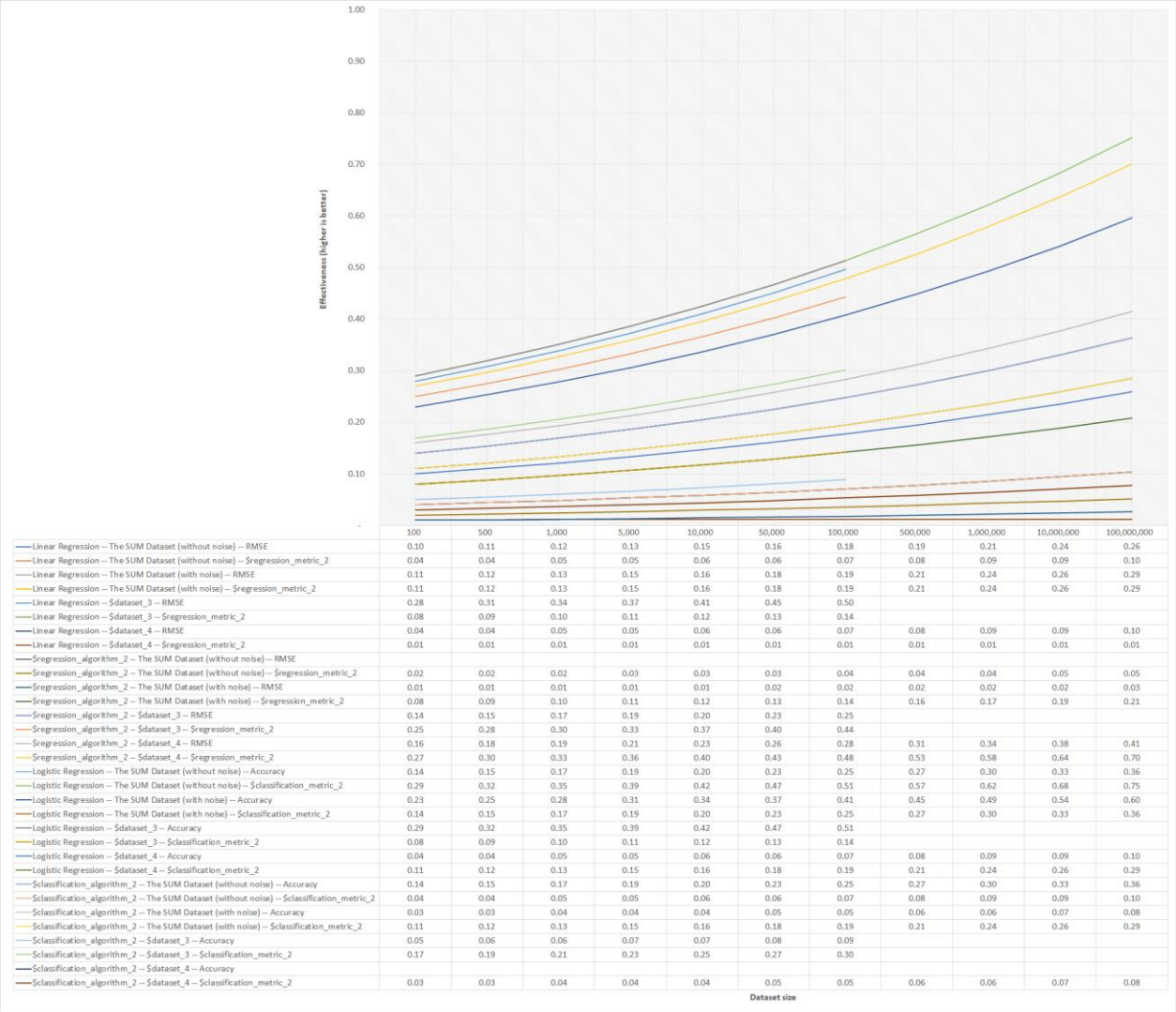
The (Un?) Reasonable Effectiveness of Data: Report

Team: $team_id
Student IDs:
Total Time Required (in hours):



	100	500	1,000	5,000	10,000	50,000	100,000	500,000	1,000,000	10,000,000	100,000,000
Linear Regression -- The SUM Dataset (without noise) -- RMSE	0.10	0.11	0.12	0.13	0.15	0.16	0.18	0.19	0.21	0.24	0.26
Linear Regression -- The SUM Dataset (without noise) -- $regression_metric_2	0.04	0.04	0.05	0.05	0.06	0.06	0.07	0.08	0.09	0.09	0.10
Linear Regression -- The SUM Dataset (with noise) -- RMSE	0.11	0.12	0.13	0.15	0.16	0.18	0.19	0.21	0.24	0.26	0.29
Linear Regression -- The SUM Dataset (with noise) -- $regression_metric_2	0.11	0.12	0.13	0.15	0.16	0.18	0.19	0.21	0.24	0.26	0.29
Linear Regression -- $dataset_3 -- RMSE	0.28	0.31	0.34	0.37	0.41	0.45	0.50				
Linear Regression -- $dataset_3 -- $regression_metric_2	0.08	0.09	0.10	0.11	0.12	0.13	0.14				
Linear Regression -- $dataset_4 -- RMSE	0.04	0.04	0.05	0.05	0.06	0.06	0.07	0.08	0.09	0.09	0.10
Linear Regression -- $dataset_4 -- $regression_metric_2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
$regression_algorithm_2 -- The SUM Dataset (without noise) -- RMSE											
$regression_algorithm_2 -- The SUM Dataset (without noise) -- $regression_metric_2	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.05
$regression_algorithm_2 -- The SUM Dataset (with noise) -- RMSE	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.03
$regression_algorithm_2 -- The SUM Dataset (with noise) -- $regression_metric_2	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.16	0.17	0.19	0.21
$regression_algorithm_2 -- $dataset_3 -- RMSE	0.14	0.15	0.17	0.19	0.20	0.23	0.25				
$regression_algorithm_2 -- $dataset_3 -- $regression_metric_2	0.25	0.28	0.30	0.33	0.37	0.40	0.44				
$regression_algorithm_2 -- $dataset_4 -- RMSE	0.16	0.18	0.19	0.21	0.23	0.26	0.28	0.31	0.34	0.38	0.41
$regression_algorithm_2 -- $dataset_4 -- $regression_metric_2	0.27	0.30	0.33	0.36	0.40	0.43	0.48	0.53	0.58	0.64	0.70
Logistic Regression -- The SUM Dataset (without noise) -- Accuracy	0.14	0.15	0.17	0.19	0.20	0.23	0.25	0.27	0.30	0.33	0.36
Logistic Regression -- The SUM Dataset (without noise) -- $classification_metric_2	0.29	0.32	0.35	0.39	0.42	0.47	0.51	0.57	0.62	0.68	0.75
Logistic Regression -- The SUM Dataset (with noise) -- Accuracy	0.23	0.25	0.28	0.31	0.34	0.37	0.41	0.45	0.49	0.54	0.60
Logistic Regression -- The SUM Dataset (with noise) -- $classification_metric_2	0.14	0.15	0.17	0.19	0.20	0.23	0.25	0.27	0.30	0.33	0.36
Logistic Regression -- $dataset_3 -- Accuracy	0.29	0.32	0.35	0.39	0.42	0.47	0.51				
Logistic Regression -- $dataset_3 -- $classification_metric_2	0.08	0.09	0.10	0.11	0.12	0.13	0.14				
Logistic Regression -- $dataset_4 -- Accuracy	0.04	0.04	0.05	0.05	0.06	0.06	0.07	0.08	0.09	0.09	0.10
Logistic Regression -- $dataset_4 -- $classification_metric_2	0.11	0.12	0.13	0.15	0.16	0.18	0.19	0.21	0.24	0.26	0.29
$classification_algorithm_2 -- The SUM Dataset (without noise) -- Accuracy	0.14	0.15	0.17	0.19	0.20	0.23	0.25	0.27	0.30	0.33	0.36
$classification_algorithm_2 -- The SUM Dataset (without noise) -- $classification_metric_2	0.04	0.04	0.05	0.05	0.06	0.06	0.07	0.08	0.09	0.09	0.10
$classification_algorithm_2 -- The SUM Dataset (with noise) -- Accuracy	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.06	0.06	0.07	0.08
$classification_algorithm_2 -- The SUM Dataset (with noise) -- $classification_metric_2	0.11	0.12	0.13	0.15	0.16	0.18	0.19	0.21	0.24	0.26	0.29
$classification_algorithm_2 -- $dataset_3 -- Accuracy	0.05	0.06	0.06	0.07	0.07	0.08	0.09				
$classification_algorithm_2 -- $dataset_3 -- $classification_metric_2	0.17	0.19	0.21	0.23	0.25	0.27	0.30				
$classification_algorithm_2 -- $dataset_4 -- Accuracy											
$classification_algorithm_2 -- $dataset_4 -- $classification_metric_2	0.03	0.03	0.04	0.04	0.04	0.05	0.05	0.06	0.06	0.07	0.08

Findings/Answer

Question 1: To what extent does the effectiveness of machine-learning algorithms depend on the size and complexity of the data? [200-300 words]

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Question 2: Looking only at the performance of your best performing algorithm on "The SUM dataset (without noise)": how well was machine-learning suitable to solve the task of predicting a) the target value and b) the target class? Consider in your assessment, how well a simple rule-based algorithm could have performed. [100 words max]

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Data, Algorithms, etc.

Algorithm 1	Linear Regression
Algorithm 2	$regression_algorithm_2
Algorithm 3	Logistic Regression
Algorithm 4	$classification_algorithm_2
Dataset 1	The SUM Dataset (without noise)
Dataset 2	The SUM Dataset (with noise)
Dataset 3	$dataset_3
Dataset 4	$dataset_4
Metric 1	RMSE
Metric 2	$regression_metric_2
Metric 3	Accuracy
Metric 4	$classification_metric_2

Contributions (max. 200 words)

$student_id_1 did Lorem dolor sit amet, consetetur sadipscing elitr, sed diam. $student_id_2, and $student_id_3 did Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod. did Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod. We all did sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod.

Additional Information

If you feel that any additional information is needed to understand your work, please provide it here.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.