Comparison Table

Include zeroes, remove 5000 outliers

Property Damage Stats	Remove zeroes	Remove zeroes, remove 513 outliers	Include zeroes, keep outliers	Include zeroes, remove 513 outliers *(Used this dataset as baseline)*	Include zeroes, remove 1,000 outliers	Include zeroes, remove 2,000 outliers
Max:	$17,900,000,000	$100,000,000	$17,900,000,000	$100,000,000	$50,000,000	$15,000,000
Mean:	$1,527,410	$312,800	$340,965	$69,746	$47,392	$30,355
Median:	$8,000	$8,000	$0	$0	$0	$0
Min:	$10	$10	$0	$0	$0	$0

Algorithm	MAE		MSE		RMSE		MAE	Comments
	Training	**Test**	**Training**	**Test**	**Training**	**Test**	**DIFF**	
Random Forest								
Baseline	10,192	27,985	2,908,505,925	19,156,222,167	53,931	138,406	175%	Evidence of overfitting
Tuned	27,157	27,739	17,531,346,006	19,479,915,784	132,406	139,571	2%	Suitable
Light GBM								
Baseline	25,751	26,708	15,935,606,039	18,901,718,418	126,236	137,484	4%	Suitable
Tuned	25,862	26,861	16,013,959,656	18,950,807,358	126,546	137,662	4%	Suitable
XGBoost								
Baseline	25,001	26,917	14,684,962,256	19,102,140,608	121,182	138,211	8%	Suitable
Tuned	25,275	26,860	15,048,715,958	19,034,888,273	122,673	137,967	6%	Suitable
Tweedie Regression								
Baseline	29,731	30,585	18,761,667,052	20,554,164,254	136,973	143,367	3%	Suitable
Tuned	28,632	29,326	21,214,757,723	22,392,103,136	145,653	149,640	2%	Suitable; power = 1
MIN	10,192	26,708	2,908,505,925	18,901,718,418	53,931	137,484	0	
IMPROVEMENT FROM BASELINE								
Random Forest								
Baseline	-74%	-77%	-99%	-99%	-90%	-91%		
Tuned	-76%	-77%	-99%	-99%	-90%	-91%		
Light GBM								
Baseline	-74%	-76%	-99%	-99%	-90%	-90%		

Tuned	-75%	-77%	-99%	-99%	-90%	-91%		
XGBoost								
Baseline	-75%	-77%	-99%	-99%	-89%	-91%		
Tuned	-74%	-77%	-99%	-99%	-89%	-91%		
Tweedie Regression								
Baseline	-75%	-75%	-99%	-99%	-90%	-90%		
Tuned	-76%	-77%	-99%	-99%	-91%	-90%		
FROM 2,000								
Random Forest								
Baseline	-46%	-53%	-86%	-87%	-63%	-64%		
Tuned	-46%	-47%	-86%	-87%	-62%	-63%		
Light GBM								
Baseline	-45%	-46%	-86%	-87%	-62%	-64%		
Tuned	-46%	-46%	-86%	-87%	-62%	-64%		
XGBoost								
Baseline	-46%	-48%	-86%	-87%	-62%	-64%		
Tuned	-44%	-47%	-85%	-87%	-61%	-64%		
Tweedie Regression								
Baseline	-80%	-80%	-95%	-95%	-78%	-78%		
Tuned	-82%	-83%	-95%	-95%	-78%	-78%		