How to (gracefully) jump start data science initiatives at a start-up

Cho-Nan Tsai, CTO @ Camino Financial

About Me

CTO @

CAMINO FINANCIAL

Fortune 500

Online lender

Start-ups

Hispanic Market

Data Science initiatives are hard!!!!

87% of data science projects never make it into production.

- Venture Beat

77% of businesses report that "business adoption" of big data and AI initiatives continues to represent a big challenge for business.

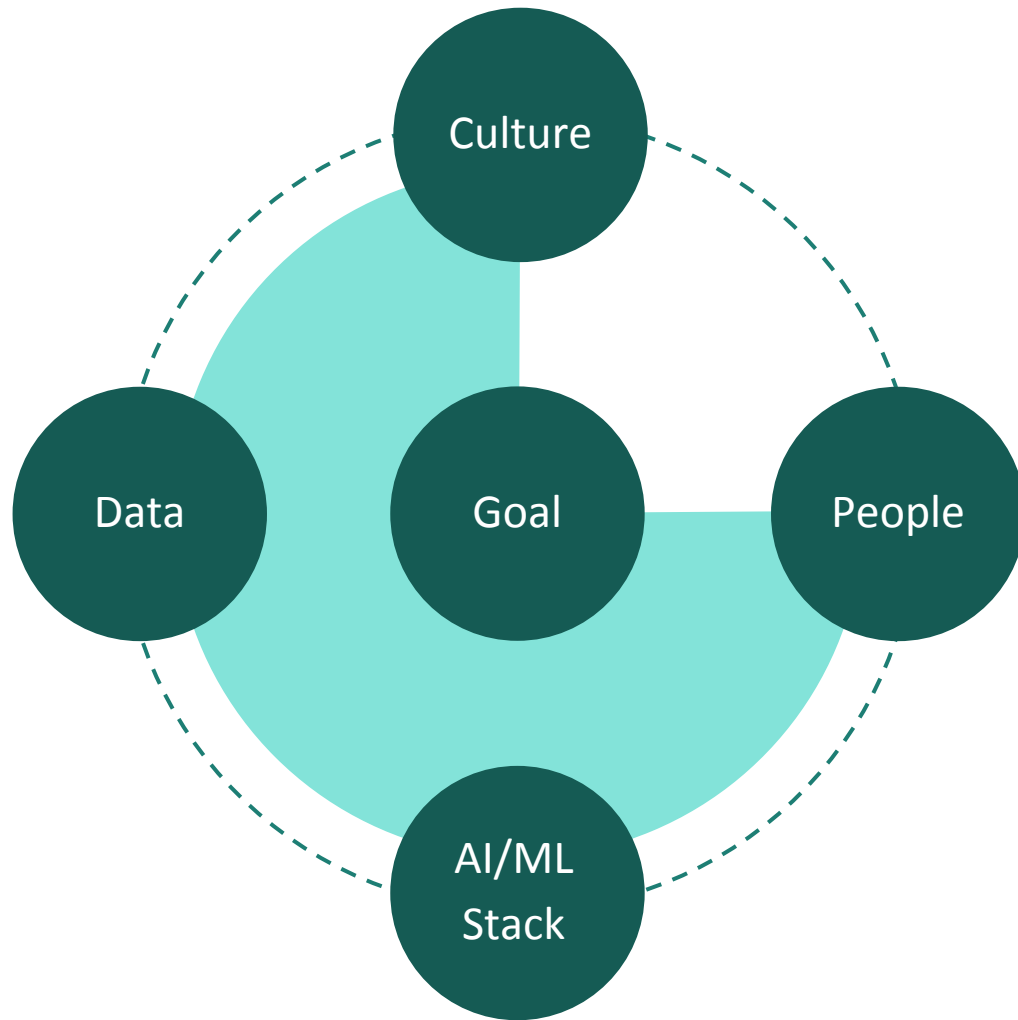*- New Advantage Survey*

85% of data science projects fail.

- Gartner

What can we do about it?

The 5 Golden Pillars of Data Science

The Future

Thank you

https://www.linkedin.com/in/michaeltsai/

https://obicho.github.io/talks/

- Don't do moonshot projects on day one
- Early initiatives should be tangible, low hanging fruits
- Leadership should have a good understanding of capabilities of ML/AI
- Iterate always but never stall
- Goals should be clearly stated and yet not too broad
- Contribute to marketing goals:
  - Decrease user acquisition cost by X% through better targeting
  - Increase funnel conversion by X% by offering actionable and personalized call-to-action messages
- Contribute to data-driven process:
  - Combine various data sources to deliver more customer insight
  - Set up BI dashboards

People

- Your first hire should
  - wear many hats
  - Have an open mind
  - Wear many hats
  - Have a good business sense
  - Can draw conclusions and
  - Do cause-effect analysis
- Setting proper expectations up front: Employer wants beautiful dashboards and useful insights right away but the data scientist wants to build large scale machine learning clusters using complex algorithms -- **almost never happens at an early stage startup**

- Treat data as your first class citizen
- Build an organizational culture that encourages to bring on ideas with the support of data
- Don't operate on a hunch
- If there is a bigger fish to fry (other than data), you must go for it (like existential threat or finding your product-market fit)
- Data is not the CEO or the President; deprioritize when necessary

- Watch out for data starvation
- Be creative where you get your data
- Open data initiatives
- Buy data may be OK, but as a last resort. Validate your source
- Embed event tracking and user tracking early on in your product
- Don't use production database to do work

- Spend the time to set up the right infrastructure first
- Approach open source alternative first:
  - R Studio - Analysis and base plots
  - R Markdown - Reports
  - R Shiny - Interactive visualization
  - Python (PANDAS) - data frames
  - d3.js - visualization
- Do's before data science analysis
  - Create data pipeline
  - Data instrumentation
  - Data catalog
  - ETL (extract-transform-load)
- Data engineering 80% ; Data Science 20% only!
- Protect your bank roll - don't turn on data science services unless you are ready