

2023年7月27日
データサイエンティスト集会 in VRC

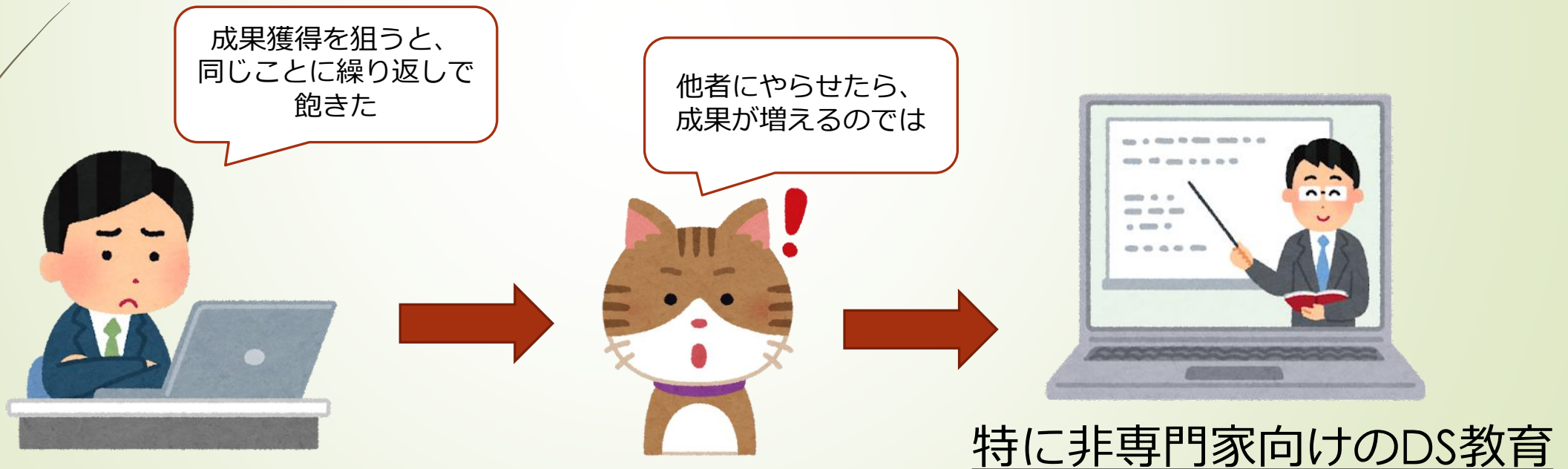
超初心者向け データサイエンス業務 最初の一歩！

ぶんちん

1

自己紹介 ぶんちゃん

- ▶ 複合経営が特徴の企業（製造業）に所属
- ▶ データ分析担当者だったが。。。



注意！！

私はスマートな業界ではなく、

泥臭い The 重厚長大 製造業

- ▶ 基本的に高度な技術の導入に体制が整い切れていない組織
- ▶ 新規技術の開発ではなく、具体的な経済効果の獲得が私の主目的
- ▶ 多くの案件を根拠にしているものの、あくまで私の経験則

が前提のお話です。

初心者が直面する壁

将来のためにデータサイエンスの勉強するぞ！
数学！機械学習！プログラミング！etc



勉強したことが仕事で使えない。
使ってみたけど成果につながらない。
実力不足。もっと勉強しないと！



Q.いつになったら
仕事に生かせるの??
成果を出せるの??

A.不足しているのは技術ではなく、
仕事の進め方（技能）かも？

よくある勘違い

- 高度な技術は基本的な技術の上位互換！
- それを活用した仕事ができるようになる！
- 技術を学べばすぐに実績を積める！



現実はそんなに甘くない

データサイエンス技術適用の構造

高度な手法は基本的な手法の
単純な上位互換ではない！

基本的な手法

- そこそこの性能
- 使いやすい

少し高度な手法

- 性能向上
- デメリット追加

高度な手法

- さらに性能向上
- さらなるデメリット

課題

技術的に高度になるほど扱いづらくなっていく
→ 案件ごとに全体最適のバランス調整が重要

いきなり使いこなせますか？

高度な技術の適用から考えてしまうと。。。。

たとえ良い結果が得られた場合であっても...

いくらの効果と費用になるの？

どのようなリスクが考えられる？

どうやって運用していくの？

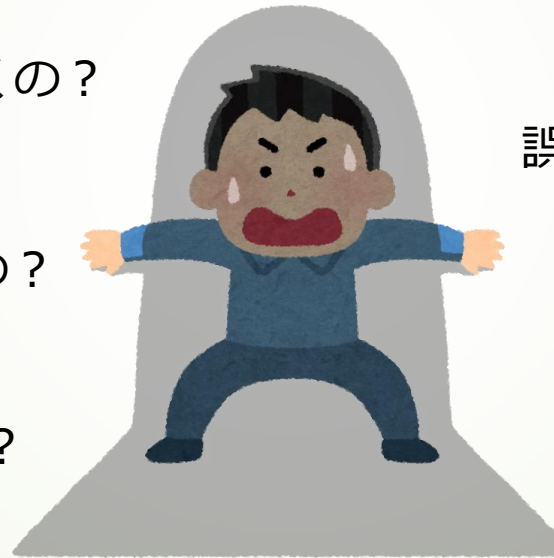
誤動作した場合の対応ってどうなの？

現場適用どうするの？

モデルの監視はどうするの？

体制は？他部署との調整は？

オーバースペックになっていない？ など



全て解決しないと効果獲得まで至らない
様々な課題の整理・対応も併せて少なくないリソースが必要
→中断判断も難しく、泥沼案件化することが多い

世の中は失敗だらけ

“POC死”と検索すると、
死屍累々なのがよくわかります

数年前にPoC（Proof of Concept；概念実証）が流行
専門家と協力してAI（≒データサイエンス）を活用したプロジェクトが乱立
そのほとんどが失敗。9割が失敗といわれているが、もっと多いのでは？
PoCが黒歴史的ワードに。

私が考えている失敗の理由

- ① 課題設定の失敗
- ② 技術課題以外の問題
 - 理想論に基づいた目標設定
 - 実現不可能な課題の泥沼化
 - 適用先の業務設計が不十分

ポイント1：適切な課題設定

技術（シーズ）から課題を探すな！！

課題解決ではなく、技術適用が目的になる
→成果につながりにくい

私の初心者におすすめの取り組み課題

- 具体的に困っている人がいる
- ちょっとした改善活動

詳細は別の機会に説明します

★ポイント2：基本的な手法で具体化

手順

- 1. データの可視化（+α）程度でできる施策を検討**
100点中20~50点の成果を最速で実現。定量評価を可能にする
- 2. 上記のアプローチが原理的に可能か、トライアルを実施。**
技術面以外も含む課題を洗い出す。

理由

- ▶ 技術課題とは別の部分の課題を明確にしたい
- ▶ 必要とされる"最低スペック"を検討できるよう、比較基準を最速で作りたい
- ▶ 見込みがなければ即撤退できるよう、投入リソースを最小限にしたい

ここまでできてから、徐々に手法を高度化していきましょう

まとめ

- 高度な手法を身に着けるだけでは成果につなげられない
- 適切な課題設定と技術課題以外の対応が超重要
- 特に初心者は、意図的に簡単な手法から順に検討を進めよう

とにかく

- **技術課題の明確化**
- **技術面以外の課題への対応** そこからスタートです。