Centralized vs Embedded Models for ML Scientists

Matteo Ruffini, Joos-Hendrik Boese
Amazon Music ML,
Berlin
ruffinim@amazon.com

Introduction

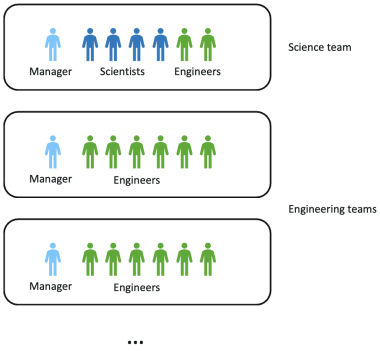
Business landscape:

- 85% of ML projects fail [1]
- Team structure has impact on likelihood of project success

This talk:

- Discuss two ML teams organization models
  - Centralized Model: Core ML team owns central ML component
  - Embedded Model: Scientists assigned to feature-owning team
- Dive deep with case studies, discuss comparisons and applications

Centralized Model



- ML scientists clustered in a single team
- The team can contain scientists and engineers
- All report to the same manager
- All work on same / team level goals
- Consults other teams on ML use cases

Business Case Study: Amazon Music Core ML Team

Amazon Music Core ML Team

- Enable personalization across Amazon Music
- Develop ML platform for personalization
  (No customer-facing experience owned)
- Support internal teams to use the platform

Way of working

- 75% of the time on Core ML projects
- 25% of the time on support to platform users
- Yearly planning cycle, with in-year iterations

Business Case Study: Amazon Music Core ML Team

Learnings

+ High Impact
  - Scalability: platform work enabled personalization across Amazon Music
  - Dramatic increase in engagement and retention

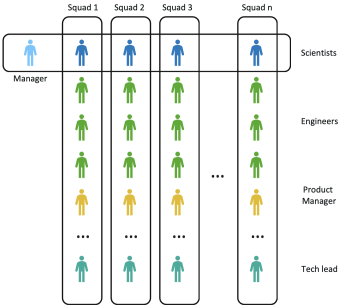
+ Employee Retention & High Standards
  - Scientists working with other scientists - Lab environment
  - Scientists working on cutting edge ML, write papers, contribute to SOTA
  - Dedicated Science Manager helps for retention.

Business Case Study: Amazon Music Core ML Team

Challenges

- Speed of innovation
  - You need to influence other teams to be successful and test new ideas
  - ..and this takes effort and time
- Acquisition of business-level expertise
  - Limited exposure to business problems
  - Depth of expertise comes at the cost of breadth

Embedded Model



- Scientists embedded in *squads*: full stack, feature-owning teams
- Squads contain all needed roles (ML, product, design...)
- Squads work in full autonomy
- Squad-lead does not manage scientists
- A central manager manages scientists across squads

Business Case Study: [A Previous Company's] Data Science Team

Way of working: ML science initiatives

- Scientists scope a project with squad leads
- This is typically not the only project of a squad
- Once project is done, scientist may change squad

Big diversity across projects:

- Advanced Analytics
- ML Products (recsys, ranking, price prediction...)

Business Case Study: [A Previous Company's] Data Science Team

Learnings

+ Some projects had high impact
  - Big wins on advanced analytics projects (with no ML in production)
+ Worked as project incubator for ML products
  - Squads allowed to test ML products in production quickly
+ Quick growth in seniority of ML scientists
  - Interactions with leadership and project scoping foster quick growth
  - Exposure to business discussions $\rightarrow$ domain expertise

Business Case Study: [A Previous Company's] Data Science Team

Challenges

- Quality of innovation
  - Low hanging fruits are often preferred to long term, high impact wins.
- Limited synergies
  - Difficult to transfer successes from one case to another
- Not all squads are ML Savvy
  - Project failed due to unclear scoping / needs
  - Negative impact on scientists retention

Summary

Centralized Model

- Promotes *Depth of Expertise*

Embedded Model

- Promotes *Breadth of Expertise*

Summary

Centralized Model

- Promotes *Depth of Expertise*
- Surrounds scientists with *scientists*

Embedded Model

- Promotes *Breadth of Expertise*
- Surrounds scientists with *Cross-Disciplinary Teams*

Summary

Centralized Model

- Promotes *Depth of Expertise*
- Surrounds scientists with *scientists*
- Works well to run and maintain *Core ML products in production*

Embedded Model

- Promotes *Breadth of Expertise*
- Surrounds scientists with *Cross-Disciplinary Teams*
- Works well for *Incubating ML projects and products*

Summary

Centralized Model

- Promotes *Depth of Expertise*
- Surrounds scientists with *scientists*
- Works well to run and maintain *Core ML products in production*
- Work on *products*

Embedded Model

- Promotes *Breadth of Expertise*
- Surrounds scientists with *Cross-Disciplinary Teams*
- Works well for *Incubating ML projects and products*
- Work on *projects*

Takeaways

Takeaway 1: Which structure do I need?

- **Embedded Model**
  - If you don't have an ML team
  - If you want to explore multiple ML applications.
- **Central Model**
  If you know what to do and can afford high-risk / high-reward investments.

Takeaways

Takeaway 2: Hire the right scientists for the right team

- Embedded is very different from a postdoc environment! Papers are also more difficult.
- For embedded, you need folks comfortable with product management and stakeholders.

Takeaways

Takeaway 3: Central teams can spin off from the embedded model

- Successful ML projects can spinoff into dedicated teams.
- Central teams works well when the company has the right level of maturity.

Takeaways

Takeaway 4: Consider a hybrid system

- Why not to have a Central core team and embedded scientists in squads?
- It has high costs, not everyone can afford it!
- Risk: All want to be in the central team - consider rotations.

Thank You

Contact information

- ruffinim@amazon.com