Supplementary Table 4. Features associated with suicidal ideation selected using machine-learning algorithms that use oversampled data

Variables	LR	RF	kNN	GB	AB	Score
Age	○	○	○	○	○	5
Sex	○	○	○	○	○	5
Town	○	○	○	○	○	5
Medical care	○	○	○	○		4
Subjective health status	○	○	○	○	○	5
Hypertension	○	○	○	○		4
Dyslipidemia	○	○	○	○	○	5
Stroke	○	○	○	○	○	5
OA	○	○	○	○	○	5
RA	○	○	○	○		4
Osteoporosis	○	○	○	○	○	5
DM	○	○	○	○	○	5
Renal failure	○	○		○		3
Liver cirrhosis				○		1
Stress awareness	○	○	○	○	○	5
Depressive mood	○	○	○	○	○	5
Subjective body type recognition	○	○	○	○	○	5
Weight change in past 1 year	○	○	○	○	○	5
Walking	○	○	○	○	○	5
Current smoker	○	○	○	○	○	5
High risk drinking	○	○	○	○		4
Household income quartile						
Lowest	○	○	○	○	○	5
2nd	○	○	○	○	○	5
3rd	○	○	○	○	○	5
Highest	○	○	○	○	○	5
Educational status						
≤Elementary school	○	○	○	○	○	5
Middle school	○	○	○	○	○	5
High school	○	○	○	○	○	5
≥University graduate	○	○	○	○	○	5
Housing type						
Single detached house	○	○	○	○	○	5
Apartment	○	○	○	○	○	5
Tenement	○	○	○	○	○	5
Others	○	○	○			3
Marital status						
Married, live together	○	○	○	○		4
Married, separated	○	○	○	○		4
Not married	○	○	○	○	○	5
Job position						
Permanent worker	○	○	○	○	○	5
Temporary worker	○	○	○	○	○	5
Self–employed	○	○	○	○	○	5
Others	○	○	○	○	○	5

LR, logistic regression; RF, random forest; kNN, k-nearest neighbors; GB, gradient boosting; AB, adaptive boosting; OA, osteoarthritis; RA, rheumatic arthritis; DM, diabetes mellitus.