Explore Recruitment Bias

Sravya Gudipudi

22/06/2022

## Summary

Employee recruitment involves the human resource management selection process to choose the most suitable candidate from a group of applicants who have applied for a certain job position. In the process of employee recruitment, unsuitable candidates are selected. However, in most cases, there is a bias in the process of employee recruitment. This can either be unconscious bias or conscious bias. Either conscious or unconscious biases bring a problem to the company since the employees selected might not be up to the task and not talented for the job position the company might require. This is because there might be wrong candidates selected, leaving out the most suitable candidates due to recruitment bias. Therefore, this project will implement a machine learning algorithm that will help in the accurate and reliable selection of the most suitable candidates from a group of applicants and also predict the employees to be shortlisted and the key drivers for shortlisting the most suitable candidates for any given job vacancy. The project will answer the research question: Do age and ethnicity influence the shortlisting of the most suitable candidates?

## Introduction

In the case where the recruitment team in the human resource department has an unchecked bias towards applicants for a given job vacancy, they perpetuate inequality and discrimination. In a workplace, biased hiring, promotion, training, or termination will always lead to inequality and disparity that affects the company negatively in terms of productivity and revenue gain. According to Kidwell et al. (2018), organizations that have practiced biased recruitment have always faced the loss of employee morale, diverse experiences and ideas, and litigation. An organization's unhealthy company culture and high turnover are always because of biased recruitment. Therefore, this paper intends to work on developing a model that can eliminate the recruitment bias for an organization and perform an analysis of the factors that majorly influence recruitment bias.

## Literature Review

In most cases, the entire recruitment team might be biased when selecting the most suitable candidates without their knowledge (Raghavan et al., 2020). According to Raghavan et al. (2020), recruitment bias affects the potential candidates for a given job position negatively in terms of promotion, compensation, and evaluation. Many organizations do not have an already defined process for recruitment. According to Moher et al. (2018), the recruitment process always depends on the job requirements, nature, and type. Organizations always adapt their recruitment process to choose the most suitable applicants for interviews from a group of applicants. This freedom of choice for the recruitment process, therefore, increases the probability of being biased, either conscious or unconscious. Therefore, this project needs to be developed to help solve the problem of recruitment bias.

## Theory

H1: The shortlisting process is influenced by gender and ethnicity.

## Data

The dataset has been retrieved from Kaggle.com website. It includes:

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4

## ✔ tibble 3.1.7 ✔ dplyr 1.0.9

## ✔ tidyr 1.2.0 ✔ stringr 1.4.0

## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──

## ✖ dplyr::filter() masks stats::filter()

## ✖ dplyr::lag() masks stats::lag()

## Loading required package: lubridate

##

## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':

##

## date, intersect, setdiff, union

## Loading required package: PerformanceAnalytics

## Loading required package: xts

## Loading required package: zoo

##

## Attaching package: 'zoo'

## The following objects are masked from 'package:base':

##

## as.Date, as.Date.numeric

##

## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':

##

## first, last

##

## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':

##

## legend

## Loading required package: quantmod

## Loading required package: TTR

## Registered S3 method overwritten by 'quantmod':

## method from

## as.zoo.data.frame zoo

## Type 'citation("pROC")' for a citation.

##

## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':

##

## cov, smooth, var

## Loading required package: bitops

## Rattle: A free graphical interface for data science with R.

## Version 5.5.1 Copyright (c) 2006-2021 Togaware Pty Ltd.

## Type 'rattle()' to shake, rattle, and roll your data.

## Loading required package: rpart

##

## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':

##

## %+%, alpha

## Loading required package: lattice

##

## Attaching package: 'caret'

## The following object is masked from 'package:purrr':

##

## lift

## Registered S3 method overwritten by 'GGally':

## method from

## +.gg ggplot2

##

## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':

##

## chisq.test, fisher.test

## corrplot 0.92 loaded

## Loading required package: carData

##

## Attaching package: 'car'

## The following object is masked from 'package:psych':

##

## logit

## The following object is masked from 'package:dplyr':

##

## recode

## The following object is masked from 'package:purrr':

##

## some

## Loading required package: maditr

##

## To aggregate several columns with one summary: take(mtcars, mpg, hp, fun = mean, by = am)

##

## Attaching package: 'maditr'

## The following objects are masked from 'package:xts':

##

## first, last

## The following objects are masked from 'package:dplyr':

##

## between, coalesce, first, last

## The following object is masked from 'package:purrr':

##

## transpose

## The following object is masked from 'package:readr':

##

## cols

##

## Attaching package: 'expss'

## The following object is masked from 'package:car':

##

## recode

## The following objects are masked from 'package:stringr':

##

## fixed, regex

## The following objects are masked from 'package:dplyr':

##

## compute, contains, na\_if, recode, vars

## The following objects are masked from 'package:purrr':

##

## keep, modify, modify\_if, when

## The following objects are masked from 'package:tidyr':

##

## contains, nest

## The following object is masked from 'package:ggplot2':

##

## vars

## Rows: 280

## Columns: 9

## $ ApplicantCode <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1…

## $ Gender <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…

## $ ATSIyn <int> 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 1, 1, 2,…

## $ ShortlistedNY <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…

## $ Interviewed <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…

## $ FemaleONpanel <int> 1, 1, 1, 2, 2, 2, 2, 1, 1, NA, NA, NA, NA, NA, NA, NA, N…

## $ OfferNY <int> 1, 1, 1, 1, 0, 0, 0, 0, 0, NA, NA, NA, NA, NA, NA, NA, N…

## $ AcceptNY <int> 1, 1, 1, 1, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …

## $ JoinYN <int> 1, 1, 1, 1, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …

First we check if there are any missing values in the dataset.

sapply(recruitment\_hr, function(x) sum(is.na(x)))

## ApplicantCode Gender ATSIyn ShortlistedNY Interviewed

## 0 0 0 0 0

## FemaleONpanel OfferNY AcceptNY JoinYN

## 225 225 252 252

There are columns with missing values. We remove rows with the missing values.

recruitmenthr <- recruitment\_hr[complete.cases(recruitment\_hr), ]

sapply(recruitmenthr, function(x) sum(is.na(x)))

## ApplicantCode Gender ATSIyn ShortlistedNY Interviewed

## 0 0 0 0 0

## FemaleONpanel OfferNY AcceptNY JoinYN

## 0 0 0 0

## Methodology

With data having been cleaned, next will involve working with the dataset for exploratory and descriptive data analysis.

First, we gain insights into the number of applicants by their Gender.

recruitment\_hr %>%

select(ApplicantCode,Gender)%>%

group\_by(Gender) %>%

summarize(n = n()) %>%

ungroup() %>%

mutate(Percent = n / sum(n)\*100) %>%

mutate(Gender = factor(Gender, labels = c("Male","Female"))) %>%

adorn\_totals("row")

## Gender n Percent

## Male 78 27.85714

## Female 202 72.14286

## Total 280 100.00000

Next, we perform analysis to identify the number of applicants by their residence in the location of the company.

recruitment\_hr %>%

select(ApplicantCode,ATSIyn)%>%

group\_by(ATSIyn) %>%

summarize(n = n()) %>%

ungroup() %>%

mutate(Percent = n / sum(n)\*100) %>%

mutate(ATSIyn = factor(ATSIyn, labels = c("Aboriginal Torres Strait Islander","General")))%>%

adorn\_totals("row")

## ATSIyn n Percent

## Aboriginal Torres Strait Islander 121 43.21429

## General 159 56.78571

## Total 280 100.00000

Next we perform analysis on the number of accepted candidates and the rejected candidates.

recruitment\_hr %>%

select(ApplicantCode,ShortlistedNY)%>%

group\_by(ShortlistedNY) %>%

summarize(n = n()) %>%

ungroup() %>%

mutate(Percent = n / sum(n)\*100) %>%

adorn\_totals("row")

## ShortlistedNY n Percent

## 0 192 68.57143

## 1 88 31.42857

## Total 280 100.00000

Next, an analysis on the number of candidates interviewed is performed.

recruitment\_hr %>%

select(ApplicantCode,Interviewed)%>%

group\_by(Interviewed) %>%

summarize(n = n()) %>%

ungroup() %>%

mutate(Percent = n / sum(n)\*100) %>%

mutate(Interviewed = factor(Interviewed, labels = c("Not Interviewed","Interviewed")))%>%

adorn\_totals("row")

## Interviewed n Percent

## Not Interviewed 225 80.35714

## Interviewed 55 19.64286

## Total 280 100.00000

To decide on whether there was fairness in gender, we perform analysis on the number of Females on the interview panel.

recruitment\_hr %>%

select(ApplicantCode,FemaleONpanel)%>%

na.omit()%>%

group\_by(FemaleONpanel) %>%

summarize(n = n()) %>%

ungroup() %>%

mutate(Percent = n / sum(n)\*100) %>%

mutate(FemaleONpanel = factor(FemaleONpanel, labels = c("Male Only","Female Panel member")))%>%

adorn\_totals("row")

## FemaleONpanel n Percent

## Male Only 33 60

## Female Panel member 22 40

## Total 55 100

Analysis on ethnicity and shortlisting of the applicants.

tbl = table(factor(recruitment\_hr$ATSIyn,labels = c("Aboriginal Torres Strait Islander","General Applicant")),factor(recruitment\_hr$ShortlistedNY,

labels = c("Not Shortlisted","Shortlisted")))

addmargins(tbl)

##

## Not Shortlisted Shortlisted Sum

## Aboriginal Torres Strait Islander 102 19 121

## General Applicant 90 69 159

## Sum 192 88 280

Analysis on Gender and shortlisting of the applicants.

tbl = table(factor(recruitment\_hr$Gender,labels = c("Male","Female")),factor(recruitment\_hr$ShortlistedNY,

labels = c("Not Shortlisted","Shortlisted")))

addmargins(tbl)

##

## Not Shortlisted Shortlisted Sum

## Male 40 38 78

## Female 152 50 202

## Sum 192 88 280

Create a logistic regression model for predicting the shortlisting of applicants.

set.seed(7)

lr <- glm(ShortlistedNY ~ Gender + ATSIyn,family=binomial("logit"),data = recruitment\_hr)

summary(lr)

##

## Call:

## glm(formula = ShortlistedNY ~ Gender + ATSIyn, family = binomial("logit"),

## data = recruitment\_hr)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -1.4371 -0.9342 -0.4764 0.9382 2.1130

##

## Coefficients:

## Estimate Std. Error z value Pr(>|z|)

## (Intercept) -1.2432 0.6777 -1.834 0.0666 .

## Gender -1.1957 0.3006 -3.978 6.94e-05 \*\*\*

## ATSIyn 1.5157 0.3096 4.895 9.83e-07 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for binomial family taken to be 1)

##

## Null deviance: 348.59 on 279 degrees of freedom

## Residual deviance: 306.61 on 277 degrees of freedom

## AIC: 312.61

##

## Number of Fisher Scoring iterations: 4

## Results

The analysis provided information regarding the objective of this paper. First, it has been identified that for the applicants who applied for the job position, there were more Female applicants than Male applicants. Female applicants occupied 72.1% of the total applicants.

## Gender n Percent

## Male 78 27.85714

## Female 202 72.14286

## Total 280 100.00000

For ethnicity distribution among the applicants, 43.2% were Aboriginal or Torres Strait Islander.

## ATSIyn n Percent

## Aboriginal Torres Strait Islander 121 43.21429

## General 159 56.78571

## Total 280 100.00000

Therefore, to ensure that there is no gender or ethnicity bias, there should be a fair portion of 72% females shortlisted and 43.2% applicants from Aboriginal or Torres Strait Islander shortlisted.

However, the analysis results for ethnicity, the shortlisted individuals from Aboriginal Torres Strait Islander accounted for 15%. That means the shortlisting was more on the general applicants.

##

## Not Shortlisted Shortlisted Sum

## Aboriginal Torres Strait Islander 102 19 121

## General Applicant 90 69 159

## Sum 192 88 280

Results also show that 56.7% of the females were shortlisted. There was a high number of rejections from 72% of the total applicant females to 56.7% of shortlisted females, which suggests a bias towards male applicants.

##

## Not Shortlisted Shortlisted Sum

## Male 40 38 78

## Female 152 50 202

## Sum 192 88 280

## Implications

To help solve the problem of recruitment bias, I would recommend from the results of this paper that the recruitment department in any organization should implement a strategy that ensures that the ethnic background, gender, and name are hidden from the recruiter reviewing the job application.

## Conclusion

From the data analysis performed in this paper, it is clear that ethnicity and gender bias exist in the recruitment process of shortlisting suitable candidates from a pool of interested applicants. Therefore, the theory that the shortlisting process is influenced by gender and ethnicity is true.

# References

Kidwell, R. E., Eddleston, K. A., & Kellermanns, F. W. (2018). Learning bad habits across generations: How negative imprints affect human resource management in the family firm. Human Resource Management Review, 28(1), 5-17.

Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. PLoS biology, 16(3), e2004089.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 469-481).