Creating Model for Training Data

Angela Zhao

2020-10-27

Contents


```
# Libraries
library(tidyverse)
library(here)
library(cvTools)
library(coefplot)
library(glmnet)
library(GGally)
library(ROCR)

file_raw_train_data <-
  here("data/train_data_speed_dating.csv")

recode_race <-
  c(
    "1" = "Black",
    "2" = "White",
    "3" = "Hispanic",
    "4" = "API",
```

```
    "5" = "Indigenous",
    "6" = "Other"
  )


recode_goal <-
  c(
    "1" = 'Seemed like a fun night out',
    "2" = "To meet new people",
    "3" = "To get a date",
    "4" = "Looking for a serious relationship",
    "5" = "To say I did it",
    "6" = "Other"
  )


recode_date <- c(
  "1" = 'Several times a week',
  "2" = "Twice a week",
  "3" = "Once a week",
  "4" = "Twice a month",
  "5" = "Once a month",
  "6" = "Several times a year",
  "7" = "Almost never"
)


recode_go_out <-
  c(
    "1" = 'Several times a week',
    "2" = "Twice a week",
    "3" = "Once a week",
    "4" = "Twice a month",
    "5" = "Once a month",
    "6" = "Several times a year",
    "7" = "Almost never"

  )
```

Reading and cleaning Data

```
train <-
  file_raw_train_data %>%
  read_csv()
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   positin1 = col_logical(),
##   field = col_character(),
```

```
##   undergra = col_logical(),
##   mn_sat = col_logical(),
##   tuition = col_logical(),
##   from = col_character(),
##   career = col_character(),
##   attr4_1 = col_logical(),
##   sinc4_1 = col_logical(),
##   intel4_1 = col_logical(),
##   fun4_1 = col_logical(),
##   amb4_1 = col_logical(),
##   shar4_1 = col_logical(),
##   attr5_1 = col_logical(),
##   sinc5_1 = col_logical(),
##   intel5_1 = col_logical(),
##   fun5_1 = col_logical(),
##   amb5_1 = col_logical(),
##   attr1_s = col_logical(),
##   sinc1_s = col_logical()
##   # ... with 56 more columns
## )
## i Use `spec()` for the full column specifications.
```

```r
# Added in the ratings of partner rating you and you rating the partner
# diff is the avg of partner rating you - avg of you rating the partner

train <-
  train %>%
  mutate_at(
    vars(
      career_c,
      field_cd,
      race,
      race_o,
      zipcode,
      go_out,
      date,
      goal,
      met,
      met_o,
      samerace,
      gender,
      match
    ),
    as.character
  ) %>%
  mutate(
    field_cd =
      case_when(
        field_cd %in% c("2", "4", "5", "10") ~ "STEM",
        field_cd %in% c("1", "3", "8", "9", "11", "13") ~ "Social_Science",
        field_cd %in% c("6", "7", "14", "15", "16", "17") ~ "Humanities",
        TRUE ~ "Other"
      ),
    race = recode(race, !!! recode_race),
```

```
    goal = recode(goal, !!! recode_goal),
    met = if_else(met == "2", "0", met),
    met_o = if_else(met_o == "2", "0", met_o),
    race_o = recode(race_o, !!! recode_race),
    gender = if_else(gender == "0", "female", "male"),
    career_c =
      case_when(
        career_c %in% c("3", "4", "5") ~ "STEM",
        career_c %in% c("1", "7", "8", "9", "10", "11", "12", "13", "17") ~ "Social_Science",
        career_c %in% c("6") ~ "Humanities",
        TRUE ~ "Other"
      )
  ) %>%
  mutate_at(
    vars(
      career_c,
      field_cd,
      race,
      race_o,
      goal,
      met,
      met_o,
      gender
    ),
    as.factor
  )
```

Normalize the preferences, if out of 100, to percentages instead.

```
train <-
  train %>%
  rowwise() %>%
  mutate(
    sum_2_1 = sum(c(attr2_1, sinc2_1, intel2_1, fun2_1, amb2_1, shar2_1)),
    sum_4_1 = sum(c(attr4_1, sinc4_1, intel4_1, fun4_1, amb4_1, shar4_1)),
  ) %>%
  mutate_at(
    vars(attr2_1, sinc2_1, intel2_1, fun2_1, amb2_1, shar2_1),
    ~ . / sum_2_1
  ) %>%
  mutate_at(
    vars(attr4_1, sinc4_1, intel4_1, fun4_1, amb4_1, shar4_1),
    ~ . / sum_4_1
  ) %>%
  dplyr::select(-c(sum_2_1, sum_4_1))

train <-
  train %>%
  rowwise() %>%
  mutate(
    rating = sum(c(attr, sinc, intel, fun, amb, shar)),
    avg_partner_rating =
```

```
      mean(c(attr, sinc, intel, fun, amb, shar), na.rm = TRUE) / 10
  ) %>%
  mutate_at(
    vars(attr, sinc, intel, fun, amb, shar),
    ~ . / rating
  ) %>%
  select(-rating)
```

Choosing Variables

```
# Selecting only the variables that determine perception of the round of speed

# Field and career are too unevenly entered, so best to use the coded versions instead.

vars_wanted <-
  train %>%
  select(
    rowname,
    gender,
    condtn,
    match:race_o,
    like_o:age,
    field_cd,
    race:imprelig,
    goal:go_out,
    career_c:exphappy,
    dec:met,
    avg_partner_rating
  ) %>%
  select(-c(met, met_o, dec)) %>%
  summarize_all(~ sum(is.na(.))) %>%
  summarize_all(~ sum(.)) %>%
  mutate_all(~ . / 4258) %>%
  pivot_longer(
    cols = everything(),
    names_to = "var",
    values_to = "percent_missing"
  ) %>%
  filter(percent_missing < 0.5)

vars_wanted <-
  vars_wanted %>%
  arrange(desc(percent_missing)) %>%
  filter(var != "income") %>%
  pull(var)

all_training_data <-
  train %>%
  select(rowname, all_of(vars_wanted))
```

Rescaling 1-10 scale to 0-1 scale

```
all_training_data <-
  all_training_data %>%
  mutate_at(
    vars(prob, prob_o, like, like_o, imprace, imprelig, sports:exphappy),
    ~ . / 10
  ) %>%
  mutate(
    match = as.double(match)
  )


cor_data <-
  all_training_data %>%
  drop_na()
```

Checking for correlations

```
# Convert data to numeric
corr <-
  cor_data %>%
  mutate_all(as.integer)

# ggcorr(corr,
#     method = c("pairwise", "spearman"),
#     nbreaks = 6,
#     hjust = 0.8,
#     label = TRUE,
#     label_size = 3,
#     color = "grey50")
```

Create a train and a validate set

```
random_rows <-
  all_training_data %>%
  pull(rowname) %>%
  as_tibble %>%
  sample_frac(0.8) %>%
  mutate(
    value = as.character(value)
  ) %>%
  pull(value)

train_model_data <-
  all_training_data %>%
  as_tibble() %>%
  filter(rowname %in% random_rows)
```

```
test_model_data <-
  all_training_data %>%
  anti_join(train_model_data, by = "rowname") %>%
  select(-rowname)


lasso_model_data <-
  train_model_data %>%
  select(-rowname) %>%
  mutate_at(vars(condtn), as.factor) %>%
  drop_na()
```

Using lasso to weed out extraneous vars

```
# convert the data to a design matrix
X = model.matrix(match ~ 0 + ., lasso_model_data)
Y = lasso_model_data %>% pull(match)


# standardize
X = scale(X)
Y = scale(Y)

set.seed(1244)

# create training and test set
train.ind = sample(nrow(X), round(nrow(X)/2))
X.train = X[train.ind,]
X.test = X[-train.ind,]
Y.train = Y[train.ind]
Y.test = Y[-train.ind]

# set lambda sequence to use for lasso and ridge
lambdas = 10^seq(-2,1.5,0.1)

#lasso
fm.lasso = glmnet(X.train, Y.train, alpha = 1, lambda = lambdas, thresh = 1e-12)

fm.lasso %>%
  summary()
```

```
##           Length Class     Mode
## a0          36   -none-    numeric
## beta      2484   dgCMatrix S4
## df          36   -none-    numeric
## dim          2   -none-    numeric
## lambda      36   -none-    numeric
## dev.ratio   36   -none-    numeric
## nulldev      1   -none-    numeric
## npasses      1   -none-    numeric
## jerr         1   -none-    numeric
```

```
## offset       1   -none-    logical
## call         6   -none-    call
## nobs         1   -none-    numeric
```

```
cv <- cv.glmnet(X.train, Y.train)

cv$lambda.min
```

```
## [1] 0.02421449
```

```
covars_wanted <-
  extract.coef(cv) %>%
  select(coeff = Coefficient) %>%
  filter(coeff != "(Intercept)") %>%
  pull(coeff)


covars_wanted <-
  covars_wanted %>%
  as_tibble() %>%
  filter(value %in% colnames(train)) %>%
  pull(value)

covars_wanted
```

```
##  [1] "attr"               "sinc"               "prob"
##  [4] "prob_o"             "like"               "like_o"
##  [7] "avg_partner_rating" "imprace"            "tvsports"
## [10] "art"                "gaming"             "clubbing"
## [13] "tv"                 "shopping"           "age"
```

```
# train_model_data <-
#   train_model_data %>%
#   select(match, all_of(covars_wanted))
```

Data used

```
# Add in field_cd and career_c b/c rep as factors in covars_wanted
train_model_data <-
  train_model_data %>%
  select(match, attr, sinc, fun, prob, prob_o, like, like_o, avg_partner_rating, imprace, tvsports, clu
```

Functions

Accuracy

```
accuracy <- function(table_mat) {
  sum(diag(table_mat)) / sum(table_mat)
}
```

Precision and

```
precision <- function(matrix) {

    tp <- matrix[2, 2]

    fp <- matrix[1, 2]
    return(tp / (tp + fp))
}

recall <- function(matrix) {
    tp <- matrix[2, 2]
    fn <- matrix[2, 1]
    return(tp / (tp + fn))
}
```

Linear Regression

With all lasso covariates

```
m1 <-
  glm(
    match ~ .,
    data = train_model_data,
    family = 'binomial'
  )



predict <- predict(m1, test_model_data, type = 'response')
# confusion matrix
table_mat <- table(test_model_data$match, predict > 0.5)
table_mat
```

```
##
##     FALSE TRUE
##   0   551   13
##   1    93   33
```

```
accuracy(table_mat)
```

```
## [1] 0.8463768
```

```
precision(table_mat)
```

```
## [1] 0.7173913
```

```
recall(table_mat)
```

```
## [1] 0.2619048
```

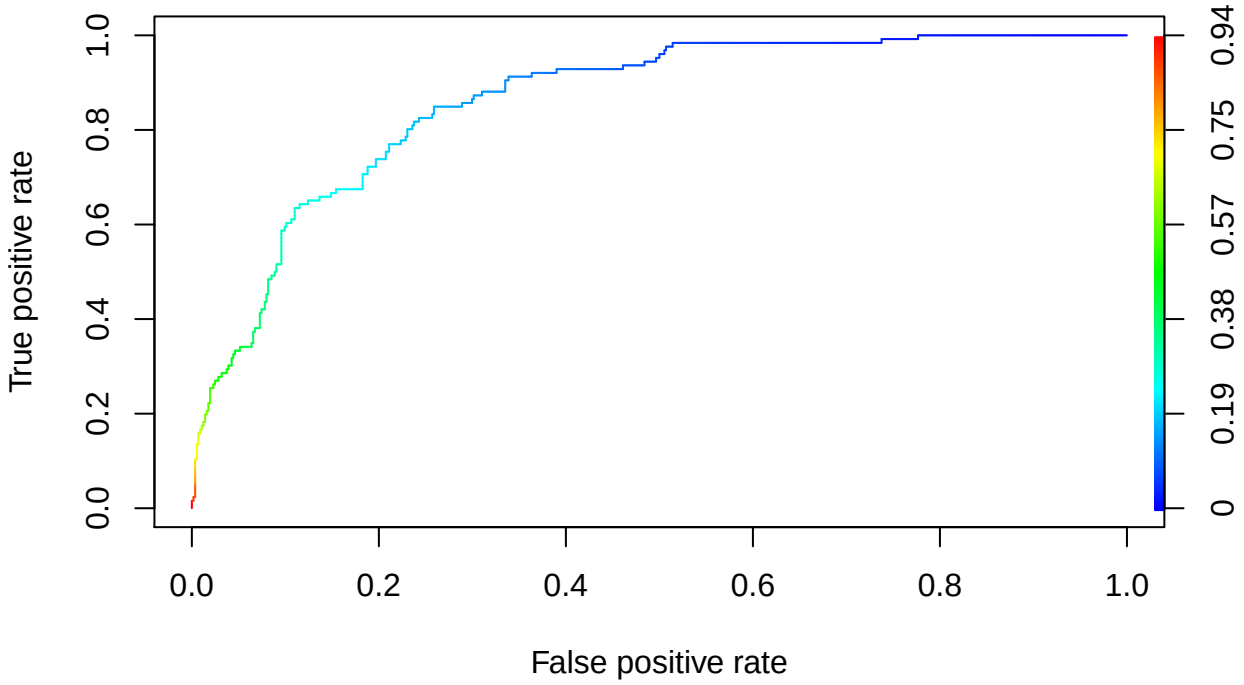
ROC curve

```
index <-
  predict[!is.na(predict)] %>%
  as.data.frame() %>%
  add_rownames() %>%
  pull(rowname)
```

```
## Warning: ‘add_rownames()‘ is deprecated as of dplyr 1.0.0.
## Please use ‘tibble::rownames_to_column()‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_warnings()‘ to see where this warning was generated.
```

```
t <-
  test_model_data %>%
  add_rownames() %>%
  filter(rowname %in% index)


ROCRpred <- prediction(predict[!is.na(predict)], t$match)
ROCRperf <- performance(ROCRpred, 'tpr', 'fpr')
plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2, 1.7))
```



Creating a model with interaction terms

```
m2 <-
  glm(
    match ~ . + like:avg_partner_rating + attr:like_o,
    data = train_model_data,
    family = 'binomial'
  )

m2 %>%
  summary()
```

```
##
## Call:
## glm(formula = match ~ . + like:avg_partner_rating + attr:like_o,
##     family = "binomial", data = train_model_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0245  -0.5810  -0.2958  -0.0780   3.3629
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -15.09438    2.40489  -6.277 3.46e-10 ***
```

```
## attr                    -16.43536    8.36017  -1.966  0.04931 *
## sinc                     -3.05433    2.44250  -1.250  0.21112
## fun                       4.53784    2.53057   1.793  0.07294 .
## prob                      1.65875    0.35865   4.625 3.75e-06 ***
## prob_o                    1.00770    0.32328   3.117  0.00183 **
## like                     14.29050    2.46943   5.787 7.17e-09 ***
## like_o                   -0.55933    1.94250  -0.288  0.77339
## avg_partner_rating       13.15929    2.59113   5.079 3.80e-07 ***
## imprace                  -0.58554    0.22462  -2.607  0.00914 **
## tvsports                 -0.58610    0.23247  -2.521  0.01170 *
## clubbing                  0.54193    0.25425   2.131  0.03305 *
## movies                   -0.12667    0.37313  -0.339  0.73425
## shopping                 -0.32508    0.24782  -1.312  0.18959
## age_o                     0.01314    0.01820   0.722  0.47047
## like:avg_partner_rating -15.52744    3.33878  -4.651 3.31e-06 ***
## attr:like_o              33.20009   11.73776   2.828  0.00468 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2495.3  on 2721  degrees of freedom
## Residual deviance: 1837.3  on 2705  degrees of freedom
##   (684 observations deleted due to missingness)
## AIC: 1871.3
##
## Number of Fisher Scoring iterations: 7
```

```
predict <- predict(m2, test_model_data, type = 'response')
# confusion matrix
table_mat <- table(test_model_data$match, predict > 0.5)
table_mat
```

```
##
##     FALSE TRUE
##   0   549   15
##   1    95   31
```

```
accuracy(table_mat)
```

```
## [1] 0.8405797
```

```
precision(table_mat)
```

```
## [1] 0.673913
```

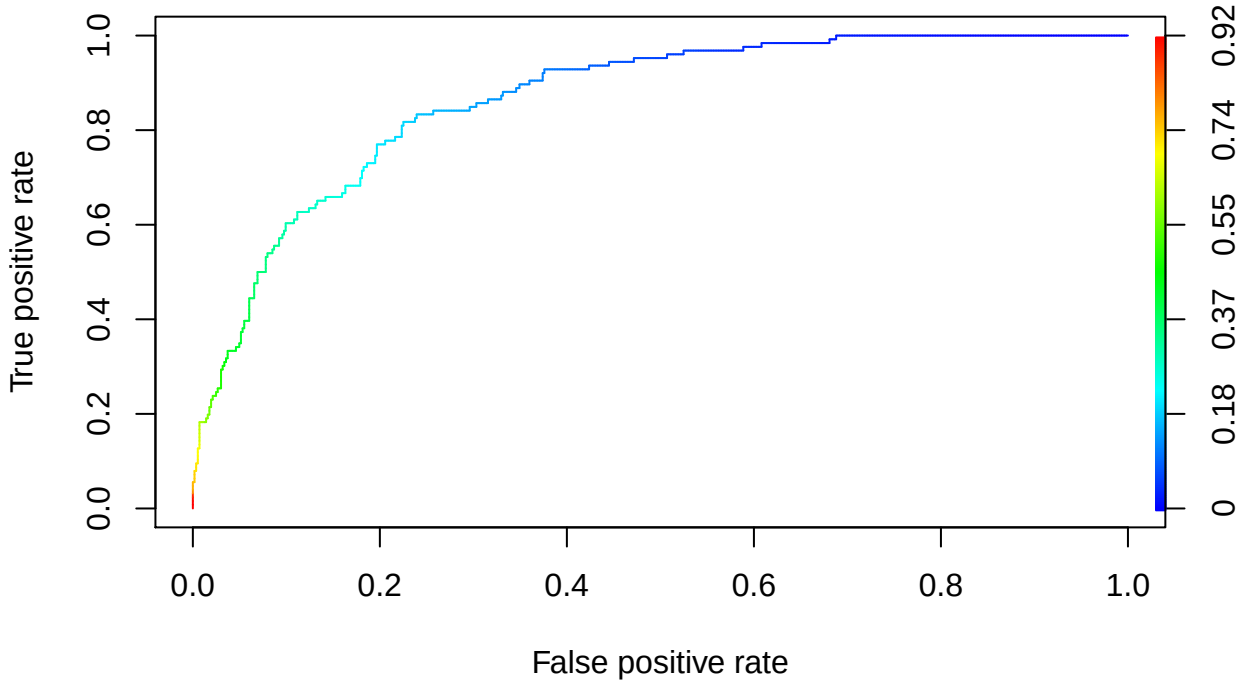
```
recall(table_mat)
```

```
## [1] 0.2460317
```

```r
index <-
  predict[!is.na(predict)] %>%
  as.data.frame() %>%
  add_rownames() %>%
  pull(rowname)


t <-
  test_model_data %>%
  add_rownames() %>%
  filter(rowname %in% index)


ROCRpred <- prediction(predict[!is.na(predict)], t$match)
ROCRperf <- performance(ROCRpred, 'tpr', 'fpr')
plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2, 1.7))
```




```r
m3 <-
  glm(
    match ~ . + .:.,
    data = train_model_data,
    family = 'binomial'
  )

summary(m3)
```

```
##
## Call:
## glm(formula = match ~ . + .:., family = "binomial", data = train_model_data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3465  -0.5523  -0.2558  -0.0569   3.2706
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  2.562844  13.054234   0.196  0.84436
## attr                       -56.841880  31.455184  -1.807  0.07075 .
## sinc                       -24.747707  34.378208  -0.720  0.47161
## fun                        -30.784516  39.747810  -0.774  0.43864
## prob                        10.019098   6.953780   1.441  0.14964
## prob_o                       9.082191   6.335384   1.434  0.15170
## like                        13.540555  11.128836   1.217  0.22371
## like_o                      -7.564276   8.549030  -0.885  0.37626
## avg_partner_rating         -15.269994  14.293980  -1.068  0.28539
## imprace                      9.519474   4.561240   2.087  0.03688 *
## tvsports                    -4.044074   4.913000  -0.823  0.41043
## clubbing                     0.050032   5.245624   0.010  0.99239
## movies                      -1.240649   7.671741  -0.162  0.87153
## shopping                    -3.726093   5.303631  -0.703  0.48233
## age_o                       -0.146225   0.349461  -0.418  0.67563
## attr:sinc                  126.684808  63.019932   2.010  0.04441 *
## attr:fun                   -37.906535  68.661470  -0.552  0.58089
## attr:prob                   -2.546082  12.293240  -0.207  0.83592
## attr:prob_o                 -7.663772  12.383565  -0.619  0.53600
## attr:like                    3.125321  19.002320   0.164  0.86936
## attr:like_o                 49.942301  16.196762   3.083  0.00205 **
## attr:avg_partner_rating     25.525084  25.618929   0.996  0.31909
## attr:imprace                -6.013046   8.549871  -0.703  0.48187
## attr:tvsports               12.997879   8.883784   1.463  0.14344
## attr:clubbing              -15.675446   9.684145  -1.619  0.10552
## attr:movies                -17.965598  12.962466  -1.386  0.16576
## attr:shopping               -0.002026   9.286829   0.000  0.99983
## attr:age_o                   0.767650   0.713843   1.075  0.28221
## sinc:fun                    -7.404374  73.300368  -0.101  0.91954
## sinc:prob                   -2.939977  14.811377  -0.198  0.84266
## sinc:prob_o                -17.444293  14.110920  -1.236  0.21637
## sinc:like                    5.214982  23.900245   0.218  0.82727
## sinc:like_o                 24.004323  18.794308   1.277  0.20153
## sinc:avg_partner_rating     12.593972  29.897225   0.421  0.67358
## sinc:imprace               -11.089449   9.755364  -1.137  0.25564
## sinc:tvsports               15.017989  10.362597   1.449  0.14727
## sinc:clubbing              -14.798400  10.817342  -1.368  0.17130
## sinc:movies                 12.073123  15.991392   0.755  0.45026
## sinc:shopping                5.786778  10.652504   0.543  0.58697
## sinc:age_o                  -0.760253   0.795965  -0.955  0.33951
## fun:prob                   -15.882649  15.202179  -1.045  0.29613
## fun:prob_o                  -0.209788  14.549697  -0.014  0.98850
## fun:like                   -12.151115  21.925781  -0.554  0.57945
## fun:like_o                   3.789919  21.031678   0.180  0.85700
```

```
## fun:avg_partner_rating       60.775067  30.163908   2.015  0.04392 *
## fun:imprace                  -7.899051  10.079435  -0.784  0.43323
## fun:tvsports                 -5.880713   9.642161  -0.610  0.54193
## fun:clubbing                 22.095171  10.854835   2.036  0.04180 *
## fun:movies                  -19.475894  16.835081  -1.157  0.24733
## fun:shopping                 -8.832325  11.132946  -0.793  0.42757
## fun:age_o                     1.203518   0.926750   1.299  0.19407
## prob:prob_o                   0.156190   1.814301   0.086  0.93140
## prob:like                    -1.286061   3.282279  -0.392  0.69519
## prob:like_o                  -4.211013   2.785797  -1.512  0.13064
## prob:avg_partner_rating      -1.042306   4.190312  -0.249  0.80356
## prob:imprace                 -0.166862   1.460711  -0.114  0.90905
## prob:tvsports                -3.642695   1.465607  -2.485  0.01294 *
## prob:clubbing                 1.328363   1.649540   0.805  0.42065
## prob:movies                   1.979507   2.366999   0.836  0.40299
## prob:shopping                -1.580659   1.596352  -0.990  0.32209
## prob:age_o                    0.003581   0.122297   0.029  0.97664
## prob_o:like                  -4.056636   4.037084  -1.005  0.31497
## prob_o:like_o                -4.185983   2.038347  -2.054  0.04001 *
## prob_o:avg_partner_rating     2.466120   4.779503   0.516  0.60587
## prob_o:imprace               -0.793209   1.334133  -0.595  0.55214
## prob_o:tvsports               1.192584   1.407076   0.848  0.39668
## prob_o:clubbing              -0.863729   1.544249  -0.559  0.57594
## prob_o:movies                -2.146037   2.357755  -0.910  0.36272
## prob_o:shopping               1.183620   1.449374   0.817  0.41413
## prob_o:age_o                  0.066822   0.101833   0.656  0.51170
## like:like_o                   8.959405   5.191781   1.726  0.08440 .
## like:avg_partner_rating     -18.116225   4.351411  -4.163 3.14e-05 ***
## like:imprace                  0.747291   2.425024   0.308  0.75796
## like:tvsports                 2.889923   2.632532   1.098  0.27230
## like:clubbing                -5.878522   2.928154  -2.008  0.04469 *
## like:movies                   0.557524   4.034072   0.138  0.89008
## like:shopping                -0.325484   2.858456  -0.114  0.90934
## like:age_o                    0.061479   0.216330   0.284  0.77627
## like_o:avg_partner_rating     2.282481   6.388288   0.357  0.72087
## like_o:imprace               -1.397034   1.816362  -0.769  0.44181
## like_o:tvsports              -1.215176   1.893804  -0.642  0.52109
## like_o:clubbing               1.418560   2.114694   0.671  0.50234
## like_o:movies                 1.130330   3.174828   0.356  0.72182
## like_o:shopping               1.034812   2.028433   0.510  0.60994
## like_o:age_o                 -0.179254   0.132567  -1.352  0.17632
## avg_partner_rating:imprace    1.081051   3.059304   0.353  0.72381
## avg_partner_rating:tvsports   2.206865   3.211487   0.687  0.49197
## avg_partner_rating:clubbing   2.389709   3.636994   0.657  0.51114
## avg_partner_rating:movies     7.133484   5.032193   1.418  0.15632
## avg_partner_rating:shopping   6.231323   3.526703   1.767  0.07725 .
## avg_partner_rating:age_o      0.018684   0.272873   0.068  0.94541
## imprace:tvsports             -0.813929   1.021921  -0.796  0.42576
## imprace:clubbing              0.417600   1.036140   0.403  0.68692
## imprace:movies               -3.545976   1.689619  -2.099  0.03584 *
## imprace:shopping             -0.945740   0.988694  -0.957  0.33879
## imprace:age_o                -0.085895   0.077685  -1.106  0.26886
## tvsports:clubbing             1.325086   1.020763   1.298  0.19424
## tvsports:movies              -1.858386   1.646502  -1.129  0.25903
```

```
## tvsports:shopping             2.052988   1.097223   1.871  0.06133 .
## tvsports:age_o               -0.067017   0.081607  -0.821  0.41152
## clubbing:movies               0.036597   1.699280   0.022  0.98282
## clubbing:shopping            -0.499216   1.172573  -0.426  0.67029
## clubbing:age_o                0.104458   0.087209   1.198  0.23100
## movies:shopping              -0.113312   1.734800  -0.065  0.94792
## movies:age_o                  0.024377   0.130500   0.187  0.85182
## shopping:age_o               -0.036149   0.086283  -0.419  0.67525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2495.3  on 2721  degrees of freedom
## Residual deviance: 1730.8  on 2616  degrees of freedom
##   (684 observations deleted due to missingness)
## AIC: 1942.8
##
## Number of Fisher Scoring iterations: 7
```

```
predict <- predict(m3, test_model_data, type = 'response')
# confusion matrix
table_mat <- table(test_model_data$match, predict > 0.5)
table_mat
```

```
##
##     FALSE TRUE
##   0   546   18
##   1    86   40
```

```
accuracy(table_mat)
```

```
## [1] 0.8492754
```

```
precision(table_mat)
```

```
## [1] 0.6896552
```

```
recall(table_mat)
```

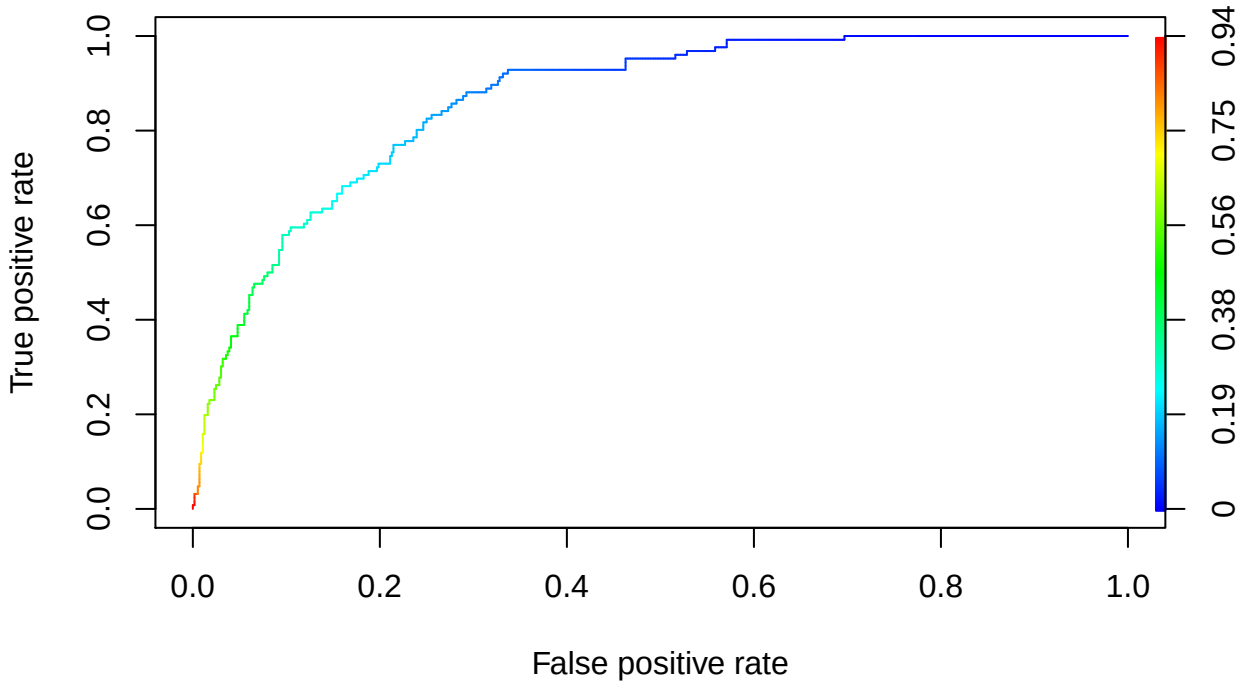
```
## [1] 0.3174603
```

```
index <-
  predict[!is.na(predict)] %>%
  as.data.frame() %>%
  add_rownames() %>%
  pull(rowname)


t <-
```

```
  test_model_data %>%
  add_rownames() %>%
  filter(rowname %in% index)


ROCRpred <- prediction(predict[!is.na(predict)], t$match)
ROCRperf <- performance(ROCRpred, 'tpr', 'fpr')
plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2, 1.7))
```



Stepwise AIC regression on model 2

```
train_model_data <-
  train_model_data %>%
  drop_na()



full_model = glm(match ~ . + .:., family = binomial(), data = train_model_data)

#summary(full_model)


null_model = glm(match ~ 1, family = binomial(), data = train_model_data)
```

```r
#summary(null_model)

steps <-
  step(null_model, list(upper = full_model), direction = 'forward')
```

```
## Start:  AIC=2497.31
## match ~ 1
##
##                       Df Deviance    AIC
## + like                 1   2183.0 2187.0
## + like_o               1   2208.4 2212.4
## + avg_partner_rating   1   2247.8 2251.8
## + prob                 1   2310.4 2314.4
## + prob_o               1   2345.9 2349.9
## + fun                  1   2458.0 2462.0
## + sinc                 1   2467.8 2471.8
## + imprace              1   2477.5 2481.5
## + attr                 1   2479.6 2483.6
## + clubbing             1   2480.2 2484.2
## <none>                     2495.3 2497.3
## + age_o                1   2493.6 2497.6
## + shopping             1   2493.8 2497.8
## + movies               1   2494.0 2498.0
## + tvsports             1   2494.2 2498.2
##
## Step:  AIC=2186.97
## match ~ like
##
##                       Df Deviance    AIC
## + like_o               1   1949.5 1955.5
## + prob_o               1   2081.0 2087.0
## + prob                 1   2145.8 2151.8
## + clubbing             1   2167.3 2173.3
## + imprace              1   2170.7 2176.7
## + avg_partner_rating   1   2171.3 2177.3
## + fun                  1   2175.4 2181.4
## + sinc                 1   2177.2 2183.2
## + tvsports             1   2180.1 2186.1
## + shopping             1   2180.3 2186.3
## + attr                 1   2180.5 2186.5
## <none>                     2183.0 2187.0
## + movies               1   2182.6 2188.6
## + age_o                1   2182.9 2188.9
##
## Step:  AIC=1955.52
## match ~ like + like_o
##
##                       Df Deviance    AIC
## + prob                 1   1924.4 1932.4
## + prob_o               1   1934.0 1942.0
## + imprace              1   1939.9 1947.9
## + sinc                 1   1940.1 1948.1
## + fun                  1   1941.6 1949.6
```

```
## + clubbing            1   1943.0 1951.0
## + attr                1   1943.7 1951.7
## + avg_partner_rating  1   1944.2 1952.2
## + like:like_o         1   1945.5 1953.5
## + tvsports            1   1945.8 1953.8
## + shopping            1   1947.5 1955.5
## <none>                    1949.5 1955.5
## + movies              1   1948.3 1956.3
## + age_o               1   1949.5 1957.5
##
## Step:  AIC=1932.4
## match ~ like + like_o + prob
##
##                      Df Deviance    AIC
## + attr                1   1912.2 1922.2
## + prob_o              1   1912.3 1922.3
## + imprace             1   1914.8 1924.8
## + sinc                1   1915.2 1925.2
## + prob:like           1   1915.9 1925.9
## + fun                 1   1916.4 1926.4
## + tvsports            1   1919.3 1929.3
## + shopping            1   1919.3 1929.3
## + like:like_o         1   1920.6 1930.6
## + clubbing            1   1921.1 1931.1
## <none>                    1924.4 1932.4
## + avg_partner_rating  1   1923.0 1933.0
## + movies              1   1923.7 1933.7
## + age_o               1   1924.4 1934.4
## + prob:like_o         1   1924.4 1934.4
##
## Step:  AIC=1922.2
## match ~ like + like_o + prob + attr
##
##                      Df Deviance    AIC
## + prob_o              1   1901.4 1913.4
## + prob:like           1   1902.8 1914.8
## + fun                 1   1903.2 1915.2
## + imprace             1   1903.2 1915.2
## + sinc                1   1906.9 1918.9
## + tvsports            1   1907.5 1919.5
## + attr:like_o         1   1907.9 1919.9
## + shopping            1   1908.3 1920.3
## + like:like_o         1   1908.7 1920.7
## + clubbing            1   1908.8 1920.8
## + avg_partner_rating  1   1908.9 1920.9
## <none>                    1912.2 1922.2
## + movies              1   1912.0 1924.0
## + attr:like           1   1912.0 1924.0
## + prob:like_o         1   1912.1 1924.1
## + age_o               1   1912.2 1924.2
## + attr:prob           1   1912.2 1924.2
##
## Step:  AIC=1913.44
## match ~ like + like_o + prob + attr + prob_o
```

```
##
##                       Df Deviance    AIC
## + prob:like            1   1891.7 1905.7
## + imprace              1   1892.3 1906.3
## + fun                  1   1893.7 1907.7
## + attr:like_o          1   1896.0 1910.0
## + tvsports             1   1896.1 1910.1
## + prob_o:like_o        1   1896.9 1910.9
## + sinc                 1   1897.1 1911.1
## + like:like_o          1   1897.6 1911.6
## + shopping             1   1898.1 1912.1
## + clubbing             1   1898.2 1912.2
## + avg_partner_rating   1   1898.3 1912.3
## <none>                     1901.4 1913.4
## + prob_o:like          1   1901.1 1915.1
## + prob:like_o          1   1901.2 1915.2
## + movies               1   1901.3 1915.3
## + age_o                1   1901.3 1915.3
## + attr:like            1   1901.3 1915.3
## + attr:prob_o          1   1901.4 1915.4
## + prob:prob_o          1   1901.4 1915.4
## + attr:prob            1   1901.4 1915.4
##
## Step:  AIC=1905.69
## match ~ like + like_o + prob + attr + prob_o + like:prob
##
##                       Df Deviance    AIC
## + imprace              1   1881.6 1897.6
## + fun                  1   1884.2 1900.2
## + attr:like_o          1   1885.5 1901.5
## + tvsports             1   1886.5 1902.5
## + sinc                 1   1887.1 1903.1
## + like:like_o          1   1887.6 1903.6
## + prob_o:like_o        1   1887.8 1903.8
## + avg_partner_rating   1   1888.2 1904.2
## + clubbing             1   1888.3 1904.3
## + shopping             1   1888.6 1904.6
## <none>                     1891.7 1905.7
## + prob:like_o          1   1891.3 1907.3
## + movies               1   1891.5 1907.5
## + prob_o:like          1   1891.5 1907.5
## + age_o                1   1891.6 1907.6
## + attr:like            1   1891.6 1907.6
## + attr:prob_o          1   1891.7 1907.7
## + attr:prob            1   1891.7 1907.7
## + prob:prob_o          1   1891.7 1907.7
##
## Step:  AIC=1897.63
## match ~ like + like_o + prob + attr + prob_o + imprace + like:prob
##
##                       Df Deviance    AIC
## + fun                  1   1873.9 1891.9
## + attr:like_o          1   1875.5 1893.5
## + avg_partner_rating   1   1876.6 1894.6
```

```
## + tvsports            1   1877.2 1895.2
## + sinc                1   1877.3 1895.3
## + like:like_o         1   1877.4 1895.4
## + prob_o:like_o       1   1877.6 1895.6
## + clubbing            1   1879.6 1897.6
## <none>                    1881.6 1897.6
## + shopping            1   1880.4 1898.4
## + like:imprace        1   1880.8 1898.8
## + like_o:imprace      1   1881.2 1899.2
## + prob:like_o         1   1881.4 1899.4
## + prob:imprace        1   1881.5 1899.5
## + age_o               1   1881.5 1899.5
## + prob_o:like         1   1881.5 1899.5
## + prob_o:imprace      1   1881.6 1899.6
## + attr:like           1   1881.6 1899.6
## + attr:prob_o         1   1881.6 1899.6
## + movies              1   1881.6 1899.6
## + attr:prob           1   1881.6 1899.6
## + prob:prob_o         1   1881.6 1899.6
## + attr:imprace        1   1881.6 1899.6
##
## Step:  AIC=1891.86
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     like:prob
##
##                       Df Deviance    AIC
## + attr:like_o          1   1867.8 1887.8
## + avg_partner_rating   1   1869.1 1889.1
## + prob_o:like_o        1   1869.3 1889.3
## + like:like_o          1   1869.7 1889.7
## + tvsports             1   1869.7 1889.7
## + attr:fun             1   1871.8 1891.8
## <none>                     1873.9 1891.9
## + sinc                 1   1872.1 1892.1
## + clubbing             1   1872.1 1892.1
## + shopping             1   1872.7 1892.7
## + like:imprace         1   1873.0 1893.0
## + fun:like             1   1873.3 1893.3
## + like_o:imprace       1   1873.4 1893.4
## + age_o                1   1873.5 1893.5
## + fun:imprace          1   1873.6 1893.6
## + prob:like_o          1   1873.7 1893.7
## + prob:imprace         1   1873.7 1893.7
## + attr:prob_o          1   1873.8 1893.8
## + attr:like            1   1873.8 1893.8
## + prob_o:like          1   1873.8 1893.8
## + prob_o:imprace       1   1873.8 1893.8
## + fun:prob             1   1873.8 1893.8
## + fun:like_o           1   1873.8 1893.8
## + prob:prob_o          1   1873.8 1893.8
## + movies               1   1873.8 1893.8
## + attr:prob            1   1873.8 1893.8
## + attr:imprace         1   1873.8 1893.8
## + fun:prob_o           1   1873.8 1893.8
```

```
##
## Step:  AIC=1887.78
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     like:prob + like_o:attr
##
##                       Df Deviance    AIC
## + avg_partner_rating  1   1862.6 1884.6
## + prob_o:like_o       1   1862.8 1884.8
## + tvsports            1   1863.6 1885.6
## + like:like_o         1   1864.5 1886.5
## <none>                    1867.8 1887.8
## + sinc                1   1865.9 1887.9
## + clubbing            1   1866.1 1888.1
## + attr:fun            1   1866.4 1888.4
## + attr:prob_o         1   1866.5 1888.5
## + shopping            1   1866.6 1888.6
## + like:imprace        1   1867.1 1889.1
## + fun:like            1   1867.2 1889.2
## + fun:imprace         1   1867.4 1889.4
## + age_o               1   1867.5 1889.5
## + like_o:imprace      1   1867.5 1889.5
## + prob:imprace        1   1867.6 1889.6
## + prob_o:like         1   1867.7 1889.7
## + attr:prob           1   1867.7 1889.7
## + fun:like_o          1   1867.7 1889.7
## + fun:prob            1   1867.8 1889.8
## + attr:like           1   1867.8 1889.8
## + prob:prob_o         1   1867.8 1889.8
## + prob_o:imprace      1   1867.8 1889.8
## + movies              1   1867.8 1889.8
## + fun:prob_o          1   1867.8 1889.8
## + attr:imprace        1   1867.8 1889.8
## + prob:like_o         1   1867.8 1889.8
##
## Step:  AIC=1884.56
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + like:prob + like_o:attr
##
##                              Df Deviance    AIC
## + like:avg_partner_rating     1   1848.6 1872.6
## + tvsports                    1   1857.7 1881.7
## + prob_o:like_o               1   1857.8 1881.8
## + like:like_o                 1   1859.3 1883.3
## + like_o:avg_partner_rating   1   1859.6 1883.6
## <none>                            1862.6 1884.6
## + clubbing                    1   1860.7 1884.7
## + sinc                        1   1861.1 1885.1
## + attr:fun                    1   1861.1 1885.1
## + shopping                    1   1861.2 1885.2
## + attr:prob_o                 1   1861.3 1885.3
## + avg_partner_rating:imprace  1   1861.5 1885.5
## + like:imprace                1   1861.6 1885.6
## + prob:avg_partner_rating     1   1861.9 1885.9
## + fun:like                    1   1862.1 1886.1
```

```
## + fun:imprace                 1   1862.2 1886.2
## + age_o                       1   1862.2 1886.2
## + like_o:imprace              1   1862.2 1886.2
## + fun:avg_partner_rating      1   1862.4 1886.4
## + prob:imprace                1   1862.4 1886.4
## + prob_o:avg_partner_rating   1   1862.4 1886.4
## + prob_o:like                 1   1862.5 1886.5
## + fun:like_o                  1   1862.5 1886.5
## + movies                      1   1862.5 1886.5
## + fun:prob                    1   1862.5 1886.5
## + attr:prob                   1   1862.5 1886.5
## + attr:avg_partner_rating     1   1862.5 1886.5
## + prob:like_o                 1   1862.6 1886.6
## + prob_o:imprace              1   1862.6 1886.6
## + fun:prob_o                  1   1862.6 1886.6
## + attr:like                   1   1862.6 1886.6
## + attr:imprace                1   1862.6 1886.6
## + prob:prob_o                 1   1862.6 1886.6
##
## Step:  AIC=1872.57
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + like:prob + like_o:attr + like:avg_partner_rating
##
##                              Df Deviance    AIC
## + prob_o:like_o               1   1843.5 1869.5
## + tvsports                    1   1844.0 1870.0
## + like:like_o                 1   1844.6 1870.6
## + like_o:avg_partner_rating   1   1845.4 1871.4
## + clubbing                    1   1846.3 1872.3
## + attr:fun                    1   1846.5 1872.5
## <none>                            1848.6 1872.6
## + sinc                        1   1846.8 1872.8
## + shopping                    1   1847.1 1873.1
## + attr:prob_o                 1   1847.3 1873.3
## + fun:avg_partner_rating      1   1847.5 1873.5
## + like:imprace                1   1847.8 1873.8
## + avg_partner_rating:imprace  1   1848.0 1874.0
## + age_o                       1   1848.2 1874.2
## + like_o:imprace              1   1848.2 1874.2
## + fun:imprace                 1   1848.3 1874.3
## + fun:like                    1   1848.3 1874.3
## + attr:avg_partner_rating     1   1848.4 1874.4
## + movies                      1   1848.4 1874.4
## + attr:like                   1   1848.4 1874.4
## + prob_o:avg_partner_rating   1   1848.5 1874.5
## + prob:imprace                1   1848.5 1874.5
## + prob:avg_partner_rating     1   1848.5 1874.5
## + prob_o:like                 1   1848.5 1874.5
## + fun:like_o                  1   1848.5 1874.5
## + attr:imprace                1   1848.5 1874.5
## + fun:prob                    1   1848.5 1874.5
## + prob:like_o                 1   1848.5 1874.5
## + attr:prob                   1   1848.6 1874.6
## + prob:prob_o                 1   1848.6 1874.6
```

```
## + fun:prob_o                  1   1848.6 1874.6
## + prob_o:imprace              1   1848.6 1874.6
##
## Step:  AIC=1869.48
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + like:prob + like_o:attr + like:avg_partner_rating +
##     like_o:prob_o
##
##                              Df Deviance    AIC
## + tvsports                    1   1839.0 1867.0
## + like:like_o                 1   1839.1 1867.1
## + like_o:avg_partner_rating   1   1839.9 1867.9
## + clubbing                    1   1841.4 1869.4
## <none>                            1843.5 1869.5
## + sinc                        1   1841.6 1869.6
## + attr:fun                    1   1841.7 1869.7
## + shopping                    1   1842.0 1870.0
## + attr:prob_o                 1   1842.3 1870.3
## + fun:avg_partner_rating      1   1842.6 1870.6
## + like:imprace                1   1842.7 1870.7
## + avg_partner_rating:imprace  1   1842.8 1870.8
## + fun:imprace                 1   1843.0 1871.0
## + like_o:imprace              1   1843.1 1871.1
## + fun:like                    1   1843.2 1871.2
## + attr:avg_partner_rating     1   1843.2 1871.2
## + prob:imprace                1   1843.3 1871.3
## + movies                      1   1843.3 1871.3
## + attr:like                   1   1843.3 1871.3
## + age_o                       1   1843.3 1871.3
## + prob_o:like                 1   1843.3 1871.3
## + prob_o:avg_partner_rating   1   1843.3 1871.3
## + fun:like_o                  1   1843.4 1871.4
## + prob:avg_partner_rating     1   1843.4 1871.4
## + attr:imprace                1   1843.4 1871.4
## + fun:prob_o                  1   1843.5 1871.5
## + fun:prob                    1   1843.5 1871.5
## + prob_o:imprace              1   1843.5 1871.5
## + prob:like_o                 1   1843.5 1871.5
## + prob:prob_o                 1   1843.5 1871.5
## + attr:prob                   1   1843.5 1871.5
##
## Step:  AIC=1867.01
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o
##
##                               Df Deviance    AIC
## + like:like_o                  1   1834.5 1864.5
## + clubbing                     1   1835.1 1865.1
## + like_o:avg_partner_rating    1   1835.3 1865.3
## + prob:tvsports                1   1835.5 1865.5
## + attr:tvsports                1   1836.7 1866.7
## + sinc                         1   1836.9 1866.9
## <none>                             1839.0 1867.0
```

```
## + attr:fun                     1   1837.0 1867.0
## + shopping                     1   1837.8 1867.8
## + attr:prob_o                  1   1837.9 1867.9
## + fun:avg_partner_rating       1   1838.0 1868.0
## + fun:tvsports                 1   1838.2 1868.2
## + like:imprace                 1   1838.5 1868.5
## + like:tvsports                1   1838.5 1868.5
## + avg_partner_rating:imprace   1   1838.5 1868.5
## + imprace:tvsports             1   1838.5 1868.5
## + fun:imprace                  1   1838.6 1868.6
## + like_o:imprace               1   1838.6 1868.6
## + fun:like                     1   1838.7 1868.7
## + attr:avg_partner_rating      1   1838.7 1868.7
## + movies                       1   1838.7 1868.7
## + attr:like                    1   1838.8 1868.8
## + prob:imprace                 1   1838.8 1868.8
## + age_o                        1   1838.8 1868.8
## + avg_partner_rating:tvsports  1   1838.9 1868.9
## + prob_o:avg_partner_rating    1   1838.9 1868.9
## + prob_o:like                  1   1838.9 1868.9
## + attr:imprace                 1   1839.0 1869.0
## + fun:like_o                   1   1839.0 1869.0
## + prob:like_o                  1   1839.0 1869.0
## + fun:prob_o                   1   1839.0 1869.0
## + like_o:tvsports              1   1839.0 1869.0
## + prob_o:imprace               1   1839.0 1869.0
## + prob:avg_partner_rating      1   1839.0 1869.0
## + prob_o:tvsports              1   1839.0 1869.0
## + attr:prob                    1   1839.0 1869.0
## + prob:prob_o                  1   1839.0 1869.0
## + fun:prob                     1   1839.0 1869.0
##
## Step:  AIC=1864.51
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o
##
##                               Df Deviance    AIC
## + clubbing                     1   1830.4 1862.4
## + prob:tvsports                1   1831.1 1863.1
## + attr:fun                     1   1832.2 1864.2
## + attr:tvsports                1   1832.2 1864.2
## + sinc                         1   1832.4 1864.4
## <none>                             1834.5 1864.5
## + attr:prob_o                  1   1833.2 1865.2
## + shopping                     1   1833.3 1865.3
## + prob_o:like                  1   1833.3 1865.3
## + fun:avg_partner_rating       1   1833.5 1865.5
## + fun:tvsports                 1   1833.7 1865.7
## + like:tvsports                1   1833.9 1865.9
## + prob:like_o                  1   1834.0 1866.0
## + imprace:tvsports             1   1834.0 1866.0
## + like_o:imprace               1   1834.0 1866.0
## + like_o:avg_partner_rating    1   1834.1 1866.1
```

```
## + fun:imprace                   1   1834.1 1866.1
## + avg_partner_rating:imprace    1   1834.1 1866.1
## + like:imprace                  1   1834.2 1866.2
## + movies                        1   1834.2 1866.2
## + age_o                         1   1834.3 1866.3
## + fun:like                      1   1834.3 1866.3
## + avg_partner_rating:tvsports   1   1834.4 1866.4
## + attr:avg_partner_rating       1   1834.4 1866.4
## + prob:prob_o                   1   1834.4 1866.4
## + prob:imprace                  1   1834.4 1866.4
## + attr:imprace                  1   1834.4 1866.4
## + attr:like                     1   1834.4 1866.4
## + prob_o:imprace                1   1834.5 1866.5
## + prob_o:tvsports               1   1834.5 1866.5
## + prob_o:avg_partner_rating     1   1834.5 1866.5
## + fun:prob_o                    1   1834.5 1866.5
## + attr:prob                     1   1834.5 1866.5
## + fun:like_o                    1   1834.5 1866.5
## + like_o:tvsports               1   1834.5 1866.5
## + fun:prob                      1   1834.5 1866.5
## + prob:avg_partner_rating       1   1834.5 1866.5
##
## Step:  AIC=1862.43
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + clubbing + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o
##
##                                Df Deviance    AIC
## + fun:clubbing                  1   1825.1 1859.1
## + attr:clubbing                 1   1825.3 1859.3
## + prob:tvsports                 1   1826.8 1860.8
## + attr:tvsports                 1   1828.0 1862.0
## + attr:fun                      1   1828.1 1862.1
## <none>                              1830.4 1862.4
## + tvsports:clubbing             1   1828.6 1862.6
## + shopping                      1   1828.7 1862.7
## + sinc                          1   1828.9 1862.9
## + attr:prob_o                   1   1829.1 1863.1
## + like:clubbing                 1   1829.3 1863.3
## + prob_o:like                   1   1829.4 1863.4
## + fun:avg_partner_rating        1   1829.5 1863.5
## + prob:clubbing                 1   1829.5 1863.5
## + fun:tvsports                  1   1829.6 1863.6
## + movies                        1   1829.8 1863.8
## + like_o:avg_partner_rating     1   1829.8 1863.8
## + avg_partner_rating:imprace    1   1829.8 1863.8
## + imprace:tvsports              1   1829.8 1863.8
## + prob:like_o                   1   1829.9 1863.9
## + avg_partner_rating:clubbing   1   1830.0 1864.0
## + like:tvsports                 1   1830.0 1864.0
## + like:imprace                  1   1830.0 1864.0
## + like_o:imprace                1   1830.0 1864.0
## + fun:imprace                   1   1830.1 1864.1
## + age_o                         1   1830.1 1864.1
```

```
## + prob:imprace                 1   1830.2 1864.2
## + attr:avg_partner_rating      1   1830.2 1864.2
## + fun:like                     1   1830.2 1864.2
## + imprace:clubbing             1   1830.3 1864.3
## + attr:like                    1   1830.3 1864.3
## + attr:imprace                 1   1830.3 1864.3
## + prob:prob_o                  1   1830.4 1864.4
## + like_o:clubbing              1   1830.4 1864.4
## + avg_partner_rating:tvsports  1   1830.4 1864.4
## + prob_o:tvsports              1   1830.4 1864.4
## + fun:prob_o                   1   1830.4 1864.4
## + fun:like_o                   1   1830.4 1864.4
## + prob_o:imprace               1   1830.4 1864.4
## + prob_o:avg_partner_rating    1   1830.4 1864.4
## + attr:prob                    1   1830.4 1864.4
## + prob:avg_partner_rating      1   1830.4 1864.4
## + prob_o:clubbing              1   1830.4 1864.4
## + like_o:tvsports              1   1830.4 1864.4
## + fun:prob                     1   1830.4 1864.4
##
## Step:  AIC=1859.06
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + clubbing + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o + fun:clubbing
##
##                               Df Deviance    AIC
## + attr:clubbing                1   1820.4 1856.4
## + prob:tvsports                1   1821.4 1857.4
## + attr:tvsports                1   1822.4 1858.4
## + attr:fun                     1   1822.8 1858.8
## <none>                             1825.1 1859.1
## + like:clubbing                1   1823.1 1859.1
## + shopping                     1   1823.3 1859.3
## + tvsports:clubbing            1   1823.4 1859.4
## + fun:avg_partner_rating       1   1823.5 1859.5
## + attr:prob_o                  1   1823.5 1859.5
## + sinc                         1   1823.9 1859.9
## + prob_o:like                  1   1823.9 1859.9
## + fun:tvsports                 1   1824.2 1860.2
## + avg_partner_rating:imprace   1   1824.3 1860.3
## + prob:clubbing                1   1824.3 1860.3
## + movies                       1   1824.5 1860.5
## + prob:like_o                  1   1824.5 1860.5
## + imprace:tvsports             1   1824.5 1860.5
## + like_o:imprace               1   1824.5 1860.5
## + like_o:avg_partner_rating    1   1824.5 1860.5
## + like:imprace                 1   1824.6 1860.6
## + like:tvsports                1   1824.6 1860.6
## + age_o                        1   1824.7 1860.7
## + prob:imprace                 1   1824.8 1860.8
## + imprace:clubbing             1   1824.8 1860.8
## + avg_partner_rating:clubbing  1   1824.9 1860.9
## + attr:imprace                 1   1824.9 1860.9
## + attr:avg_partner_rating      1   1824.9 1860.9
```

```
## + attr:like                    1   1824.9 1860.9
## + like_o:clubbing              1   1824.9 1860.9
## + fun:imprace                  1   1825.0 1861.0
## + fun:like_o                   1   1825.0 1861.0
## + prob_o:tvsports              1   1825.0 1861.0
## + fun:like                     1   1825.0 1861.0
## + avg_partner_rating:tvsports  1   1825.0 1861.0
## + prob:prob_o                  1   1825.0 1861.0
## + prob_o:imprace               1   1825.0 1861.0
## + prob:avg_partner_rating      1   1825.0 1861.0
## + attr:prob                    1   1825.0 1861.0
## + fun:prob                     1   1825.0 1861.0
## + prob_o:avg_partner_rating    1   1825.0 1861.0
## + fun:prob_o                   1   1825.1 1861.1
## + like_o:tvsports              1   1825.1 1861.1
## + prob_o:clubbing              1   1825.1 1861.1
##
## Step:  AIC=1856.39
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + clubbing + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o + fun:clubbing +
##     attr:clubbing
##
##                               Df Deviance    AIC
## + prob:tvsports                1   1816.4 1854.4
## + attr:tvsports                1   1816.9 1854.9
## <none>                             1820.4 1856.4
## + attr:fun                     1   1818.5 1856.5
## + shopping                     1   1818.6 1856.6
## + attr:prob_o                  1   1818.7 1856.7
## + fun:avg_partner_rating       1   1818.8 1856.8
## + tvsports:clubbing            1   1819.0 1857.0
## + like:clubbing                1   1819.0 1857.0
## + sinc                         1   1819.1 1857.1
## + prob_o:like                  1   1819.2 1857.2
## + avg_partner_rating:imprace   1   1819.4 1857.4
## + fun:tvsports                 1   1819.4 1857.4
## + movies                       1   1819.7 1857.7
## + prob:like_o                  1   1819.8 1857.8
## + like_o:imprace               1   1819.8 1857.8
## + attr:imprace                 1   1819.8 1857.8
## + like_o:avg_partner_rating    1   1819.9 1857.9
## + imprace:tvsports             1   1819.9 1857.9
## + like:imprace                 1   1819.9 1857.9
## + like:tvsports                1   1819.9 1857.9
## + age_o                        1   1820.0 1858.0
## + imprace:clubbing             1   1820.0 1858.0
## + prob:imprace                 1   1820.1 1858.1
## + prob:clubbing                1   1820.2 1858.2
## + attr:avg_partner_rating      1   1820.3 1858.3
## + avg_partner_rating:tvsports  1   1820.3 1858.3
## + like_o:clubbing              1   1820.3 1858.3
## + fun:imprace                  1   1820.3 1858.3
## + fun:like_o                   1   1820.3 1858.3
```

```
## + attr:like                     1   1820.3 1858.3
## + avg_partner_rating:clubbing  1   1820.3 1858.3
## + prob_o:imprace               1   1820.3 1858.3
## + prob_o:clubbing              1   1820.4 1858.4
## + fun:like                     1   1820.4 1858.4
## + attr:prob                    1   1820.4 1858.4
## + prob:prob_o                  1   1820.4 1858.4
## + prob_o:tvsports              1   1820.4 1858.4
## + fun:prob                     1   1820.4 1858.4
## + prob:avg_partner_rating      1   1820.4 1858.4
## + like_o:tvsports              1   1820.4 1858.4
## + fun:prob_o                   1   1820.4 1858.4
## + prob_o:avg_partner_rating    1   1820.4 1858.4
##
## Step:  AIC=1854.44
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + clubbing + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o + fun:clubbing +
##     attr:clubbing + prob:tvsports
##
##                               Df Deviance    AIC
## + attr:tvsports                1   1813.8 1853.8
## + like:tvsports                1   1814.2 1854.2
## <none>                             1816.4 1854.4
## + tvsports:clubbing            1   1814.4 1854.4
## + attr:fun                     1   1814.7 1854.7
## + attr:prob_o                  1   1814.8 1854.8
## + sinc                         1   1814.9 1854.9
## + fun:avg_partner_rating       1   1815.0 1855.0
## + shopping                     1   1815.0 1855.0
## + avg_partner_rating:tvsports  1   1815.1 1855.1
## + prob_o:like                  1   1815.3 1855.3
## + avg_partner_rating:imprace   1   1815.3 1855.3
## + like:clubbing                1   1815.5 1855.5
## + fun:tvsports                 1   1815.5 1855.5
## + movies                       1   1815.8 1855.8
## + prob:like_o                  1   1815.8 1855.8
## + like_o:avg_partner_rating    1   1815.9 1855.9
## + like:imprace                 1   1815.9 1855.9
## + attr:imprace                 1   1815.9 1855.9
## + imprace:tvsports             1   1815.9 1855.9
## + like_o:imprace               1   1815.9 1855.9
## + prob:clubbing                1   1815.9 1855.9
## + age_o                        1   1816.0 1856.0
## + imprace:clubbing             1   1816.0 1856.0
## + prob:imprace                 1   1816.1 1856.1
## + avg_partner_rating:clubbing  1   1816.3 1856.3
## + attr:avg_partner_rating      1   1816.3 1856.3
## + fun:imprace                  1   1816.3 1856.3
## + like_o:clubbing              1   1816.4 1856.4
## + fun:like_o                   1   1816.4 1856.4
## + prob:avg_partner_rating      1   1816.4 1856.4
## + attr:like                    1   1816.4 1856.4
## + prob_o:imprace               1   1816.4 1856.4
```

```
## + prob_o:tvsports              1   1816.4 1856.4
## + attr:prob                    1   1816.4 1856.4
## + prob:prob_o                  1   1816.4 1856.4
## + prob_o:clubbing              1   1816.4 1856.4
## + fun:like                     1   1816.4 1856.4
## + fun:prob                     1   1816.4 1856.4
## + like_o:tvsports              1   1816.4 1856.4
## + fun:prob_o                   1   1816.4 1856.4
## + prob_o:avg_partner_rating    1   1816.4 1856.4
##
## Step:  AIC=1853.8
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + clubbing + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o + fun:clubbing +
##     attr:clubbing + prob:tvsports + attr:tvsports
##
##                               Df Deviance    AIC
## + tvsports:clubbing            1   1811.4 1853.4
## <none>                             1813.8 1853.8
## + attr:prob_o                  1   1812.2 1854.2
## + sinc                         1   1812.2 1854.2
## + avg_partner_rating:tvsports  1   1812.4 1854.4
## + like:tvsports                1   1812.4 1854.4
## + attr:fun                     1   1812.4 1854.4
## + fun:avg_partner_rating       1   1812.5 1854.5
## + shopping                     1   1812.5 1854.5
## + prob_o:like                  1   1812.6 1854.6
## + avg_partner_rating:imprace   1   1812.8 1854.8
## + like:clubbing                1   1812.9 1854.9
## + prob:like_o                  1   1813.1 1855.1
## + attr:imprace                 1   1813.1 1855.1
## + fun:tvsports                 1   1813.1 1855.1
## + like_o:imprace               1   1813.2 1855.2
## + movies                       1   1813.2 1855.2
## + imprace:tvsports             1   1813.2 1855.2
## + like:imprace                 1   1813.3 1855.3
## + prob:clubbing                1   1813.3 1855.3
## + like_o:avg_partner_rating    1   1813.4 1855.4
## + imprace:clubbing             1   1813.4 1855.4
## + age_o                        1   1813.4 1855.4
## + prob:imprace                 1   1813.5 1855.5
## + like_o:clubbing              1   1813.7 1855.7
## + fun:imprace                  1   1813.7 1855.7
## + fun:like_o                   1   1813.7 1855.7
## + prob_o:imprace               1   1813.7 1855.7
## + avg_partner_rating:clubbing  1   1813.8 1855.8
## + prob:prob_o                  1   1813.8 1855.8
## + prob:avg_partner_rating      1   1813.8 1855.8
## + prob_o:tvsports              1   1813.8 1855.8
## + fun:like                     1   1813.8 1855.8
## + attr:prob                    1   1813.8 1855.8
## + attr:like                    1   1813.8 1855.8
## + attr:avg_partner_rating      1   1813.8 1855.8
## + fun:prob                     1   1813.8 1855.8
```

```
## + fun:prob_o                    1   1813.8 1855.8
## + prob_o:avg_partner_rating     1   1813.8 1855.8
## + prob_o:clubbing               1   1813.8 1855.8
## + like_o:tvsports               1   1813.8 1855.8
##
## Step:  AIC=1853.41
## match ~ like + like_o + prob + attr + prob_o + imprace + fun +
##     avg_partner_rating + tvsports + clubbing + like:prob + like_o:attr +
##     like:avg_partner_rating + like_o:prob_o + like:like_o + fun:clubbing +
##     attr:clubbing + prob:tvsports + attr:tvsports + tvsports:clubbing
##
##                               Df Deviance    AIC
## <none>                             1811.4 1853.4
## + attr:prob_o                  1   1809.5 1853.5
## + avg_partner_rating:tvsports  1   1809.7 1853.7
## + like:tvsports                1   1809.8 1853.8
## + sinc                         1   1809.8 1853.8
## + attr:fun                     1   1810.1 1854.1
## + prob_o:like                  1   1810.1 1854.1
## + fun:avg_partner_rating       1   1810.1 1854.1
## + shopping                     1   1810.2 1854.2
## + avg_partner_rating:imprace   1   1810.3 1854.3
## + fun:tvsports                 1   1810.4 1854.4
## + like:clubbing                1   1810.5 1854.5
## + attr:imprace                 1   1810.7 1854.7
## + prob:like_o                  1   1810.7 1854.7
## + movies                       1   1810.7 1854.7
## + like_o:avg_partner_rating    1   1810.8 1854.8
## + like:imprace                 1   1810.9 1854.9
## + imprace:clubbing             1   1810.9 1854.9
## + like_o:imprace               1   1810.9 1854.9
## + age_o                        1   1811.0 1855.0
## + prob:imprace                 1   1811.0 1855.0
## + prob:clubbing                1   1811.0 1855.0
## + imprace:tvsports             1   1811.3 1855.3
## + fun:imprace                  1   1811.3 1855.3
## + prob:avg_partner_rating      1   1811.3 1855.3
## + fun:like_o                   1   1811.4 1855.4
## + like_o:clubbing              1   1811.4 1855.4
## + prob_o:imprace               1   1811.4 1855.4
## + avg_partner_rating:clubbing  1   1811.4 1855.4
## + attr:like                    1   1811.4 1855.4
## + attr:prob                    1   1811.4 1855.4
## + prob:prob_o                  1   1811.4 1855.4
## + fun:prob_o                   1   1811.4 1855.4
## + prob_o:tvsports              1   1811.4 1855.4
## + fun:like                     1   1811.4 1855.4
## + fun:prob                     1   1811.4 1855.4
## + like_o:tvsports              1   1811.4 1855.4
## + prob_o:clubbing              1   1811.4 1855.4
## + prob_o:avg_partner_rating    1   1811.4 1855.4
## + attr:avg_partner_rating      1   1811.4 1855.4
```

```
best_model <-
  glm(formula = match ~ like + like_o + prob + attr + imprace +
    prob_o + fun + tvsports + avg_partner_rating + clubbing +
    movies + like:prob + like_o:attr + like:avg_partner_rating +
    like:like_o + like_o:prob_o + attr:tvsports + fun:clubbing +
    attr:clubbing + imprace:movies + like:movies + clubbing:movies +
    attr:prob_o + tvsports:clubbing + prob:tvsports + tvsports:avg_partner_rating +
    like:clubbing, family = binomial(), data = train_model_data)




predict <- predict(m2, test_model_data, type = 'response')
# confusion matrix
table_mat <- table(test_model_data$match, predict > 0.5)
table_mat
```

```
##
##     FALSE TRUE
##   0   549   15
##   1    95   31
```

```
accuracy(table_mat)
```

```
## [1] 0.8405797
```

```
precision(table_mat)
```

```
## [1] 0.673913
```

```
recall(table_mat)
```

```
## [1] 0.2460317
```

```
index <-
  predict[!is.na(predict)] %>%
  as.data.frame() %>%
  add_rownames() %>%
  pull(rowname)

summary(best_model)
```

```
##
## Call:
## glm(formula = match ~ like + like_o + prob + attr + imprace +
##     prob_o + fun + tvsports + avg_partner_rating + clubbing +
##     movies + like:prob + like_o:attr + like:avg_partner_rating +
##     like:like_o + like_o:prob_o + attr:tvsports + fun:clubbing +
##     attr:clubbing + imprace:movies + like:movies + clubbing:movies +
##     attr:prob_o + tvsports:clubbing + prob:tvsports + tvsports:avg_partner_rating +
```

```
##     like:clubbing, family = binomial(), data = train_model_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.32202  -0.56597  -0.27440  -0.07125   3.03377
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -11.9769     3.3388  -3.587 0.000334 ***
## like                          7.8042     3.7413   2.086 0.036984 *
## like_o                       -4.1701     3.0942  -1.348 0.177754
## prob                          4.5629     1.8108   2.520 0.011743 *
## attr                         -7.6546    11.1708  -0.685 0.493196
## imprace                       2.2230     1.1231   1.979 0.047788 *
## prob_o                        6.2972     2.1720   2.899 0.003741 **
## fun                          -6.4612     5.9831  -1.080 0.280183
## tvsports                     -5.0877     2.2382  -2.273 0.023020 *
## avg_partner_rating           10.8923     2.9731   3.664 0.000249 ***
## clubbing                      1.3230     2.5739   0.514 0.607249
## movies                       -3.5385     2.1326  -1.659 0.097069 .
## like:prob                    -2.3876     2.2707  -1.051 0.293031
## like_o:attr                  42.3154    13.8853   3.047 0.002308 **
## like:avg_partner_rating     -14.3853     3.7821  -3.804 0.000143 ***
## like:like_o                   6.3184     3.1729   1.991 0.046440 *
## like_o:prob_o                -4.0751     1.8991  -2.146 0.031890 *
## attr:tvsports                13.5330     7.7181   1.753 0.079530 .
## fun:clubbing                 21.4531     9.1531   2.344 0.019088 *
## attr:clubbing               -18.5520     8.3521  -2.221 0.026335 *
## imprace:movies               -3.6129     1.3858  -2.607 0.009130 **
## like:movies                   5.6808     2.6073   2.179 0.029348 *
## clubbing:movies               0.5148     1.3284   0.388 0.698338
## attr:prob_o                 -13.9883    10.0082  -1.398 0.162208
## tvsports:clubbing             1.5868     0.9038   1.756 0.079121 .
## prob:tvsports                -2.6118     1.2600  -2.073 0.038184 *
## tvsports:avg_partner_rating   4.0675     2.3560   1.726 0.084275 .
## like:clubbing                -3.3326     2.0098  -1.658 0.097277 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2495.3  on 2721  degrees of freedom
## Residual deviance: 1793.9  on 2694  degrees of freedom
## AIC: 1849.9
##
## Number of Fisher Scoring iterations: 7
```

```
t <-
  test_model_data %>%
  add_rownames() %>%
  filter(rowname %in% index)


ROCRpred <- prediction(predict[!is.na(predict)], t$match)
```

```
ROCRperf <- performance(ROCRpred, 'tpr', 'fpr')
plot(ROCRperf, colorize = TRUE, text.adj = c(-0.2, 1.7))
```