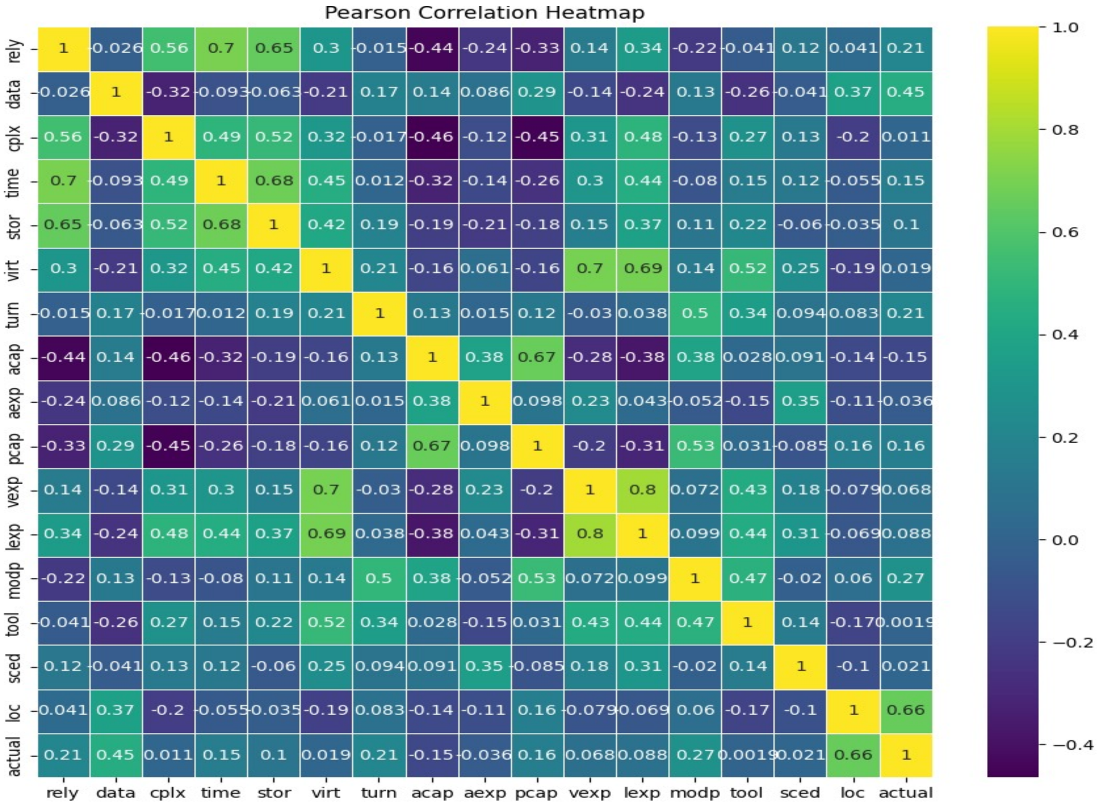


Measures	Without Feature Selection		With Feature Selection	
	Mean	**SD**	**Mean**	**SD**
MSE	18476460.7450	234433.2416	20838461.3713	171352.7200
MAE	2101.3670	24.1657	2314.0458	11.0766
RMSE	54298.3388	27.2824	4564.8778	18.7588
MMRE	0.6773	0.0078	0.7458	0.0242

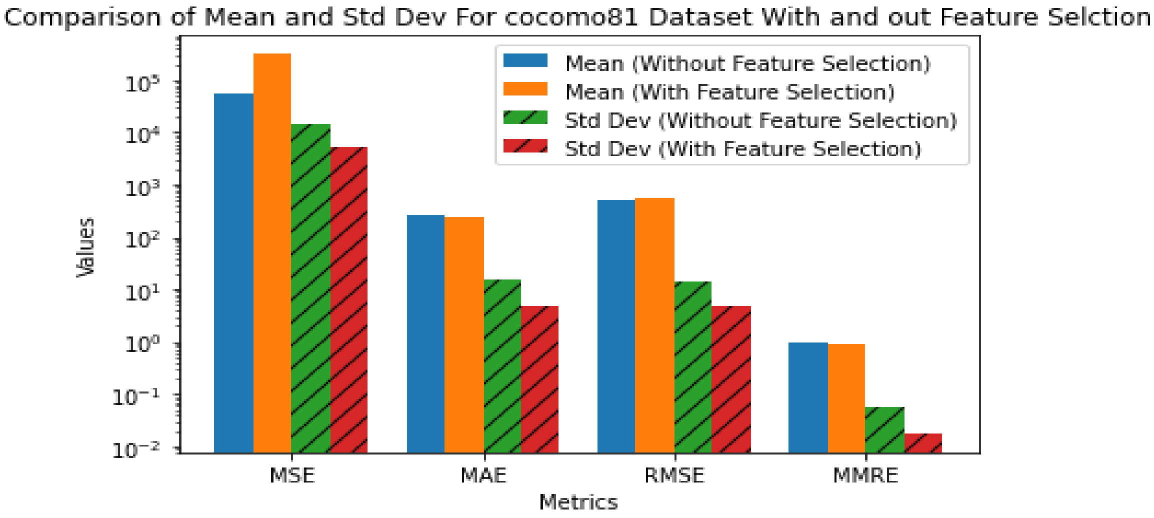
4. cocomo81 Dataset

cocomo81 consist of 63 samples each sample has 17 attributes.

feature selection



The selected features are loc which are highly correlated features whose correlation coefficient is greater than 0.5

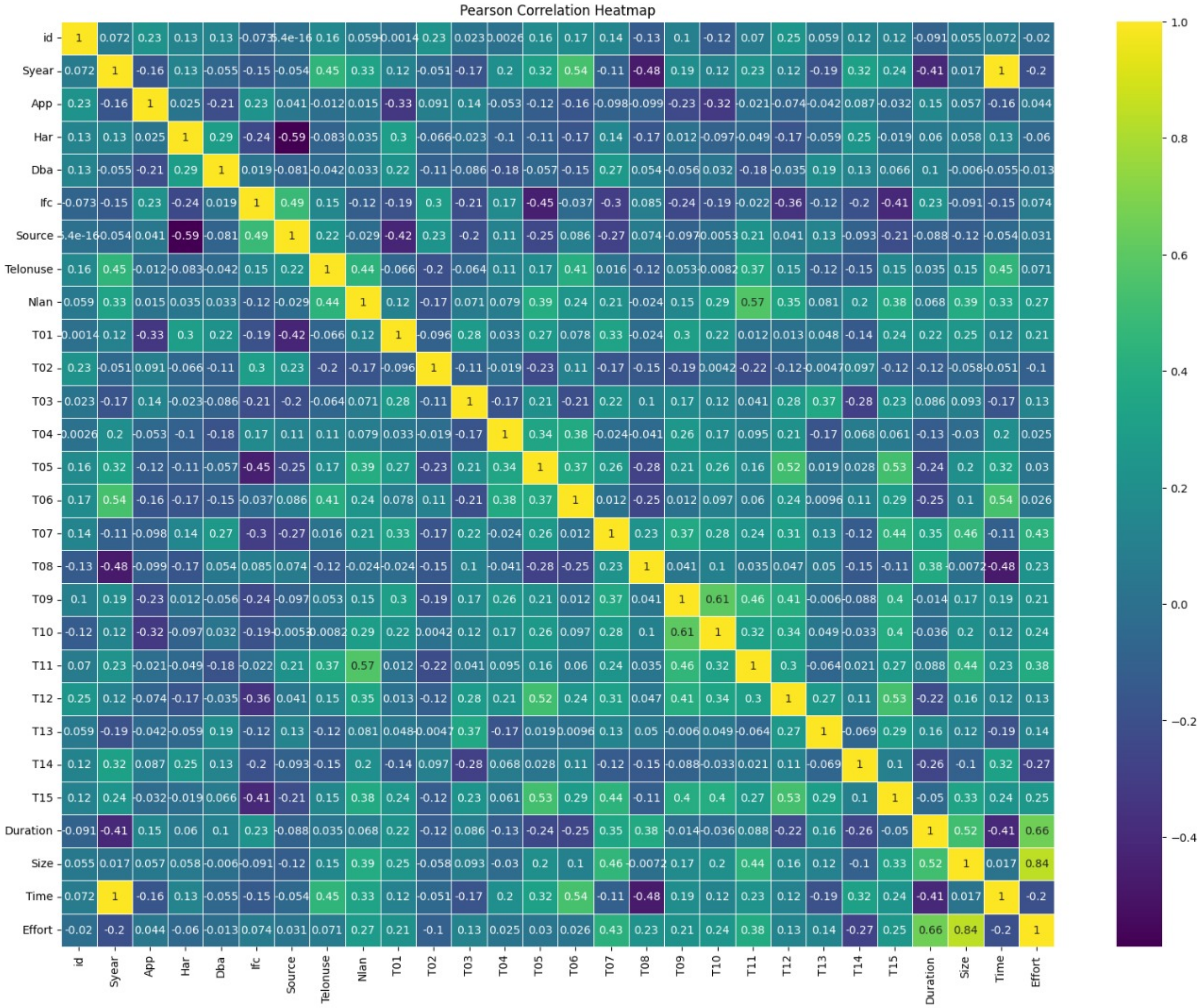


Measures	Without Feature Selection		With Feature Selection	
	Mean	SD	Mean	SD
MSE	53921.8849	14041.7981	309291.2282	5368.2188
MAE	256.7712	15.4838	236.7179	4.7079
RMSE	503.7120	14.0038	556.1186	4.8267
MMRE	0.9826	0.0593	0.9059	0.0242

5.Maxwell Dataset

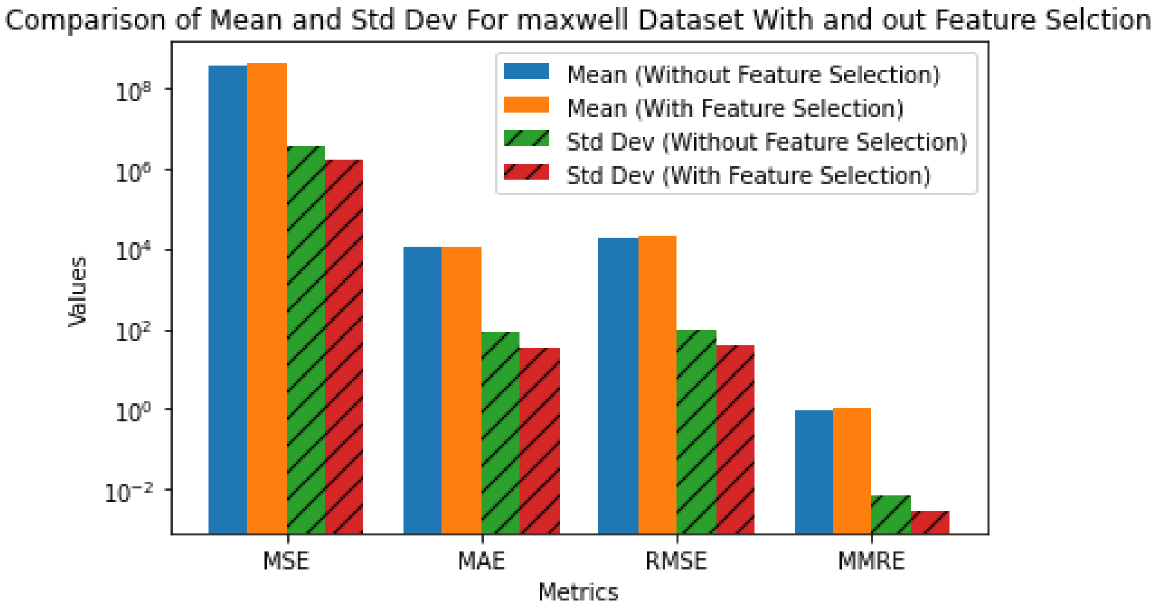
Maxwell consist of 62 samples each sample has 28 attributes.

feature selection



	id	Syear	App	Har	Dba	Ifc	Source	Telonuse	Nlan	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12	T13	T14	T15	Duration	Size	Time	Effort
id	1	0.072	0.23	0.13	0.13	-0.073	5.4e-16	0.16	0.059	-0.0014	0.23	0.023	0.0026	0.16	0.17	0.14	-0.13	0.1	-0.12	0.07	0.25	0.059	0.12	0.12	-0.091	0.055	0.072	-0.02
Syear	0.072	1	-0.16	0.13	-0.055	-0.15	-0.054	0.45	0.33	0.12	-0.051	-0.17	0.2	0.32	0.54	-0.11	-0.48	0.19	0.12	0.23	0.12	-0.19	0.32	0.24	-0.41	0.017	1	-0.2
App	0.23	-0.16	1	0.025	-0.21	0.23	0.041	-0.012	0.015	-0.33	0.091	0.14	-0.053	-0.12	-0.16	-0.098	-0.099	-0.23	-0.32	-0.021	-0.074	-0.042	0.087	-0.032	0.15	0.057	-0.16	0.044
Har	0.13	0.13	0.025	1	0.29	-0.24	-0.59	-0.083	0.035	0.3	-0.066	-0.023	-0.1	-0.11	-0.17	0.14	-0.17	0.012	-0.097	-0.049	-0.17	-0.059	0.25	-0.019	0.06	0.058	0.13	-0.06
Dba	0.13	-0.055	-0.21	0.29	1	0.019	-0.081	-0.042	0.033	0.22	-0.11	-0.086	-0.18	-0.057	-0.15	0.27	0.054	-0.056	0.032	-0.18	-0.035	0.19	0.13	0.066	0.1	-0.006	-0.055	-0.013
Ifc	-0.073	-0.15	0.23	-0.24	0.019	1	0.49	0.15	-0.12	-0.19	0.3	-0.21	0.17	-0.45	-0.037	-0.3	0.085	-0.24	-0.19	-0.022	-0.36	-0.12	-0.2	-0.41	0.23	-0.091	-0.15	0.074
Source	.4e-16	-0.054	0.041	-0.59	-0.081	0.49	1	0.22	-0.029	-0.42	0.23	-0.2	0.11	-0.25	0.086	-0.27	0.074	-0.097	-0.0053	0.21	0.041	0.13	-0.093	-0.21	-0.088	-0.12	-0.054	0.031
Telonuse	0.16	0.45	-0.012	-0.083	-0.042	0.15	0.22	1	0.44	-0.066	-0.2	-0.064	0.11	0.17	0.41	0.016	-0.12	0.053	-0.0082	0.37	0.15	-0.12	-0.15	0.15	0.035	0.15	0.45	0.071
Nlan	0.059	0.33	0.015	0.035	0.033	-0.12	-0.029	0.44	1	0.12	-0.17	0.071	0.079	0.39	0.24	0.21	-0.024	0.15	0.29	0.57	0.35	0.081	0.2	0.38	0.068	0.39	0.33	0.27
T01	0.0014	0.12	-0.33	0.3	0.22	-0.19	-0.42	-0.066	0.12	1	-0.096	0.28	0.033	0.27	0.078	0.33	-0.024	0.3	0.22	0.012	0.013	0.048	-0.14	0.24	0.22	0.25	0.12	0.21
T02	0.23	-0.051	0.091	-0.066	-0.11	0.3	0.23	-0.2	-0.17	-0.096	1	-0.11	-0.019	-0.23	0.11	-0.17	-0.15	-0.19	0.0042	-0.22	-0.12	-0.0047	0.097	-0.12	-0.12	-0.058	-0.051	-0.1
T03	0.023	-0.17	0.14	-0.023	-0.086	-0.21	-0.2	-0.064	0.071	0.28	-0.11	1	-0.17	0.21	-0.21	0.22	0.1	0.17	0.12	0.041	0.28	0.37	-0.28	0.23	0.086	0.093	-0.17	0.13
T04	0.0026	0.2	-0.053	-0.1	-0.18	0.17	0.11	0.11	0.079	0.033	-0.019	-0.17	1	0.34	0.38	-0.024	-0.041	0.26	0.17	0.095	0.21	-0.17	0.068	0.061	-0.13	-0.03	0.2	0.025
T05	0.16	0.32	-0.12	-0.11	-0.057	-0.45	-0.25	0.17	0.39	0.27	-0.23	0.21	0.34	1	0.37	0.26	-0.28	0.21	0.26	0.16	0.52	0.019	0.028	0.53	-0.24	0.2	0.32	0.03
T06	0.17	0.54	-0.16	-0.17	-0.15	-0.037	0.086	0.41	0.24	0.078	0.11	-0.21	0.38	0.37	1	0.012	-0.25	0.012	0.097	0.06	0.24	0.0096	0.11	0.29	-0.25	0.1	0.54	0.026
T07	0.14	-0.11	-0.098	0.14	0.27	-0.3	-0.27	0.016	0.21	0.33	-0.17	0.22	-0.024	0.26	0.012	1	0.23	0.37	0.28	0.24	0.31	0.13	-0.12	0.44	0.35	0.46	-0.11	0.43
T08	-0.13	-0.48	-0.099	-0.17	0.054	0.085	0.074	-0.12	-0.024	-0.024	-0.15	0.1	-0.041	-0.28	-0.25	0.23	1	0.041	0.1	0.035	0.047	0.05	-0.15	-0.11	0.38	-0.0072	-0.48	0.23
T09	0.1	0.19	-0.23	0.012	-0.056	-0.24	-0.097	0.053	0.15	0.3	-0.19	0.17	0.26	0.21	0.012	0.37	0.041	1	0.61	0.46	0.41	-0.006	-0.088	0.4	-0.014	0.17	0.19	0.21
T10	-0.12	0.12	-0.32	-0.097	0.032	-0.19	-0.0053	-0.0082	0.29	0.22	0.0042	0.12	0.17	0.26	0.097	0.28	0.1	0.61	1	0.32	0.34	0.049	-0.033	0.4	-0.036	0.2	0.12	0.24
T11	0.07	0.23	-0.021	-0.049	-0.18	-0.022	0.21	0.37	0.57	0.012	-0.22	0.041	0.095	0.16	0.06	0.24	0.035	0.46	0.32	1	0.3	-0.064	0.021	0.27	0.088	0.44	0.23	0.38
T12	0.25	0.12	-0.074	-0.17	-0.035	-0.36	0.041	0.15	0.35	0.013	-0.12	0.28	0.21	0.52	0.24	0.31	0.047	0.41	0.34	0.3	1	0.27	0.11	0.53	-0.22	0.16	0.12	0.13
T13	0.059	-0.19	-0.042	-0.059	0.19	-0.12	0.13	-0.12	0.081	0.048	-0.0047	0.37	-0.17	0.019	0.0096	0.13	0.05	-0.006	0.049	-0.064	0.27	1	-0.069	0.29	0.16	0.12	-0.19	0.14
T14	0.12	0.32	0.087	0.25	0.13	-0.2	-0.093	-0.15	0.2	-0.14	0.097	-0.28	0.068	0.028	0.11	-0.12	-0.15	-0.088	-0.033	0.021	0.11	-0.069	1	0.1	-0.26	-0.1	0.32	-0.27
T15	0.12	0.24	-0.032	-0.019	0.066	-0.41	-0.21	0.15	0.38	0.24	-0.12	0.23	0.061	0.53	0.29	0.44	-0.11	0.4	0.4	0.27	0.53	0.29	0.1	1	-0.05	0.33	0.24	0.25
Duration	-0.091	-0.41	0.15	0.06	0.1	0.23	-0.088	0.035	0.068	0.22	-0.12	0.086	-0.13	-0.24	-0.25	0.35	0.38	-0.014	-0.036	0.088	-0.22	0.16	-0.26	-0.05	1	0.52	-0.41	0.66
Size	0.055	0.017	0.057	0.058	-0.006	-0.091	-0.12	0.15	0.39	0.25	-0.058	0.093	-0.03	0.2	0.1	0.46	-0.0072	0.17	0.2	0.44	0.16	0.12	-0.1	0.33	0.52	1	0.017	0.84
Time	0.072	1	-0.16	0.13	-0.055	-0.15	-0.054	0.45	0.33	0.12	-0.051	-0.17	0.2	0.32	0.54	-0.11	-0.48	0.19	0.12	0.23	0.12	-0.19	0.32	0.24	-0.41	0.017	1	-0.2
Effort	-0.02	-0.2	0.044	-0.06	-0.013	0.074	0.031	0.071	0.27	0.21	-0.1	0.13	0.025	0.03	0.026	0.43	0.23	0.21	0.24	0.38	0.13	0.14	-0.27	0.25	0.66	0.84	-0.2	1

The selected features are rely,data,turn,modp,loc which are highly correlated features whose correlation coefficient is greater than 0.2

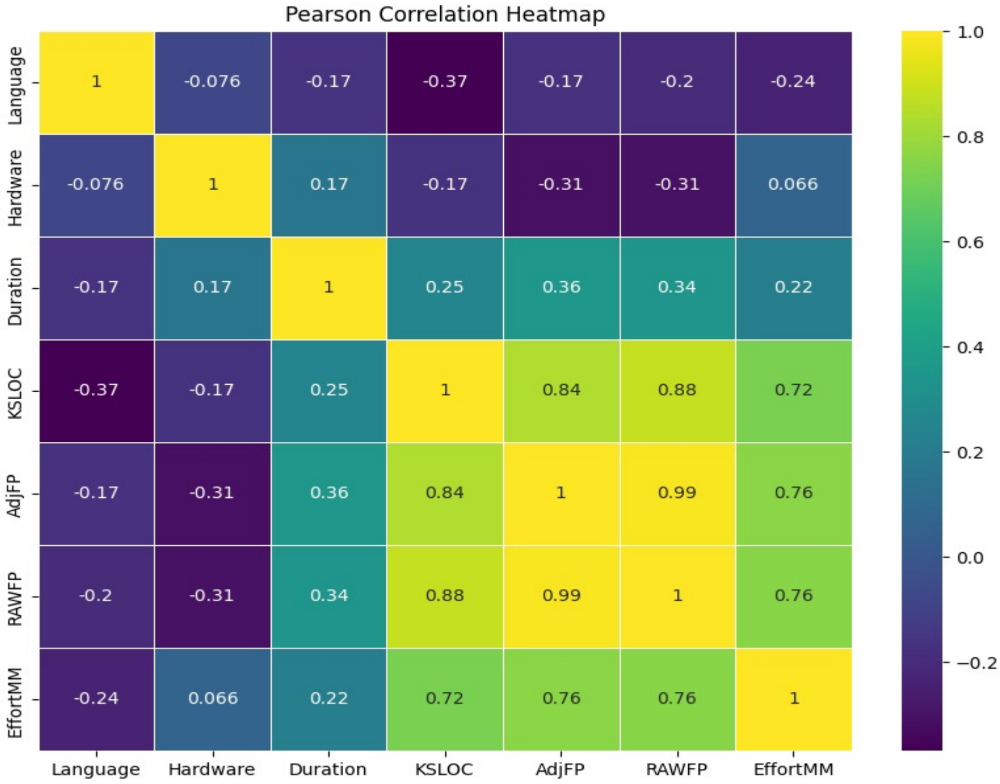


Measures	**Without Feature Selection**		**With Feature Selection**	
	Mean	**SD**	**Mean**	**SD**
MSE	382273116.6848	3597535.7906	395283100.3094	1624851.2171
MAE	11109.6831	83.1689	11412.3168	34.0108
RMSE	19551.5897	91.9514	19881.6858	40.8785
MMRE	0.9577	0.0072	0.9838	0.0242

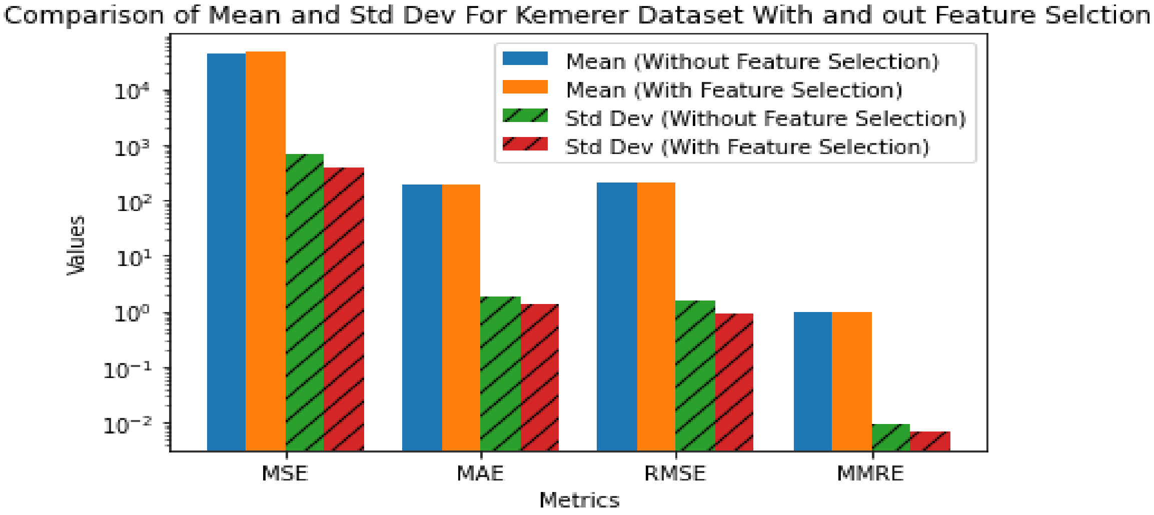
6.Kemerer Dataset

Kermer consist of 15 samples each sample has 7 attributes.

feature selection



The selected features are KSLOC,AdjFP,RAWFP which are highly correlated features whose correlation coefficient is greater than 0.2



Measures	**Without Feature Selection**		**With Feature Selection**	
	Mean	**SD**	**Mean**	**SD**
MSE	44398.5961	672.1044	45375.6067	397.3796
MAE	189.7734	1.8009	191.9485	1.3858
RMSE	210.7037	1.5969	213.0135	0.9342
MMRE	0.9654	0.0092	0.9765	0.0070