Table 2. Machine-learning random forest (RF) models outperformed the naïve baseline models in predicting the impact percentile scores of articles in the 'unseen' test set. Model performance was evaluated using the mean absolute error (MAE), and by relative differences and effect sizes in MAE between the R and baseline models, for each impact metric.

Impact metric	Model	MAE (95% CI)	Difference in MAE (%)	Cohen's d (95% CI)	Effect size
Attention score	Random forest	0.08 (0.08 - 0.09)	-68.5%	1.50 (1.31 - 1.70)	large
	Baseline	0.26 (0.25 - 0.28)			
Downloads	Random forest	0.15 (0.14 - 0.16)	-40.0%	0.78 (0.66 - 0.90)	medium
	Baseline	0.25 (0.24 - 0.26)			
Citations	Random forest	0.17 (0.16 - 0.18)	-34.5%	0.65 (0.54 - 0.77)	medium
	Baseline	0.26 (0.25 - 0.27)			