What causes Job/Career satisfaction in Computer Science and would you write unethical software?

William Pembleton

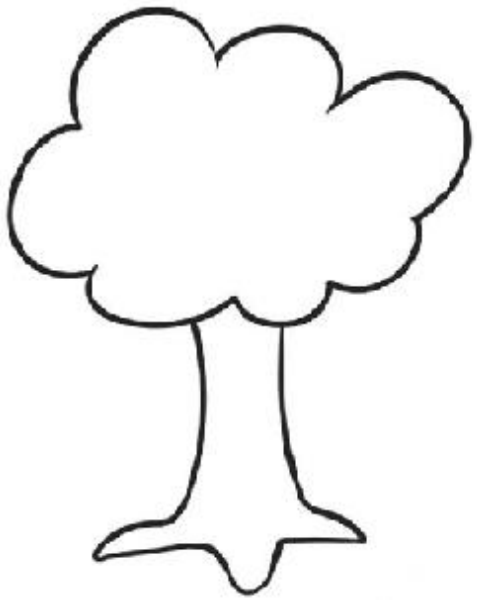
Abstract

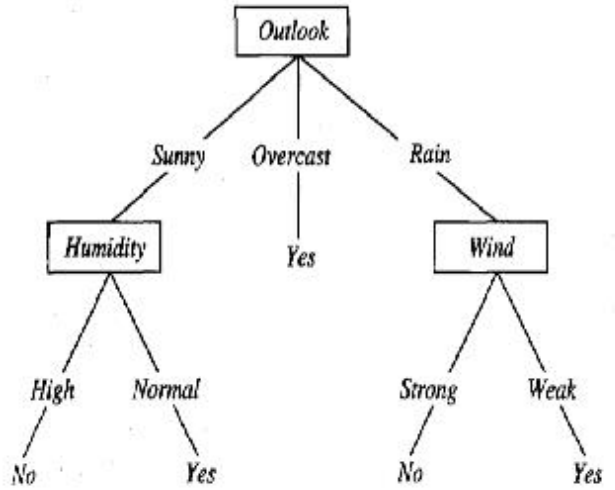
What's the most interesting is what didn’t show up in the trees

- Age
- Open source
- How long coding (professionally/non-professionally)
- What time you wake up
- Hours spent on the computer in a day
- Exercise
- How many meals do you skip in a week

What’s J48?

J48 is a decision tree learning algorithm. It can take in a dataset and produce trees





Where’d I get my data from?

Stack Overflow is a Q&A website where developers can learn from each other

Every year they release a survey that asks developers question about their job

What do they ask about?

- What kind of developer are you? (Full stack, back end, mobile)
- How satisfied are you with your current job?
- What languages do you use? (Java, Python, Pascal)
- What IDE do you use? (Eclipse, IDLE, Sublime)

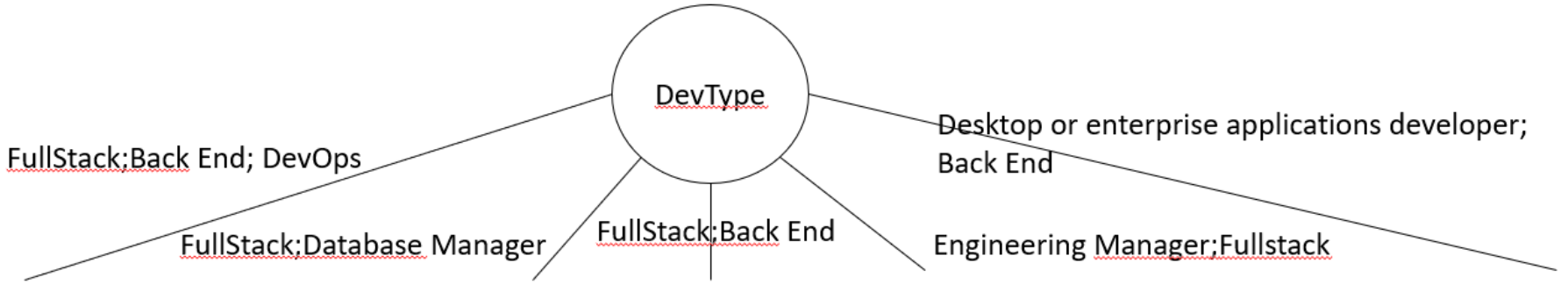
Methodological (Preprocessing)

- They asked questions like “Which of the following developer types describe you? Please select all that apply.”
- The way Stack Overflow stored this is something like “Full Stack; Database Administrator; DevOps”
- Separated each of these out into their own column

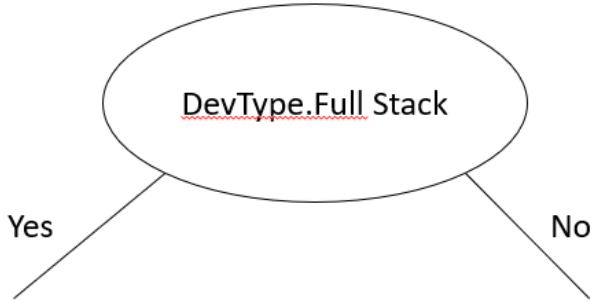
DevType.Full Stack	DevType.Database Adminstrator	DevType.Devops
Yes	Yes	Yes
Yes	No	Yes

Methodological (Preprocessing)

- This is so that the tree doesn't turn into something like this with a ton of nodes

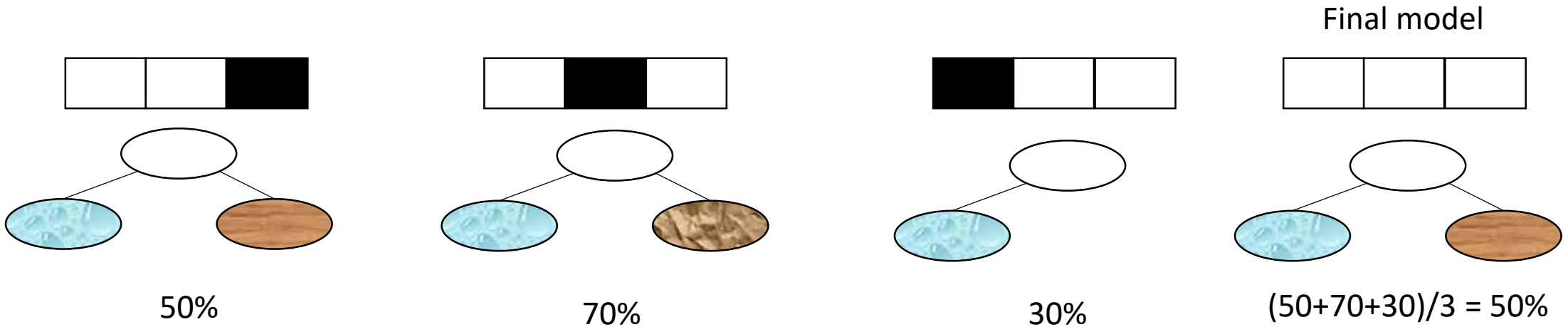


- Now the tree will look like this



K-Fold Cross Validation

- K-Fold cross validation is a way to grade how well a model did at it's given task without spoiling any data
- Split the data into k sections (Folds), save one fold for testing and use the others to train the model on
- Repeat k times until you have k models and get an accuracy for each

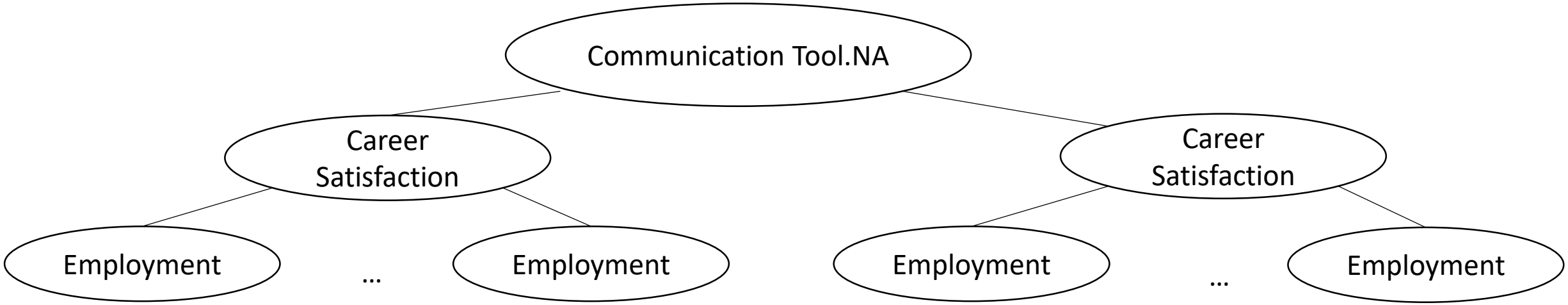


Confusion Matrix

- A confusion matrix is an output that you can build with supervised learning algorithms
- It is a way to tell what an algorithm thinks of a particular example vs what it actually is

```
=== Confusion Matrix ===

     a     b     c     d   <-- classified as
 36932  4184   108   217 |     a = No
 17568  8114    68   181 |     b = Depends on what it is
    67    21 27985     0 |     c = NA
  1936   984    17   473 |     d = Yes
```

Results (Job Satisfaction) Simplified



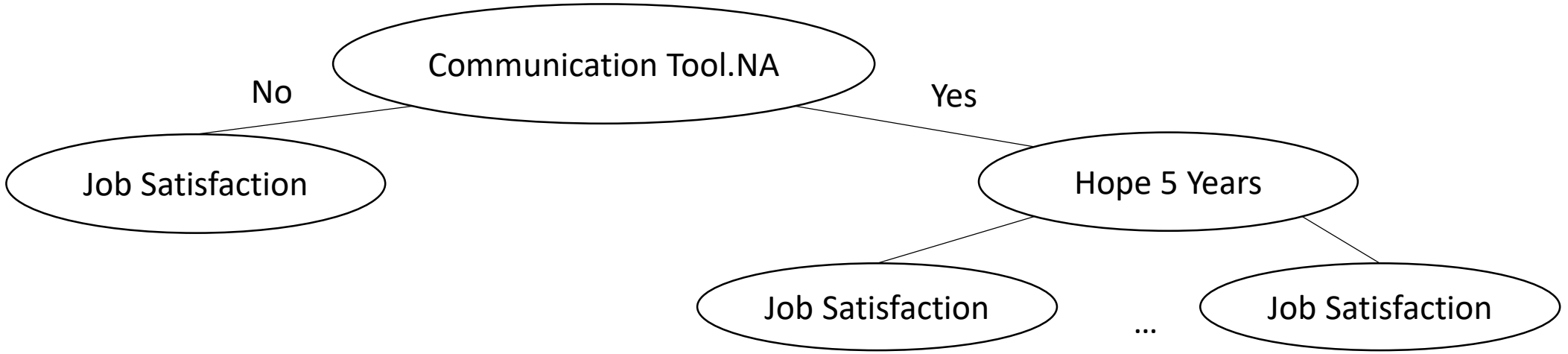

```
=== Confusion Matrix ===

    a     b     c     d     e     f     g     h   <-- classified as
 4383   146  7566   112    20     1   108   100 |     a = Extremely satisfied
   58  2212  1855   442   104     3  1060   584 |     b = Moderately dissatisfied
 2095   547 21402   788   144     2   861   166 |     c = Moderately satisfied
   80   394  2270  1232   112     0   791    87 |     d = Neither satisfied nor dissatisfied
  211   457  7108   878   246     1  1012    99 |     e = Slightly satisfied
   73    50   273    60    16 29028    42    37 |     f = NA
  103   935  3240   718   153     0  1729   179 |     g = Slightly dissatisfied
   60   602   537    90    65     4   245   879 |     h = Extremely dissatisfied
```

Confidence Factor: 1/2^15
Accuracy: 62%

Job Satisfaction

- As mentioned in the abstract it’s the stuff that isn’t in the tree that’s interesting
- Age
- Open source
- How long coding (professionally/non-professionally)
- What time you wake up
- Hours spent on the computer in a day
- Exercise
- How many meals do you skip in a week

Results (Career satisfaction) Simplified




```
=== Confusion Matrix ===

    a     b     c     d     e     f     g     h   <-- classified as
 6482   610  6561    97    98    59   281   128 |     a = Extremely satisfied
  148  2054  3063   229   159    90   479    94 |     b = Neither satisfied nor dissatisfied
 4381  1641 19718   365   458    95   975   293 |     c = Moderately satisfied
  110  1066  3437   525   164    23  1023   239 |     d = Slightly dissatisfied
  704  1556  9177   375   579    50   814   229 |     e = Slightly satisfied
    0     0     3     0     0 22348     0     0 |     f = NA
  145   570  1761   254    48    32  1924   528 |     g = Moderately dissatisfied
  106   305   674    57    28    20   588   835 |     h = Extremely dissatisfied
```

Confidence Factor: 1^-7
Accuracy: 55%

Career Satisfaction

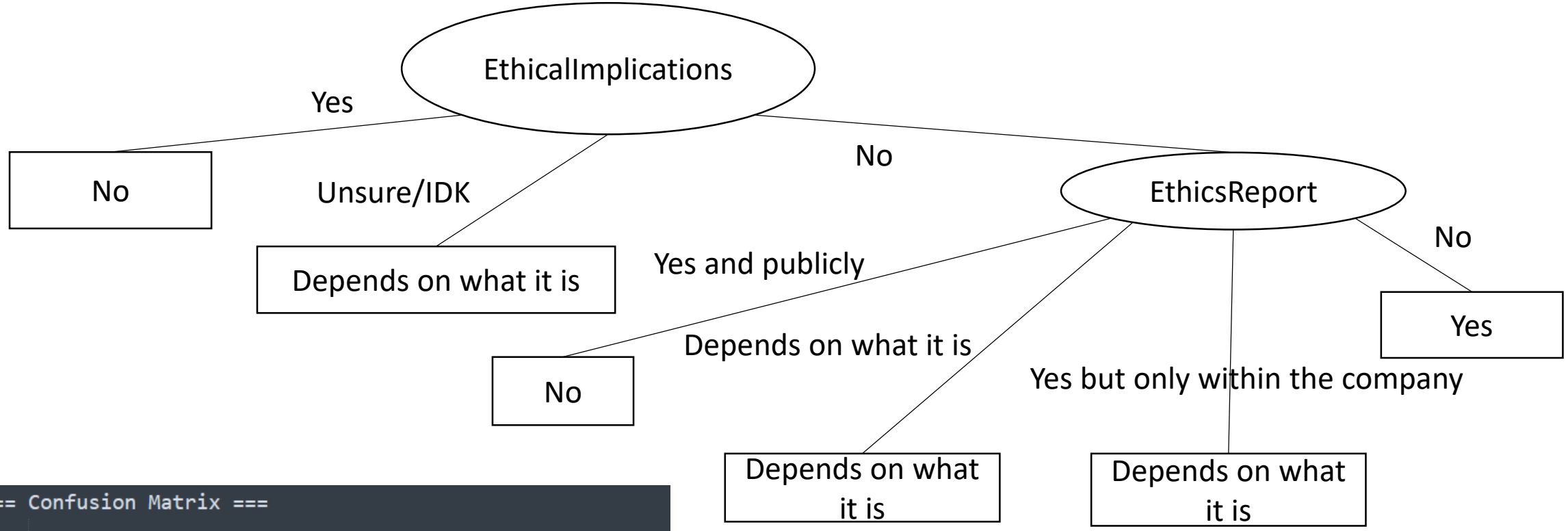
Once again, it’s the stuff that didn’t show up that’s the most interesting

- Age
- Language shows up this time but only in circumstances where the number of examples is small (usually < 300).
- Open source
- How long coding (professionally/non-professionally)
- Wake time
- Hours on the computer
- Exercise
- How many meals do you skip in a week

Ethics Choice variables

- EthicsChoice: Imagine that you were asked to write code for a purpose or product that you consider extremely unethical. Do you write the code anyway?
- EthicsReport: Do you report or otherwise call out the unethical code in question?
- EthicalImplications: Do you believe that you have an obligation to consider the ethical implications of the code that you write?

Results (Ethics choice) Simplified




```
=== Confusion Matrix ===

     a     b     c     d   <-- classified as
 36932  4184   108   217 |     a = No
 17568  8114    68   181 |     b = Depends on what it is
    67    21 27985     0 |     c = NA
  1936   984    17   473 |     d = Yes
```

Confidence Factor: 1/2^14
Accuracy: 74%

Future Research

- Look at the times where people are satisfied with their career and not satisfied with their job and vice versa and see if there is something that causes a person to be happy with their career but not their job.
- Build a tree without career satisfaction for the job satisfaction tree and seeing how much accuracy changes. Likewise, do so for the career satisfaction tree.
- Do a full study and ask participants what they think causes their job/career satisfaction.