What Makes a Great Hire

And can a computer help tell us?

Machine Learning and Resumes

Agenda

How important are resume skills versus other parts of the resume?

Can we use machine learning to identify great hire attributes?

- Do we have time to properly analyze and evaluate each resume?
- Are there high performing candidate features that might be non-obvious to us?
- Are we limited by the rules we can (manually) create that identifies a great hire?
  › Can we input all high performing insurance agent resumes and let the computer help decide which features are important and which features we should avoid in a new hire?
- As always, recruiter and hiring team has the last word
  › This shortens the process by only sending the best candidates to recruiters for review
  › This should uncover highly desirable candidate features and help recruiters recruit the candidates

Is machine learning for us?

- Is it mainstream enough? Do we need math degrees? Data science employees? Can we leverage cloud computing? Are there privacy concerns? EEO concerns?

Can we use machine learning for lead generation?

- Further segment our leads and build more custom messaging (improve conversion rates)
- Further support our agents in communities with local information (conversation starters)

Is Skill the Overall Driver for Making a Great Hire?

Indeed skills classifications

- Skills weighted, based on density of keywords and skill location in resume
  - More desirable skill combinations/mixes identified by project team
  - Feeder skills in addition to insurance skills identified by project team
  - Undesirable skills identified by project team

Types of skills identified

- Industry skills: Insurance and/or feeder (banking, retail, etc...)
- Personal skills: Independent worker, self starter, time management, etc...
- Work skills: Customer service, sales, etc...
- Character skills: We or I, sentiment, language level, etc...

Are we running out of insurance candidates?

- Do we need to start looking for more candidates in feeder positions/companies?
- Are candidates with previous insurance experience actually the best candidates?

The Nature of a Candidate

How did the candidate evolve?

- Did candidate find oneself early or late?
  - › Changing industries, changing jobs, changing skills - malleable? advantage for us?
  - › One industry, one type of job, stable – will be committed? handle adversity?

How is the candidate connected?

- Did the candidate stay in one town/region for entire career?
  - › Go to college in local community – know many alumni
  - › All career changes in the same town – use connections to sell insurance

Is the candidate successful?

- Would insurance represent a big jump in prestige/salary? Deepen commitment?
- Work for Fortune 500 company or small businesses? Is small better?
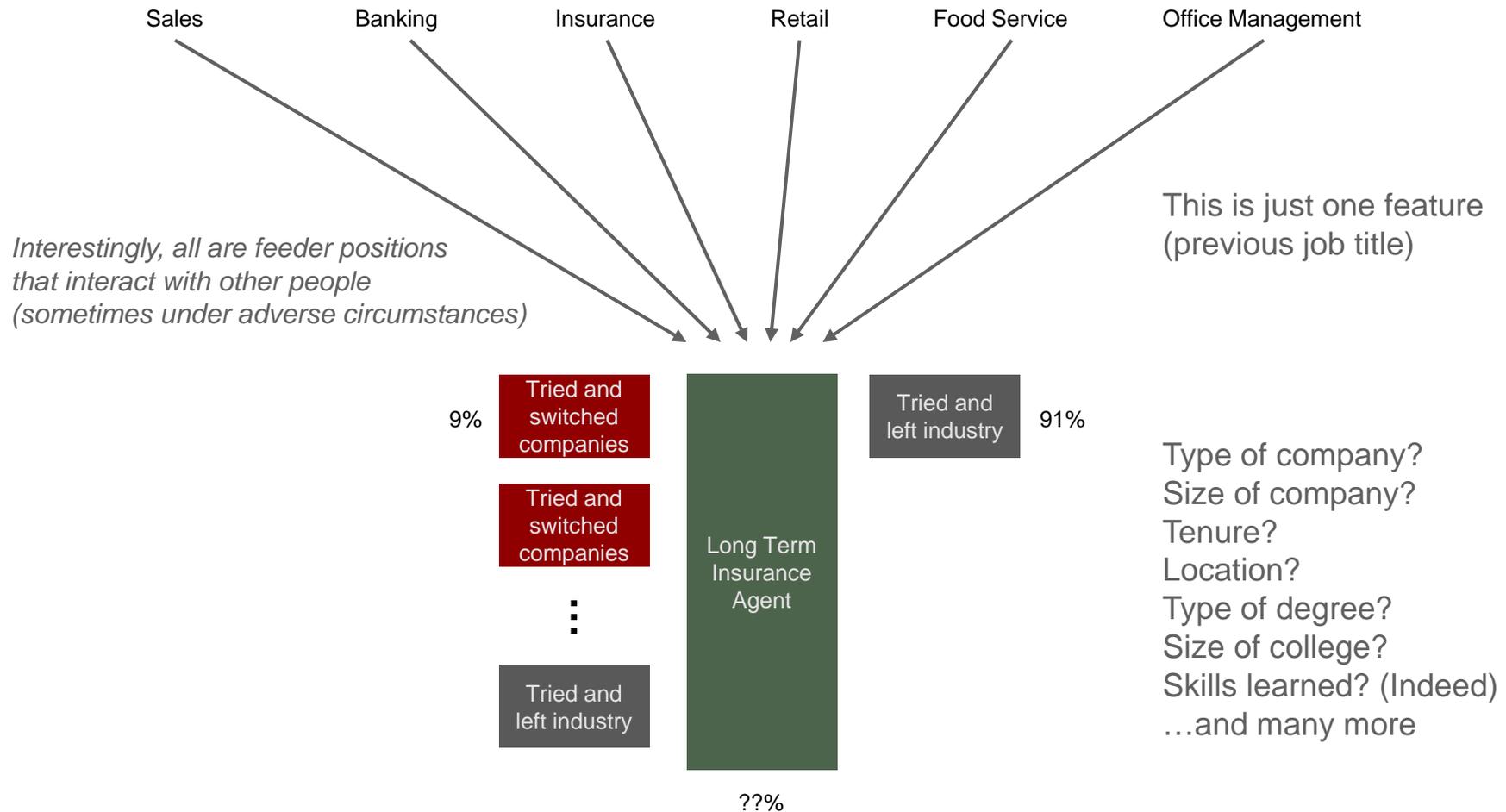- Multiple positions at the same company? Promotions?

Are there factors outside the resume that might be important?

How important is candidate personality? As important as skills?

Do we know all that makes a great hire? Can we reduce turnover through better candidate selection?

Which of These Occupations Makes for a Great Hire?

Limited supply of candidates with exact insurance qualifications and 90% of those leave profession

Sales   Banking   Insurance   Retail   Food Service   Office Management

Interestingly, all are feeder positions that interact with other people (sometimes under adverse circumstances)

This is just one feature (previous job title)

9%   **Tried and switched companies**

Tried and switched companies

⋮

Tried and left industry

Long Term Insurance Agent

Tried and left industry   91%

??%

Type of company?
Size of company?
Tenure?
Location?
Type of degree?
Size of college?
Skills learned? (Indeed)
…and many more

Deductive or Inductive

Candidate selection currently is rules-based – a recruiter reviews the resume

- For example, previously worked as an insurance agent (but 90% leave after first job)
- Banking, finance, sales, customer service skills a plus
- The limits of rules
  › How do we write a rule for selling door-to-door? Ability to get appointments?
  › How do we write a rule for prospect rudeness on the front porch?
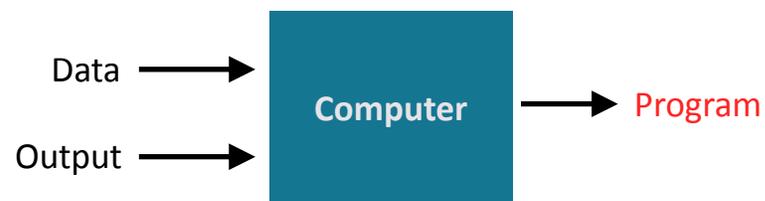 › How do we write a rule for sticking with it after early discouragements?

What if we could infer a great candidate?

- Use resumes of high performers to go find more high performing resumes?

Rules-based (deductive)

Data ⟶ **Computer** ⟶ Output

Program ⟶

Machine Learning (inductive)

Data ⟶ **Computer** ⟶ Program

Output ⟶

Machine Learning

50+ features extracted from resumes

- Feature selection/extraction from unstructured text the hardest part of the project
- Skills, including weighted skills and skill combinations are some of the model features
- Both high performers (e.g. 3+ years as agent) and non-performers identified
- Look for correlations between features and high performance as an agent
  - Multiple models used - decision trees, classifiers (beyond skills), neural networks
  - Some features may be calculated – e.g. average time at employer

Both metadata and resume text analyzed for features

- Positions, length of time in position, company, degree, etc… all metadata
- Written text holds clues on leadership, skills, self-starter, skills, domain knowledge

Training set and test sets used to determine model effectiveness

- Once model is chosen, simply a matter of feeding it new resumes
- Model outputs the probability new candidate will be a high performing agent

Machine learning is non-discriminatory

- Features based on race, gender and disability never fed to the models

Isn't Machine Learning High Tech Only?

Cloud computing is overtaking client/server

- Cloud companies are mainstream
  - › Amazon has over 1 million cloud customers
  - › Microsoft, IBM and Google all large cloud companies and all have machine learning built-in
- Many Fortune 500+ companies are moving entire infrastructures to cloud
  - › Scale of efficiencies are staggering, upgrades, security, privacy, etc all centrally managed

Cloud computing is now more that renting disk space

- High powered database software now in the cloud
- Machine learning models now reside in the cloud
  - › Very simple to use interfaces with no math or programming knowledge needed
  - › Different providers offer different models (hundreds of options) - hard not to find solution

Hardware as important as software in machine learning

- Need very fast computers and parallel processing
- "Flex" as needed - early on, we need a lot of power, later less (pay as you go)

We can be sure that our competitors will not pass up using machine learning

Using Machine Learning to Market Position to Candidate

Parts of resume holds clues to what makes a great hire

- Auto-generate an email to prospect and mention resume specifics
  - › You are great at working with people, especially under adverse circumstances
  - › You have lived in the same location for a long time and know everyone
  - › This would represent a huge upside in salary and prestige
  - › You have skills handling financial data
- Candidate starts to believe s/he can really make the jump - has the skills
  - › Improves candidate response rate to position posting
  - › Strong marketing message and follow-up on best candidates
- Invest more in top-tier candidates
  - › Special programs for those who can go the distance - mentors? more check-ins?
  - › Perhaps increased training? More patience with early setbacks?

Other Machine Learning Applications

Lead Generation

- Demographic and environmental data about prospects
  - Open government data - property taxes, home sales, etc…
  - Resumes - what makes a great hire becomes what makes a great lead
  - Local news that may drive a heightened awareness/need for insurance (local fire)
- Custom messaging
  - Message to prospect living in $500K home different from $100K home
  - Use machine learning to further segment our leads into smaller groups
  - Let machine learning pick the right message for each customer segment
- Digital advertising
  - Use machine learning to analyze cookies and determine best spots for ads
  - Local news events drive "bursts" of ad buys in local markets
- Insurance agent support
  - News and local events analyzed and sent to agents as conversation starters
  - Daily email on need-to-know events going on in local/regional market

Next Steps

Leverage and expand upon Indeed resume project
- All skills identified by Indeed become features fed to machine learning models

Run an experiment to see if model can be found
- Data science (outside) team to do feature extraction on 10MM current resumes
- Set thresholds for identifying high performing agents (time in position?)
- Create Amazon cloud account and load dataset
  › Only features go to the cloud, not entire resumes
  › Data can be blinded (no contact information), but not necessary (secure environment)
  › Need to load equal amounts of high and low performers so models can distinguish
  › Both a training set and test set are used to determine model effectiveness

Recruiting or lead generation as first project?
- Recruiting can decide go/no go, based on Indeed work
- If wait, highly recommend we use it for lead generation and agent support