Littler

International Association of Rehabilitation Professionals

Data-Driven Hiring
The hype, the hope, the promises, and the pitfalls

November 1, 2019

1

Matthew U. Scherer, J.D.
Littler Mendelson P.C.
Portland, Oregon
mscherer@littler.com
503.889.8881

Littler
Employment & Labor Law Solutions Worldwide®

2

Agenda

- Technological Background (10-15 minutes)
  - Three Concepts: Big Data, Algorithms, and Machine Learning (ML)
- ML in Hiring (20-25 minutes)
  - Basic concept
  - How's it's being used now
  - Promises and pitfalls
  - Comparison to traditional employee selection
- Legal considerations (15 minutes)
  - Privacy laws
  - Antidiscrimination laws
- Questions (30 minutes)

3

Background: Overview

Three related concepts: Big Data, Algorithms, Machine Learning

4

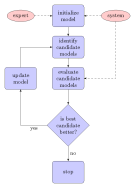
Background: Big Data

- **A field of data science relating to data sets that are too large to be analyzed using traditional data-processing software**
  - Can be thought of as spreadsheets with potentially thousands of columns and millions of rows
  - Microsoft Excel has a limit of ~1 million rows per sheet
- **Can refer to:**
  - The data sets themselves
  - Tools that can be used to analyze such data sets
  - Insights can be drawn from them

5

Background: Algorithm

- **Any process or set of rules designed to solve a problem or perform a task**
- **Example: A decision tree (see right)**
- **Most commonly used to describe computer programs that perform a series of calculations and output something that a human can use (a prediction, recommendation, assessment, etc)**



6

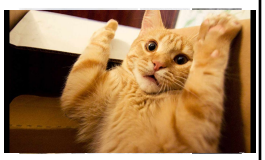
Background: Machine Learning

- **A type of computer program (using that term loosely) that leverages algorithms and statistical techniques to "learn" from data**
  - "Learn" does not mean it "understands" in the way a human does—just that as it gets exposed to more and more data, it gets better and better at the analytical task it is asked to do
- **"Deep learning" is a type of machine learning that uses neural networks**
  - As name implies, neural networks are inspired by the way neurons in the brain are thought to interact with each other
- **Type of machine learning at issue here is called "supervised learning"**
  - The machine "learns" by being fed labeled examples and looking for correlations between the data and the labels

7

Background: Machine Learning

- **Example: Program designed to recognize cats in photographs**
  - Two labels:
    - "This image has a cat in it"
    - "This image does not have a cat in it"
  - Over time, starts noticing things that tend to occur more frequently (e.g., certain colors, textures, and shapes) in photos with cats
- **But—what if it's only fed pictures of orange tabby cats?**



8

A Motivating Example: Big Data in Politics

- **The best way to figure out a person's political views would be to ask them directly**
- **But it's not possible to interrogate every single voter**
- **So what do they do instead? Sample.**
  - Ask some voters a series of questions--who they've voted for in the past, who they intend to vote for next year
  - Also ask those voters for demographic information--age, gender, race, income, education level, etc
- **Even though you can't ask every voter directly, you can ask enough of them to get a pretty good idea—a model—of how those demographic characteristics connect to voting patterns**
- **So you could then say that a person in zip code 97211 with a college degree and an income between $50k and $100k per year has a 60% probability of voting for the Democratic nominee next year.**
- **You can use those predictions to decide who to send political mailers to, who could be drawn into the district, etc**

9

A Motivating Example: Big Data in Politics

- The best way to figure out a person's political views would be to ask them directly
- But it's not possible to interrogate every single voter
- So what do they do instead? Sample.
  - Ask some voters a series of questions--who they've voted for in the past, who they intend to vote for next year
  - Also ask those voters for demographic information--age, gender, race, income, education level, etc
- Even though you can't ask every voter directly, you can ask enough of them to get a pretty good idea—a model—of how those demographic characteristics connect to voting patterns
- So you could then say that a person in zip code 97211 with a college degree and an income between $50k and $100k per year has a 60% probability of voting for the Democratic nominee next year.
- You can use those predictions to decide who to send political mailers to, who could be drawn into the district, etc

10

Data-Driven Hiring: Basic Idea

- Take large sets of data with information about past candidates and current employees
- Use machine learning to build a model that relates different characteristics of those candidates to job performance
- Use that model to predict how well future job applicants will perform in a particular position

11

Data-Driven Hiring: Basic Idea

- Slide: Look at resumes alone
- Presents the fewest risks in terms of privacy
- Limits the amount of potentially irrelevant information considered
- Slide: Resumes + other existing/static data
- Resumes and publicly available data sources
- Resumes and publicly available data sources and proprietary data obtained from data vendors
- Each additional layer of data raises privacy risk and also risk of relying on irrelevant information
- At the same time, uses potentially relevant information that a candidate did not include in their application
- Slide: Look at some combination of the above and video interview assessments
- Software that claims to be able to perform assessments of candidate characteristics based on visual and audio signals during a video interview
- Usually proprietary software--not easy to know how well they actually work
- Slide: "Chatbots"
- Allows employers to elicit specific information from a candidate
- E.g., determine whether the candidate possesses the bare minimum qualifications for a position
- Often used in combination with the other forms of ML-based learning discussed above

12

Data-Driven Hiring: Resume/Application-Based

- **Presents the fewest risks in terms of privacy**
- **Limits the amount of potentially irrelevant information considered**

13

Data-Driven Hiring: Resume + Other Existing Data

- **Options:**
  - Resumes and publicly available data sources
  - Resumes and publicly available data sources and proprietary data obtained from data vendors
- **Each additional layer of data raises privacy risk and also risk of relying on irrelevant information**
- **At the same time, uses potentially relevant information that a candidate did not include in their application**

14

Data-Driven Hiring: Chatbots

- **Allows employers to elicit specific information from a candidate**
- **E.g., determine whether the candidate possesses the bare minimum qualifications for a position**
- **Often used in combination with the other forms of ML-based learning discussed above**

15

Data-Driven Hiring: Video Interviews

- **Software that claims to be able to perform assessments of candidate characteristics based on visual and audio signals during a video interview**
- **Usually proprietary software—not easy to know how well they actually work**
- **As we'll see, that could pose some legal risk**

16

Data-Driven Hiring: Promise

- **Insights from non-obvious factors—challenging conventional wisdom**
- **Reducing human bias**
  - Machines can't be bigots
  - They are never lazy--will always look at all available information on an applicant, not rely on stereotypes

17

Data-Driven Hiring: Potential Pitfalls

- **Only as good as the data it's trained on**
  - If you're relying on past hiring decisions and performance evaluations made by biased humans, those biases will essentially be "baked into" the data
- **Don't have data on people who don't get hired**
  - Cannot know how well those people would have performed
  - Makes it risky to rely on hire status alone—but often, that's all you have!

18

Data-Driven Hiring: Comparison to Traditional Employment Tests

- **Traditional employment tests**
  - Could be paper-and-pencil, oral assessment, or task-based
  - Types
    - Job-specific tests (ideal)
    - Aptitude tests that test general skills
    - Knowledge tests
- **Employment tests are built around KSAOs—Knowledge, Skills, Abilities and Other Characteristics**
  - Idea is you identify the KSAOs important to a particular position and then test to see if the candidate possesses them
- **Of course, the process of designing such a test can be expensive, and no test can completely capture all the traits relevant to a particular job**

19

Data-Driven Hiring: Comparison to Traditional Employment Tests

- **It is very difficult to use existing data to directly test for KSAOs**
  - Resumes can tell you whether a candidate says they have a particular KSAO
  - Other data may tell you whether the candidate has characteristics that correlate with having a particular KSAO
  - Might even have some usable data that sheds light on whether a candidate possessed a particular KSAO in the past
    - E.g., passing Board exams tells you that a doctor had the requisite knowledge of the state-of-the-art in a particular medical field as of the date it was taken
- **But it's very difficult to adequately capture KSAOs for a particular job from existing data alone**

20

Legal Considerations: Privacy

- **Fair Credit Reporting Act (FCRA)**
  - One of the circumstances it applies to is using information "for employment purposes"
- **General Data Protection Regulation (GDPR)**
  - Limitations on what info can be obtained
  - Extensive requirements of consent
  - Disclosure upon request
  - "Right to an explanation"
    - Highly problematic
- **State-level legislation: California Consumer Privacy Act (CCPA)**
  - Many similar laws proposed in legislatures across the country
- **This is a fast-moving field. Bottom line: consult with counsel before collecting or using data on a candidate outside resumes and other voluntarily provided application materials.**

21

Legal Considerations: Disparate Treatment Discrimination

- **Disparate treatment means explicitly treating members of one protected group better than another**
  - E.g., "we only hire men for this position"
- **Easy to remove *explicit* race and gender (etc) info when building a model**
- **But disparate treatment can come in through a side door**
  - Encoding biases of past recruiters/supervisors
  - Redundant encoding of protected characteristics
  - "Norming" to correct disparate impact

22

Legal Considerations: Disparate Impact

- **Disparate impact is when you use facially neutral criteria that have the *effect* of disproportionately excluding members of a protected class**
  - E.g., imposing physical fitness requirements may disfavor women or some individuals with disabilities
- **Three-step analysis**
  - Is there a disparate impact in a statistical sense? (if no, employer wins)
  - If so, is the hiring practice that created the disparate impact job-related and consistent with business necessity? (if no, employee wins)
  - If so, did the employer have a comparably effective and less discriminatory alternative available to it? (if yes, employee still wins)

23

Legal Considerations: Disparate Impact

- **Disparate impact is when you use facially neutral criteria that have the *effect* of disproportionately excluding members of a protected class**
  - E.g., imposing physical fitness requirements may disfavor women or some individuals with disabilities
- **Three-step analysis**
  - Is there a disparate impact in a statistical sense? (if no, employer wins)
  - If so, is the hiring practice that created the disparate impact job-related and consistent with business necessity? (if no, employee wins)
  - If so, did the employer have a comparably effective and less discriminatory alternative available to it? (if yes, employee still wins)

24

Legal Considerations: Disparate Impact

- **If a test does have a disparate impact, employer must demonstrate job-relatedness to escape liability**
  - This generally requires a formal validation study showing that the selection procedure actually measures traits that are relevant to job performance
- **Validation studies have *always* been expensive and difficult—with Big Data and ML, they will often be wholly impractical**

25

Legal Considerations: Disparate Impact

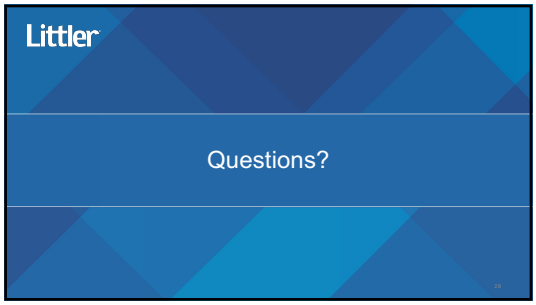
- **Disparate impact is when you use facially neutral criteria that have the *effect* of disproportionately excluding members of a protected class**
  - E.g., imposing physical fitness requirements may disfavor women or some individuals with disabilities
- **Three-step analysis**
  - Is there a disparate impact in a statistical sense? (if no, employer wins)
  - If so, is the hiring practice that created the disparate impact job-related and consistent with business necessity? (if no, employee wins)
  - If so, did the employer have a comparably effective and less discriminatory alternative available to it? (if yes, employee still wins)

26

Antidiscrimination Laws: Threading the Needle

- **The big challenge: Avoiding disparate impact without engaging in disparate treatment**
- **It's actually really easy to tell an algorithm "make sure that the model you give me has at least 30% non-white candidates"**
- **The problem is that such quotas might be unlawful disparate treatment**
- **Likewise, if you design an algorithm that systematically rates non-whites 10% lower than whites, it's easy to program the algorithm to adjust the scores so that the different groups have the same average score**
- **But that kind of "race-norming" is also probably a form of unlawful disparate treatment**

27



28

29