Table for Understanding Results

Legend	
FL_LR	Federated Learning using Logistic Regression
FL_SD_LR	Federated Learning using Logistic Regression and Synthetic Data
k	Number of Iterations
t	Total (Computation) time in seconds
n	Number of instances/records in a given dataset
p	Number of features in a given dataset
s	Number of clients in a given dataset
LL	Value of Logistic Loss Function, whose range is from 0 (least margin of error) to 1 (highest margin of error)
LLH	Value of Logarithmic Likelihood Function, whose range is from $-\infty$ (worst model fit) to 0 (best model fit)
AA	Approximate Accuracy of final model, given by e^{LL}
E (*)	Average of *
R (*)	Range of *
C (*)	Confidence interval of * at a 95% confidence level

Dataset Descriptions

Datasets	Symbol	Description	n	p	s	Response Variable	Source
Adult Income	D1	Adult's annual salaries in US Dollars	48,842	14	5	If annual income is >$50K or ≤ $50K	UCI
Automobile	D2	Car Specifications	205	21	22	If car price (in US Dollars) is within one of the 11 equally divided intervals from [5117, 45400]	UCI
Heart Disease	D3	Main heart disease precursors	303	75	4	0,1,2 indicate no heart disease and 3,4 indicate heart disease	UCI
Diabetes	D4	Main diabetes symptoms	101,767	49	439	≥ 30% of readmission, < 30% of readmission, or no readmission	UCI
Student Performances	D5	Students' academic performances in Portuguese and Mathematics	1,044	30	2	If a student passed (50%) or failed (≤ 50%)	UCI

University's Quality of Life	D6	University metrics that are used to determine average quality of its student body	285	17	38	Bad quality of life 1,2 or good quality of life 3,4,5	UCI

Tabular Results

Total Number of Iterations for Model Convergence

Datasets	FL_SD_LR			FL_LR
	E(k)	R(k)	C(k)	**k**
Adult Income	6	4 - 9	(5, 7)	108
Automobile	8	5 - 12	(6, 10)	168
Heart Disease	5	3 - 8	(4, 6)	125
Diabetes	55	32 - 79	(52, 58)	753
Student Performances	7	5 - 9	(6, 8)	252
University's Quality of Life	9	6 - 12	(7, 11)	137

The smallest factor of which the number the number of iterations changed from FL_LR to FL_SD_LR was a factor of approximately 14, which was in the Diabetes (D4) dataset.

The biggest factor of which the number the number of iterations changed from FL_LR to FL_SD_LR was a factor of 25, which was in the Student Performance (D5) dataset.

Total (Computation) Time Taken for Model Convergence

Datasets	FL_SD_LR			FL_LR
	E(t)	R(t)	C(t)	**t**
Adult Income	29.96839	17.09602 – 43.52939	(23.45782, 36.47896)	272.13762
Automobile	18.09699	10.826 – 26.8390	(15.43236, 20.76162)	610.23627
Heart Disease	14.93217	9.23702 – 21.74821	(11.67568, 18.18866)	213.14715
Diabetes	735.47428	413.7019 – 1057.59356	(710.76486, 760.1837)	3735.32932
Student Performances	27.88597	29.24998 – 34.71514	(24.84305, 30.92889)	426.14308
University's Quality of Life	33.42473	28.42473 – 38.92473	(29.65964, 37.18982)	411.62555

The smallest factor of which the total (computation) time changed from FL_LR to FL_SD_LR was a factor of approximately 5, which was in the Diabetes (D4) dataset.

The biggest factor of which the total (computation) time changed from FL_LR to FL_SD_LR was a factor of 34, which was in the Automobile (D2) dataset.

Logistic Loss Values of Final Models

Datasets	FL_SD_LR			FL_LR
	E(LL)	R(LL)	C(LL)	**LL**
Adult Income	0.25153	0.22348 – 0.26735	(0.24811, 0.25495)	0.25253
Automobile	0.29734	0.27024 – 0.29999	(0.29511, 0.29957)	0.2958
Heart Disease	0.30411	0.27386 – 0.31903	(0.30184, 0.30638)	0.30432
Diabetes	0.27508	0.23235 – 0.28324	(0.27274, 0.27742)	0.27502
Student Performances	0.25322	0.23235 – 0.28324	(0.25043, 0.25601)	0.25268
University's Quality of Life	0.27178	0.23235 – 0.28324	(0.27161, 0.27195)	0.27190

The factors of which the logistic loss value changed from FL_SD_LR to FL_LR were all factors of approximately 1, which implies that the final model's error margins were almost identical between FL_LR and FL_SD_LR.

Logarithmic Likelihood of Final Models

Datasets	FL_SD_LR			FL_LR
	E(LLH)	R(LLH)	C(LLH)	**LLH**
Adult Income	- 7418.84086	-7679.35477 – -7158.32695	(-7535.75048, -7301.93124)	-7418.04494
Automobile	-809.93967	-896.5046 – -723.37474	(-836.29098, -783.58836)	-810.19597
Heart Disease	-3455.38416	-3826.03508 – -3084.73324	(-3568.90049, -3341.86783)	-3455.03993
Diabetes	-15460.10069	-15980.44403 – -14939.75735	(-15830.19322, -15090.00816)	-15460.46884
Student Performances	-4753.44224	-4942.70078 – -4564.1837	(-4862.86706, -4644.01742)	-4754.18452
University's Quality of Life	-737.64251	-774.83249 – -700.45253	(-750.03115, -725.25387)	-737.12735

The factors of which the logarithmic likelihood value changed from FL_SD_LR to FL_LR were all factors of approximately 1, which implies that the final model's fits to the global data were almost identical between FL_LR and FL_SD_LR.

Accuracy of Final Models

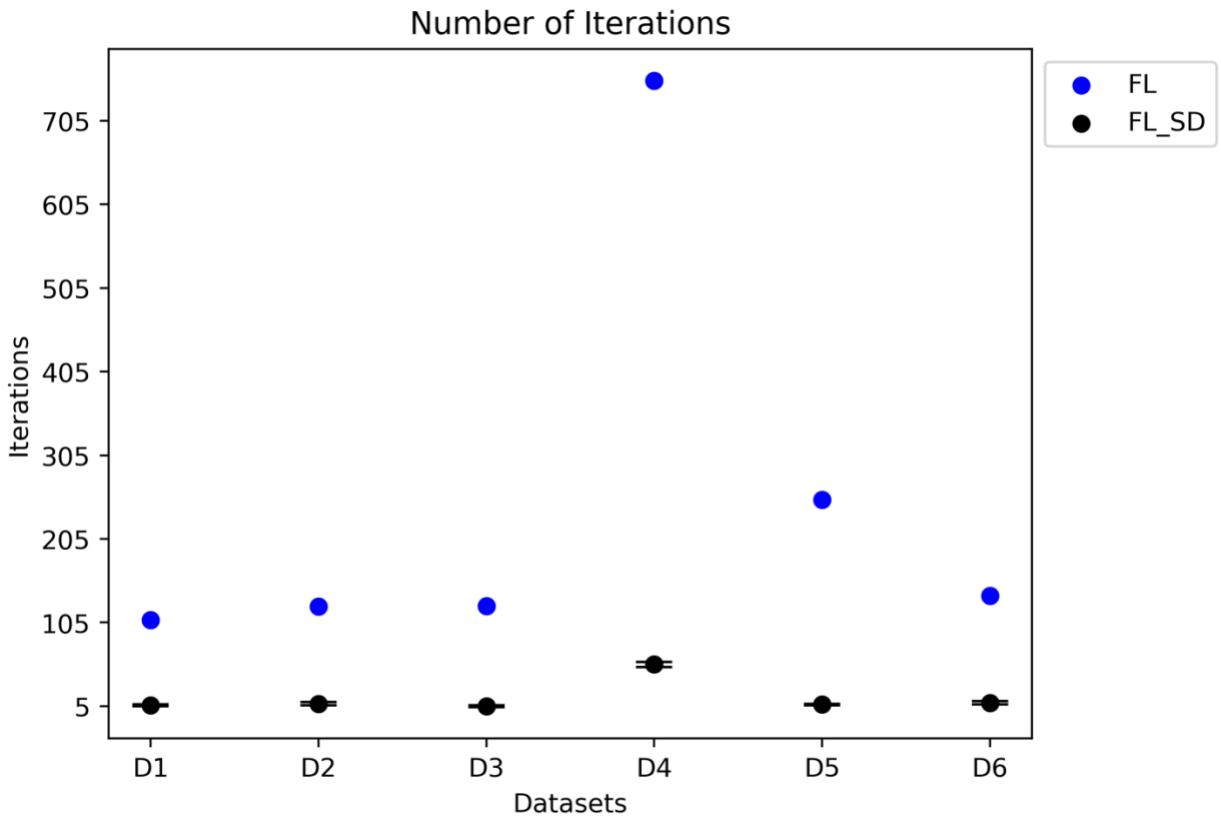
Datasets	FL_SD_LR			FL_LR
	E(AA)	R(AA)	C(AA)	**AA**
Adult Income	0.77761	0.76540 – 0.79973	(0.77495, 0.78027)	0.77683
Automobile	0.74279	0.74082 – 0.76319	(0.74113, 0.74444)	0.74393
Heart Disease	0.73777	0.72685 – 0.76043	(0.73610, 0.73945)	0.73762
Diabetes	0.75951	0.75333 – 0.76854	(0.75773, 0.76129)	0.75955
Student Performances	0.77629	0.75333 – 0.79266	(0.77413, 0.77846)	0.77671
University's Quality of Life	0.76202	0.75333 – 0.79266	(0.76189, 0.76215)	0.76193

The factor of which the accuracy of the final model changed from FL_SD_LR to FL_LR were all factors of approximately 1, which implies that the final model almost identical in accuracy between FL_LR and FL_SD_LR.

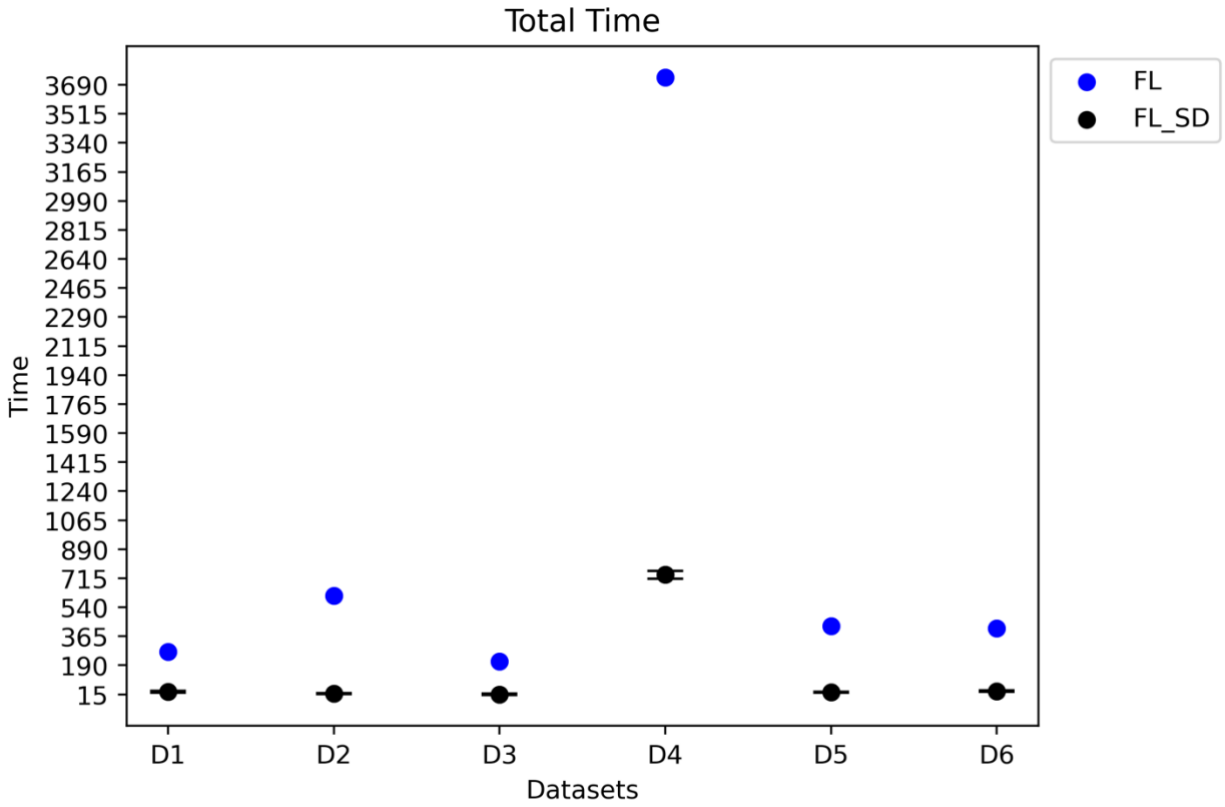
Graphical Results

Error Bar Graphs

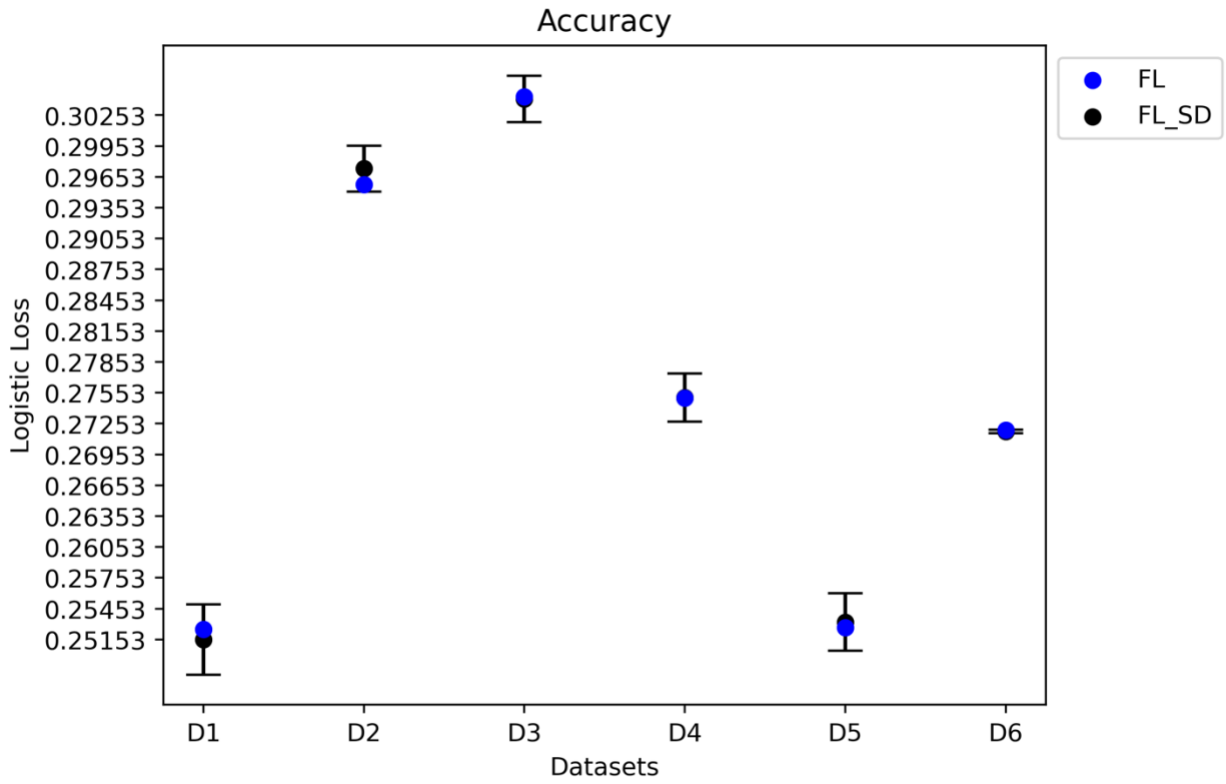
Total Number of Iterations



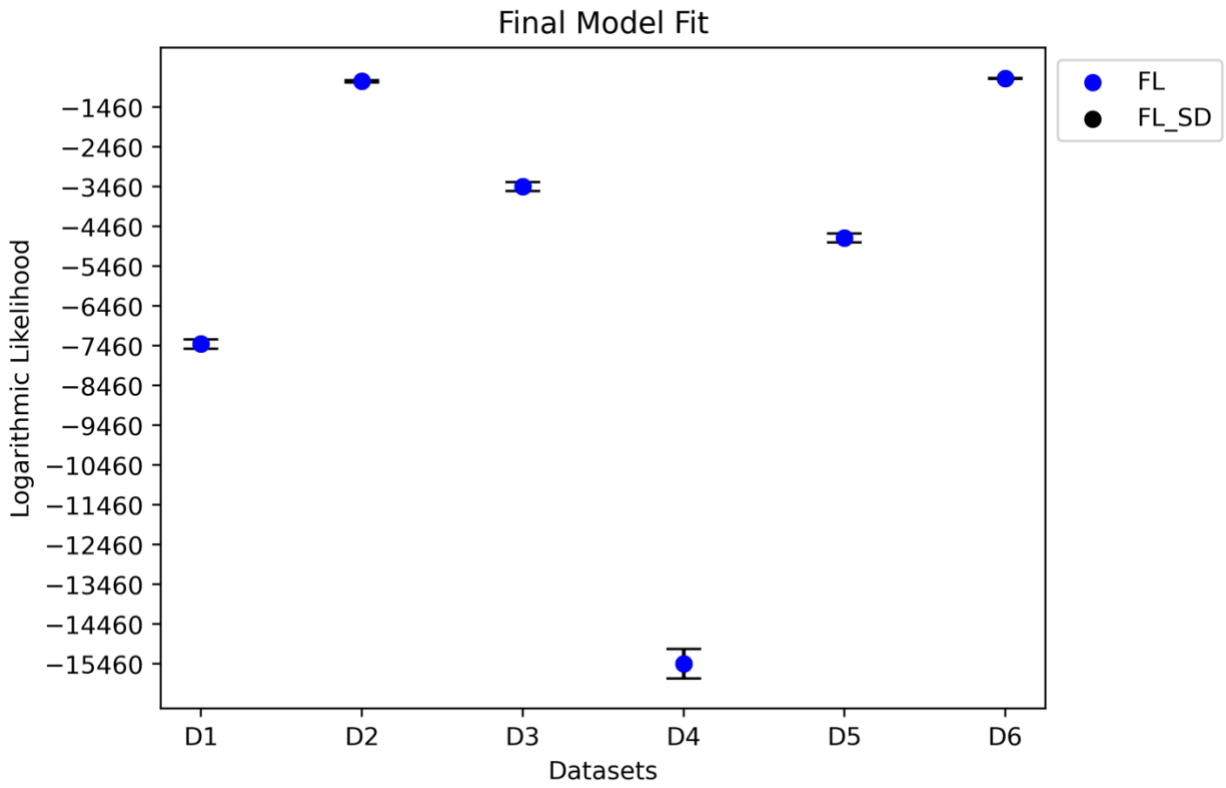
Total (Computation) Time



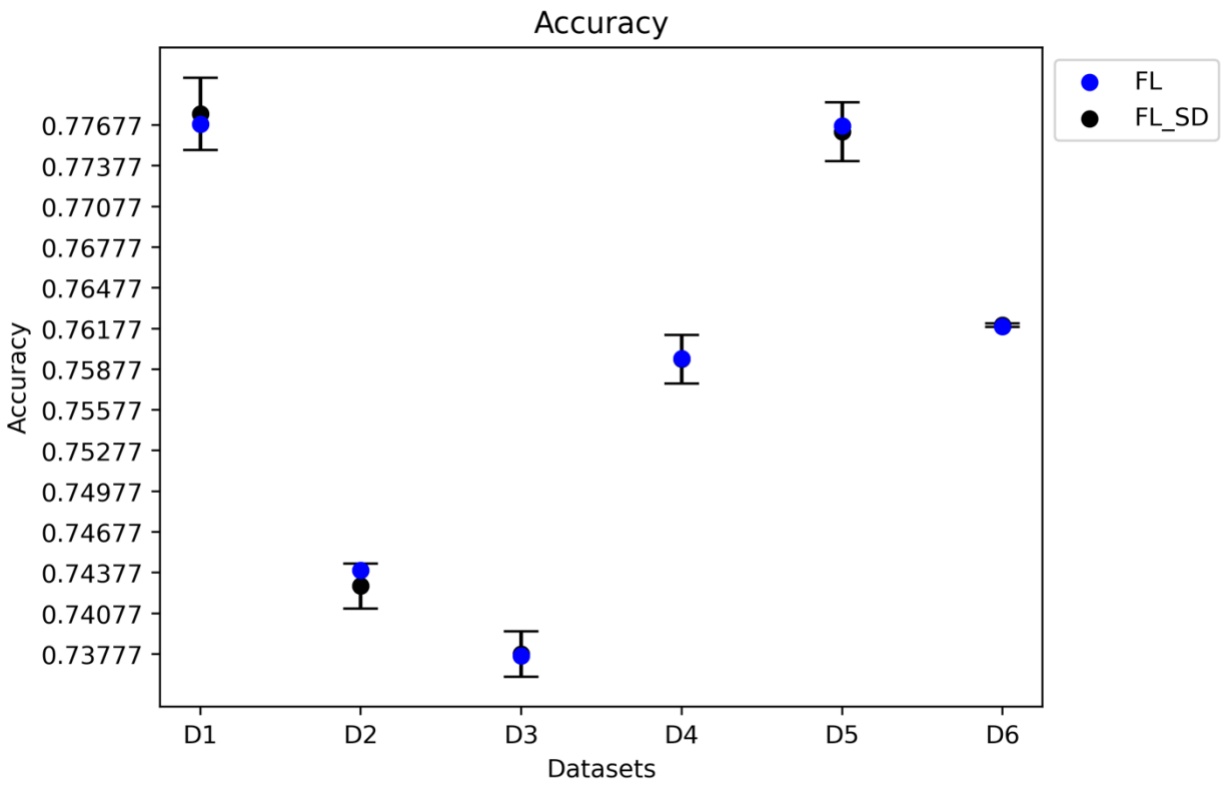
Logistic Loss Value of Final Models



Logarithmic Likelihood Of Final Models

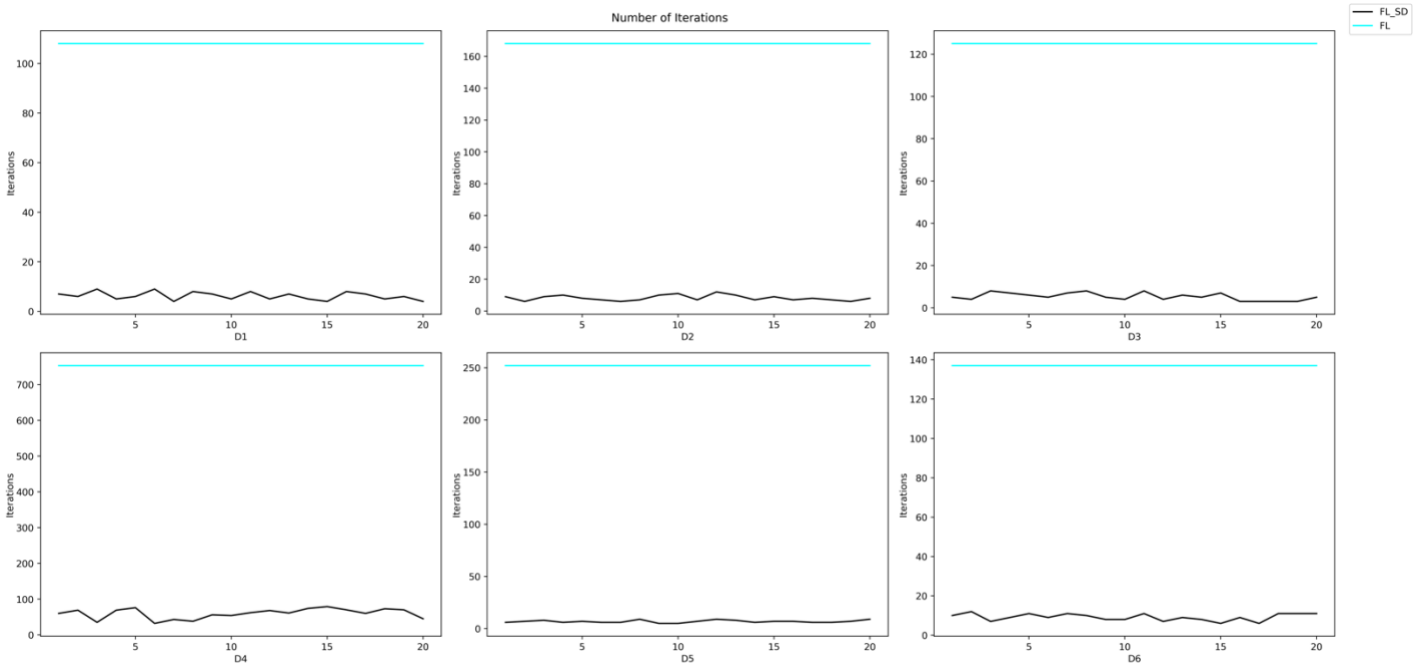


Approximate Accuracy of Final Models

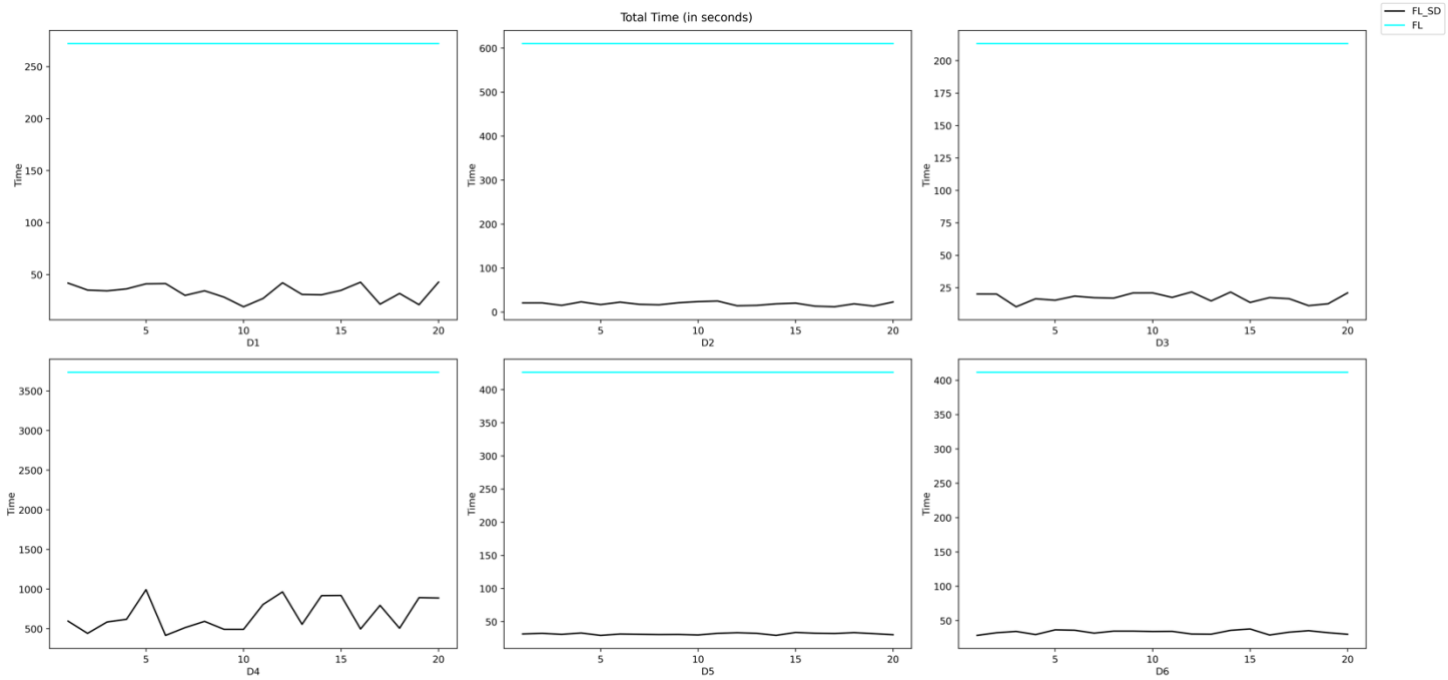


Linear Graphs

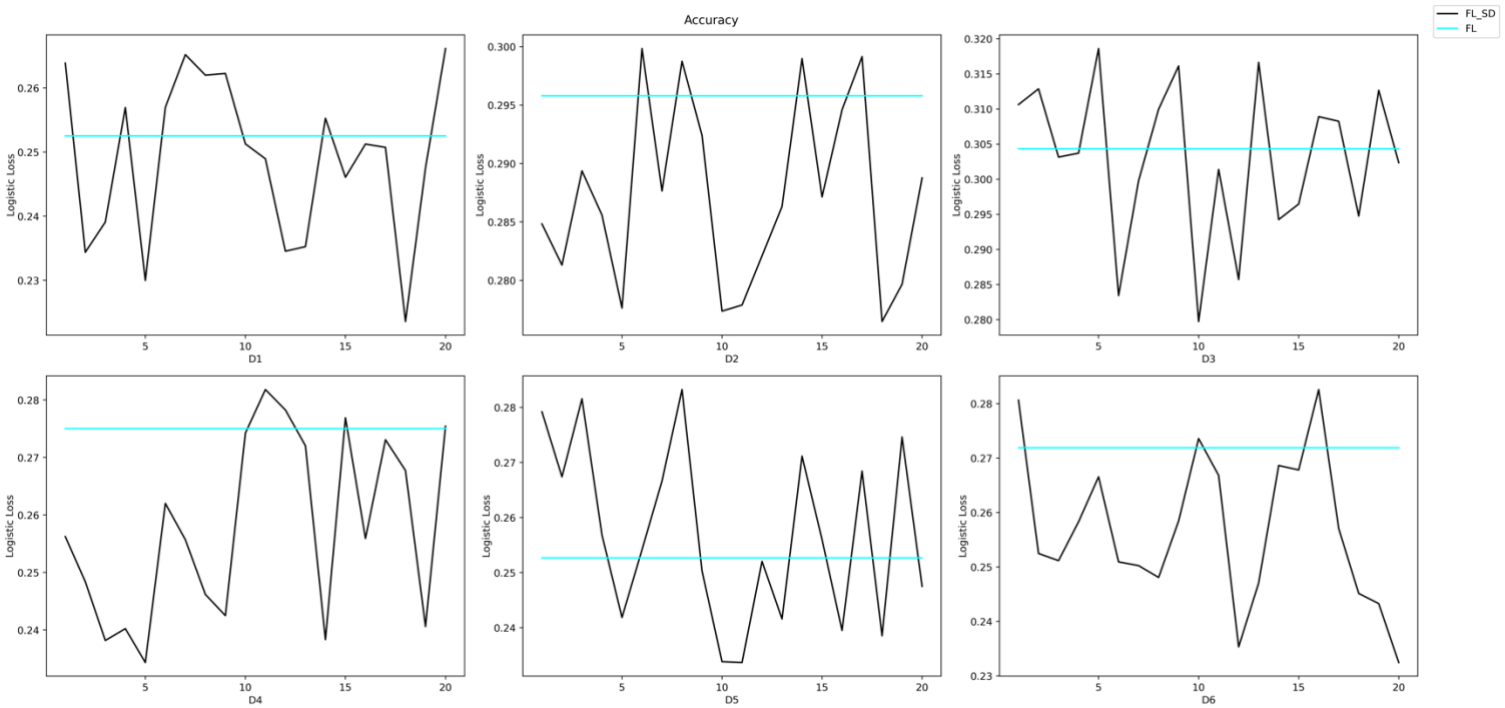
Total Number of Iterations



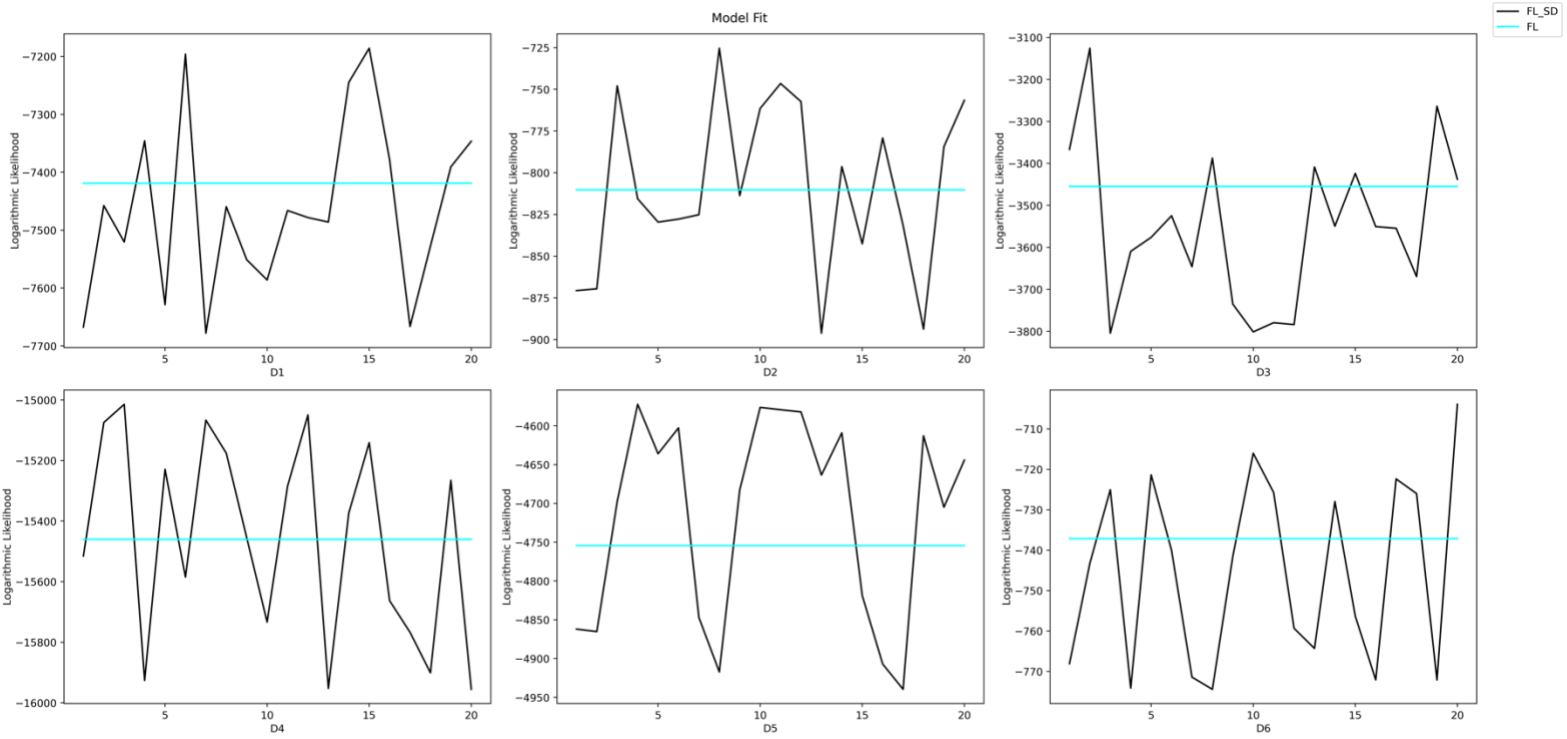
Total (Computation) Time



Logistic Loss of Final Models



Logarithmic Likelihood of Final Models



Approximate Accuracies of Final Model