Understanding the Effect of Certain Socioeconomic Factors on Mental Health Outcomes

Medha Pappula & Kade Yen

Project Goal

- Mental health => leading global concern
- Individuals with lower socioeconomic status => exhibits higher rates of mental disorders and vice versa for higher socioeconomic status individuals
- Leads to our question:

What specific socioeconomic factors among adults in the United States have a **profound impact** on mental health **severity**?

Dataset Description

- https://www.cdc.gov/nchs/nhis/2023nhis.htm
- 29522 instances
- 647 attributes
- Class Label
  - Combines PHQ41_A, PHQ42_A, PHQ44_A, and PHQ44_A attributes

Code	Description
1	Not at all
2	Several days
3	More than half the days
4	Nearly every day
7	Refused
8	Not Ascertained
9	Don't Know

MHA: Brief mental health assessment

Variable #	Question #	Variable Name	Source Variables	Description	Type	Location	Length
1	MHA.0020.00.4	PHQ41_A		How often little interest in things, past 2 weeks	Num	518	1
2	MHA.0030.00.4	PHQ42_A		How often feeling down, past 2 weeks	Num	519	1
3	Recode	PHQ2SCREEN_A	PHQ41_A; PHQ42_A	PHQ-2 screener result	Num	520	1
4	MHA.0040.00.4	PHQ43_A		How often felt nervous/anxious/on edge, past 2 weeks	Num	521	1
5	MHA.0050.00.4	PHQ44_A		How often can't stop/control worrying, past 2 weeks	Num	522	1
6	Recode	GAD2SCREEN_A	PHQ43_A; PHQ44_A	GAD-2 screener result	Num	523	1

Pre-Processing

Removal of all attributes with >70% missing because:

- Lack of information - little value
- Curse of dimensionality - increases complexity of model with little improvement
- Leads to overfitting

We removed these by deleting the attribute all together.

Pre-Processing - Continued

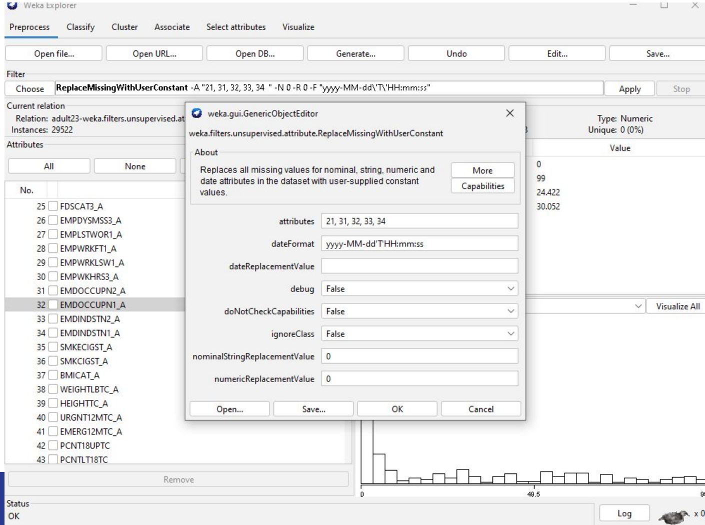
Filling in default values within the attribute because:

- Not everyone willing to answer questions
- Refusal to answer
- Participants don’t know if they have or don’t have

Replaced default value using Weka’s

ReplaceMissingWithUserConstant

7	Refused
8	Not Ascertained
9	Don't Know



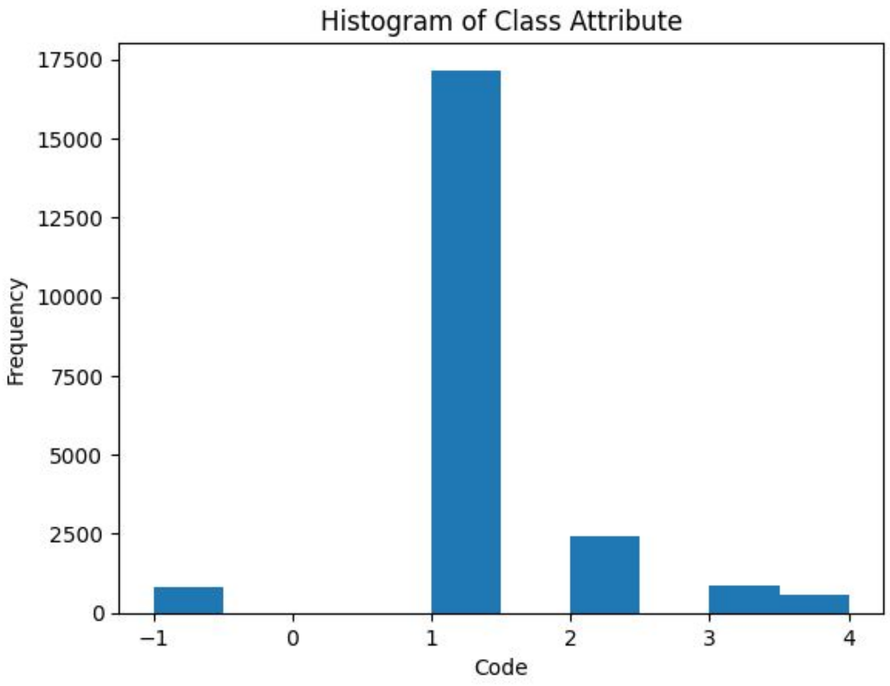
Pre-Processing - Continued

Engineering the class attribute involved combining the 4 previously mentioned into a class variable. This was completed using a python script.

Overall left with 336 attributes.

```
def engineer_attribute(row):

 """
 Engineers a new attribute based on PHQ41_A, PHQ42_A, PHQ43_A, PHQ44_A.
 row: A pandas Series representing a row in the DataFrame.
 Returns: The engineered attribute value.
 """

 phq_values = [row['PHQ41_A'], row['PHQ42_A'], row['PHQ43_A'], row['PHQ44_A']]
 if all(1 <= val <= 1 and not np.isnan(val) for val in phq_values):
   return 1
 elif all(7 <= val <= 9 and not np.isnan(val) for val in phq_values):
   return -1
 else:
   average = np.nanmean(phq_values)
   if 2 <= average <= 4:
     return round(average)
   else:
     return np.nan  # Or another default value if needed

# Apply the function to create a new column
df['engineered_attribute'] = df.apply(engineer_attribute, axis=1)

# Print the DataFrame with the new column
print(df[['PHQ41_A', 'PHQ42_A', 'PHQ43_A', 'PHQ44_A', 'engineered_attribute']])
```

Spread of Class Variable



Manual Attribute Removal

Removed all attributes related to mental health or explicitly mentioned depression/anxiety

Unweighted frequencies:

DEPEV_A Ever had depression

Code	Description	Frequency	Per
1	Yes	5769	
2	No	23708	
7	Refused	28	
8	Not Ascertained	0	
9	Don't Know	17	

Frequency Missing:

Unweighted frequencies:

ANXMED_A Take medication for worried/nervous/anxious feelings

Code	Description	Frequency	Percent
1	Yes	4329	14.66
2	No	24470	82.89
7	Refused	44	0.15
8	Not Ascertained	657	2.23
9	Don't Know	22	0.07

Frequency Missing:

Training / Validation / Testing Dataset

Completed using Python sklearn's train_test_split function. Used stratified sampling

- Train/test: 70%
- Validation: 30%

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
```

Attribute Selection Algorithms

CfsSubsetEval with BestFirst

- Good subset of features contains features highly correlated with the target class but uncorrelated with each other.
- Best First
  - Forward Selection: Empty set of features and add features incrementally
  - Backward Elimination: Start will all features and remove one by one
  - Bidirectional Search: Combine both forward and backward elimination

Selected attributes: 10,63,107,134,140,150,154,159,160,189

DISAB3_A
MLTFAMFLG_A
EVRMARRIED_A
SMKNOW_A
PAITOOTH3M_A
VIGIL4_A
DISCRIM5_A
MHTHND_A
MHTHDLY_A
HYSTEV2_A

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}}$$

Attribute Selection Algorithms - Continued

Handpicked

- Calculating the Spearman rank correlation index of each attribute in relation to the class “engineered attribute”.
- Taking the absolute value of the correlation coefficients
- Take the top 10, around a threshold of 0.6 on both the positive and negative end.

0.634879	121	HRTESTLAST_A
0.610201	26	EMPDYSMSS3_A
0.596229	1	URBRRL
0.593039	70	REGION
0.592348	7	PSTRAT
0.590748	224	LONGCOVD1_A
0.587731	150	VIGIL4_A
0.58613	134	SMKNOW_A
0.582264	71	INTV_QRT
0.579129	151	VIGIL3_A

```
# Calculate Spearman rank correlations
correlations = df.corr(method='spearman')['engineered_attribute'].drop('engineered_attribute')

# Take the absolute value of the correlations
absolute_correlations = correlations.abs()

# Get the top 10 most correlated attributes
top_10_correlations = absolute_correlations.nlargest(10)

print(top_10_correlations)
```

Attribute Selection Algorithms - Continued

CorrelationAttributeEval with Ranker

- CorrelationAttributeEval evaluates attributes based on their correlation with the class label.
- Measures the relationship between each feature and target class.
- Uses the Pearson correlation.

0.633966	261	SOCSCLPAR_A
0.633344	266	COGMEMDFF_A
0.568578	311	LSATIS4_A
0.538037	312	PHSTAT_A
0.511334	262	SOCERRNDS_A
… [omitted for length]		
-0.703882	154	DISCRIM5_A
-0.724287	150	VIGIL4_A
-0.745768	160	MHTHDLY_A
-0.747032	159	MHTHND_A

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}}$$

Attribute Selection Algorithms - Continued

GainRatioAttributeEval with Ranker

Ranked attributes:

0.697292	159	MHTHND_A
0.6864	160	MHTHDLY_A
0.632153	261	SOCSCLPAR_A
0.510287	136	TBIHLSBMC_A
0.506297	318	WTFA_A
0.505223	137	TBILCDCMG_A
0.50145	161	HOMEHC12M_A
0.593591	203	RXDG12M_A
0.590783	214	MEDNG12M_A
0.589613	215	MEDDL12M_A

[... omitted for length]

$$\mathrm{GainRatio(A)} = \mathrm{Gain(A)}/\mathrm{SplitInfo}_A\mathrm{(D)}$$

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$$

$$Gain(A) = Info(D) - Info_A(D)$$

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

Attribute Selection Algorithms - Continued

SymmetricUncertAttributeEval

- SymmetricUncertAttributeEval is an attribute evaluator based on the concept of Symmetrical Uncertainty (SU)
- Entropy is a measure of the uncertainty or randomness in data.
- Symmetrical uncertainty is a normalized form

of information gain, designed to remove biases towards attributes with many values.

0.564391	322	FAM_A
0.192102	318	WTFA_A
0.19129	159	MHTHND_A
0.088267	160	MHTHDLY_A
0.083587	261	SOCSCLPAR_A
0.078292	150	VIGIL4_A
0.09597	311	LSATIS4_A
0.0568936	156	DISCRIM3_A
0.064919	266	COGMEMDFF_A
0.063711	154	DISCRIM5_A
0.061754	153	VIGIL1_A

Model Selection

Decision Tables

- Decision tables are concise visual representations of which actions to perform based on a given dataset.
- Each row responds to a combination of feature values,
- Rules which are the specific values/conditions in which the decision is made
- This model is good for interpretation and creates rules that might be valuable for analysis.
- Not suitable for continuous variables and may be impractical for complex models.

Age Group	Income Level	Purchase (Yes/No)
Young	Low	No
Young	High	Yes
Middle-aged	Low	Yes
Middle-aged	High	Yes
Senior	Low	No
Senior	High	No

Model Selection - Continued

J48:

- This classifier is a subset of existing decision tree algorithms.
- It is an open-source Java implementation of the C4.5 decision tree algorithm.
- Uses a recursive process to build the tree.
- Uses information gain to measure how much a feature reduces the uncertainty for the class label.
- J48 is great for larger datasets as well as handling both categorical and continuous variables.
- Prone to overfitting and bias towards certain features due to the use of gain ratios.

Model Selection - Continued

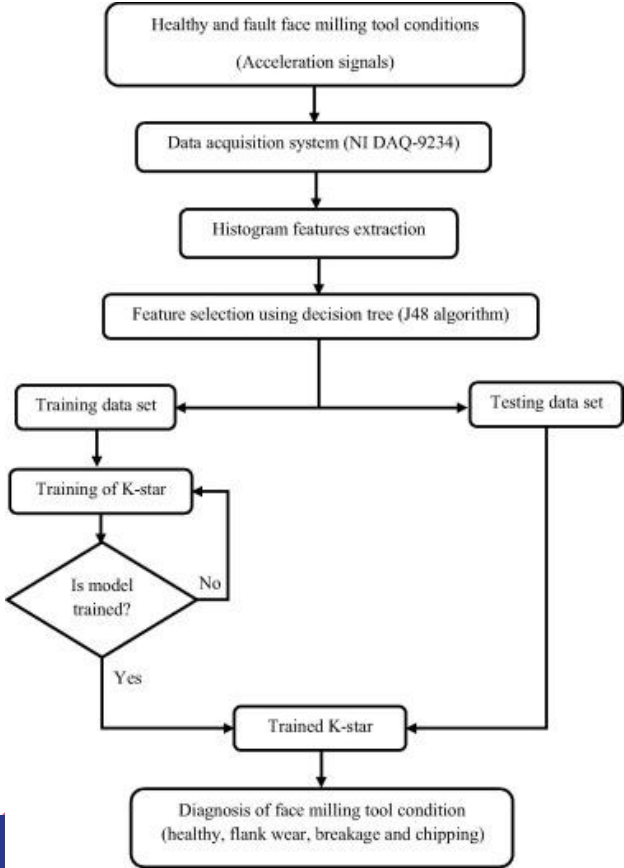
Bagging:

- Bagging involves training multiple models independently on different subsets of the data.
- The data will be randomly sampled an n amount of times with replacement.
- The model will train it on each of the data samples which create predictions.
- Models' predictions will be combined through simple averaging to make an overall prediction.
- Bagging can reduce variance and improve stability because it trains on multiple different models with different subsets of data.
- Increased computation because requires training of multiple models as well as not being useful for low variance models.

Model Selection - Continued

KStar

- Stores all the training data and makes predictions only when a new instance is classified.
- Makes decisions based on distances between new instances and the stored training data.
- Instances that are closer to the new instance have more influence and more weight on final prediction.
- K star is good for complex distributions of data as well as categorical and continuous data.
- Complex and may be slow for large datasets as well as being memory intensive because it saves all the training instances.



Why Not Others? (AdaBoostM1)

```
=== Summary ===

Correctly Classified Instances        5250               74.7225 %
Incorrectly Classified Instances      1776               25.2775 %
Kappa statistic                          0.5718
Mean absolute error                      0.2556
Root mean squared error                  0.3307
Relative absolute error                 98.8777 %
Root relative squared error             91.7491 %
Total Number of Instances             7026

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.922    0.232    0.600      0.922    0.727      0.623    0.845     0.574     -1
                 0.973    0.172    0.854      0.973    0.910      0.812    0.901     0.845     1
                 0.000    0.000    ?          0.000    ?          ?        0.712     0.176     2
                 0.000    0.000    ?          0.000    ?          ?        0.617     0.075     3
                 0.000    0.000    ?          0.000    ?          ?        0.607     0.078     4
Weighted Avg.    0.747    0.151    ?          0.747    ?          ?        0.833     0.614

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1773  151    0    0    0 |    a = -1
   95 3477    0    0    0 |    b = 1
  573  142    0    0    0 |    c = 2
  253  142    0    0    0 |    d = 3
  261  159    0    0    0 |    e = 4
```

Example Statistics

```
=== Summary ===

Correctly Classified Instances       18343               88.7636 %
Incorrectly Classified Instances      2322               11.2364 %
Kappa statistic                          0.8051
Mean absolute error                      0.074
Root mean squared error                  0.1876
Relative absolute error                 32.3704 %
Root relative squared error             55.5117 %
Total Number of Instances            20665

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.875    0.039    0.900      0.875    0.887      0.843    0.958     0.904     -1
                 0.919    0.093    0.932      0.919    0.926      0.824    0.963     0.970     1
                 0.856    0.034    0.694      0.856    0.766      0.748    0.960     0.747     2
                 0.635    0.007    0.717      0.635    0.674      0.666    0.963     0.638     3
                 0.639    0.007    0.645      0.639    0.642      0.635    0.962     0.631     4
Weighted Avg.    0.888    0.069    0.891      0.888    0.889      0.815    0.961     0.917

=== Confusion Matrix ===

     a     b     c     d     e   <-- classified as
  5199   459   182    52    47 |     a = -1
   453 11046   415    51    53 |     b = 1
    66   114  1458    38    27 |     c = 2
    24   153    25   380    16 |     d = 3
    37    79    22     9   260 |     e = 4
```

Summary of Accuracies

	Model Type				
Attribute Group	Decision Table	J48	Bagging	KStar	Averages
cfsEVAL	88.7636	88.9475	88.6422	86.0233	**88.09415**
reliefEVAL	88.0444	88.0159	86.2226	75.6618	84.486175
corrEVAL	85.8098	86.4098	85.8098	82.3086	85.0845
gainEVAL	88.6991	85.8098	85.8098	85.3829	86.4254
symmEVAL	88.9268	89.4677	86.2226	85.454	**87.517775**
Averages	**88.04874**	**87.73014**	86.5414	82.96612	

cfsEVAL

```
=== Summary ===

Correctly Classified Instances       18343               88.7636 %
Incorrectly Classified Instances      2322               11.2364 %
Kappa statistic                          0.8051
Mean absolute error                      0.074
Root mean squared error                  0.1876
Relative absolute error                 32.3704 %
Root relative squared error             55.5117 %
Total Number of Instances            20665

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.875    0.039    0.900      0.875    0.887      0.843    0.958     0.904     -1
                 0.919    0.093    0.932      0.919    0.926      0.824    0.963     0.970     1
                 0.856    0.034    0.694      0.856    0.766      0.748    0.960     0.747     2
                 0.635    0.007    0.717      0.635    0.674      0.666    0.963     0.638     3
                 0.639    0.007    0.645      0.639    0.642      0.635    0.962     0.631     4
Weighted Avg.    0.888    0.069    0.891      0.888    0.889      0.815    0.961     0.917

=== Confusion Matrix ===

    a     b     c     d     e   <-- classified as
 5199   459   182    52    47 |     a = -1
  453 11046   415    51    53 |     b = 1
   66   114  1458    38    27 |     c = 2
   24   153    25   380    16 |     d = 3
   37    79    22     9   260 |     e = 4
```

Decision Table

```
== Summary ===

orrectly Classified Instances       18381               88.9475 %
ncorrectly Classified Instances      2284               11.0525 %
appa statistic                          0.8059
ean absolute error                      0.0686
oot mean squared error                  0.1879
elative absolute error                 30.0186 %
oot relative squared error             55.5871 %
otal Number of Instances            20665

== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.877    0.041    0.895      0.877    0.886      0.841    0.946     0.865     -1
                0.932    0.109    0.922      0.932    0.927      0.824    0.950     0.955     1
                0.776    0.023    0.754      0.776    0.765      0.743    0.943     0.743     2
                0.664    0.009    0.694      0.664    0.679      0.669    0.945     0.630     3
                0.636    0.006    0.673      0.636    0.654      0.648    0.917     0.533     4
eighted Avg.    0.889    0.077    0.889      0.889    0.889      0.814    0.948     0.894

== Confusion Matrix ===

    a     b     c     d     e   <-- classified as
 5208   471   148    59    53 |     a = -1
  475 11195   244    65    39 |     b = 1
   72   250  1322    39    20 |     c = 2
   24   139    24   397    14 |     d = 3
   38    82    16    12   259 |     e = 4
```

J48

symmEVAL

```
=== Summary ===

Correctly Classified Instances        6248               88.9268 %
Incorrectly Classified Instances       778               11.0732 %
Kappa statistic                          0.806
Mean absolute error                      0.0759
Root mean squared error                  0.1883
Relative absolute error                 33.1669 %
Root relative squared error             55.6112 %
Total Number of Instances             7026

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.875    0.045    0.888      0.875    0.881      0.834    0.960     0.905     -1
                 0.932    0.107    0.923      0.932    0.928      0.827    0.964     0.972     1
                 0.806    0.022    0.765      0.806    0.785      0.766    0.965     0.782     2
                 0.620    0.008    0.705      0.620    0.660      0.652    0.966     0.634     3
                 0.606    0.007    0.676      0.606    0.639      0.633    0.970     0.615     4
Weighted Avg.    0.889    0.077    0.889      0.889    0.889      0.815    0.963     0.920

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1773  163   49   20   22 |    a = -1
  167 3795   83   16   10 |    b = 1
   25   63  462   13   10 |    c = 2
   20   46    7  124    3 |    d = 3
   11   44    3    3   94 |    e = 4
```

Decision Table

```
=== Summary ===

Correctly Classified Instances        6286               89.4677 %
Incorrectly Classified Instances       740               10.5323 %
Kappa statistic                          0.817
Mean absolute error                      0.0677
Root mean squared error                  0.1853
Relative absolute error                 29.6203 %
Root relative squared error             54.7355 %
Total Number of Instances             7026

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.874    0.039    0.900      0.874    0.887      0.842    0.951     0.874     -1
                 0.931    0.092    0.933      0.931    0.932      0.839    0.952     0.956     1
                 0.815    0.024    0.750      0.815    0.781      0.761    0.948     0.722     2
                 0.725    0.009    0.704      0.725    0.714      0.706    0.943     0.670     3
                 0.723    0.008    0.675      0.723    0.698      0.691    0.934     0.662     4
Weighted Avg.    0.895    0.067    0.896      0.895    0.895      0.826    0.951     0.899

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1772  152   54   23   26 |    a = -1
  160 3790   86   23   12 |    b = 1
   19   61  467   13   13 |    c = 2
    9   35    8  145    3 |    d = 3
    9   24    8    2  112 |    e = 4
```

J48

Winner - symmEVAL with J48

```
=== Summary ===

Correctly Classified Instances        6286               89.4677 %
Incorrectly Classified Instances       740               10.5323 %
Kappa statistic                          0.817
Mean absolute error                      0.0677
Root mean squared error                  0.1853
Relative absolute error                 29.6203 %
Root relative squared error             54.7355 %
Total Number of Instances             7026

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.874    0.039    0.900      0.874    0.887      0.842    0.951     0.874     -1
                 0.931    0.092    0.933      0.931    0.932      0.839    0.952     0.956     1
                 0.815    0.024    0.750      0.815    0.781      0.761    0.948     0.722     2
                 0.725    0.009    0.704      0.725    0.714      0.706    0.943     0.670     3
                 0.723    0.008    0.675      0.723    0.698      0.691    0.934     0.662     4
Weighted Avg.    0.895    0.067    0.896      0.895    0.895      0.826    0.951     0.899

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 1772  152   54   23   26 |    a = -1
  160 3790   86   23   12 |    b = 1
   19   61  467   13   13 |    c = 2
    9   35    8  145    3 |    d = 3
    9   24    8    2  112 |    e = 4
```

Recreating Our Model

1. Access preprocessed adult23_train+test.csv
2. Tranform to Nominal datatypes
3. Go to the “Select Attributes” tab and choose the correct class “engineered_attribute”
4. Select SymmetricUncertAttributeEval (Symmetric Uncertainty Attribute Evaluation Selection) as the Attribute Evaluator, and Ranker as the Search Method
5. Hit Start and wait for the program to finish
6. Take note of top 11 (threshold 0.55) features; keep the index values for these features
7. Go back to the Preprocess tab and click Remove other attributes except this
8. Click on the Classify tab and click “Percentage Split” under Test Options, write 70%
9. Select the J48 model under trees
10. Click start and wait for it to complete

So, which features are important?

cfsEVAL	reliefEVAL	corrEVAL	gainEVAL	symmEVAL
DISAB3_A	HRTESTLAST_A	SOCSCLPAR_A	**MHTHND_A**	FAM_A
MLTFAMFLG_A	EMPDYSMSS3_A	COGMEMDFF_A	**MHTHDLY_A**	WTFA_A
EVRMARRIED_A	URBRRL	LSATIS4_A	SOCSCLPAR_A	**MHTHND_A**
SMKNOW_A	REGION	PHSTAT_A	TBIHLSBMC_A	**MHTHDLY_A**
PAITOOTH3M_A	PSTRAT	SOCERRNDS_A	WTFA_A	SOCSCLPAR_A
VIGIL4_A	LONGCOVD1_A	DISCRIM5_A	FAM_A	VIGIL4_A
DISCRIM5_A	VIGIL4_A	VIGIL4_A	TBILCDCMG_A	LSATIS4_A
MHTHND_A	SMKNOW_A	**MHTHDLY_A**	HOMEHC12M_A	DISCRIM3_A
MHTHDLY_A	INTV_QRT	**MHTHND_A**	RXDG12M_A	COGMEMDFF_A
HYSTEV	VIGIL3_A		MEDNG12M_A	DISCRIM5_A
			MEDDL12M_A	VIGIL1_A

Delay Due to Cost

Attribute	Description
MHTHND_A	Needed counseling, therapy but did not get it due to cost, past 12 months
MHTHDLY_A	Delayed counseling, therapy due to cost, past 12 months
MEDNG12M_A	Needed medical care but did not get it due to cost, past 12 months
MEDDL12M_A	Delayed medical care due to cost, past 12 months

Health

Attribute	Description
DISAB3_A	The Washington Group Short Set Composite Disability Indicator
SMKNOW_A	Now smoke cigarettes
PAITOOTH3M_A	Toothache or jaw pain
HRTESTLAST_A	A How long since hearing test
LONGCOVD1_A	Had COVID-19 symptoms for 3 or more months
SMKNOW_A	Now smoke cigarettes
PHSTAT_A	General health status
TBIHLSBMC_A	Headache, sensitivities, balance problems or mood change, past 12 months
WTFA_A	Weight - Final Annual

Personal Specifics

Attribute	Description
VIGIL4_A	Avoid certain situations and places
DISCRIM5_A	You are threatened or harassed
VIGIL3_A	Watch what you say and how you say it
SOCSCLPAR_A	Language socially
COGMEMDFF_A	Difficulty remembering/concentrating
SOCERRNDS_A	Difficulty doing errands alone
TBILCDCMG_A	A Lost consciousness, dazed or confused, or had gap in memory, past 12 months
VIGIL1_A	Prepare for possible insults before leaving home

Family Structure

Attribute	Description
FAM_A	Number of Emergency Contacts
HOMEHC12M_A	Received care at home, past 12 months
MLTFAMFLG_A	Indicator for multifamily households
EVRMARRIED_A	Sample adult has ever been married

Odd Ones Out

Attribute	Description
EMPDYSMSS3_A	Days missed work, past 12 months (top-coded)
LONGCOVD1_A	Had COVID-19 symptoms for 3 or more months

Discussion Part 1

Some attributes don’t fit into the primary category, such as:

- Having COVID-19 for three or more months.
- Days missed at work in the past 12 months.

Data is from a 2023 survey, a period where COVID-19 effects were still relevant.

- This could introduce bias toward these attributes being significant.

A future study with more recent data could better assess COVID-19's impact on mental health.
These attributes are NOT included in the SymmetricUncertAttributeEval attribute group, meaning:

- The final model selected doesn't include these attributes.
- The model could be more generalizable to non-COVID-19 periods, but external testing is needed to confirm.

Discussion Part 2

- General trends indicate that factors contributing to mental health severity include:
    a. Health
    b. Financial status
    c. Family support
    d. Internal thoughts
- Three of these factors can be assessed in non-psychological settings:
    a. Hospitals can check a patient's family support, general health, and how long they waited to seek care.
    b. These checks can guide recommendations for psychological evaluations, ensuring timely care.
- As mental health becomes more accepted, this information can help reduce barriers to support for vulnerable individuals despite lingering taboos.

Conclusions

- Limitations:
    a. Year of data collection
    b. Uneven class distribution

- Great model performance 88% accuracy. This is well suited for medical application

Any Questions?