Unemployment and Mental Illness

The Yellow Team

Anastasiia Altunina, Alexandra Romanenko, Ivan Skvortsov, Maksim Peshkov, Mikhail Grigoryan

Research questions and hypotheses

During the COVID-19 pandemic due to self-isolation, lack of social relations many people became depressed and anxious. Also many people had employment gap.

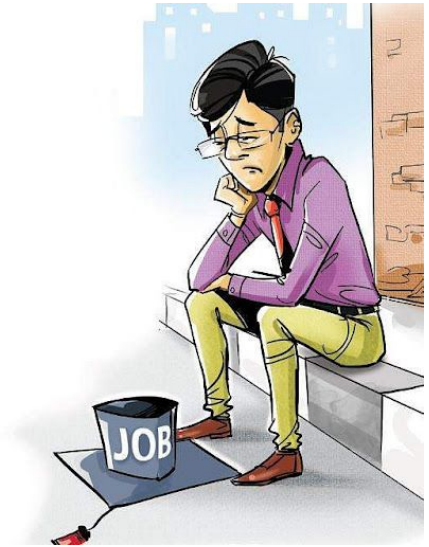
Do mental health problems influence job leaving?

Is the length of employment gap significant for becoming unemployed?

Hypotheses:

- Having mental illnesses significantly and positively affects the probability of becoming an unemployed.
- The length of the resume gap is important for unemployment probability.

Dataset



Source: “Unemployment and mental illness survey”

- 334 observations (296 after cleaning), 31 variables

Demographics	Employment	Mental	Additional
Age, gender, region, education, income, h/h income	Employment status, resumé gap	Mental illness, hospitalizations, disorders (i. e. depression, anxiety, PAs)	Internet access, section 8 housing, food stamps, etc.

Descriptive statistics

Almost third of sample are anxious.

27% people are depressed.

31% people are tired.

87% have their own PC and 96% have internet access.

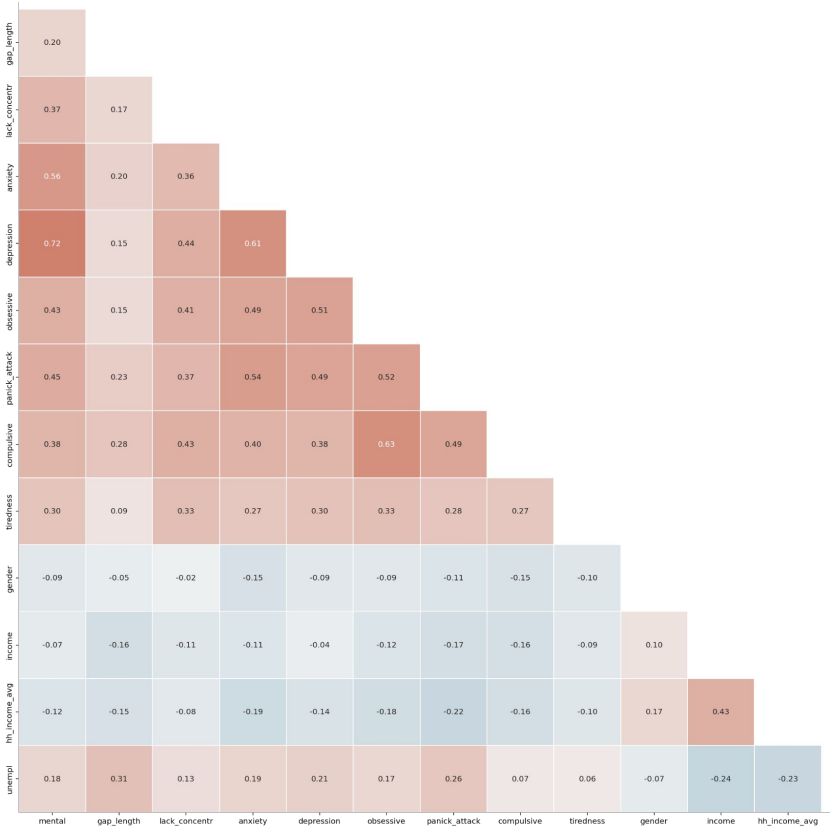
69% are employed.

	Mean	Dtype
mental	0.253	Binary
disabled	0.108	Binary
age	-	Categorical
gap_length	9.034	Numeric
days_hosp	3.010	Numeric
internet_access	0.966	Binary
food_stamps	0.068	Binary
section_8	0.020	Binary
own_pc	0.878	Binary
lack_concentr	0.166	Binary
anxiety	0.304	Binary
depression	0.277	Binary
obsessive	0.135	Binary
swings	0.125	Binary
panick_attack	0.152	Binary
compulsive	0.091	Binary
tiredness	0.311	Binary
gender	0.489	Binary
hh_income_avg	71950	Numeric
empl	0.695946	Binary

Data Preprocessing

- Missing values treated with kNN Imputer
- Categorical variables: One-Hot Encoding
- Numerical variables: Normalization (x-E(x))/sd(x)
- Balanced train/test split (target variable = ‘empl’)

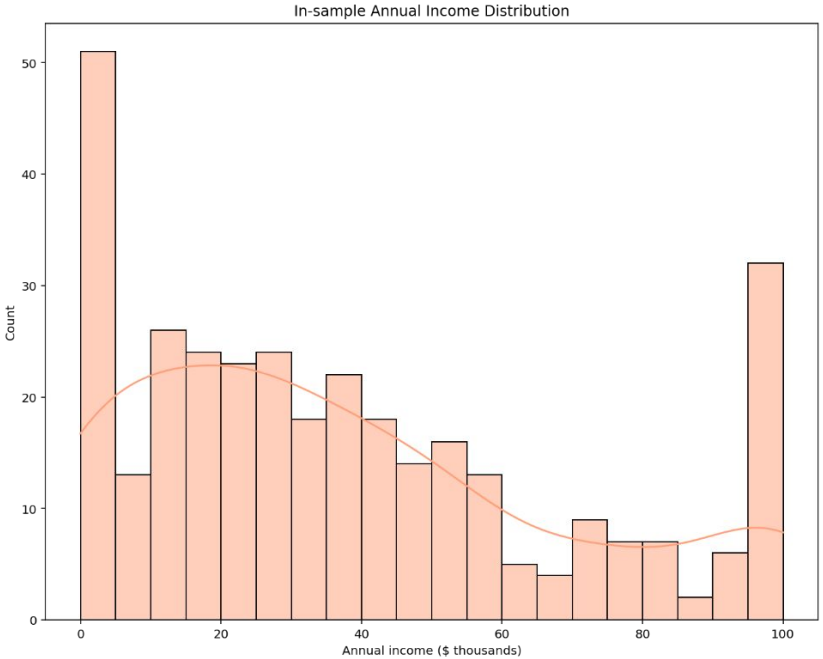
Correlation Matrix

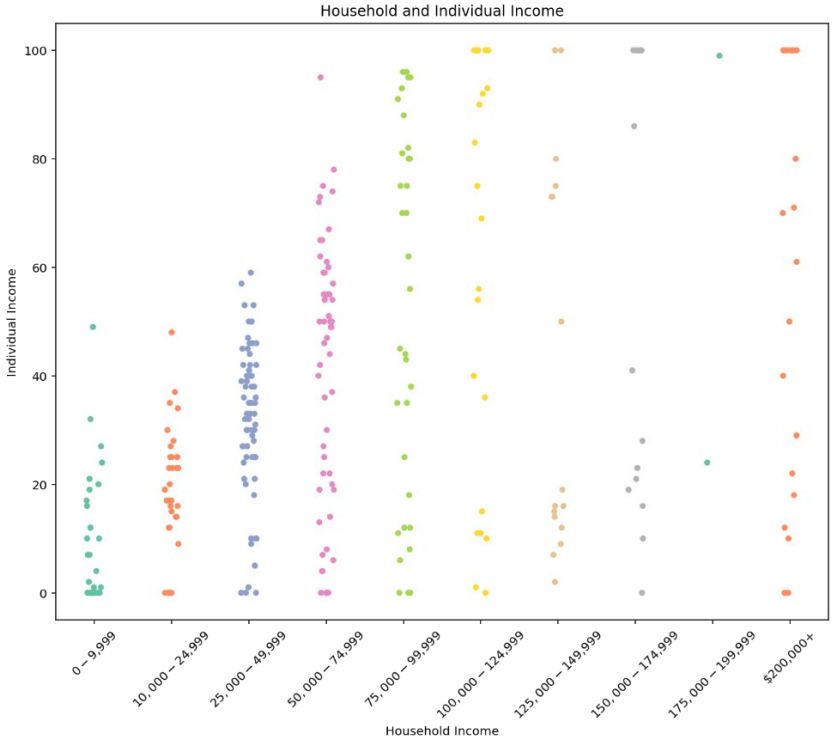


- Indicators of mental health are correlated.
- *Gap length* is correlated with unemployment (=> using control).
- *Slight correlates for unemployment:* mental illness, depression, PAs, anxiety.

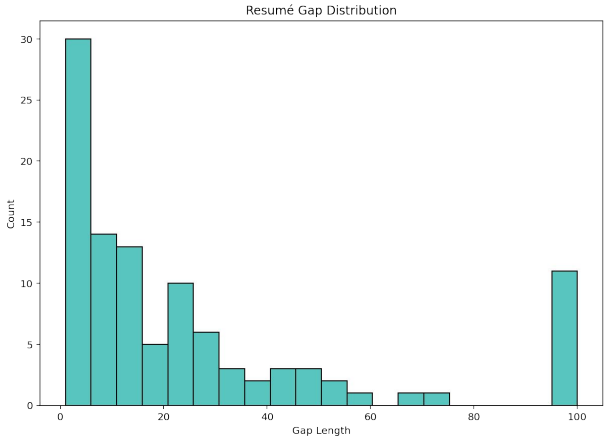
These relationships are the reason to investigate the importance of mental health for employment.

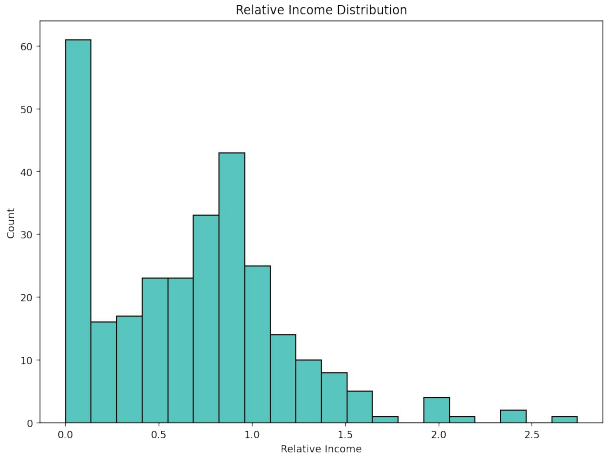
Income

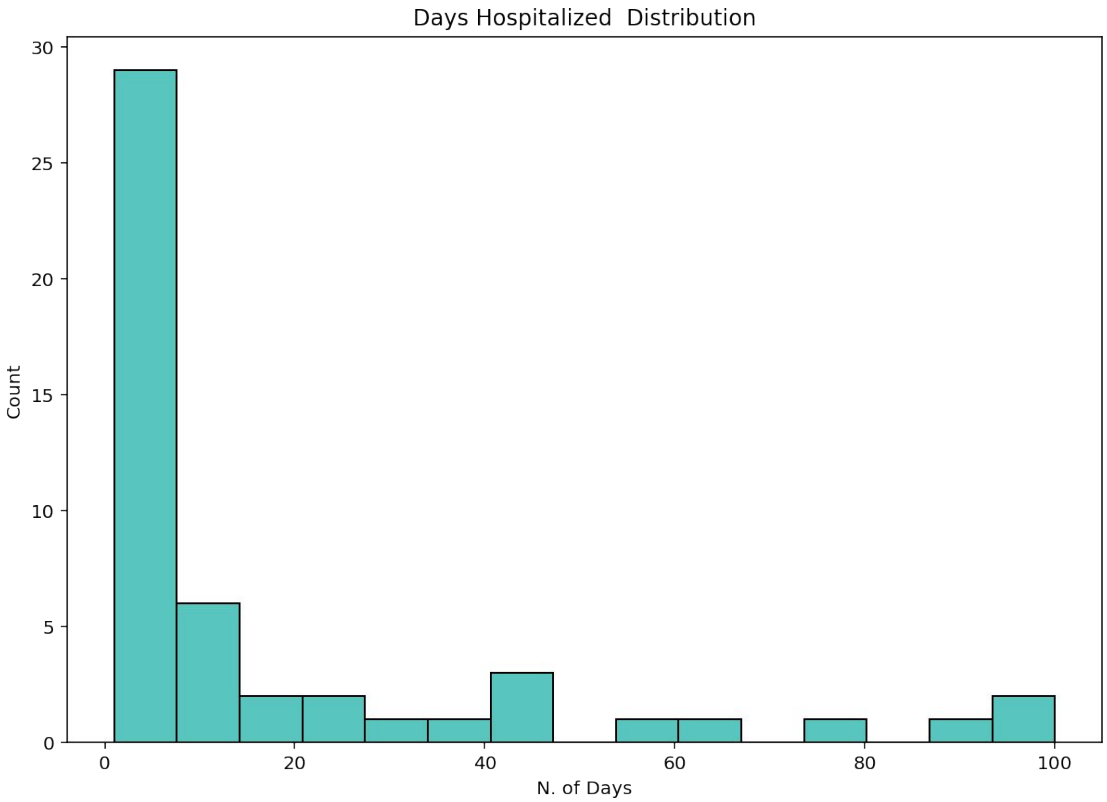

In-sample Annual Income Distribution
Count
Annual income ($ thousands)
0
10
20
30
40
50
0
20
40
60
80
100


Household and Individual Income
Individual Income
Household Income
0
20
40
60
80
100
0 – 9,999
10,000 – 24,999
25,000 – 49,999
50,000 – 74,999
75,000 – 99,999
100,000 – 124,999
125,000 – 149,999
150,000 – 174,999
175,000 – 199,999
$200,000+

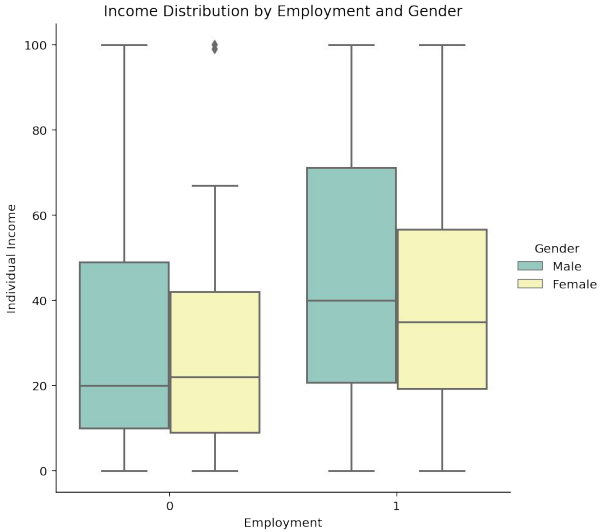
Distributions of numerical variables



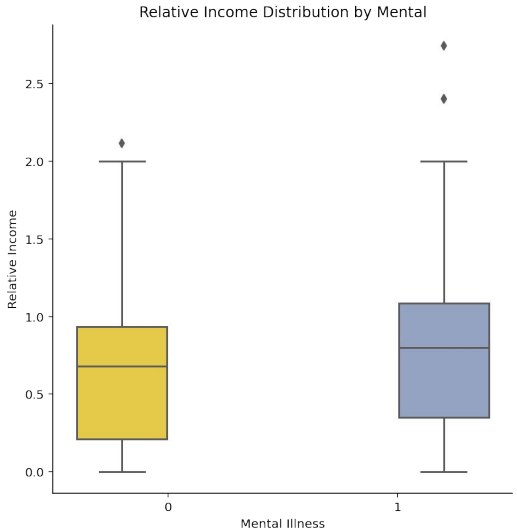




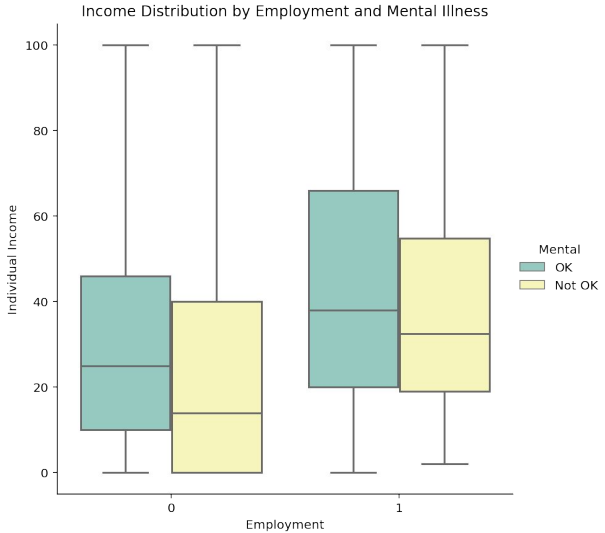
Conditional Distributions



Women have lower income that men and employed persons have higher income than unemployed.



People with mental illnesses have the same distribution of relative income.



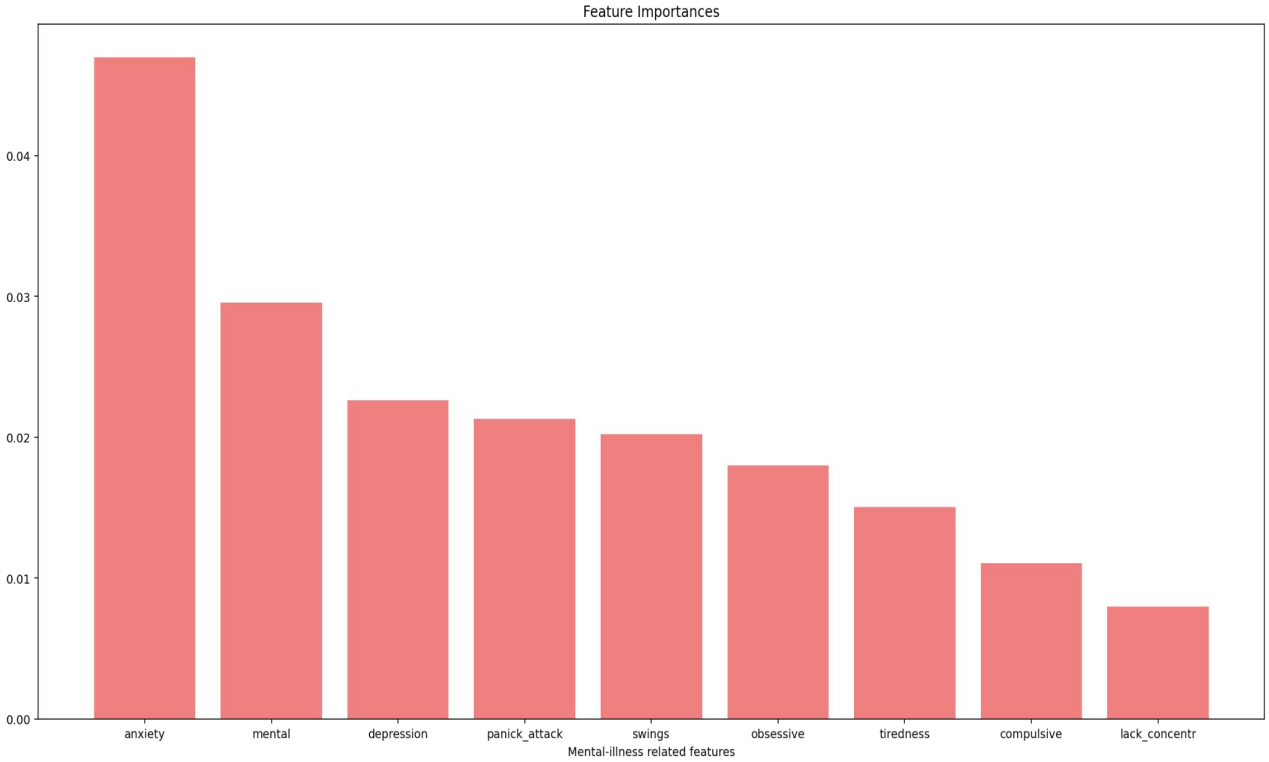
Persons with mental illnesses are having less income than people without them.

Model Comparison

	Accuracy	ROC AUC	F1	CV params
Random Forest	0.771	0.8249	0.8595	*max_depth* = 4
XGBoost	0.770	0.7947	0.8496	*depth* = 2, *iterations* = 100
Logit	0.743	0.7005	0.843	*C* = 10.0, *l1_ratio* = 0.0
kNN	0.703	0.7426	0.807	*n_neighbors* = 8, *weights* = “distance”

*see more models and CV params in work notebook

Random Forest



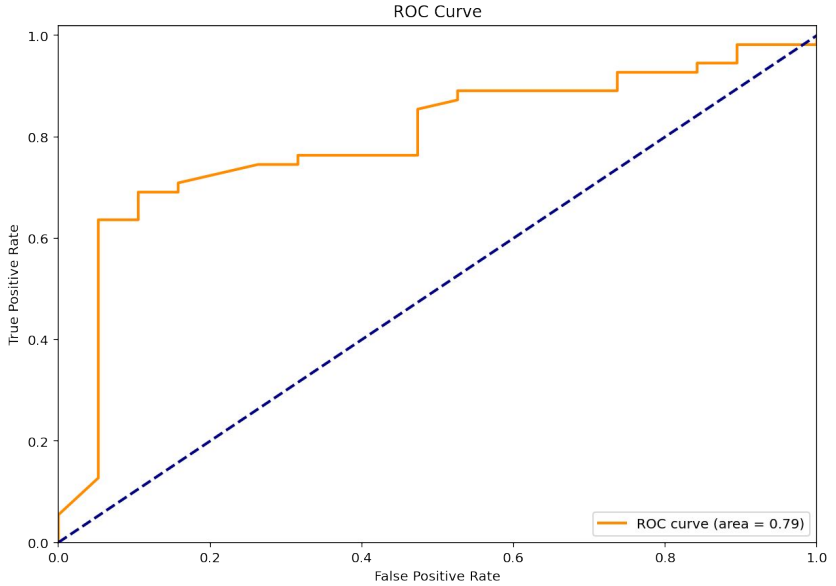
Feature importance — average decrease in entropy caused by splitting tree by this feature

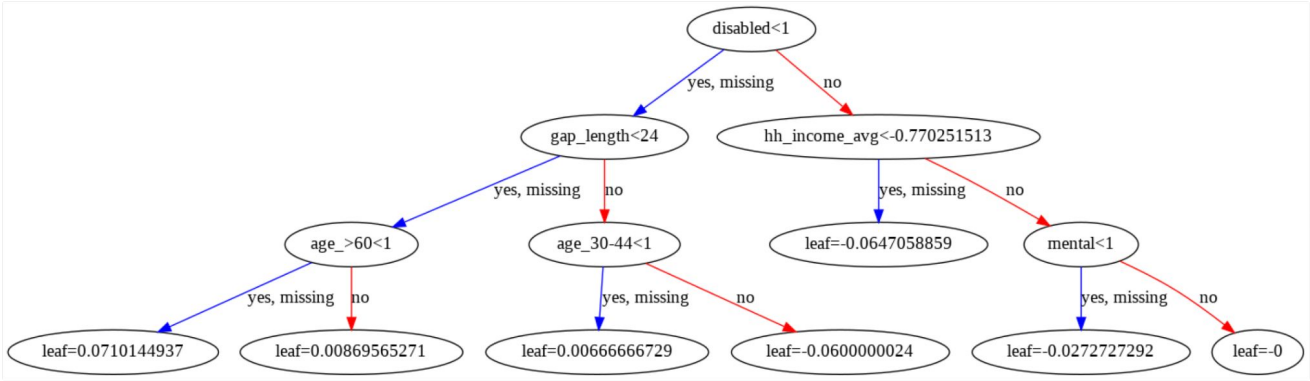
Impossible to estimate direction of the effects due to non-linearity.

XGBoost

XGBoost demonstrates one of the best accuracy and ROC-AUC.

XGBoost feature tree shows that the most important determinants of employment are “disabled”, “gap length” and household income.





Logistic Regression

Logit Regression Results

Dep. Variable:	**empl**	No. Observations:	222
Model:	Logit	Df Residuals:	201
Method:	MLE	Df Model:	20
Date:	Thu, 14 Oct 2021	Pseudo R-squ.:	0.2682
Time:	15:04:02	Log-Likelihood:	-101.82
converged:	True	LL-Null:	-139.13
Covariance Type:	nonrobust	LLR p-value:	3.129e-08

	coef	std err	z	P>|z|	[0.025	0.975]
mental	0.2815	0.620	0.454	0.650	-0.933	1.496
disabled	-2.3681	0.690	-3.432	0.001	-3.721	-1.016
gap_length	-0.0366	0.010	-3.485	0.000	-0.057	-0.016
days_hosp	0.0105	0.018	0.594	0.553	-0.024	0.045
internet_access	1.5258	0.558	2.735	0.006	0.432	2.619
food_stamps	-0.6616	0.724	-0.914	0.361	-2.081	0.757
section_8	0.5910	1.421	0.416	0.677	-2.194	3.376
own_pc	0.2709	0.512	0.529	0.597	-0.733	1.274
lack_concentr	-0.5045	0.580	-0.869	0.385	-1.642	0.633
anxiety	-0.9585	0.545	-1.760	0.078	-2.026	0.109
depression	0.1424	0.643	0.221	0.825	-1.118	1.403
obsessive	-0.2102	0.802	-0.262	0.793	-1.781	1.361
swings	-0.0051	0.634	-0.008	0.994	-1.248	1.238
panick_attack	-0.4186	0.631	-0.663	0.507	-1.656	0.819
compulsive	1.2500	0.899	1.390	0.165	-0.513	3.013
tiredness	-0.0035	0.439	-0.008	0.994	-0.865	0.858
gender	0.4105	0.377	1.088	0.277	-0.329	1.150
hh_income_avg	0.1625	0.211	0.769	0.442	-0.252	0.577
age_30-44	0.2933	0.513	0.572	0.567	-0.712	1.299
age_45-60	0.5274	0.608	0.867	0.386	-0.664	1.719
age_>60	-1.5017	0.570	-2.634	0.008	-2.619	-0.384

The length of employment gap significantly and negatively affects to employment

Disabled persons have more probability to be unemployed

Most of mental illnesses are not significant (or anxiety is significant but works in an opposite way)

We assume that absence of significance and weird effects can be caused by endogeneity and sample bias

Results of our research

We use different methods to data preprocessing (kNN Imputer, One-Hot Encoding, Normalization)

Analyze descriptive statistics of data and different relations in box-plots, histograms and scatter plots

Build ML methods to investigate research question and make carefully prediction of unemployment (Random Forest, XGBoost, LDA,CatBoost, Logit, kNN)

Estimate metrics of classification and choose best model (Random Forest) with more than 77% accuracy.

Check the importance of mental health problems features to unemployment and find that they are **insignificant.**

But employment gap is indeed **significant**.

Anyway, we wish everyone to find new job after work leaving as soon as possible ❤️


Thank you for your attention!
Do not ask difficult questions please ...
Telefoner.ru
meme-arsenal.ru