

# Analysis - Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline

Salim Salmi

12/17/2020

## Contents

<b>Introduction</b>	<b>1</b>
<b>Data Files</b>	<b>1</b>
<b>Expert labeling</b>	<b>2</b>
Confusion Matrix . . . . .	2
Multilevel analysis . . . . .	3
No support and Expert comments . . . . .	3
No support and Support system . . . . .	3
Expert comments and Support system . . . . .	3
<b>Counsellor perception of utility</b>	<b>4</b>
<b>Evaluation of Algorithm</b>	<b>10</b>
<b>R version information</b>	<b>11</b>

## Introduction

This document presents statistical analyses, for a recommendation support system used in suicide crisis counseling, of recommendation performance, noticeable difference in counselor reactions, and the system's perceived utility by counselors as reported in the paper:

*Content-Based Recommender Support System for Counselors in a Suicide Prevention Chat Helpline: Design and Evaluation Study*

## Data Files

Data files are stored in a CSV file format.

Table 1: Fields and description from csv file expert-data.csv

variable	description
expert	the ID of the expert judge, with 8 different experts
chat	the chat ID, of the chat transcript the counselor received, with 3 different possible chats

variable	description
participant	the participating counselor who was labeled by the expert, 24 participating counselors
expert_label	the label given by the expert
true_label	the support condition

Table 2: Fields and description from csv file utility-data.csv

variable	description
user	the ID of the counselor, with 24 different participants
value	the rating (-3 to 3) the user gave of the perceived utility
condition	the support condition
chat	the chat ID, of the chat transcript the counselor received, with 3 different possible chats

Table 3: Fields and description from csv file algorithm-data.csv

variable	description
user	the ID of the counselor, with 24 different participants
chat	the chat ID, of the chat transcript the counselor received, with 3 different possible chats
rating	the rating (1-7) the user gave of the relatedness of a recommendation to the chat
condition	one of two types of recommendation: either randomly selected from the corpus or found by the algorithm to be semantically highly related to the chat.

## Expert labeling

### Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  1  2  3
##           1 74 66 85
##           2 65 76 46
##           3 39 36 47
##
## Overall Statistics
##
##           Accuracy : 0.3689
##           95% CI : (0.3279, 0.4114)
```

```
##      No Information Rate : 0.3333
##      P-Value [Acc > NIR] : 0.0455712
##
##              Kappa : 0.0534
##
## Mcnemar's Test P-Value : 0.0003829
##
## Statistics by Class:
##
##              Class: 1 Class: 2 Class: 3
## Sensitivity          0.4157   0.4270   0.26404
## Specificity          0.5758   0.6882   0.78933
## Pos Pred Value       0.3289   0.4064   0.38525
## Neg Pred Value       0.6634   0.7061   0.68204
## Prevalence           0.3333   0.3333   0.33333
## Detection Rate       0.1386   0.1423   0.08801
## Detection Prevalence 0.4213   0.3502   0.22846
## Balanced Accuracy     0.4958   0.5576   0.52669
```

## Multilevel analysis

Multilevel analysis with the expert label predicting the condition. Random effects for experts and counselors were used.

### No support and Expert comments

Table 4: Anova

	Chisq	Df	Pr(>Chisq)
<b>expert_label</b>	1.433	2	0.4885

### No support and Support system

Table 5: Anova

	Chisq	Df	Pr(>Chisq)
<b>expert_label</b>	4.722	2	0.09434

### Expert comments and Support system

Table 6: Anova

	Chisq	Df	Pr(>Chisq)
<b>expert_label</b>	11.05	2	0.003988

Summary of model for the significant result:

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: true_label ~ expert_label + (1 | expert) + (1 | participant)
```

```
## Data: counsellor_difference.data
##
##      AIC      BIC   logLik deviance df.resid
##    492.2    511.6   -241.1    482.2     351
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.1426 -1.1348  0.0486  0.8812  1.2854
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
## participant (Intercept) 0          0
## expert      (Intercept) 0          0
## Number of obs: 356, groups: participant, 24; expert, 8
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.25300    0.16406   1.542  0.12305
## expert_label2 -0.75509    0.24862  -3.037  0.00239 **
## expert_label3  0.01363    0.27563   0.049  0.96055
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) expr_2
## expert_lbl2 -0.660
## expert_lbl3 -0.595  0.393
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

(Intercept)	expert_label2	expert_label3
0.253	-0.7551	0.01363

## Counsellor perception of utility

T test results for support system utility score:

```
##
## One Sample t-test
##
## data: generated$value
## t = -0.67905, df = 23, p-value = 0.5039
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.8430015  0.4263348
## sample estimates:
## mean of x
## -0.2083333
```

T test results for senior counselor advice utility score:

```
##
## One Sample t-test
##
```

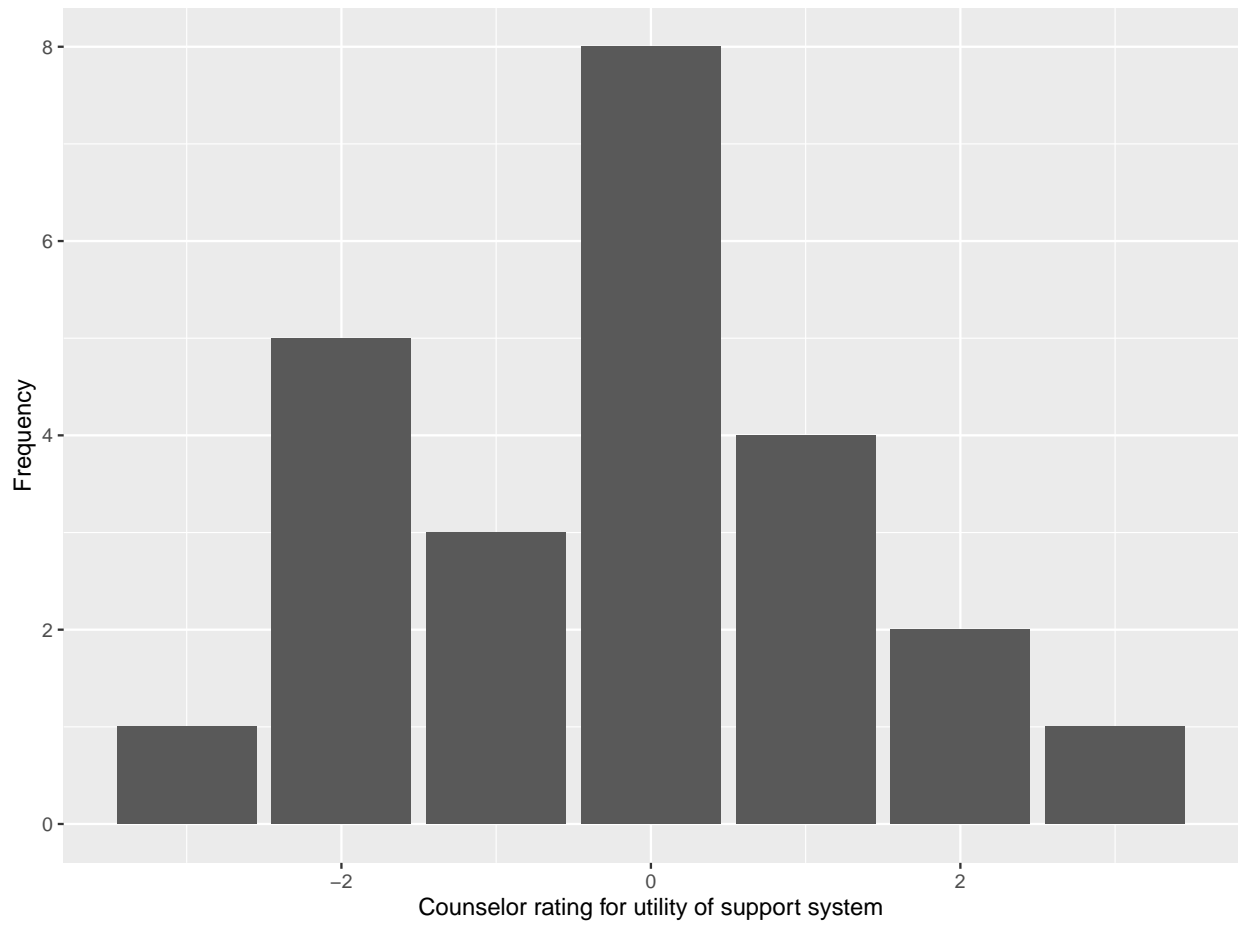
```
## data: expert$value
## t = 5.1678, df = 23, p-value = 3.079e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 0.8745661 2.0421005
## sample estimates:
## mean of x
## 1.458333
```

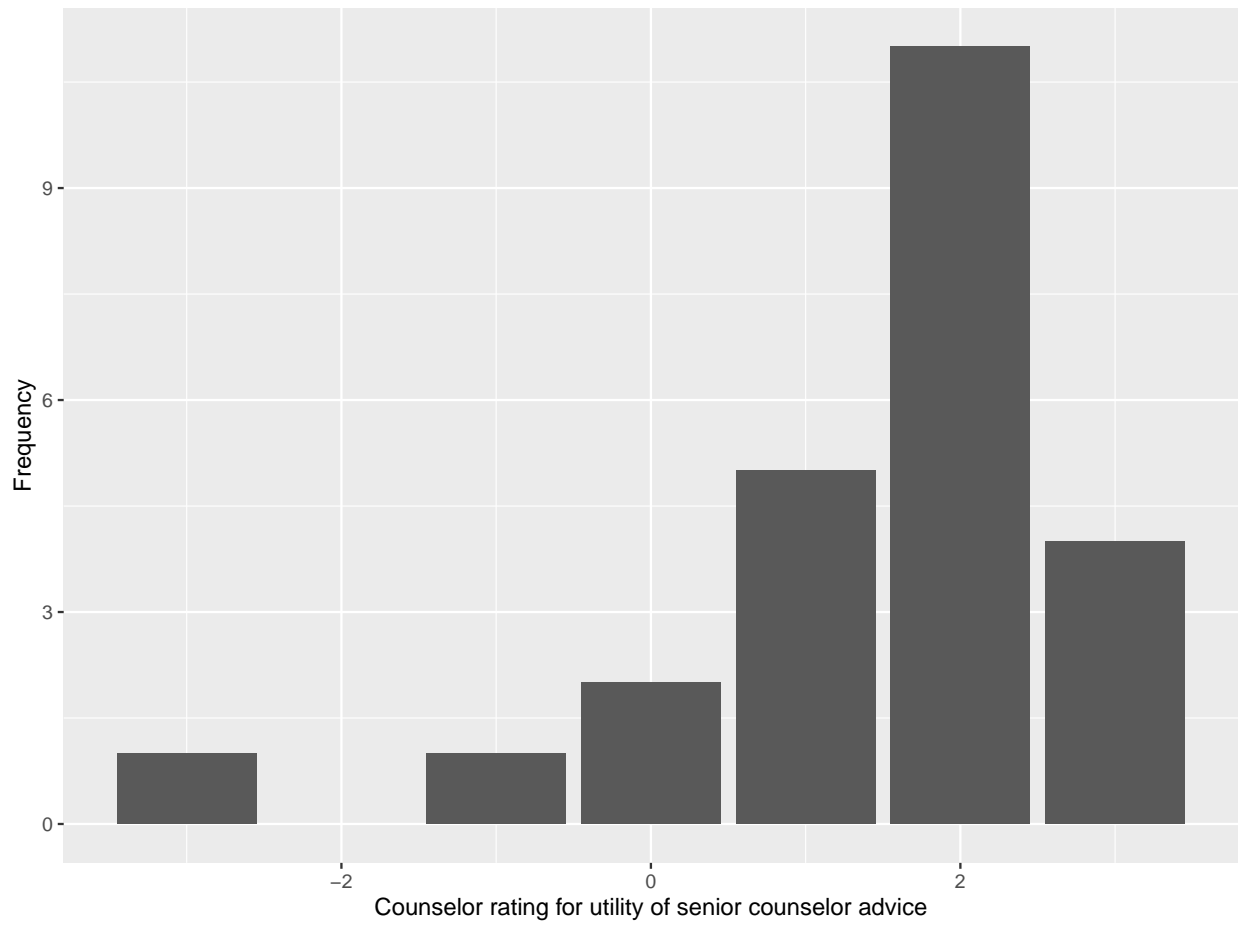
T test results for no support utility score:

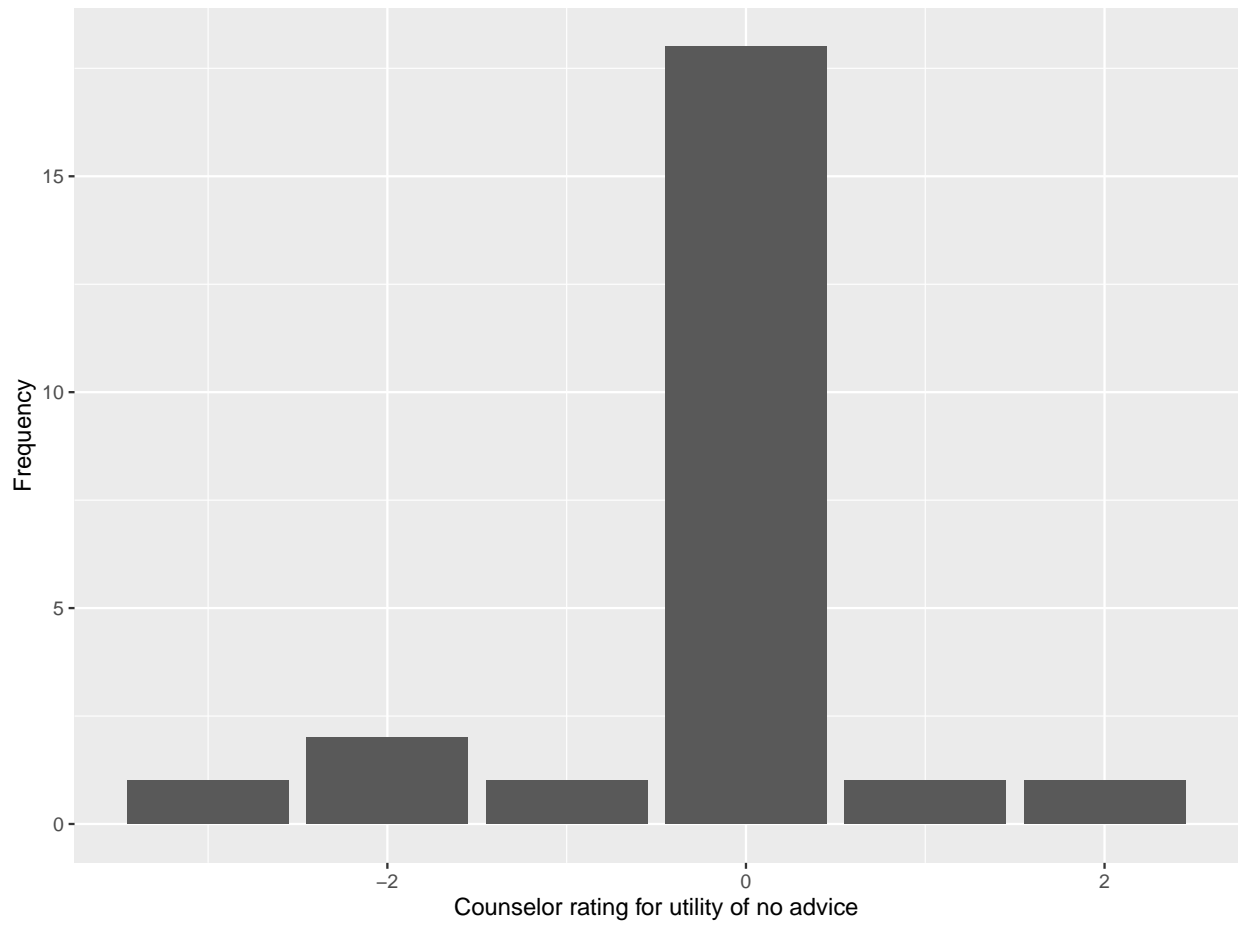
```
##
## One Sample t-test
##
## data: none$value
## t = -1.0445, df = 23, p-value = 0.3071
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.6209234 0.2042567
## sample estimates:
## mean of x
## -0.2083333
```

Histograms for of counselor utility scores fo each condition:

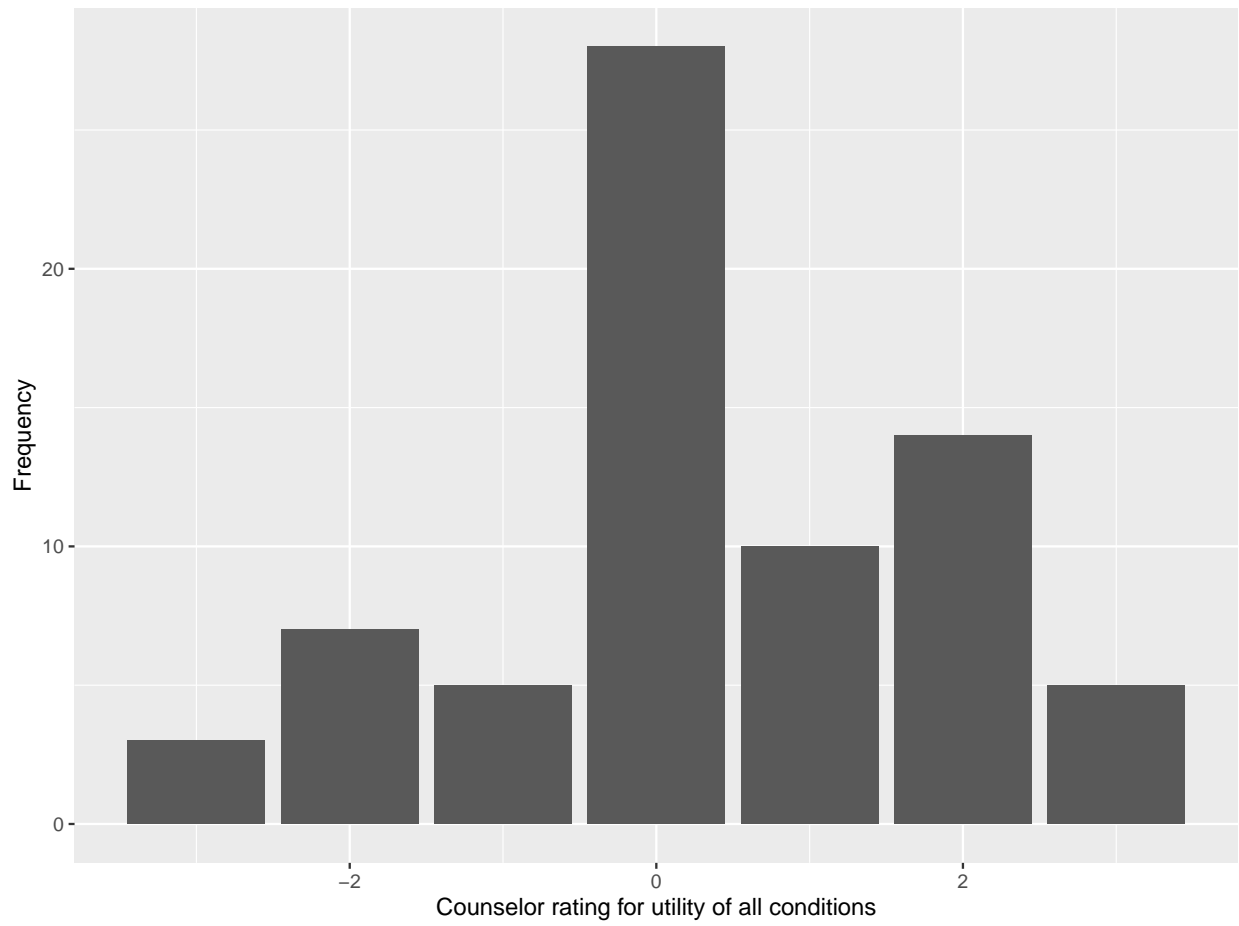
```
##
## F test to compare two variances
##
## data: value by condition
## F = 2.3662, num df = 23, denom df = 23, p-value = 0.04411
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 1.023612 5.469859
## sample estimates:
## ratio of variances
## 2.366224
```



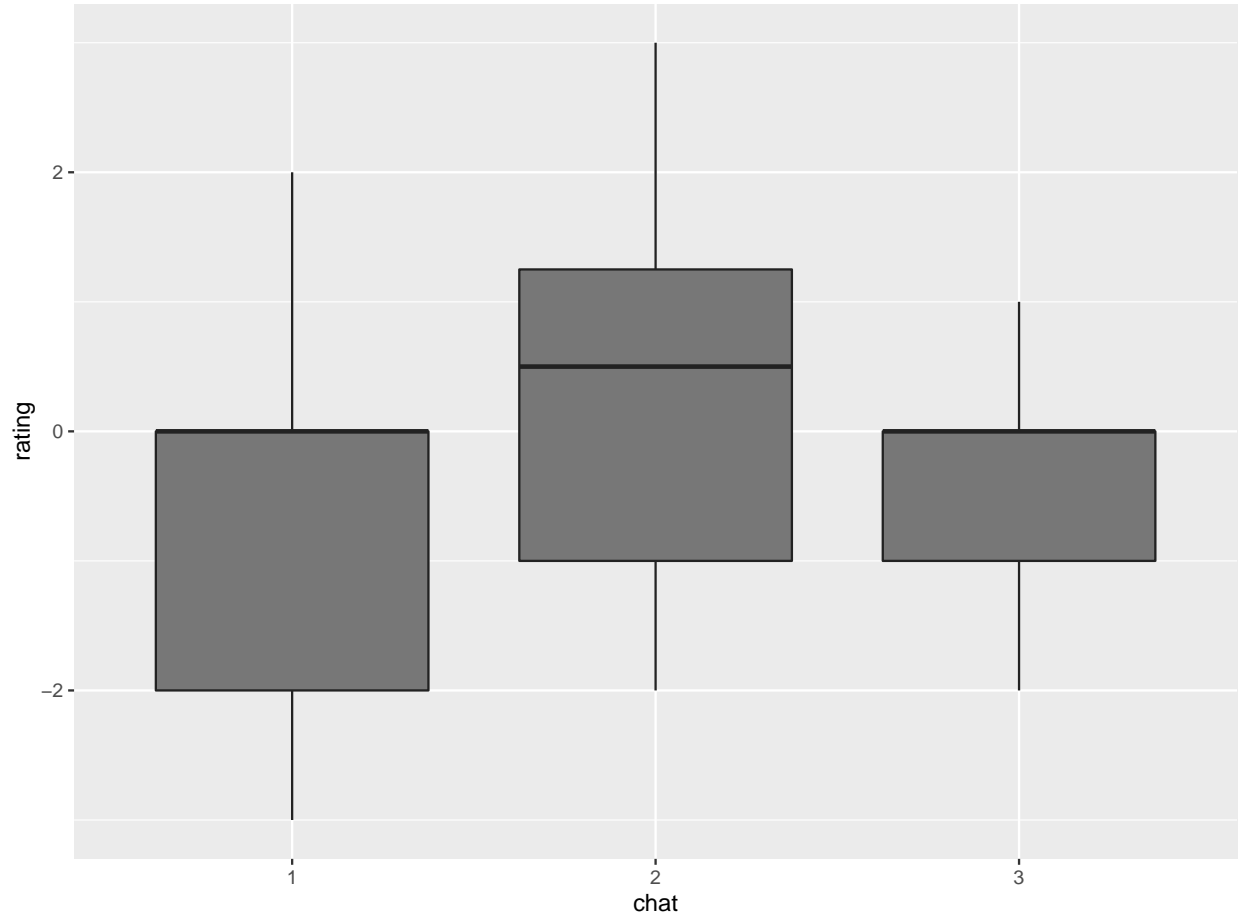








Box plot for visual comparison of difference in rating between chats for the support system condition.



## Evaluation of Algorithm

Multilevel analysis on counselor level of agreement on relatedness of the problem found in a chat segment and an ongoing chat, for the chat segments that were randomly selected from the corpus and chat segments based on semantic similarity.

Anova of the multilevel model:

Table 8: Multilevel model for prediction user rating of relevance  
(continued below)

	npar	AIC	BIC	logLik	deviance	Chisq	Df
<b>algorithm.null</b>	5	2758	2781	-1374	2748	NA	NA
<b>algorithm.model</b>	6	2671	2698	-1329	2659	89.1	1

	Pr(>Chisq)
<b>algorithm.null</b>	NA
<b>algorithm.model</b>	3.758e-21

The beta values of the model:

```
##               chat2               chat3 conditionAlgorithm
##            -0.05888024          -0.02888465          0.30322570
```

## R version information

This analysis has been compiled with following R version.

**R version 4.0.3 (2020-10-10)**

**Platform:** x86\_64-w64-mingw32/x64 (64-bit)

**locale:** LC\_COLLATE=English\_United\_States.1252, LC\_CTYPE=English\_United\_States.1252, LC\_MONETARY=English\_United\_States.1252, LC\_NUMERIC=C and LC\_TIME=English\_United\_States.1252

**attached base packages:** stats, graphics, grDevices, utils, datasets, methods and base

**other attached packages:** caret(v.6.0-86), lattice(v.0.20-41), sjPlot(v.2.8.6), car(v.3.0-10), carData(v.3.0-4), pander(v.0.6.3), ggplot2(v.3.3.2), lme4(v.1.1-26) and Matrix(v.1.2-18)

**loaded via a namespace (and not attached):** nlme(v.3.1-149), lubridate(v.1.7.9.2), insight(v.0.11.1), tools(v.4.0.3), backports(v.1.2.0), R6(v.2.5.0), sjlabelled(v.1.1.7), rpart(v.4.1-15), colorspace(v.2.0-0), nnet(v.7.3-14), withr(v.2.3.0), tidyselct(v.1.1.0), emmeans(v.1.5.3), curl(v.4.3), compiler(v.4.0.3), performance(v.0.6.1), sandwich(v.3.0-0), labeling(v.0.4.2), bayestestR(v.0.8.0), scales(v.1.1.1), mvtnorm(v.1.1-1), stringr(v.1.4.0), digest(v.0.6.27), foreign(v.0.8-80), minqa(v.1.2.4), rmarkdown(v.2.6), rio(v.0.5.16), pkgconfig(v.2.0.3), htmltools(v.0.5.0), rlang(v.0.4.9), readxl(v.1.3.1), farver(v.2.0.3), generics(v.0.1.0), zoo(v.1.8-8), ModelMetrics(v.1.2.2.2), dplyr(v.1.0.2), zip(v.2.1.1), magrittr(v.2.0.1), parameters(v.0.10.1), Rcpp(v.1.0.5), munsell(v.0.5.0), abind(v.1.4-5), lifecycle(v.0.2.0), pROC(v.1.16.2), stringi(v.1.5.3), multcomp(v.1.4-15), yaml(v.2.2.1), MASS(v.7.3-53), plyr(v.1.8.6), recipes(v.0.1.15), grid(v.4.0.3), sjmisc(v.2.8.5), forcats(v.0.5.0), crayon(v.1.3.4), ggeffects(v.1.0.1), haven(v.2.3.1), splines(v.4.0.3), sjstats(v.0.18.0), hms(v.0.5.3), knitr(v.1.30), pillar(v.1.4.7), boot(v.1.3-25), estimability(v.1.3), effectsize(v.0.4.1), stats4(v.4.0.3), reshape2(v.1.4.4), codetools(v.0.2-16), glue(v.1.4.2), evaluate(v.0.14), data.table(v.1.13.4), modelr(v.0.1.8), vctrs(v.0.3.5), nloptr(v.1.2.2.2), foreach(v.1.5.1), cellranger(v.1.1.0), gtable(v.0.3.0), purrr(v.0.3.4), tidyr(v.1.1.2), xfun(v.0.19), openxlsx(v.4.2.3), gower(v.0.2.2), prodlim(v.2019.11.13), xtable(v.1.8-4), broom(v.0.7.3), e1071(v.1.7-4), class(v.7.3-17), survival(v.3.2-7), timeDate(v.3043.102), tibble(v.3.0.4), iterators(v.1.0.13), lava(v.1.6.8.1), statmod(v.1.4.35), TH.data(v.1.0-10), ellipsis(v.0.3.1) and ipred(v.0.9-9)