Table A1

Accuracy metrics of the classifiers using crowd evaluation as the gold standard (at least 50% of crowd coders perceived incivility).

Classifiers	Accuracy	F1-score	Precision	Recall
NB	0.61	0.42	0.33	0.61
NB+TFIDF	0.60	0.40	0.33	0.52
LR	0.60	0.45	0.39	0.54
LR+TFIDF	0.58	0.42	0.35	0.54
SVM	0.59	0.47	0.44	0.51
SVM+TFIDF	0.58	0.40	0.32	0.54
RF	0.61	0.32	0.22	0.59
RF+TFIDF	0.62	0.28	0.19	0.67
XGB	0.59	0.46	0.41	0.54
XGB+TFIDF	0.58	0.44	0.38	0.54
Coder	0.63	0.67	0.91	0.53
Dictionary	**0.63**	**0.67**	**0.87**	**0.54**

Note. NB: naïve Bayes; LR: logistic regression; SVM: support-vector machines; RF: random forest; XGB: extreme gradient boosting classifier; TF-IDF: term frequency inverse document frequency. Coder: annotated by human coders.

Table A2

Accuracy metrics of the *Word2Vec* classifiers using crowd evaluation as the gold standard (at least 50% of crowd coders perceived incivility).

Classifiers	Accuracy	F1-score	Precision	Recall
NB	0.65	0.61	0.67	0.56
LR	0.60	0.54	0.55	0.52
SVM	0.57	0.51	0.52	0.50
RF	0.64	0.50	0.44	0.62
XGB	0.65	0.56	0.52	0.62
CNN	0.58	0.50	0.54	0.52
LSTM	0.59	0.45	0.41	0.58
RCNN	0.60	0.53	0.59	0.57
Coder	0.63	0.67	0.91	0.53
Dictionary	**0.63**	**0.67**	**0.87**	**0.54**

Note. NB: naïve Bayes; LR: logistic regression; SVM: support-vector machines; RF: random forest; XGB: extreme gradient boosting classifier; CNN: convolutional neural network; LSTM: long short-term memory recurrent neural network; RCNN: recurrent convolutional neural network. Coder: annotated by human coders.

Table B1

Accuracy metrics of the classifiers using crowd evaluation as the gold standard (at least 40% of crowd coders perceived incivility).

Classifiers	**Accuracy**	**F1-score**	**Precision**	**Recall**
NB	0.63	0.74	0.83	0.67
NB+TFIDF	0.62	0.73	0.83	0.65
LR	0.62	0.70	0.73	0.68
LR+TFIDF	0.62	0.73	0.84	0.64
SVM	0.62	0.69	0.68	0.70
SVM+TFIDF	0.60	0.73	0.88	0.62
RF	0.61	0.65	0.61	0.73
RF+TFIDF	0.60	0.65	0.62	0.69
XGB	0.56	0.67	0.73	0.62
XGB+TFIDF	0.56	0.66	0.71	0.64
Coder	0.77	0.83	0.90	0.77
Dictionary	**0.75**	**0.81**	**0.85**	**0.77**

Note. NB: naïve Bayes; LR: logistic regression; SVM: support-vector machines; RF: random forest; XGB: extreme gradient boosting classifier; TF-IDF: term frequency inverse document frequency. Coder: annotated by human coders.

Table B2

Accuracy metrics of the *Word2Vec* classifiers using crowd evaluation as the gold standard (at least 40% of crowd coders perceived incivility).

Classifiers	**Accuracy**	**F1-score**	**Precision**	**Recall**
NB	0.67	0.72	0.70	0.74
LR	0.66	0.72	0.71	0.72
SVM	0.64	0.70	0.70	0.71
RF	0.72	0.79	0.87	0.73
XGB	0.72	0.79	0.83	0.75
CNN	0.62	0.69	0.75	0.68
LSTM	0.65	0.76	0.89	0.66
RCNN	0.65	0.74	0.86	0.67
Coder	0.77	0.83	0.90	0.77
Dictionary	**0.75**	**0.81**	**0.85**	**0.77**

Note. NB: naïve Bayes; LR: logistic regression; SVM: support-vector machines; RF: random forest; XGB: extreme gradient boosting classifier; CNN: convolutional neural network; LSTM: long short-term memory recurrent neural network; RCNN: recurrent convolutional neural network. Coder: annotated by human coders.

Table C1

Accuracy metrics of the classifiers using crowd evaluation as the gold standard (at least 35% of crowd coders perceived incivility).

Classifiers	Accuracy	F1-score	Precision	Recall
NB	0.71	0.82	0.93	0.73
NB+TFIDF	0.71	0.82	0.94	0.72
LR	0.67	0.78	0.84	0.72
LR+TFIDF	0.70	0.82	0.96	0.71
SVM	0.63	0.74	0.74	0.73
SVM+TFIDF	0.71	0.83	0.98	0.72
RF	0.63	0.72	0.71	0.76
RF+TFIDF	0.64	0.75	0.77	0.75
XGB	0.67	0.79	0.90	0.70
XGB+TFIDF	0.66	0.79	0.86	0.72
Coder	0.81	0.87	0.88	0.86
Dictionary	**0.79**	**0.85**	**0.83**	**0.86**

Note. NB: naïve Bayes; LR: logistic regression; SVM: support-vector machines; RF: random forest; XGB: extreme gradient boosting classifier; TF-IDF: term frequency inverse document frequency. Coder: annotated by human coders.

Table C2

Accuracy metrics of the *Word2Vec* classifiers using crowd evaluation as the gold standard (at least 35% of crowd coders perceived incivility).

Classifiers	Accuracy	F1-score	Precision	Recall
NB	0.70	0.77	0.73	0.82
LR	0.69	0.78	0.77	0.79
SVM	0.66	0.75	0.74	0.76
RF	0.76	0.85	0.94	0.78
XGB	0.74	0.83	0.91	0.76
CNN	0.72	0.80	0.85	0.78
LSTM	0.72	0.82	0.96	0.72
RCNN	0.71	0.80	0.86	0.77
Coder	0.81	0.87	0.88	0.86
Dictionary	**0.79**	**0.85**	**0.83**	**0.86**

Note. NB: naïve Bayes; LR: logistic regression; SVM: support-vector machines; RF: random forest; XGB: extreme gradient boosting classifier; CNN: convolutional neural network; LSTM: long short-term memory recurrent neural network; RCNN: recurrent convolutional neural network. Coder: annotated by human coders.