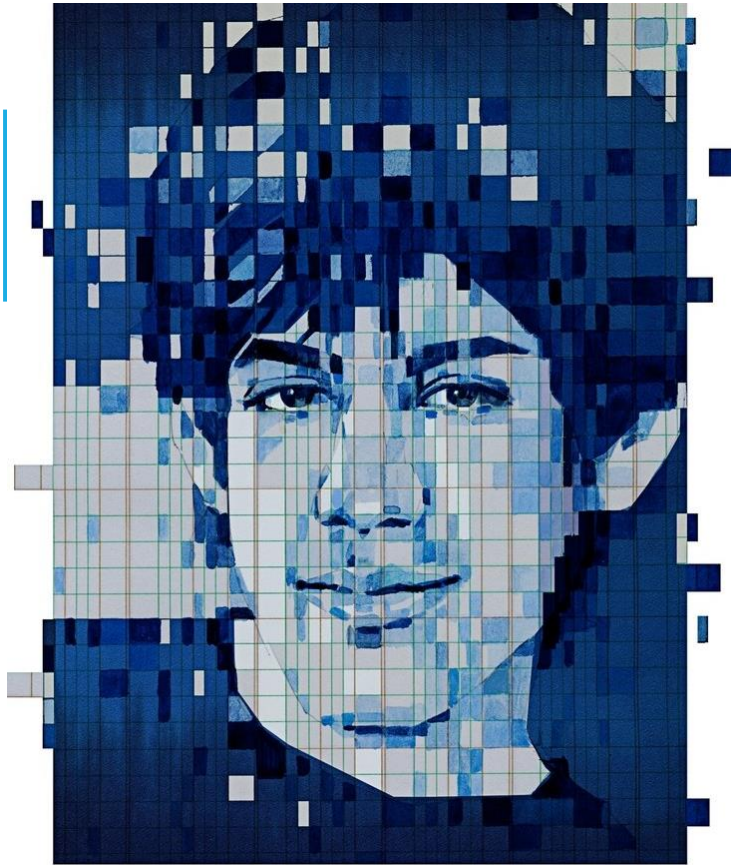


USING PREDICTIVE MODELS TO DETERMINE AMERICAN TECH WORKERS' LIKELIHOOD TO DISCUSS MENTAL HEALTH AT WORK

Miriam Callahan

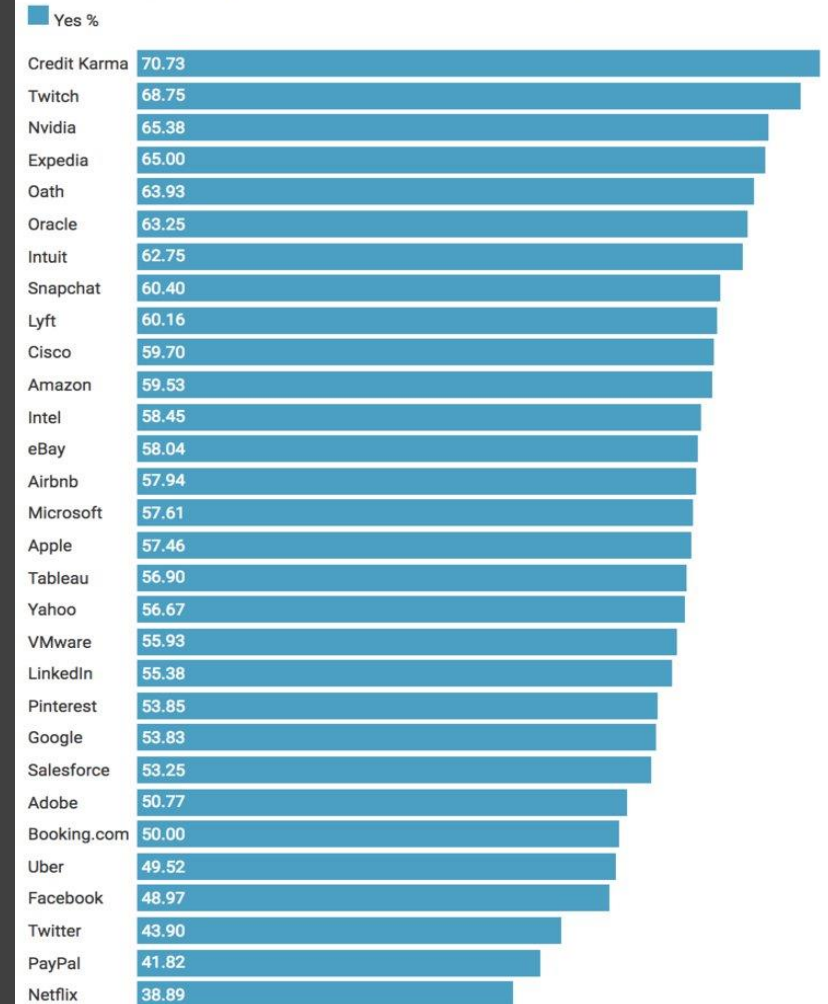


“This, I suppose, is the actual problem,” Swartz wrote, long before his suicide. “I feel my existence is an imposition on the planet.”

WHY IT MATTERS

Tech Workers Suffering From Burnout By Company

Users on Blind were asked the following survey question: Are you currently suffering from job burnout? Participants were able to answer YES or NO. This chart shows the results for the 30 companies that had the most employee responses.



Blind is an anonymous social app for tech employees. The survey ran from May 12, 2018 through May 21, 2018. A total of 11,487 Blind app users participated. At least 40 employees had to submit a response in order for a company to be considered for the top 30 list.

Source: [TeamBlind.com](https://www.teamblind.com)

Data was collected from 1 259 tech workers internationally in between August 27-28th, 2014 answering a total of 26 questions (features).

Only the 746 participants who lived in the US were used for model training and testing since many countries worldwide were severely underrepresented or not represented.

Of the 26 original features, only 15 & 21 remained after feature selection.

The likelihood a participant would discuss their mental health with coworkers and/or their supervisor were used as the target features.

ABOUT THE DATASET

GOAL OF DEPLOYMENT

Since mental health issues have such a large effect on the tech industry within the US and having a support system is associated with having a better well-being within a work environment, I decided to test how willing employees would be to discuss mental health issues with their coworkers and supervisors. This model aims to determine levels of transparency within discussions of mental health in the workplace and factors that may play into increasing this transparency and subsequently diminishing mental health issues, improving workplace satisfaction, and improving employee retention.



```
graph LR; A[Employer promotes a 'culture of health'] --> B[Increased employee satisfaction]; B --> C[Benefits for company]
```

Employer
promotes a
'culture of
health'

Increased
employee
satisfaction

Benefits for
company

LINKS IN PROCESS

ALGORITHMS USED

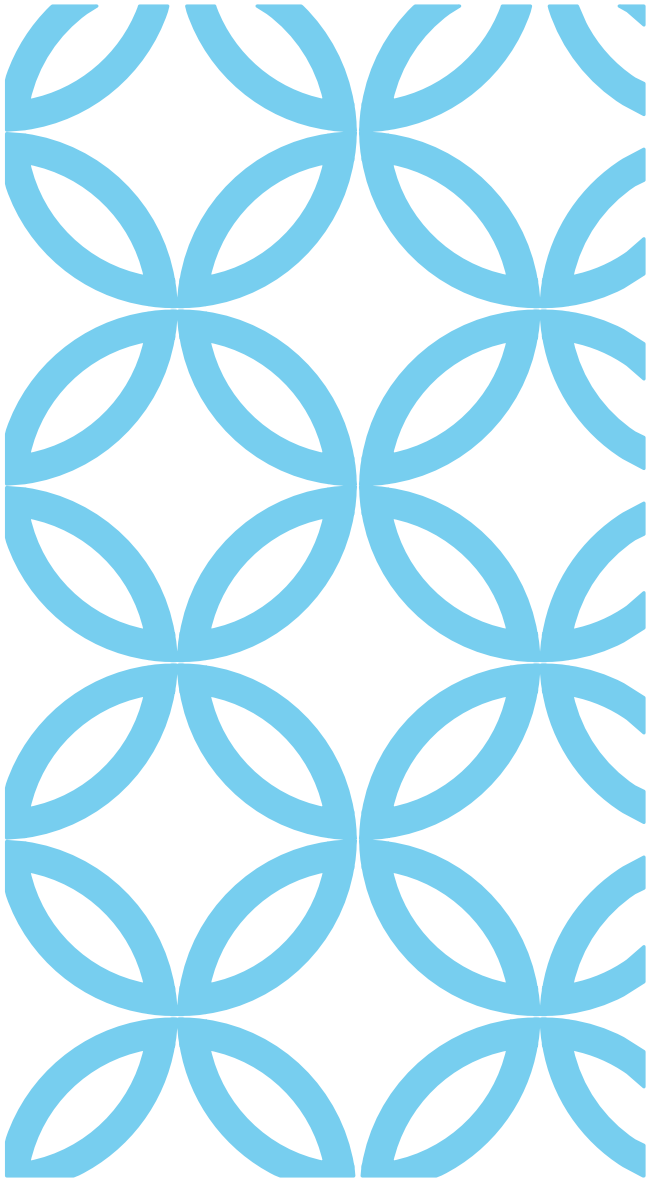
ONE R

NAÏVE BAYES

REPEATED INCREMENTAL PRUNING TO
PRODUCE ERROR REDUCTION (RIPPER)

C4.5

RANDOM FOREST



`remote_work`: Do you work remotely (outside of an office) at least 50% of the time?

`mental_health_interview`: Would you bring up a mental health issue with a potential employer in an interview?

`physical_health_interview`: Would you bring up a physical health issue with a potential employer in an interview?

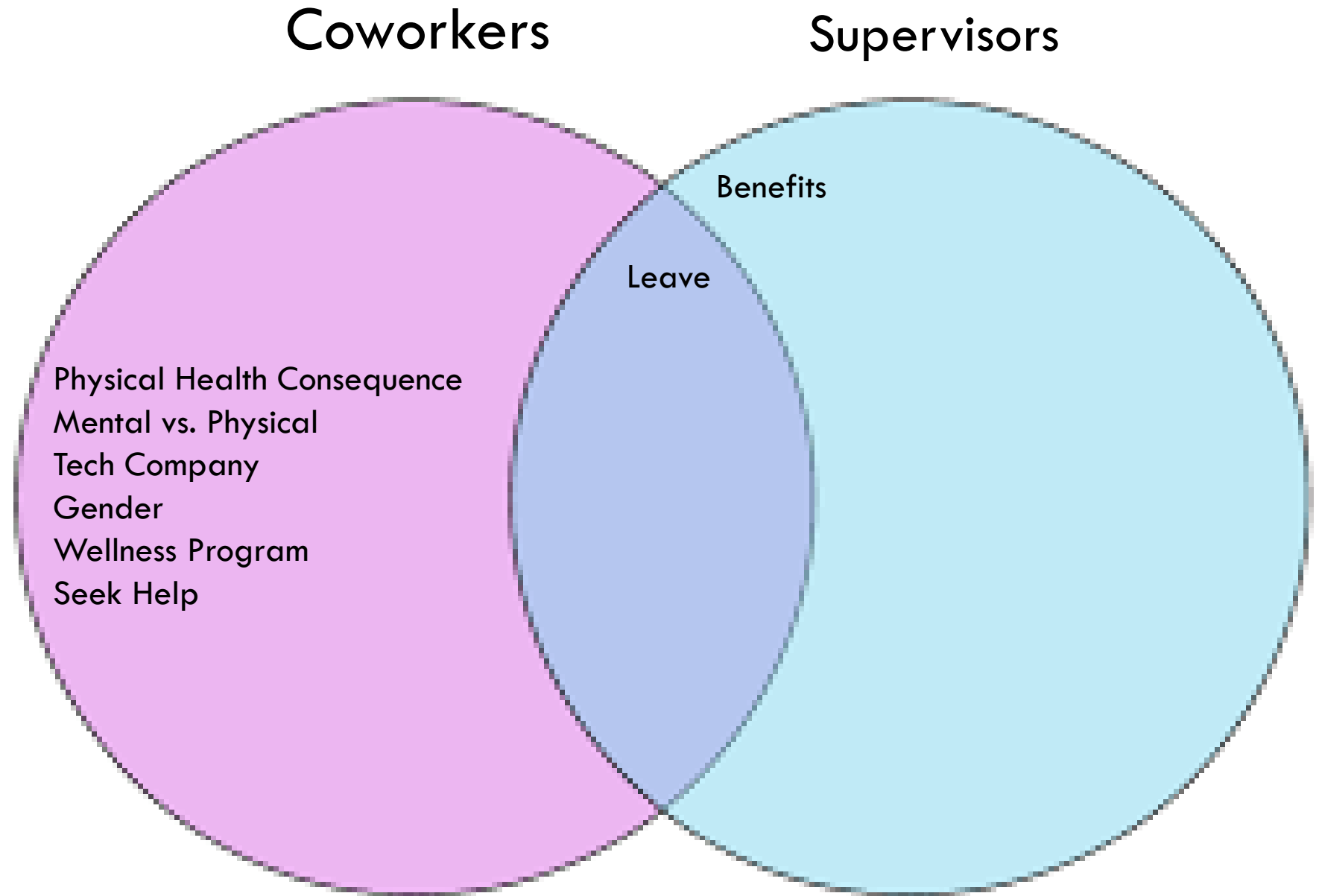
`mental_vs_physical`: Do you feel that your employer takes mental health as seriously as physical health?

`obs_consequences`: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?

FEATURES IN DATASET

DATA PREPARATION

Data was run through all five algorithms with a 60/40 split to test for which features were most important in model selection. The following features were removed from the datasets. Testing was done with 10 fold-cross validation.



BEST SUPERVISOR MODEL: C4.5



Accuracy: 58.16%



Number of Leaves: 94



Size of Tree: 120

BEST COWORKER MODEL: RIPPER

Accuracy: 61.63%

Rules: 5

(mental_health_consequence = No) and
(mental_health_interview = Maybe) and (treatment = Yes)
=> coworkers=Yes (29.0/10.0)

(mental_health_consequence = No) and (Age >= 39) and
(benefits = Yes) => coworkers=Yes (18.0/6.0)

(mental_health_consequence = Yes) and (no_employees =
More than 1000) => coworkers=No (55.0/21.0)

(mental_health_consequence = Yes) and (remote_work = No)
and (Age >= 33) => coworkers=No (24.0/10.0)

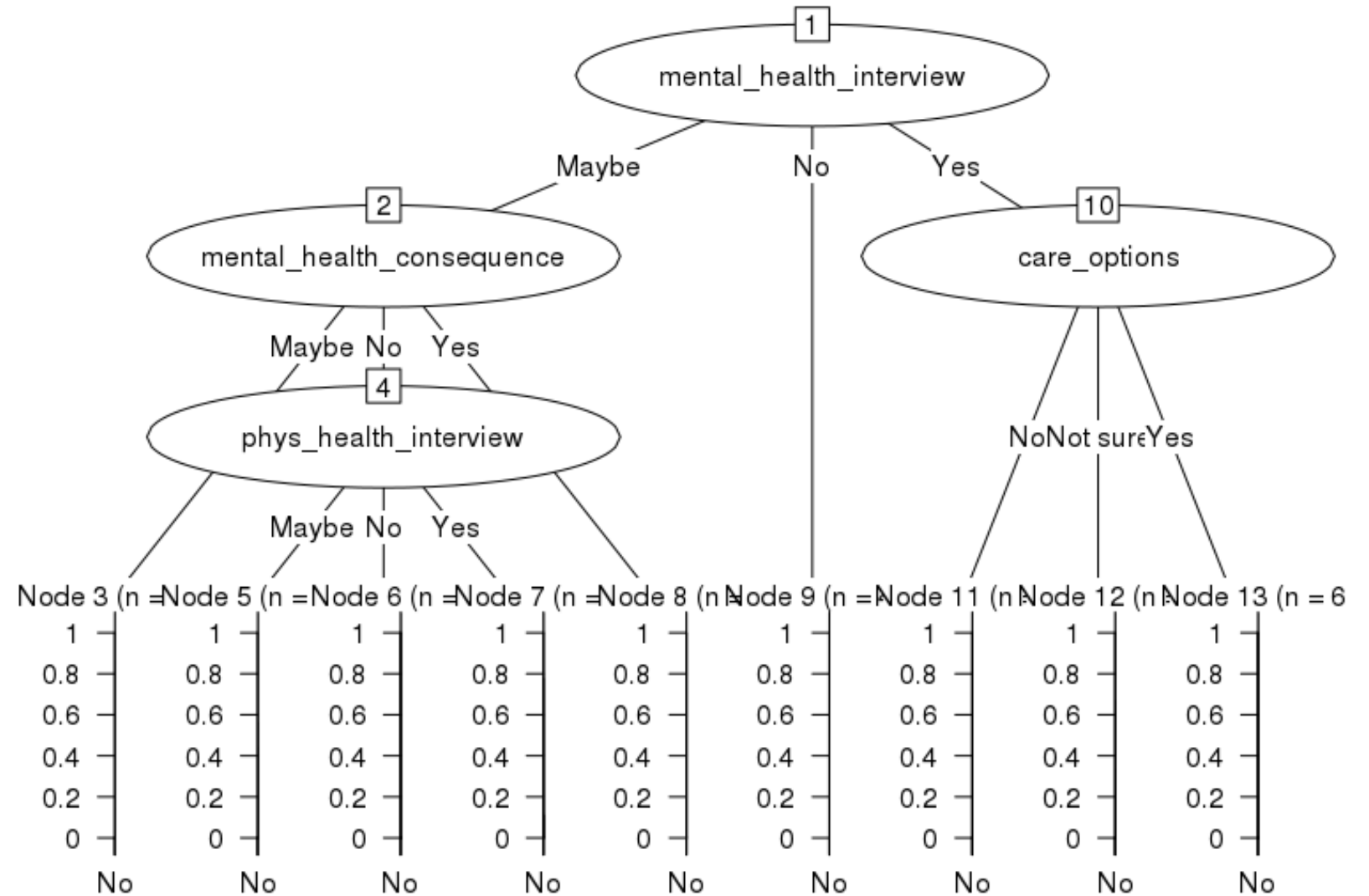
=> coworkers=Some of them (458.0/134.0)

BEST COWORKER MODEL: C4.5

Accuracy: 63.60%

Number of Leaves: 9

Tree Size: 13



IMPORTANCE OF FEATURE SELECTION

ALGORITHM - SUPERVISOR	BEFORE	AFTER
OneR	57.53%	56.63%
Naïve Bayes	57.86%	56.17%
RIPPER	54.62%	55.11%
C4.5	61.91%	58.16%

ALGORITHM - COWORKERS	BEFORE	AFTER
OneR	53.85%	57.45%
Naïve Bayes	41.81%	62.15%
RIPPER	73.67%	63.61%
C4.5	75.63%	63.60%

Overall, feature selection appeared to have an adverse effect on the supervisor data. While it did improve the coworker data for OneR and Naïve Bayes (likely due to there being fewer variables to consider), accuracy was significantly decreased with the C4.5 algorithm and RIPPER.

Look

Look into why feature selection had an adverse effect & test different feature combinations

Use

Use a later year (this survey is given out annually)

Examine

Examine other tree algorithms/
use greedy search

FUTURE CONSIDERATIONS

REFERENCES

Fagan, K. (2018, June 04). Employee Burnout Is Real. A Survey of More Than 11,000 Tech Workers Reveal Where It's Worst. Retrieved from <https://www.inc.com/business-insider/employee-burnout-survey-blind-app>

MacFarquhar, L., & MacFarquhar, L. (2017, June 19). The Darker Side of Aaron Swartz. Retrieved from <https://www.newyorker.com/magazine/2013/03/11/requiem-for-a-dream>