Technical Report: Mental Health Model

Elmeri Tikkanen, Selena Verha, Shafayet Islam

October 28, 2024

University of Helsinki
Faculty of Science

Contents

1. Introduction

Mental health includes a person's emotional, psychological, and social well-being. Access to proper employment, housing, and education, positive social connections, positive coping skills, and a safe community can improve an individual's mental health and protect them from mental health disorders [1]. While mental health disorders are generally linked to genetics, brain chemistry, and prenatal environmental exposure, several risk factors increase the likelihood that someone may develop a mental health condition. These factors include stressful situations, chronic medical conditions, brain damage, traumatic experiences, substance abuse, social isolation, and childhood history of trauma and neglect [2, 1].

Mental health conditions can affect all aspects of an individual's life. Due to the large variance in symptoms, the effect that mental disorders have is also varied. Some symptoms include negative or confused thoughts and emotions, mood changes, delusions, paranoia, hallucinations, substance abuse, aggressive behavior, and suicidal thinking. These symptoms can lead to difficulties in personal relationships, problems related to work or education, legal and financial difficulties, self-harm and harm to others, and poverty and homelessness [2]. Mental disorders and the behavioral risk factors associated with them have been linked with both non-communicable and infectious diseases. For example, depression and anxiety have been linked with an increased occurrence of cardiovascular disease and diabetes, and communicable diseases such as HIV and AIDS have been associated with serious chronic conditions such as schizophrenia. Serious mental health disorders are also associated with physical injury due to deliberate or unconscious behaviors [4].

OECD reports that the yearly cost of mental health issues in Finland is around 11 billion euros. About 55 % of all disability pensions are granted due to mental health reasons. At least every fifth Finn experiences a disturbance in mental health, and the same number of people experience depression during their lifetime. Despite this, the funding of psychiatry and mental health services has dropped. It is estimated that only half of those who have a mental health condition are receiving health care related to their condition [3].

The goal of our project is to make a tool that helps public health administrators

reduce mental health problems based on data. The mental health model has multiple explanation variables for mental health. It helps to make decisions on factors that affect the mental health most in a municipality. Mental health is divided into three dimensions, each describing different kinds of mental health issues. The explaining factors are modeled for all three dimensions of mental health, such that the end user can see which kind of mental health problems are the most prevalent and which factors have the most effect on that kind of mental health.

2. Data

We collected data for features and targets from the Finnish Institute for Health and Welfare (THL), Statistics Finland (Tilastokeskus), and The Social Insurance Institution of Finland (Kela). All the data was yearly data for each municipality, and if the municipality data wasn't available for a variable, the well-being county data was used for municipalities in that county. We selected the most important variables for our purpose and made sure that they were per capita, and if not, we divided the data by population. We reformulated the data tables such that the first two columns are municipality and year, and the rest of the columns are the variables. That way, we have a data point for each municipality-year pair. For all the variables, we had data available from years 2008-2022, and from 218 municipalities, so there are in total $15 \cdot 218 = 3270$ data points.

All data used is publicly available without permission. THL and Kela's data are under Creative Commons license BY 4.0, and Tilastokeskus has various licenses for different datasets.

2.1 Data preprocessing

Missing values were filled with values from neighbor years. Also, missing values in the influencing data frame are imputed using sklearn's "SimpleImputer" class. We use the mean of the column to impute. Also, the dimension of the dataset is normalized for all independent parameters and mental health-dependent parameters. After preprocessing, we get a table of 3270x49 datasets of dependent and independent variables. Among them, 42 columns hold numerical values to make some useful satirical assumptions.

3. Methods

We used random forest regression as a machine learning model to explain the mental health categories with the various features.

3.1 Mental health index

Because there is no straightforward way to quantify mental health, we had to create an index describing it. The Finnish Institute for Health and Welfare (THL) provides many indicators for mental health. From those, we chose the data for the whole population, and if only available for age groups, the age groups were combined with appropriate weighting. Some variables that were mostly missing values or less suitable for our index were dropped.

After combining the variables, there were, in total, 14 variables indicating mental health (appendix A). Many of these variables had quite high correlations with each other (fig. A.1), so a simple standardized/normalized average of the variables wouldn't be a fair index in terms of weighting: many variables describing the same phenomenon would give more weight to that phenomenon, and less weight to a phenomenon that is described by just one variable.

This was solved by principal component analysis, for which all the variables were first standardized by subtracting the mean and dividing by standard deviation. The first three principal axes explain $0.24165425 + 0.18971329 + 0.10119375 = 0.5325612863779248 \approx 53\%$ of the variance. That is not much, but by looking at the components of the first three principal axes (appendix A), we could separate all 14 variables into roughly three categories of mental health by taking the components with the highest value into the category. We named the final three categories of mental health "psychotic", "depression" and "involuntary care". The variables that are included in those categories are in appendix A.

This way, the 14 variables fit surprisingly well into three categories: only a few variables didn't fit clearly into a single category. The variables that didn't fit clearly into a single category are `'Alcohol mortality per 100 000 inhabitants'`, which could belong to both "psychotic" and "involuntary care" categories, and the same for `'Psychiatric`

`inpatient care, patients per 1000 inhabitants'`. These could have also been split into both categories with half-weighting. The component `'Suicides per 100 000 inhabitants'` didn't have that high value in the "psychotic" principal axis compared to other variables, but in the other principal axes, its value was deficient.

Now, one can also make a 1-dimensional index with fairer weighting from this standardized 3-dimensional mental health index by just taking an average of each of those three dimensions. This 3-dimensional index is, however, left for use, as it gives more information about the quality of mental health issues because correlations differ for each dimension. The standardization of these indices means that value 0 of the index is the average of all data points, value -1 is one standard deviation less than the average, and value 1 is one standard deviation higher than the average, and so on.

3.2 Detecting important mental health factors

We use the RandomForrestRegressor class from the sklearn package to determine the importance of the independent feature (socio-economic factors) over the three dependent mental health indexes. After that, we successfully identified 22 parameters (16 of which contain useful information). These parameters influence the Mental Health index factors over the municipalities and years.

3.3 Ranking Mental Health influencing factors

We took these 16 factors to rank the parameters for a single municipality. Again, we use the RandomForestRegressor to determine the importance of the value. The values are scaled from 0 to 1 for all the factors.

4. End product and Usage

The product for the end user is a web page that visualizes the data in various ways. The main function is that a user can choose a Finnish municipality and then see which of the three mental health categories is the most prevalent and which factors affect that kind of mental health the most. It also shows which factors affect the municipality's total mental health score the most.

The web page also includes an HTML-based interactive map that shows how different categories of mental health problems are distributed regionally. The map displays mental health index data from the latest year for a municipality upon the user selecting the municipality by clicking it. The colors on the map indicate the combined effect of all of the mental health Appendix. All the variables in the mental health factor are described in the index. The usage of the webapp is self-explanatory. The website is hosted at https://shafayetrahat.eu.pythonanywhere.com/. All codes and datasets are provided in the https://github.com/shafayetrahat/mental_health_data_analysis_finland and https://helsinkifi-my.sharepoint.com/:u:/g/personal/sjverha_ad_helsinki_fi/ES5GhwBvFXVBvmtE5MRZ9zEBER2TZfOB5d1g40_ML-94EQ?e=7D4RPQ in

5. Conclusions

This model could be a real and useful tool/product with some adjustments and improvements. Public health administrators and decision-makers could really use this kind of visualized interactive data to help decision-making. If this kind of model were more accurate and had more features, it could be an essential tool for allocating resources. There could have been much more explaining variables to catch the most important factors affecting mental health, but unfortunately, there weren't many datasets in the yearly regional form, which would be necessary for this kind of model to have enough data points for training.

5.1 Improvements

The model could be further improved by trying to take into account causality and not just correlations. The factors that affect mental health the most, that are given to end users, may not actually be the most important factors to change in order to reduce mental health issues: they might not be causing the mental health issues, but some other common factor might be causing the mental health issues and the factors that correlate with mental health issues. Additionally, the data features several risk factors that may also be considered to be symptoms or consequences of mental health issues, which highlights the importance of a robust data set and taking into account causality.

Now, the model gives the most important features that affect mental health. It could be improved such that it would show in which direction a municipality needs to change in order to reduce mental health issues.

Bibliography

[1] Center for Disease Control (CDC). About mental health. Accessed: 2024-10-26.

[2] Mayo Clinic. Mental illness. Accessed: 2024-10-26.

[3] Mieli ry. Tilastotietoa mielenterveydestÃ€ (in finnish). Accessed: 2024-10-26.

[4] M. Prince, V. Patel, S. Saxena, M. Maj, J. Maselko, M. R. Phillips, and A. Rahman. No health without mental health. *The Lancet*, 370(9590):859–877, Sep 2007.

Appendix A. Mental health index details

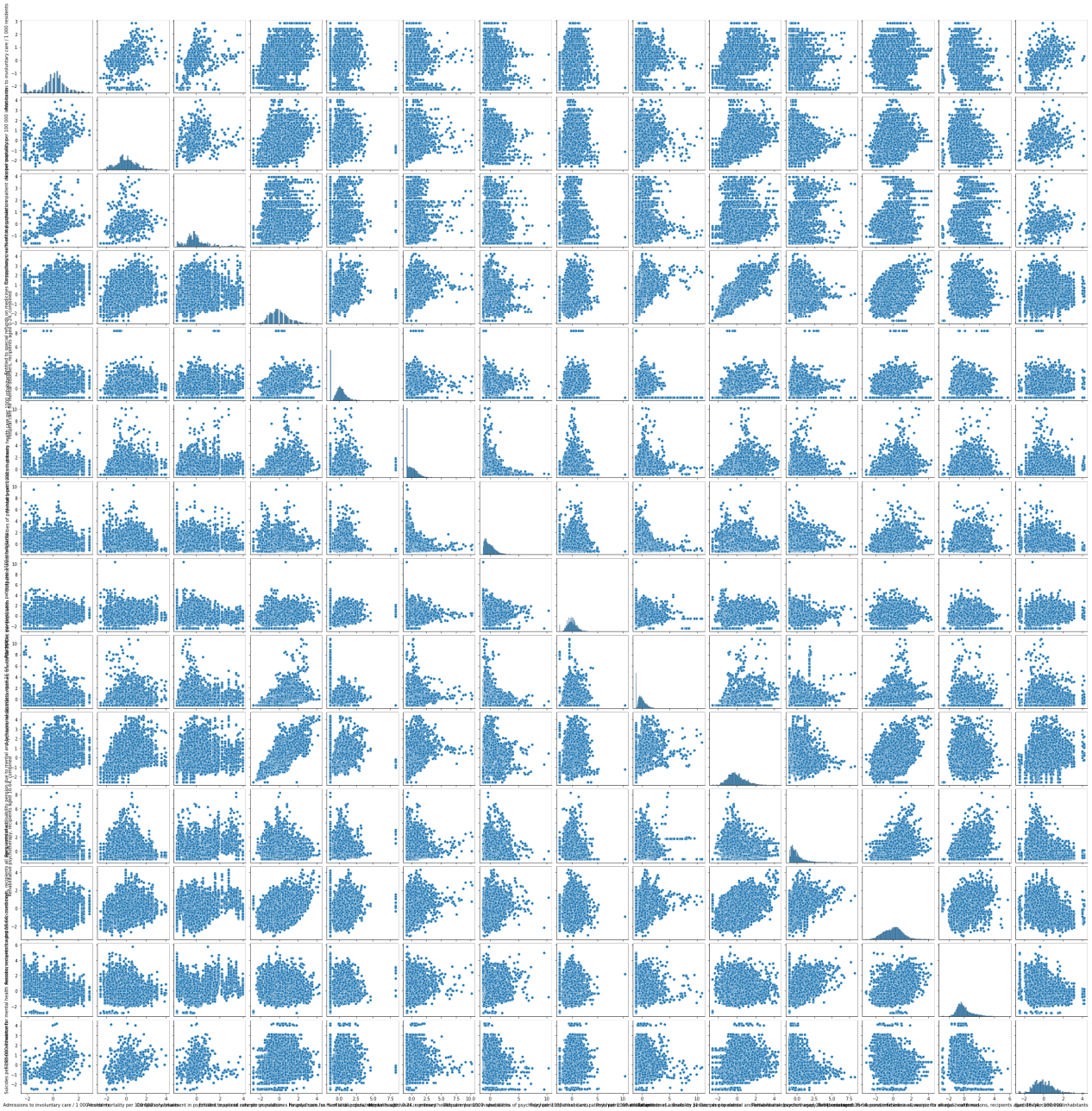

Figure A.1: Scatter matrix of the 14 mental health indicators. The order of the variables is the same as in the list below.

The 14 variables from which the mental health index is made are

```
'Admissions to involuntary care / 1 000 residents',
'Alcohol mortality per 100 000 inhabitants',
'Compulsory treatment in psychiatric inpatient care per population',
```

```
'Entitled to special refunds on medicines for psychosis,
as \% of total population',
'Hospital care for mental disorders, recipients aged 0-24, combined',
'Mental health visits in primary health care per 1000 inhabitants',
'Outpatient visits in specialities of psychiatry per 1000 inhabitants',
'Psychiatric inpatient care, patients per 1000 inhabitants',
'Psychiatric rehabilitation homes, clients on 31 Dec per population',
'Recipients of a disability pension due to mental and behavioural disorders
aged 16-64, combined',
'Rehabilitative psychotherapy, recipients aged 16-64, combined',
'Reimbursements for depression medicines, recipients all ages, combined',
'Sickness allowance for mental health reasons, recipients aged 18-64,
combined',
'Suicides per 100 000 inhabitants'
```

Components of the first 3 principal axis of the mental health index data:
First principal axis:

```
[(0.4609424765841399,
'Recipients of a disability pension due to mental and behavioural disorders
aged 16-64, combined'),
(0.4563251190355294,
'Entitled to special refunds on medicines for psychosis, as % of total
population'),
(0.31965275569813834,
'Psychiatric rehabilitation homes, clients on 31 Dec per population'),
(0.3181010447771621,
'Mental health visits in primary health care per 1000 inhabitants'),
(0.31059500869971257, 'Alcohol mortality per 100 000 inhabitants'),
(0.22669838597532244,
'Psychiatric inpatient care, patients per 1000 inhabitants'),
(0.22187600448434167, 'Admissions to involuntary care / 1 000 residents'),
(0.2139182851228127,
'Reimbursements for depression medicines, recipients all ages, combined'),
(0.19780173600830167,
'Hospital care for mental disorders, recipients aged 0-24, combined'),
(0.1972330390774692, 'Suicides per 100 000 inhabitants'),
(0.18817074912756898,
'Compulsory treatment in psychiatric inpatient care per population'),
```

```
(0.040373652296786716,
'Rehabilitative psychotherapy, recipients aged 16-64, combined'),
(0.0016983946420060514,
'Sickness allowance for mental health reasons, recipients aged 18-64,
combined'),
(-0.1342173712756169,
'Outpatient visits in specialities of psychiatry per 1000 inhabitants')]
```

Second principal axis:

```
[(0.449927343485963,
'Sickness allowance for mental health reasons, recipients aged 18-64,
combined'),
(0.44897274604807663,
'Rehabilitative psychotherapy, recipients aged 16-64, combined'),
(0.4345787854158432,
'Reimbursements for depression medicines, recipients all ages, combined'),
(0.17462638019491142,
'Outpatient visits in specialities of psychiatry per 1000 inhabitants'),
(0.1484318011527161,
'Mental health visits in primary health care per 1000 inhabitants'),
(0.1476870475261575,
'Psychiatric rehabilitation homes, clients on 31 Dec per population'),
(0.049722244684580155,
'Recipients of a disability pension due to mental and behavioural
disorders aged 16-64, combined'),
(0.04164676883432486,
'Compulsory treatment in psychiatric inpatient care per population'),
(0.04020579577530478,
'Entitled to special refunds on medicines for psychosis, as % of total
population'),
(-0.014044065231494762,
'Hospital care for mental disorders, recipients aged 0-24, combined'),
(-0.1009992192094119, 'Alcohol mortality per 100 000 inhabitants'),
(-0.17699538195978584,
'Psychiatric inpatient care, patients per 1000 inhabitants'),
(-0.3596524061667962, 'Admissions to involuntary care / 1 000 residents'),
(-0.3948338186696529, 'Suicides per 100 000 inhabitants')]
```

Third principal axis:

```
[(0.5013828235115052,
'Outpatient visits in specialities of psychiatry per 1000 inhabitants'),
(0.4214599272448657,
'Hospital care for mental disorders, recipients aged 0-24, combined'),
(0.39934145580321484,
'Compulsory treatment in psychiatric inpatient care per population'),
(0.2895007568574595, 'Admissions to involuntary care / 1 000 residents'),
(0.23859734091553292,
'Psychiatric inpatient care, patients per 1000 inhabitants'),
(0.20331650422055988, 'Alcohol mortality per 100 000 inhabitants'),
(0.16603494900471638,
'Rehabilitative psychotherapy, recipients aged 16-64, combined'),
(0.13348413696499056,
'Reimbursements for depression medicines, recipients all ages, combined'),
(0.05428342410606145,
'Sickness allowance for mental health reasons, recipients aged 18-64,
combined'),
(-0.004217587400071776, 'Suicides per 100 000 inhabitants'),
(-0.15399651419934904,
'Entitled to special refunds on medicines for psychosis, as % of total
population'),
(-0.15948204432889748,
'Recipients of a disability pension due to mental and behavioural disorders
aged 16-64, combined'),
(-0.23938991294231574,
'Psychiatric rehabilitation homes, clients on 31 Dec per population'),
(-0.27318334307572467,
'Mental health visits in primary health care per 1000 inhabitants')]
```

The three categories of mental health are (order of relative importance):
First principal axis, "psychotic."

```
'Recipients of a disability pension due to mental and behavioural disorders
aged 16-64, combined'
'Entitled to special refunds on medicines for psychosis, as % of total
population'
'Psychiatric rehabilitation homes, clients on 31 Dec per population'
'Mental health visits in primary health care per 1000 inhabitants'
'Alcohol mortality per 100 000 inhabitants'
```

'Suicides per 100 000 inhabitants'

The second principal axis, "depression,"

'Sickness allowance for mental health reasons, recipients aged 18-64, combined'
'Rehabilitative psychotherapy, recipients aged 16-64, combined'
'Reimbursements for depression medicines, recipients all ages, combined'

Third principal axis, "involuntary care,"

'Outpatient visits in specialties of psychiatry per 1000 inhabitants'
'Hospital care for mental disorders, recipients aged 0-24, combined'
'Compulsory treatment in psychiatric inpatient care per population'
'Admissions to involuntary care / 1 000 residents'
'Psychiatric inpatient care, patients per 1000 inhabitants'

mean_debt_euro_buisiness= mean debt for business
income = total household income
Pensioner_housing_allowance = pensioner's housing allowance
School_travel_allowance = school travel allowance
Rehabilitation_allowance = Rehabilitation allowance annually
Child_benefit = Child_benefit annually
Medical_benefits = medical benefits in euro(per anum)
Unemployment_insurance = per annum Unemployment_insurance
mean_debt_euro_total = household mean debt per annum

Disability_benefits, Temporary_power_support, General_housing_allowance, mean_debt_eu