SMM639 – People Analytics
Overview of the People Analytics

Dr Simone Santoni Simone.Santoni.1@city.ac.uk and Prof Vali Asimit asimit@city.ac.uk, Business School, City, University of London, UK

- Formalise People Analytics (PA)
- PA Applications Taxonomy
- Some PA Ethical Issues
- More Concrete PA Applications
- Algorithmic Bias
- Algorithmic Fairness

PA Definition

PA Definition 1:

PA is the use of people-related data to improve and inform various layers of management that aims to improve the overall effectiveness of an organisation and use of its very own social capital.

PA Definition (cont'd)

PA Definition 2:

PA is the field of people-related data that aims to enhance the decision-making within an organisation with respect to its past, present and future workforce in order to achieve the objectives of an organisation such as financial objectives, brand culture, (strategic) talent acquisition/retention etc.

A more colloquial *PA* definition would be adding business value to talent management beyond standard HR metrics, benchmarking & reporting.

Beyond any PA Definition

1. What makes people join, perform well in, and stay with an organisation?
2. Identify, diversify, and reallocate talent within an organisation
3. Improve employee engagement and empowerment
4. Workforce planning
5. Use internal and external sources of data to make prediction of ‘success‘
6. Decisions beyond ‘gut feelings‘ that are robust and fair, legally and morally correct

Trends and Related Applications

1. HR/Workforce/Talent Analytics – an excellent use case on job transfers within the IBM corporation and sport analytics
2. *Quantified Employee* approach
    - wearable devices
    - fitness trackers
    - sociometric badges
3. Understanding the (latent) organisation silos
4. Assessing the human dimension of organizational outcomes and processes – e.g., performance, engagement, leadership, workplace dynamics, organizational developmental support, and learning and knowledge creation
5. Taking stock of regulatory and legal implications

Deliverables

The usual deliverable are as follows

D1) Descriptive insights

D2) Diagnostic

D3) Predictive analytics

D4) Prescriptive analytics

D5) Actionable insights

D6) Limitations and Disclaimers

Challenges

1. Identify identifiable issues and use ‘good‘ metrics
2. Effective communication of unsettling issues
3. Build trust during the implementation
4. PA successful projects require engagement with multiple functions of an organisation
5. Human-based decisions – that profoundly affect the lives and livelihoods of people – cannot be eliminated
6. Scarce adoption of PA – *Understand the market so that the market indeed starts understanding the PA capabilities!*

Just because something could be measured, it does not mean that it should be inadequately quantified & used

Three ethical implementations phases: data collection, model building and deployment of a model on people

- Granular information vs privacy
- Does it have any direct & indirect impact?
- *Ethics values*: **transparency**, **fairness**, **privacy**, **responsibility**, **social good**, etc.
- *Algorithmic reproducibility*
- *Algorithmic explainability*
  - **knowledge and control** of goals, internal processes and outcomes
  - **transparency** of goals, internal processes and outcomes
  - **accountability** of goals, internal processes and outcomes

Beyond Bold Statements and Factoids

1. **Moneyball** and **Liverpool FC**
2. Online career assistant
3. Footfall analysis and front desk talent acquisition
4. Assistance for online contract negotiations in hospitality industry
5. Recruitment modelling – ML-based recruitment model, pre-hiring recruitment model and NLP-based generic recruitment model
6. Organizational culture analysis & communication

PA is not even a *factoid* – at large scale – nor a *buzzword*, but becomes closer to

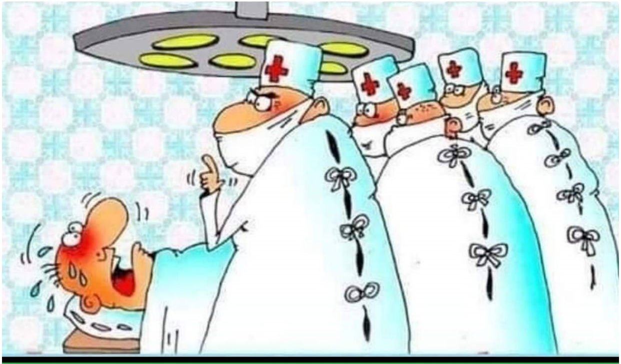
Equality, Equity & Diversity

Analytics for Social Good

1. Recidivism prediction – an interesting open source use case
2. Child welfare
3. Score prediction – loan/mortgage
4. High-school dropout & A-level dropout & University dropout
5. Predicting future (student) enrolment
6. Targeted student advising

Analytics for Social Good (cont'd)

7. Adaptive learning and recommendation systems for students



Doc, are all of them trainee surgeons?

It is true my dear patient, but don't worry, negative grading will be their punishment, so sleep tight!

Definition

Best case scenario: Understand and acknowledge the lack of ‘similar‘ treatment of an algorithm for seemingly similar pieces of data.

Worst case scenario: Understand and mystify the systemic mistreatment of an algorithm for obvious similar pieces of data.

Everything else is possible.

What could one achieve? **Answer:** Detect and adjust

Algorithmic Bias/Discrimination vs **Algorithmic Robustness**

Evaluation

Start with reasonable and appropriate assumptions

- Training data are ‘true‘ and ‘fair‘ representation of reality
- Eliminate vicious circles in data collection
- Use ‘good practice‘ for feature selection and feature engineering
- Keep in mind the true value proposition
- Envelope calculations may lure the modeller into future unconscious biases

Bias by Design

Example 1

Due to the operational issues, the Education Regulatory Body (ERB) cancels the standardised assessment and higher education prediction performance is now evaluated via teachers' assessments and high school/college features. ERB produces scores for individual students via a ML algorithm that aims to not create any disparity between state and private education, so this feature is removed from the feature space, and relies on features such as *class size, historical school's academic performance, school's postal code etc.*

Discuss each choice! Which feature should be kept and why?

Bias amongst Sub-populations

Assume that the feature space contains at least one categorical feature that is socially sensitive. Then, some formal definitions of bias measurements are as follows:

- **Statistical parity** – same predictions (amongst each sub-population)
- **Equal mistreatment/Equalised Odds** – parity amongst those from the same class
- **Predictive parity** – same proportion within each sub-population (amongst 'correct' predictions)
- **Proportional parity** – take into account the imbalance amongst sub-populations
- **Fair opportunity** – context related

Bias amongst Sub-populations (cont'd)

Definition with Supervised Learning

Assume that we observe a sample of size n

$$\left\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_n, y_n)\right\} \subseteq (\mathcal{X}, \mathcal{Y}),$$

where $\mathcal{X} \in \Re^p$ is the feature space and $\mathcal{Y}$ is the target variable set. Note that $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})$ are also known as *independent/input variables* or *covariates* or *attributes* or *predictors*. Moreover, y_i's are also known as *dependent/label/output/response/target variables*.

Linear regression: $y \sim \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p$

y is a continuous variable, i.e. infinitely many possible values;

Classification: $y \sim f(x_1, x_2, \cdots, x_p)$ and $y \in \{0, 1\}$.

Bias amongst Sub-populations (cont'd)

The estimation aim depends on the modelling tool:

1. *Statistics* – fit the response $\hat{y}$ for any given feature $\mathbf{x}$, but based on the entire observed sample of size n;
2. *Machine Learning* – fit the response $\hat{y}$ for any given feature $\mathbf{x}$, but based on a sub-sample of size $n_{train} < n$ known as *training data* and 'validate' and 'test' the model on the remaining data.

Assume that the feature space records two features $x_1 = age$ and $x_2 = gender$, which are continuous and binary variable, respectively.

Bias amongst Sub-populations: Continuous Target Variable

If we model the accident rate for insurance premium calculations, then y is continuous and a linear regression model would be appropriate.

Statistical parity means that $\hat{y}|F$ and $\hat{y}|M$ do not differ.
This is very much related to the the gender neutrality insurance pricing principle that is reinforced by law within the European Union.

Equal mistreatment means that

$$\hat{y}|(y, F) \text{ and } \hat{y}|(y, M) \text{ do not differ for any given response value } y.$$

Predictive parity is well-defined only if the target variable is categorical.

Bias amongst Sub-populations: Categorical Target Variable

Assume that y is a binary variable, i.e. a standard binary classification or the logistic regression is in place.

Statistical parity means that $\hat{y}|F$ and $\hat{y}|M$ do not differ, i.e.

$$\Pr\left(\hat{y} = 1|F\right) = \Pr\left(\hat{y} = 1|M\right).$$

Equal mistreatment means that $\hat{y}|(y, F)$ and $\hat{y}|(y, M)$ do not differ for any given possible response y, i.e.

$$\Pr\left(\hat{y} = 0|y = 1, F\right) = \Pr\left(\hat{y} = 0|y = 1, M\right)$$
$$\Pr\left(\hat{y} = 1|y = 0, F\right) = \Pr\left(\hat{y} = 1|y = 0, M\right)$$

Bias amongst Sub-populations: Categorical Target Variable

Equal opportunity is a simplified version of *equal mistreatment* and means that

$$\Pr\left(\hat{y}=0|y=1,F\right)=\Pr\left(\hat{y}=0|y=1,M\right)$$

where the event $y=1$ is the "advantaged" outcome, which is not suitable for the F/M dichotomy, but suitable for "admission to an university", "not defaulting on a loan" and "receiving a promotion".

Predictive parity means that

$$\Pr\left(y=1|\hat{y}=1,F\right)=\Pr\left(y=1|\hat{y}=1,M\right)$$
$$\Pr\left(y=0|\hat{y}=0,F\right)=\Pr\left(y=0|\hat{y}=0,M\right),$$

where the second equality could be discarded if the the event $y=1$ is the "advantaged" outcome.

Bias amongst Sub-populations: Categorical Target Variable

Assume that y is a binary variable, i.e. a standard binary classification or the logistic regression is in place, **but the ‘socially-sensitive‘ or ‘protected‘ feature is a (non-binary) categorical variable has now three possible states** $\{A, B, C\}$.

Statistical parity means that $\hat{y}|F$ and $\hat{y}|M$ do not differ, i.e.

$$\Pr(\hat{y} = 1|A) = \Pr(\hat{y} = 1|B) = \Pr(\hat{y} = 1|C).$$

Equal mistreatment means that $\hat{y}|(y, A)$, $\hat{y}|(y, B)$ and $\hat{y}|(y, C)$ do not differ for any given possible response y, i.e.

$$\Pr(\hat{y} = 0|y = 1, A) = \Pr(\hat{y} = 0|y = 1, B) = \Pr(\hat{y} = 0|y = 1, C)$$
$$\Pr(\hat{y} = 1|y = 0, A) = \Pr(\hat{y} = 1|y = 0, B) = \Pr(\hat{y} = 1|y = 0, C)$$

Bias amongst Sub-populations: Categorical Target Variable

Equal opportunity is a simplified version of *equal mistreatment* and means that

$$\Pr\left(\hat{y}=0|y=1,A\right)=\Pr\left(\hat{y}=0|y=1,B\right)=\Pr\left(\hat{y}=0|y=1,C\right)$$

where the event $y=1$ is the "advantaged" outcome.

Predictive parity means that

$$\Pr\left(y=1|\hat{y}=1,A\right)=\Pr\left(y=1|\hat{y}=1,B\right)=\Pr\left(y=1|\hat{y}=1,C\right)$$
$$\Pr\left(y=0|\hat{y}=0,A\right)=\Pr\left(y=0|\hat{y}=0,B\right)=\Pr\left(y=0|\hat{y}=0,C\right).$$

The second equality is discarded if $y=1$ is the "advantaged" outcome.

Example 2

A binary classification algorithm is trained over a dataset with a social sensitive feature that indicates the *final level of education* of an individual defined as a categorical variable with three states as follows:

- High-school at most
- Undergraduate studies only
- Postgraduate studies

Question: Discuss which of the four algorithmic bias measurements, i.e. *equal mistreatment*, *equal opportunity*, *predictive parity* and *statistical parity*, would be suitable to consider if the aim is to

C1) Identify which individual taxpayer, i.e. non-corporate taxpayer, is selected for tax auditing **or**

C2) Decide whether or not a defendant is likely to become a recidivist, i.e. criminal re-offender.

Further, expand your analysis for C2) by taking into account the input data horizon.

Example 3

An algorithm is trained to purge CVs of potential candidates for hiring a school bus driver. The target is to exclude those drivers who are more likely to drive while drunk. Data about the candidates' driving history and internet browsing are available. The training data are representative of different human populations with a balance between the Muslim and Christian religion sub-populations.

Results: The lowest mis-classification error, i.e. 'best' classifier, is achieved when the algorithm considers whether a person has visited a website with alcohol-like content. It is assumed that there is no bias selection practice for breathalyzer checks as a result of their facial traits.

Question: Discuss which of the four algorithmic bias measurements, i.e. *equal mistreatment*, *equal opportunity*, *predictive parity* and *statistical parity*, would be suitable to consider for this example.

Indirect discrimination or disparate impact example

Problem statement

Train an HR Analytics tool to advertise software engineering job adverts within a large technology savvy US corporation. The **target** is to attract internal candidates with various programming skills by displaying the advert to the 'right' candidate based on her/his/their browse history.

Those that clicked `stackexchange.com` or `scikit-learn.org` have a higher than average interest and skills for the advertised jobs, while those that clicked `pinterest.com` have a lower than average interest and skills for the advertised jobs. It turns out that 85% of the candidates were male. Why?

Question: Discuss which of the four algorithmic bias measurements, i.e. *equal mistreatment*, *equal opportunity*, *predictive parity* and *statistical parity*, would be suitable to consider for this example.