What causes job satisfaction in Computer Science?

William Pembleton

Abstract

- To be written at the end of the project

Abstract (cont.)

- To be written at the end of the project

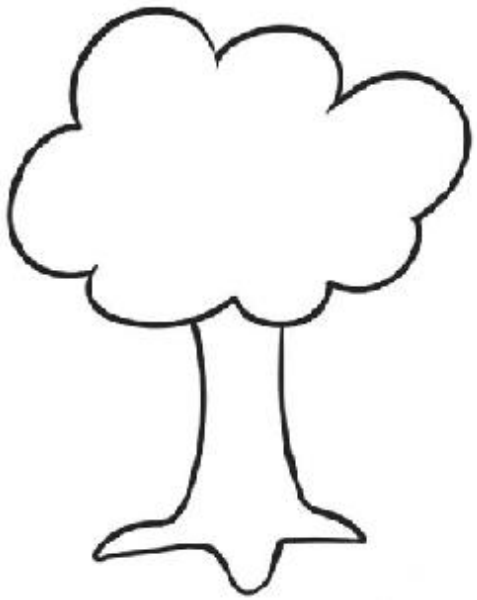
Background

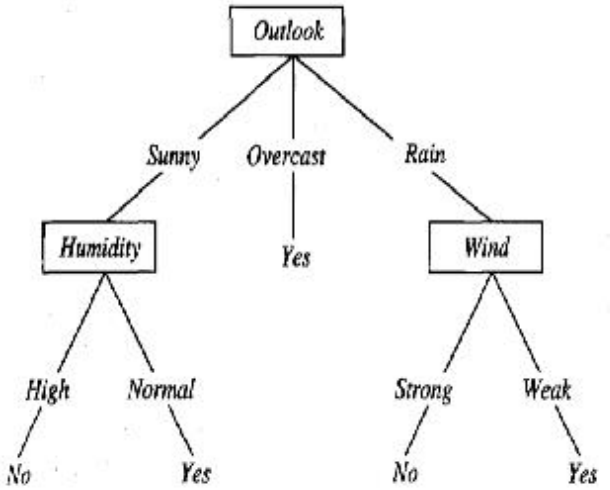
- This question is a supervised learning problem, meaning that I can use an algorithm like ID3.

- Supervised learning solves problems that answer the question of why is my data like this

What’s ID3?

ID3 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

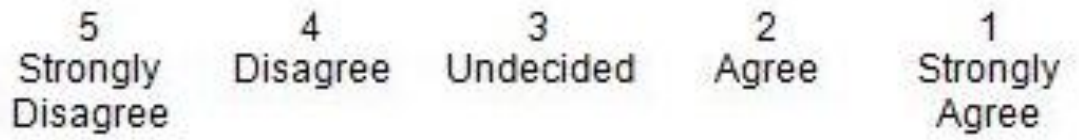
Every year they release a survey that asks developers question about their job

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?
- What IDE do you use?

Methodological (Preprocessing)

- Likert scale



- For instance take a question like “Are you extremely satisfied with your current job?” The answers they gave answers similar to what’s above
- I converted the answers the dataset gave into numeraical values so that if an answer is undecided I made it into a 3
- Mostly did this to make the tree easier to read and easier to calculate averages later on

Methodological (Preprocessing)

- They also 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 Adminstrator;DevOps”

- Separated each of these out into their own column

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

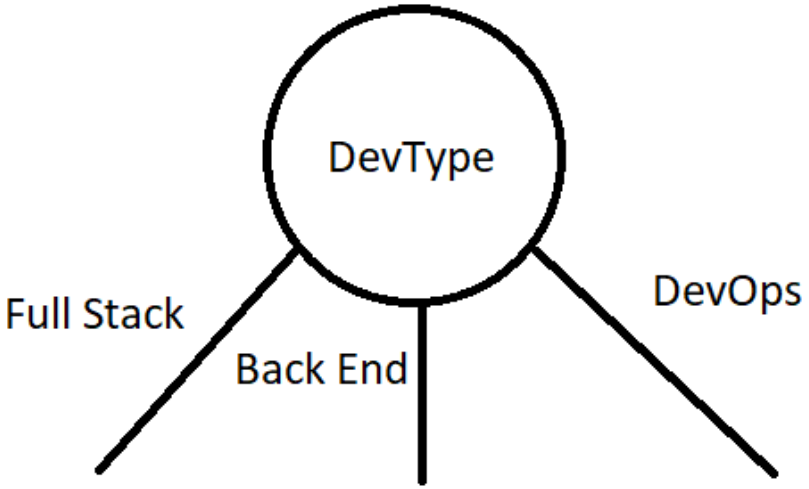
Methodological (Preprocessing)

- There were a few questions on the survey that asked participants to rank (for instance what benefits are important to you in a job) on a scale from 1-10
- I coded this similar to the last slide Benefits.Industry

- Lots of people didn’t answer this question so I made their values NaN

Methodological (Algorithm changes)

- Within the code I took in stuff like DevType.FullStack and made it so that when the tree is asked to print itself it outputs something like the picture below



Methodological (Algorithm changes)

- I added support for the Likert scale when a class is chosen that has a Likert scale for its values
- Meaning that when you get to a leaf node it will tell you what it thinks your job satisfaction will be from 1-5
- Another small change I made to the algorithm is allowing it to handle values of NaN (So I don’t have to throw away these rows)

Results (Job Satisfaction)

- Imagine a very pretty tree

Results (Language)

- Pretty tree #2

Results (Open Source)

- Pretty tree #3

Results (Student)

- I think get the idea