People Analytics

Assignment 5:
- Part 1
- Part 2
- Part 3
- Part 4

Tania Cajal (MBDS 2021/2022)

Does Human Capital Matter? A Meta- Analysis of the Relationship Between Human Capital and Firm Performance

- At both the micro and macro levels, theory predicts that investing in excellent human capital leads to improved company performance.
- However, developing or acquiring superior human capital costs time and money, which could counteract its beneficial effects.
- Managers should invest in initiatives that increase and retain firm-specific human capital, according to our findings.
- Firms with better human capital should perform better.
- RBT is particularly focused with long-term, or sustainable, benefits.

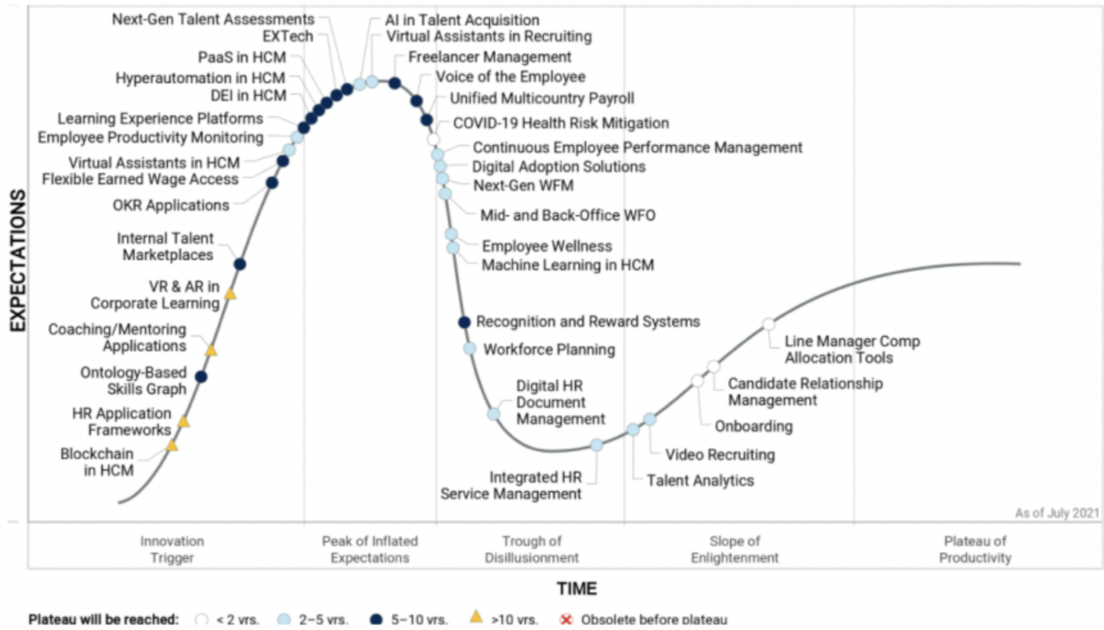
How to be great at Advanced analytics is transforming the HR landscape. Interviews with leading people analytics teams reveal how

- HR data analytics are skills for talent management and evidence-based personnel choices.
- Even the most advanced are at the early stages of building their people analytics.
- Challenges:
    - cleaning data and streamlining reporting

    - lack the ability to embed data analytics in day-to-day HR processes
    - disorganized landscape of HR technology
- McKinsey has outlined five steps to improve this.
- Key: continually iterating—retracing their steps and climbing the same stairs again—at every level of the journey to the top
- Ingredients for success:
    - Significant and dedicated data-engineering resources.
    - Breadth and depth of data sources.
    - Beyond the core HR systems to use several additional internal sources of data.
    - Advanced network analysis.
    - Survey strategy for monitoring employee sentiment (like a weekly Pulse Survey)
- Client and corporate use cases are clearly aligned with feedback loops that allow for iterative creation and ongoing learning.
- Culture of trust, empowerment, and ownership.
- Client wants and inquiries that are urgent (and often confusing), highly sensitive data, and difficulty in extrapolating meaningful and actionable insights to steer business decisions.
- Take into account population specific needs.
- COVID19 example: concerns of isolation, remote work, childcare, and work-life balance.

Hype cycle for human capital management technology, 2021

Hype Cycle for Human Capital Management Technology, 2021



Source: Gartner (July 2021)
747502

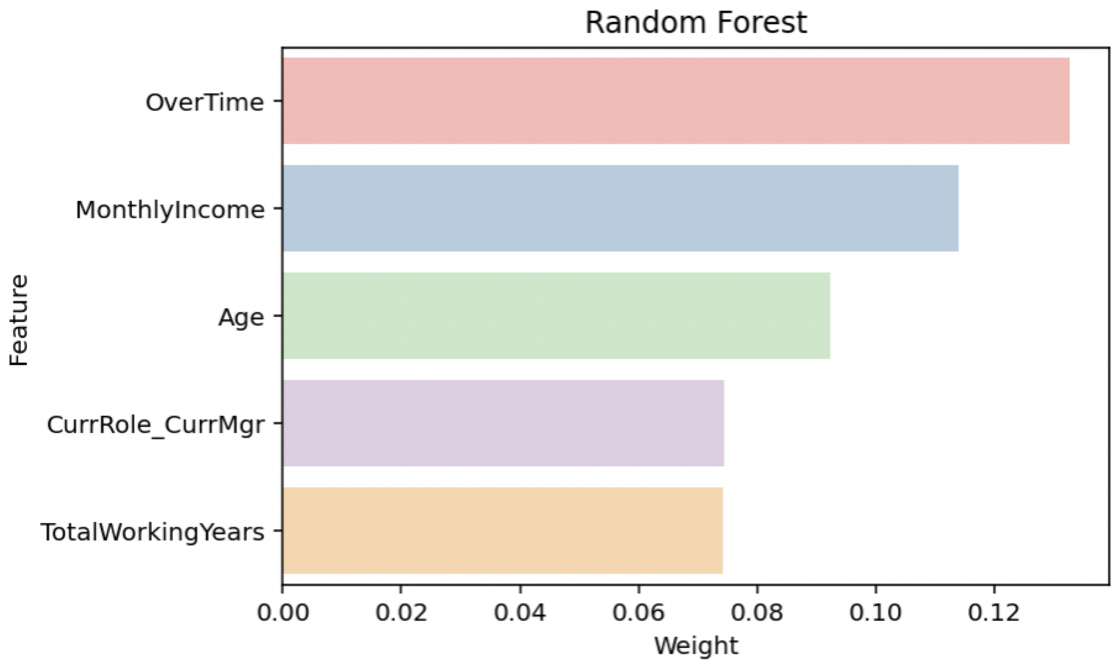
Three points and why they are key to data-driven companies:

1. **Coaching/Mentoring Applications**. It's clear why this point was at the Innovation Trigger phase. If the quality of talent is proportional to company success, technology that would enhance a process of improving current talent is expected to be hyped. At the plateau of productivity, regardless of what tech is dominant at the time, these applications should be maintained as a long-term strategy to boost employee motivation, welfare, and internal career growth.
2. **AI in Talent Acquisition**. An undeniably popular tech even in 2022 was a the Peak of Inflated Expectations for obvious reasons. The biggest companies receive thousands of applications from all around the world, making bots a great relief to the selection process. Candidates are being encouraged to create CVs for bots and not for humans. But as always, the data mining and cleaning process is likely an obstacle that needs to be tackled when processing applications. Finding the best match through AI is as useful as it is risky- a better candidate might get lost in the algorithm due to poor formatting choices. It's important for firms to keep investing resources on refining this technology.
3. **COVID-19 Health Risk Monitoring**. At the Through of Disillusionment stands the COVID-19 monitoring trend. A life-changing event in 2020 is quickly being forgotten in many countries. Mask mandates removed, full vaccination taking

place, why is it important for companies to keep this tech up? A great deal of literature and tests were produced in a short amount of time given the initial urgency the pandemic created. And while a virus has a particular behaviour, the same models and frameworks used to analyse its spread can be applied to other spreadable phenomena (i.e. the Great Resignation, team morale, perception of the company), making it extremely useful for HR as much as it is for health.

Case study: Employee attrition random forest

This project is based on a hypothetical dataset downloaded from IBM HR Analytics Employee Attrition & Performance. It has 1,470 data points (rows) and 35 features (columns) describing each employee's background and characteristics; and labelled (supervised learning) with whether they are still in the company or whether they have gone to work somewhere else. Machine Learning models can help to understand and determine how these factors relate to workforce attrition.



Feature importance graph!

(Code book attached separately)

Leo Breiman

Who is Leo Breiman? Why he is one of the most important people for ML?

He turned the Statistical Laboratory, which had only one modest computer at the time, into one of the most sophisticated computing centers in the country by focusing on applying statistics to computer science.

His practical application of statistics has, above all, bridged the gap between statistics and computer science throughout his life. In addition, it cleared the path for improvements in machine learning and data mining.

In 2001, he published two landmark papers, one introducing the concept of Random Forests (wow!), and another encouraging professionals to apply algorithms to real-life problems being solved using big data, which was not the concept it is now at the time, but still.

What are the main differences between CART and Random Forest?

Random Forest creates multiple CART trees based on "bootstrapped" samples of data and then combines the predictions. Usually, the combination is an average of all the predictions from all CART models. A bootstrap sample is a random sample conducted with replacement. Their key differences when applied are:

- Random forest has a very high accuracy.
- A RF uses many different characteristics to make a prediction.
- With RF, the risk of overfitting is higher.
- With CART, the risk of oversimplification is higher.
- Some of CART's advantages include that the rules are easily interpretable and that it offers automatic handling of variable selection, missing values, outliers, local effect modeling, variable interaction, and non-linear relationships.
- Unlike the CART model, Random Forest's rules are not easily interpretable.
- CART models have better stability, which sometimes leads data scientists to lean towards them.
- CART models are easier to explain and use to justify business decisions.