Important Notes

The size of the feature vector is 20. We can think of these 20 numbers as the minimum requirements of ‘information’ needed to get a job in a particular company/college/research group etc (for an Economics student).

I have handpicked these criterions keeping in mind as to what counts the most towards an Economics student getting a job. The ideas used here to make a personalized job classifier setting her presented for an Economics student can be leveraged by people from different fields to make their own personalized job classifiers.

I Criteria description

Criteria-1 (10/10): All criteria descriptions are in reference to this criterion. It has the following blocks in positional order (e.g. line 1 comes before line 2 according to the positional order):

1. If
2. else if
3. else
   1. if
   2. else

‘1’: If( number of 1’s in the current row in the training set is >= 18) -> label = 1.

‘2’: If( number of 1’s in the current row in the training set is <= 14) -> label = 0.

‘3.1’: If(all of these columns(2,4,6,10,12,19) in the current row equal to 1) -> label = 1

‘3.2’: label -> 0

I will refer to these numbers (1, 2, 3, 3.1, 3.2) in all other criteria in single quotes and then further convey what has changed in it in comparison to Criteria1.

Criteria-2(9/10): ‘1’: threshold decreased by 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘one’ condition (Column 19 in tmpAppDatasetConstrutor/ Column ‘T’ in excel file)

Criteria-3(8/10): ‘1’: threshold decreased by 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘two’ conditions (Column 19(T), Column12(M))

Added ‘three’ conditions (Column11(L), Column14(O), Column15(P))

Criteria-4(7/10): ‘1’: threshold decreased by 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘two’ conditions (Column 19(T), Column12(M))

Added ‘two’ conditions (Column11(L)), Column14(O))

Criteria-5(6/10): ‘1’: threshold decreased by 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘two’ conditions (Column 19(T), Column12(M))

Added ‘one’ condition (Column11(L))

Criteria-6 (5/10): ‘1’: threshold decreased by 1, condition added as follows,

* Column4(E) is 1 AND Column6(G) is 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘four’ conditions (Column2(C), Column10(K), Column 19(T), Column12(M))

Added ‘two’ conditions (Column9(J), Column17(R))

Criteria-7 (4/10): ‘1’: threshold decreased by 1, condition added as follows,

* Column3(D) is 1 AND Column6(G) is 1 AND Column9(J) is 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘4’ conditions (Column2(C), Column4(E), Column 19(T), Column12(M))

Added ‘1’ condition (Column3(D))

Criteria-8(3/10): ‘1’: threshold decreased by 1, condition added as follows,

* Column3(D) is 1 AND Column6(G) is 1 AND Column12(M) is 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘4’ conditions (Column2(C), Column4(E), Column10(K) ,Column 19(T))

Added ‘2’ conditions (Column3(D), Column16(Q))

Criteria-9(2/10): ‘1’: threshold decreased by 1, condition as follows,

* Column4(E) is 1 AND Column6(G) is 1 AND Column17(R) is 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘3’ conditions(Column2(C), Column 19(T), Column12(M))

Added ‘1’ condition (Column17(R))

Criteria-10(1/10): ‘1’: threshold decreased by 1, condition added as follows,

* Column3(D) is 1 AND Column6(G) is 1 AND Column12(M) is 1 AND Column16(Q) is 1

‘2’: threshold decreased by 1

‘3.1’: omitted ‘3’ conditions(Coulmn2(C), Coulmn4(E), Coulmn10(K))

Added ‘3’ conditions(Column3(D), Column16(Q), Column14(O) )

References

#--references

# (Strongly Recommended) Use ‘installr’ package to install packages, and a lot of other stuff (e.g. MikTeX (using installr::install.MikTeX(), etc). Don’t install manually (as it is time consuming + due to R’s idiosyncracies): https://cran.r-project.org/web/packages/installr/index.html

# (Recommended)Caret Package vignette https://cran.r-project.org/web/packages/caret/vignettes/caret.pdf

# (Recommended)Caret Package All in one guide: http://topepo.github.io/caret/index.html

# Caret Package webinar slides (quickstart): https://static1.squarespace.com/static/51156277e4b0b8b2ffe11c00/t/5310d223e4b0d21c529a0814/1393611299730/webinar.pdf

# Caret CRAN extensive guide: https://cran.r-project.org/web/packages/caret/caret.pdf

# Caret package, Notes on reproducibility: http://topepo.github.io/caret/model-training-and-tuning.html#notes-on-reproducibility

# Best Machine Learning pacakges in R: https://www.r-bloggers.com/what-are-the-best-machine-learning-packages-in-r/

# Logistic regression in R: https://www.r-bloggers.com/how-to-perform-a-logistic-regression-in-r/

# glmnet package vignette: https://web.stanford.edu/~hastie/glmnet/glmnet\_alpha.html

# neuralnet package: https://www.r-bloggers.com/fitting-a-neural-network-in-r-neuralnet-package/

# ROC and AUC: https://www.r-bloggers.com/illustrated-guide-to-roc-and-auc/

# Bootstraping: https://en.wikipedia.org/wiki/Bootstrapping\_(statistics)

# Resampling Statistics/Cross Validation: https://en.wikipedia.org/wiki/Resampling\_(statistics)#Cross-validation

# Notes on reproducibility: http://topepo.github.io/caret/model-training-and-tuning.html#notes-on-reproducibility

# R tutorial on family of apply functions: https://www.r-bloggers.com/r-tutorial-on-the-apply-family-of-functions/

# Yet another tutorial on the family of apply functions in R: <https://www.r-bloggers.com/apply-lapply-rapply-sapply-functions-in-r/>

# Library paths in R: https://stat.ethz.ch/R-manual/R-devel/library/base/html/libPaths.html

# Google's R style guide: https://google.github.io/styleguide/Rguide.xml

# Hadley Wickham's style guide: <http://adv-r.had.co.nz/Style.html>

# How to set a seed? (Stata): <https://www.stata.com/manuals13/rsetseed.pdf>

# R random number generation (very important information in “Note ” section): https://stat.ethz.ch/R-manual/R-devel/library/base/html/Random.html

# Saving the state of a Random Number Generator in R:

http://www.cookbook-r.com/Numbers/Saving\_the\_state\_of\_the\_random\_number\_generator/

# Interesting documentation Stata(good explanation for various topics): stata.com; <https://www.statalist.org/forums/>; https://www.stata.com/links/

# timing your code: <https://stats.idre.ucla.edu/r/faq/how-can-i-time-my-code/>

# Unit testing of code: <https://en.wikipedia.org/wiki/Unit_testing>

# Legacy code: <https://en.wikipedia.org/wiki/Legacy_code>

# R Markdown + Working with images + figures + tips: <http://www.zevross.com/blog/2017/06/19/tips-and-tricks-for-working-with-images-and-figures-in-r-markdown-documents/#bonus-knitr-and-r-markdown-functionality>

# Use **nbsp;** for putting a vertical space in R markdown documents.

# R shiny refresher (Zev Ross) : http://zevross.com/blog/2016/04/19/r-powered-web-applications-with-shiny-a-tutorial-and-cheat-sheet-with-40-example-apps/

# R Shiny (Reactivity, theory + some e.g’s) : https://shiny.rstudio.com/articles/reactivity-overview.html

# Software testing: <https://en.wikipedia.org/wiki/Software_testing>

# Different Ways to set figure size in R Markdown: https://sebastiansauer.github.io/figure\_sizing\_knitr/