HR Analytics
Optimization of HR Processes Using Data and Analytics
Speaker’s Introduction

Filip Trojan
Advanced Analytics
Deloitte Advisory s.r.o.

Professional data scientist with broad experience across multiple industries including banking, finance, telco, manufacturing, retail, FMCG and e-commerce.
Strong math and statistical background and always looking for application of advanced algorithms into business problems.

Skills: marketing, scorecard development, consumer lending, credit policy, bad debt provisions, supply chain, project management, six sigma, DSS, Python, R, Matlab, SAS, XML, Perl, LaTeX, AS400, Oracle, Teradata, MSSQL
Events and ways to reach out to us.

Challenge 2019
We will soon be launching our annual competition. All students and recent graduates are welcome to apply. You will have a chance to solve a real business case, benefit from workshops, work with data from one of our partners and win a cash prize. For more details follow us on Facebook.

FACEBOOK
www.facebook.com/advancedanalyticsCZ

EMAIL
ceczanalytics@deloittece.com

We look forward to hearing from you.
Presentation Content

1  (Evidence-Based) HR Management

2  HR Analytics Introduction

3  HR Analytics Case Studies

4  Small Data Problem

5  HR Analytics Resources

6  Q&A
HRM’s Added Value

- **Employees can be considered an organization’s most valuable asset**—only through employees’ knowledge, skills, and abilities company can achieve its business and strategic goals (Boselie, 2014; Paauwe & Farndale, 2017).

- **Effective and/or efficient people management** (the way companies hire, deploy, develop, motivate, and retain its employees) **is thus a must** (Barney, 2001; Baron & Armstrong, 2007; Huselid & Becker, 2011; Wright et al., 1994).

- **Positive impact of the HRM function and its policies and practices** (sophisticated selection and training practices, participation programs, formal performance appraisals, contingent pay schemes among others) **on the operational and financial performance of organizations** has been supported by several studies with both cross-sectional and longitudinal research design (see Combs, Liu, Hall, & Ketchen, 2006; Crook, Todd, Combs, Woehr, & Ketchen, 2011; Huselid, 1995; Jiang, Lepak, Hu, & Baer, 2012; Subramony, 2009).

Source: van der Laken (2018)
HRM Value Chain

Model of mechanism by which HRM practices and policies do have impact on companies’ financial and operational performance.

Source: Blumberg (2018), Cantrell et al. (2016)
Misleading Intuition & Believes in HRM

Unfortunately, HRM professionals’ decisions are too often based on intuition, experience, and beliefs that are under influence of various fads and hypes.

Source: Rynes, Colbert, & Brown (2002)
### Management Practices

On average, encouraging employees to participate in decision making is more effective for improving organizational performance than setting performance goals.

- **True**
- **False**

### General Employment Practices

Most people overevaluate how well they perform on the job.

- **True**
- **False**

### Training & Employee Development

Training for simple skills will be more effective if it is presented in one concentrated session than if it is presented in several sessions over time.

- **True**
- **False**

### Staffing

On average, applicants who answer job advertisements are likely to have higher turnover than those referred by other employees.

- **True**
- **False**

### Compensation & Benefits

Talking about salary issues during performance appraisals tends to hurt morale and future performance.

- **True**
- **False**

---

**Source:** Rynes, Colbert, & Brown (2002)
Evidence-Based HRM

- Even though HR professionals base their decisions on some sort of evidence (outcome of scientific research, organizational facts & data, benchmarking, best practices, collective experience, personal experience, intuition...), many of them pay little attention to the quality of the evidence they base their decisions on.

- Evidence-based practice is about making decisions through, the conscientious, explicit and judicious use of the best available evidence from multiple sources.

Source: https://scienceforwork.com/blog/evidence-based-management-training/
4 Sources of Evidence-Based HRM

- Evidence-based HRM combines 4 sources of information:
  1. practitioners’ professional expertise,
  2. stakeholders’ values and concerns,
  3. scientific evidence, and
  4. reliable and valid organizational metrics.

- The biggest blind spot of the HRM function usually lies in the fourth source of information: it often lacks the capability to measure and quantify the strategic contribution of its HRM activities, its bottom-line impact, and any progress therein in its own, local organizational context…

- … due to missing analytical mindset among HR professionals and absence of reliable and relevant HR data.

Public Interest in HR Analytics

Monthly Google search interest on “people analytics” and related terms over time. Values are proportional to the maximum value and fit by locally weighted regression lines (LOESS).

Source: van der Laken (2018)
HR Analytics as Data-Driven/ Evidence-Based HRM

The key to driving business performance is understanding which competencies drive employee performance, and then ensuring that these competencies are available in the workforce by creating people processes around these competencies.

In practice, this means creating:

- Recruitment processes to hire the right competencies
- Learning and development processes to train the right competencies
- Career development and compensation processes to retain the right competencies

Source: Blumberg (2018)
HR Analytics Agenda

HR analytics helps to *optimise mechanism behind HRM value chain* by allowing us to *find answers to certain key questions*, such as…

- Which channels bring us the best candidates?
- What characteristics differentiate successful candidates from unsuccessful ones?
- What factors contribute to successful onboarding?
- Which KPIs have the strongest link to the company’s financial results?
- Which training sessions are most likely to lead to improvement of work performance?
- Which interventions have the biggest impact on well-being or work-life balance perceived by employees?
- What increases or decreases the employees’ engagement level?
- Who represents hidden talent that needs to be detected and further developed?
- Where can resistance be expected with respect to planned changes in the company and who can instead be their ambassador or catalyst?
- Which factors contribute to employee turnover?
- etc.
HR Analytics Maturity Model

Level 4: Predictive Analytics
- Development of Predictive Models, Scenario Planning, Integration with Strategic Workforce Planning, Risk Analysis and Mitigation

Level 3: Advanced Analytics
- Statistical Analysis to solve Business Problems, Development of Models, Actionable solutions, Centralised Staffing & Integrated data

Level 2: Advanced Reporting
- Proactive reporting for Benchmarking and Decision making, Multi-dimensional analysis and Dashboards

Level 1: Operational Reporting
- Reactive operational reporting, Focus on data accuracy, Consistency & Timeliness

Source: Van Vulpen (2016)
HR Analytics Project Workflow

CRISP-DM

Eight Step Model for Purposeful Analytics

Why undertake the project?
- Step 1: Frame Business Questions
- Step 2: Build Hypotheses

How should the project be carried out?
- Step 3: Gather Data
- Step 4: Conduct Analyses
- Step 5: Reveal Insights
- Step 6: Determine Recommendations

What will result from the project?
- Step 7: Get Your Point Across
- Step 8: Implement and Evaluate

Source: Guenole, Ferrar & Feinzig (2017)
HR Analytics Skillsets

What skillsets are required for successful HR Analytics projects and what happens when one of them is missing?

Source: Van Vulpen (2016)
Case Study 1

Leadership of a professional services firm division engaged Deloitte to address the issue of high voluntary attrition. The objective was to provide the leadership with data-driven insights into why employees leave, identify which segments and individuals are at a higher risk of leaving in the near future and propose a plan to retain the key individuals.

Note: The initial point includes all new entrants in a given year who started working between September and December. The next indication is taken in September of the following year. Last period measured constitutes 9 months until the end of May 2016.
Case Study 1

Project business case

- Gets insights from your data
- Retain top performers
- Retain knowledge
- Cut fluctuation costs
Case Study 1

Project business case - Inputs
## Case Study 1

### Project Business Case - Outputs

<table>
<thead>
<tr>
<th>Estimated Turnover Cost Per Employee</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Direct Costs</td>
<td>CZK 55,000</td>
</tr>
<tr>
<td>2. Lost Productivity</td>
<td>CZK 290,278</td>
</tr>
<tr>
<td>3. Savings of Salary + Benefits</td>
<td>CZK 100,875</td>
</tr>
<tr>
<td><strong>Total (Calculation: 1. + 2. - 3.)</strong></td>
<td>CZK 244,403</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Cost of Employee Turnover</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Employee Churn</td>
<td>78</td>
</tr>
<tr>
<td><strong>Total Cost</strong></td>
<td>CZK 19,161,178</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Saved Costs With Attrition Being Reduced By...</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>CZK 684,328</td>
</tr>
<tr>
<td>2%</td>
<td>CZK 1,368,656</td>
</tr>
<tr>
<td>3%</td>
<td>CZK 2,052,983</td>
</tr>
<tr>
<td>4%</td>
<td>CZK 2,737,311</td>
</tr>
<tr>
<td>5%</td>
<td>CZK 3,421,639</td>
</tr>
</tbody>
</table>
Case Study 1

Target variable

- **Target date** = date when employee left company
- **Target period** = date window when target date could occur, to define binary target
- **Observation date** = monthly snapshot of employee data
- **Blackout period** = employee notice period, when we do not observe any data about employee to avoid data leakage

Model works with each employee and his 12 months history, notice period is 2 months and projection of leave up to 6 months.
Case Study 1

Workforce Analytics relies on a **wide range of data from various sources** and combines them in a single database. Both traditional and non-traditional HR data are used.
Case Study 1

Types of predictors

Standard predictors
Monthly snapshot of employee data, E.g. Monthly average salary base
Two types (inspired by banking industry)
1. Static data – not very variable over time, e.g. demographics
2. Transaction data – changing in time, e.g. monthly performance of employee

Trend predictors
Time evolution of standard transactional predictors over several months (e.g. last month, 6months, 9months, 12months).
To indicate changes in employee behavior, e.g. trend in performance in last 6M (increasing, decreasing, same)

Types of trend predictors:
1. Trend curve
2. Volatility
3. Difference now vs. last 1/3/6/12M

Peer predictors
Comparison of standard predictors with peer average, i.e. group with similar working conditions
Is employee better, worse or same as peer, e.g. peer monthly salary

Types of peer dimension:
1. Region
2. Team, unit
3. Service, client segment
4. Position level
5. Length in company
Peer is defined based on industry specifics
Case Study 1

Binning of variables – treating missing values, outliers, and non-linearities
Case Study 1

Steps in feature selection

1. Predictors pool (500-1000)
   Consolidated all calculated different types of predictors from various areas

2. Gini and Correlation Selection (200-300)
   First selection of predictors, which have at least some relevance to target (churn)

3. Variable Clustering (100)
   To avoid multicollinearity among predictors in regression, group similar predictors and within each cluster select the most significant ones

4. Automatic and Manual Selection (30)
   Automatic selection of best predictors (e.g., glmnet with LASSO and ridge penalties or stepwise regression) + manual selection (business interpretation, appropriate shape of event rate curve in binning)

Best predictors in final model

FINAL 5 to 15
Case Study 1

Performance vs. interpretability trade-off

<table>
<thead>
<tr>
<th>High performing</th>
<th>Low performing</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>Naive Bayes</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>Decision tree</td>
</tr>
<tr>
<td>Linear regression</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Black box</th>
<th>Transparent</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>xgboost</td>
</tr>
<tr>
<td>random forest</td>
<td>ensemble</td>
</tr>
<tr>
<td>NNs</td>
<td></td>
</tr>
<tr>
<td>ensemble</td>
<td>KNN</td>
</tr>
</tbody>
</table>
Case Study 1
Resolving performance vs. interpretability trade-off?

- Variable importance measures
- Partial dependence plots
- Centralized individual conditional expectation
- Local interpretation of individual predictions
Case Study 1

Why should we improve our understanding of ML models?

Improving our models
Generalisability
“Sanity Check”
Prevent wrong conclusions & potentially adversarial attacks

Trust and transparency
Can I trust my model’s decisions?
Why does my model make the predictions it makes?

Prevent Bias
Fairness
Identify and prevent bias

Source: Glander (2018)
Case Study 1

Preventing biases in ML models

Amazon ditched AI recruiting tool that favored men for technical jobs

Specialists had been building computer programs since 2014 to review résumés in an effort to automate the search process


https://www.youtube.com/watch?v=59bMh59JQDo
Case Study 1

Predictors in the final model

- No major changes in utilization
- Slightly above average billable hours
- Average training hours
- First promotion
- Fast track
- Skipping levels
- Earning less than peers
- Low salary in previous year
- Promotion in line with or faster than peers

Higher risk of attrition

- Long time spent on C and D clients
- Working on the same AB clients each year
- More days spent travelling than in the past
- Longer projects than in the past
- Less than 3 level distance between counsellors and counselees
- Less senior counsellors (M-)
- No long-term absence
## Case Study 1

Employees at risk of leaving with probability more than 18% to leave in next 6 months. Immediate action is required!

<table>
<thead>
<tr>
<th>Employee</th>
<th>Prob. of leaving</th>
<th>Status</th>
<th>No. 1 Destabilizing factor</th>
<th>No. 2. Destabilizing factor</th>
<th>No. 3 Destabilizing factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.71%</td>
<td>Active</td>
<td>Junior counsel or (In Charge or Senior)</td>
<td>Growth in minimum team size from previous year</td>
<td>Minor loss of AB clients since previous year</td>
<td>n/a</td>
</tr>
<tr>
<td>50.09%</td>
<td>Active</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Stable utilization overtime</td>
<td>n/a</td>
</tr>
<tr>
<td>48.74%</td>
<td>Active</td>
<td>Growth in minimum team size from previous year</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Same AD clients since previous year</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>46.65%</td>
<td>Active</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Stable utilization overtime</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>43.29%</td>
<td>Active</td>
<td>Junior counsel or (In Charge or Senior)</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Average training utilization (4-12%)</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>39.73%</td>
<td>Active</td>
<td>Junior counsel or (In Charge or Senior)</td>
<td>Growth in minimum team size from previous year</td>
<td>Stable utilization overtime</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>36.84%</td>
<td>Active</td>
<td>Junior counsel or (In Charge or Senior)</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Average training utilization (4-12%)</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>30.29%</td>
<td>Active</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Minor loss of AD clients since previous year</td>
<td>Average training utilization (4-12%)</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>29.28%</td>
<td>Active</td>
<td>Minor loss of AB clients since previous year</td>
<td>Stable utilization overtime</td>
<td>Junior counsel or (In Charge or Senior)</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>26.56%</td>
<td>Active</td>
<td>Junior counsel or (In Charge or Senior)</td>
<td>Growth in minimum team size from previous year</td>
<td>Stable utilization overtime</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>26.10%</td>
<td>Active</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Average training utilization (4-12%)</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>20.07%</td>
<td>Active</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>Promoted In line with peers Earning less than peers</td>
<td>Average training utilization (4-12%)</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>19.13%</td>
<td>Active</td>
<td>Minor loss of AD clients since previous year</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>n/a</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td>18.31%</td>
<td>Active</td>
<td>Same AD clients since previous year</td>
<td>Stable utilization overtime</td>
<td>n/a</td>
<td>Average training utilization (4-12%)</td>
</tr>
</tbody>
</table>
Case Study 1

Employees with probability of leaving higher than the average churn rate 6 months later.

<table>
<thead>
<tr>
<th>Employee Name</th>
<th>Probability of leaving</th>
<th>Status</th>
<th>No. 1 Destabilizing factor</th>
<th>No. 2. Destabilizing factor</th>
<th>No. 3 Destabilizing factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>67.71%</td>
<td>Active</td>
<td>Junior counselor (In Charge or Senior)</td>
<td>Growth in minimum team size from previous year</td>
<td>Minor loss of AB clients since previous year</td>
</tr>
<tr>
<td></td>
<td>50.99%</td>
<td>Left</td>
<td>Growth in minimum team size from previous year</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>48.74%</td>
<td>Left</td>
<td>Growth in minimum team size from previous year</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Stable utilization over time</td>
</tr>
<tr>
<td></td>
<td>46.83%</td>
<td>Active</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Same AB clients since previous year</td>
</tr>
<tr>
<td></td>
<td>43.29%</td>
<td>Left</td>
<td>Junior counselor (In Charge or Senior)</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td></td>
<td>39.73%</td>
<td>Active</td>
<td>Junior counselor (In Charge or Senior)</td>
<td>Growth in minimum team size from previous year</td>
<td>Stable utilization over time</td>
</tr>
<tr>
<td></td>
<td>36.04%</td>
<td>Left</td>
<td>Junior counselor (In Charge or Senior)</td>
<td>Promoted in line with peers Earning less than peers Long time on CD projects</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td></td>
<td>30.29%</td>
<td>Left</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Minor loss of AB clients since previous year</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td></td>
<td>29.26%</td>
<td>Left</td>
<td>Minor loss of AB clients since previous year</td>
<td>Stable utilization over time</td>
<td>Junior counselor (In Charge or Senior)</td>
</tr>
<tr>
<td></td>
<td>26.56%</td>
<td>Active</td>
<td>Junior counselor (In Charge or Senior)</td>
<td>Growth in minimum team size from previous year</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td></td>
<td>26.10%</td>
<td>Active</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>Promoted in line with peers Earning less than peers Long time on CD projects</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td></td>
<td>20.87%</td>
<td>Left</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>Promoted in line with peers Earning less than peers</td>
<td>Average training utilization (4-12%)</td>
</tr>
<tr>
<td></td>
<td>19.13%</td>
<td>Left</td>
<td>Minor loss of AB clients since previous year</td>
<td>Higher than average billable utilization (75-79%)</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>18.31%</td>
<td>Active</td>
<td>Same AB clients since previous year</td>
<td>Stable utilization over time</td>
<td>Average training utilization (4-12%)</td>
</tr>
</tbody>
</table>
Case Study 1
Examples of recommendations

01 VACATIONS
The long term leave or unpaid vacation is a stabilizing factor

02 COUNSELLING
Distance should be at least 3-4 levels, no junior counselors

03 SALARY & PROMOTION
Employees are motivated by pay raise and not by promotions

04 HIRING
Hire more local candidates

and many more...
Case Study 2

- Professional-services firm started to **change their sales model to achieve its growth goals**.

- To **support adoption of new sales model** across the organization the firm wanted to build a **profile of high performers** around those sales professionals who were excelling in the current model.

- The goal was to use this profile for redesigning selection and training & development systems and aligning them with business goals for sales professionals to ensure...

  1) **that the current sales professionals were being trained on and dedicated attention to those key weaknesses that were leading to underperformance**, and

  2) **that new hires to the sales team had the characteristics critical to success in the organization**.

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 2

The four-step process that is based on research and data but tailored specifically to the organization, creating a customized solution that selects employees who fit into the given context and develops the current workforce on the key drivers most strongly linked to business success:

1) Define greatness,
2) Assess and compare the current workforce to greatness,
3) Develop the current workforce toward greatness, and
4) Hire greatness.

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 2

- Defining the greatness using outputs from several interviews conducted with partners and top performing sales professionals - gathering information about the knowledge, skills, abilities, and behaviors that make a sales employee successful in the organization.

- Identification of most frequent themes/competencies + their behavioral examples/indicators (e.g., Communication competency included behaviors such as “quickly builds rapport through speech and action when first meeting with decision-makers” or “uses listening as a strategy to gather information and build trust.”)

- Definition of high performance (company tries to positively impact by implementing new selection and development systems) using combination of three common sales outcomes:
  1) sales goal attainment,
  2) average win rate, and
  3) average win size.

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 2

Assessing and comparing the current workforce to greatness using battery of methods by which each sales professional can be assessed on his/her behaviors, attitudes, knowledge, and personality traits:

- 360° feedback to assess behavior,
- sales-climate survey to assess work attitudes,
- situational-judgment test to assess job knowledge, and
- personality assessment to assess personality traits.

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 2

Using structural equation modeling to identify and model key drivers of sales outcomes.

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 2
Building team and/or individual development programs around eight strongest sales drivers and weakest performance areas (Focus quadrant)

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 2

Designing multihurdle selection process in combination with predictive algorithm weighting individual factors according to their impact on sales outcome.

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \]

Source: Mondore, Spell, Betts & Douthitt (2018)
Case Study 3

Moneyball as a good example of HR Analytics in action

Small Data Problem

HR datasets usually do not provide enough observations to train the model. We have to be careful to avoid overfitting. We are addressing the problem in two ways:

A. Use the right algorithm: logistic regression

B. Use the right binning: ctree (Conditional Inference Trees)
Small Data Problem

ctree versus EqualFrequency - number of bins

Duration represents a strong predictor

Sturges rule was used to determine number of bins

ctree suggests fewer bins compared to Sturges rule.
Small Data Problem

ctree versus EqualFrequency - number of bins

Age represents a weak predictor

ctree suggests much fewer bins (often just one) compared to Sturges rule.

Sturges rule was used to determine number of bins
Small Data Problem

ctree versus EqualFrequency - Spearman correlation between train and test predicted probabilities

Duration represents a strong predictor

ctree gives bins which are consistent between train and test for >= 1000 observations, EqualFrequency is prone to overfit for < 2000 observations.
Small Data Problem

ctree versus EqualFrequency - Spearman correlation between train and test predicted probabilities

Age represents a weak predictor

ctree gives bins which are consistent between train and test for >= 1000 observations, EqualFrequency is prone to overfit even for the highest number of observations.
Small Data Problem

Conclusions

• Use algorithms which work well on small datasets, e.g. logistic regression

• Use coarse bins. ctree works for small samples, EqualFrequency is prone to overfitting.

• Be extremely cautious for datasets smaller than 1000 observations.
HR Analytics Resources

Follow these people…
HR Analytics Resources (cont.)

Check these websites…

https://scienceforwork.com/

https://www.analyticsinhhr.com/

https://www.hranalytics101.com/

HR Analytics Resources (cont.)

Read these books…

1. **The Power of People**
2. **The Basic Principles of People Analytics**
3. **Predictive HR Analytics**
HR Analytics Resources (cont.)

Read these books…
HR Analytics Resources (cont.)

Attend following courses...

https://www.coursera.org/learn/wharton-people-analytics

https://university.business-science.io/p/hr201-using-machine-learning-h2o-lime-to-predict-employee-turnover
HR Analytics Resources (cont.)

Learn some Data Science basics…

https://www.datacamp.com/
https://www.kaggle.com/learn/overview
https://www.udemy.com/courses/search/?src=ukw&q=Data+Science
HR Analytics Resources (cont.)

Learn following analytical technologies…

SQL  R  Python

Tableau  Spark  Hadoop
Thank you.

Filip Trojan
Advanced Analytics