HITACHI
Inspire the Next

実務に役立つ分析コンペ

なぜ発表者が分析コンペで機械学習に取り組むか

2022/04/28

株式会社 日立製作所
Lumada Data Science Lab.
諸橋 政幸

自己紹介

HITACHI
Inspire the Next

- 株式会社 日立製作所
  Lumada Data Science Lab.
  データサイエンス・エキスパート

- 1999年に日立製作所へ入社。
- 2012年にデータ分析部署（その年度に新設）に異動し、データ分析を使って顧客課題を解決する業務に従事。金融・小売など多種多様な分野のプロジェクトを担当
  分析経験ゼロからスタートし、約10年間の実務経験を経て今に至る。

- 分析コンペ歴は約6年。Kaggle Master。

ところで分析コンペって知ってますか？

HITACHI
Inspire the Next

- 学生や若者がやっているもの
- 興味はあるけど参加の敷居が高そう
- そもそも聞いたことがない

１．コンペとの出会い

ゼロからの分析人生スタート

HITACHI
Inspire the Next

- 1999年入社して情報セキュリティ業務(8年) ＞ 金融事業部(4年)

- 2012年に現在のデータ分析／AI推進部署へ異動
  当時はビッグデータブーム真っ最中

- その後もAI、ディープラーニングと
  名称を変えつつも
  「データ利活用」の機運が続く

- 一見花形部署だけども、
  自身は当時30後半かつ未経験の分野、、、

こんな仕事してます

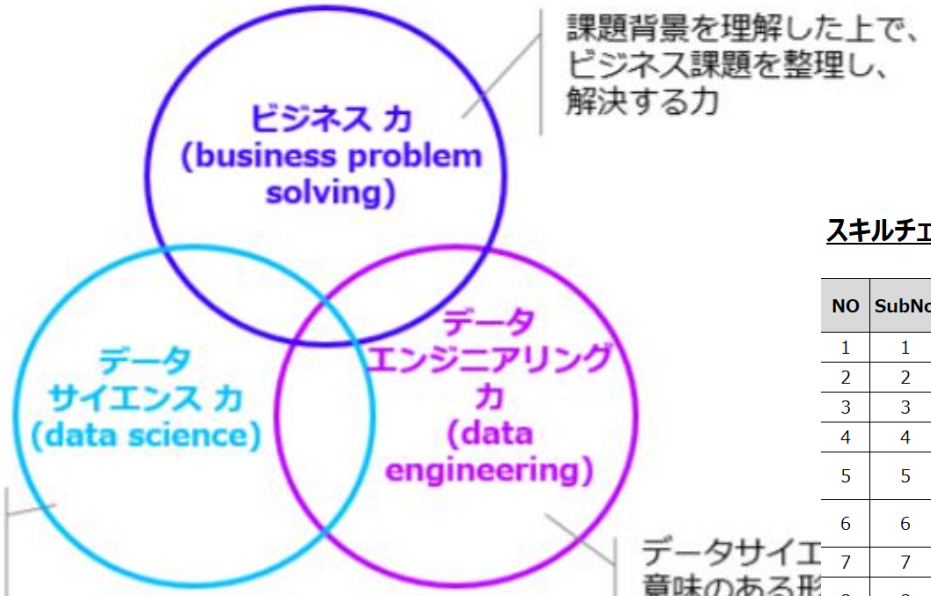
- ざっくり言うと、
  顧客の課題をデータを使って解決する仕事

- 顧客が解決したい課題の明確化
- どんなデータを活用してどうやって解決するのかを具体化
- （大規模）データの前処理加工
- データの集計・可視化
- 機械学習やディープラーニングを用いたモデル学習
- 報告書作成
- 顧客との打ち合わせ・報告

スキルセットの多さに絶望

HITACHI
Inspire the Next

- 大きくは３つのスキル



課題背景を理解した上で、
ビジネス課題を整理し、
解決する力

情報処理、人工知能、統計学などの情報科学系の知恵を理解し、使う力

データサイエ
意味のある形
ようにし、実
できるように

- 各スキルの広さと深さ

スキルチェックリスト　2021年版　<ビジネス力>

NO	SubNo	スキルカテゴリ	スキルレベル	サブカテゴリ	チェック項目
1	1	行動規範	★	ビジネスマインド	ビジネスにおける「論理とデータの重要性」を認識し、分析的でデータドリブンな考え方に基づき行動できる
2	2	行動規範	★	ビジネスマインド	「目的やゴールの設定がないままデータを分析しても、意味合いが出ない」ことを理解している
3	3	行動規範	★	ビジネスマインド	課題や仮説を言語化することの重要性を理解している
4	4	行動規範	★	ビジネスマインド	現場に出向いてヒアリングするなど、一次情報に接することの重要性を理解している
5	5	行動規範	★★	ビジネスマインド	社会における変化や技術の進化など、外的要因による分析プロジェクトへの影響をある程度見通し、柔軟に行動できる
6	6	行動規範	★★	ビジネスマインド	ビジネスではスピード感がより重要であることを認識し、時間と情報が限られた状況下でも、言わば「ザックリ感」を持って素早く意思決定を行うことができる
7	7	行動規範	★★	ビジネスマインド	作業ありきではなく、本質的な問題（イシュー）ありきで行動できる
8	8	行動規範	★★	ビジネスマインド	分析で価値ある結果を出すためには、しばしば仮説検証の繰り返しが必要であることを理解し、粘り強くタスクを完遂できる
9	9	行動規範	★★★	ビジネスマインド	プロフェッショナルとして、作業量ではなく生み出す価値視点で常に判断・行動でき、真に価値あるアウトプットを生み出すことにコミットできる
10	10	行動規範	★	データ・AI倫理	データを取り扱う人間として相応しい倫理を身に着けている（データのねつ造、改ざん、盗用を行わないなど）
11	11	行動規範	★	データ・AI倫理	データ、AI、機械学習の意図的な悪用（フェイクニュース、Botの悪用など）があり得ることを勘案し、技術に関する適切な知識と倫理を身につけている
12	12	行動規範	★★	データ・AI倫理	AI・機械学習がもたらす現在の倫理課題を説明できる（ディープフェイクによるプライバシーの侵害、バイアスによる人種差別、学習済みモデルのリバースエンジニアリングによる知的財産権の侵害など）
13	13	行動規範	★★★	データ・AI倫理	会社や組織全体におけるデータの取り扱いに関する倫理を維持・向上させるために、必要な制度や仕組みを策定し、その運営を主導することができる
14	14	行動規範	★	コンプライアンス	直近の個人情報に関する法令（個人情報保護法、EU一般データ保護規則：GDPRなど）や、匿名加工情報の概要を理解し、守るべきポイントを説明できる
15	15	行動規範	★★	コンプライアンス	担当するビジネスや業界に関係する直近の法令・ガイドラインを理解しており、データの保持期間や運用ルールに

報の概要を理解し、守るべきポイントを説明できる
担当するビジネスや業界に関係する直近の法令・ガイドラインを理解しており、データの保持期間や運用ルールに
担当するビジネスや業界に関係する直近の法令・ガイドラインを理解しており、データの保持期間や運用ルールに

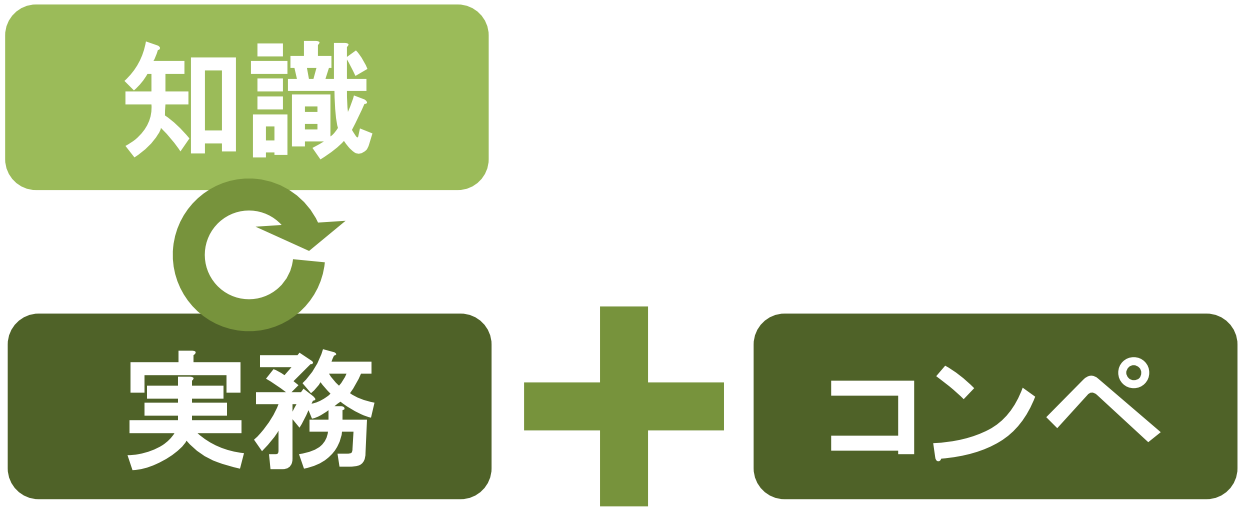
とにかく独学の日々

- とりあえず書籍で勉強の日々（100冊以上、本だらけ）
- 何が正解か分からないまま現場での実務
- いっそのこと大学で学び直そうか。。。

取り組み方に気付く

- 知ってるだけでは結局は不十分
  磨いてはじめて生きたスキルとなる

- 経験から得てしまうのが最短ルート
  それが実務と『コンペ』



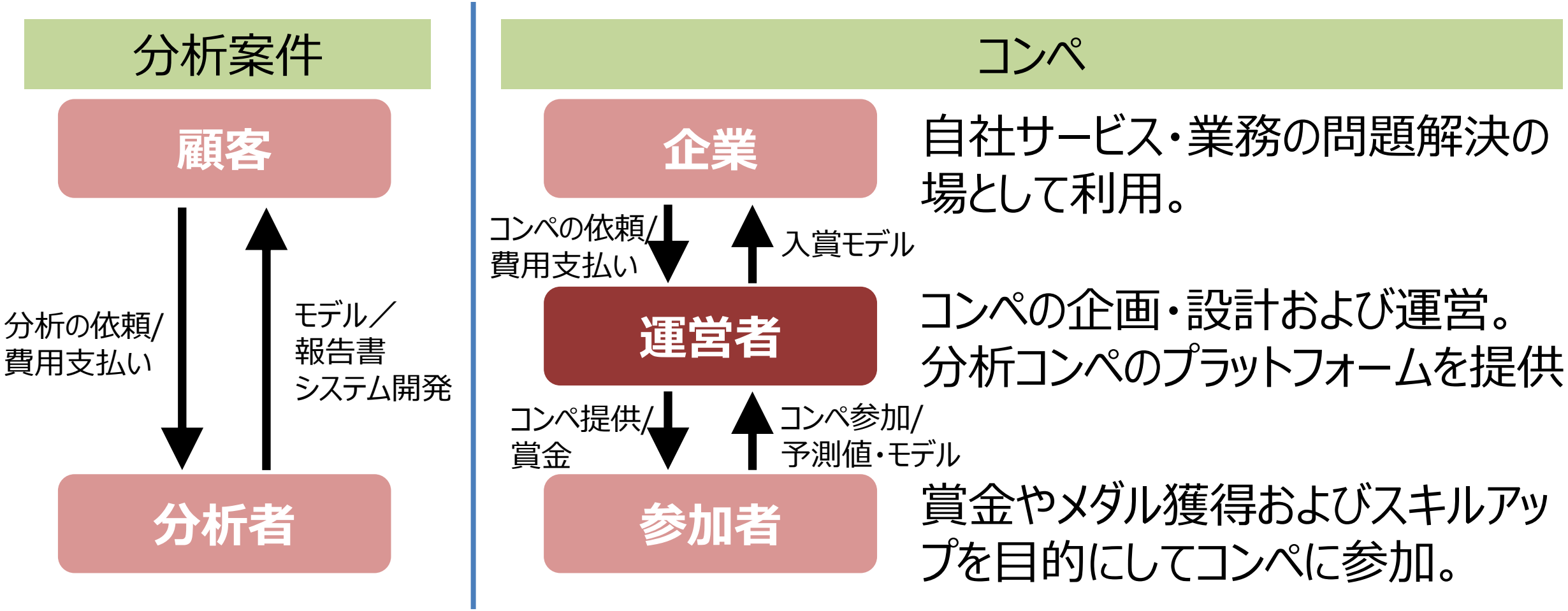
HITACHI
Inspire the Next

2．分析コンペとは

分析コンペのプラットフォーム

HITACHI
Inspire the Next

- 企業とデータサイエンティストをつなぐ仕組み



国内外のコンペサイト

- 日本国内にも有用なサイトがいくつかある
- 世界的にはKaggleが規模・知名度ともにトップ

コンペサイト	URL
Kaggle	https://www.kaggle.com/
SIGNATE	https://signate.jp/
Nishika	https://www.nishika.com/
ProbSpace	https://comp.probspace.com/
atmaCup	https://www.guruguru.science/

- 基本的には予測精度を競う
  - 店舗別の商品の販売個数を予測
  - 画像の分類ラベルを予測

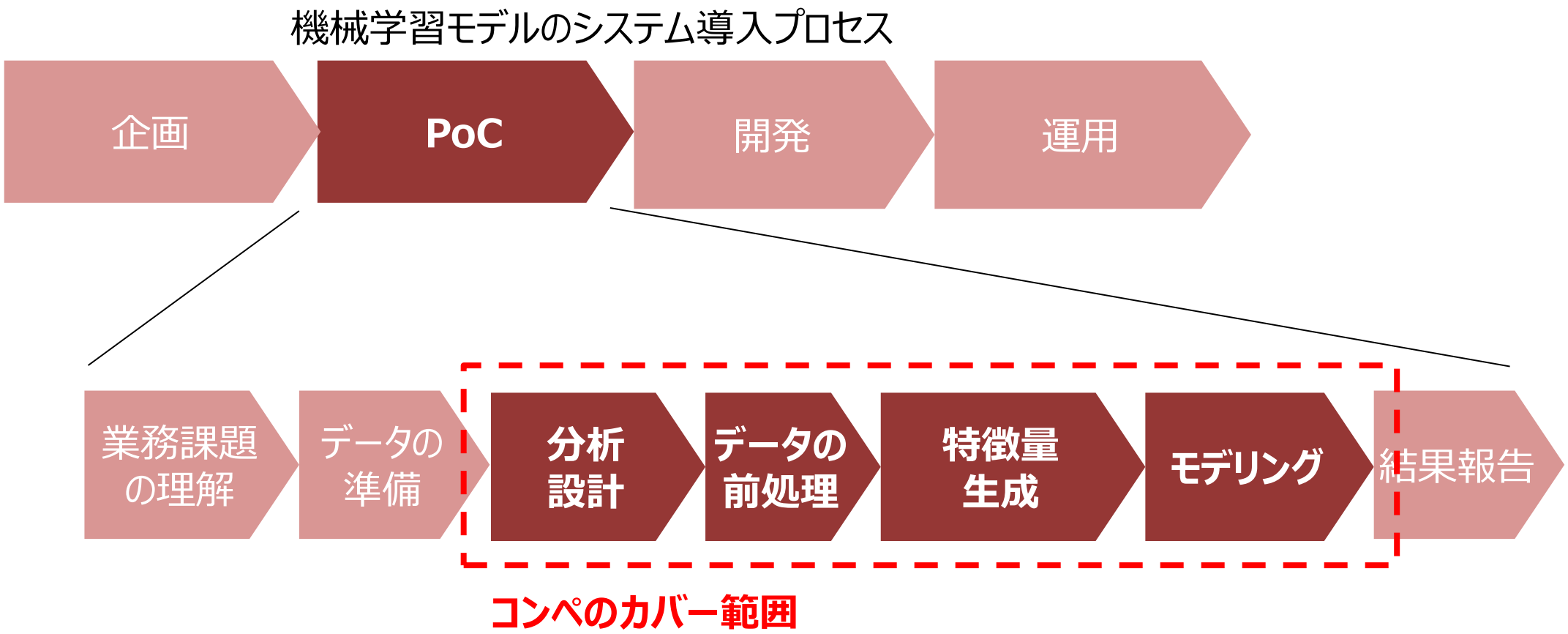
- 参加者のモチベーションは
  - 順位、メダル、称号 ★ゲーム性が高い
  - スキルアップ、力試し
  - 趣味として

コンペのカバー範囲：技術

- 教師あり学習
- 強化学習
- ディープラーニング
  - 画像分類、領域検出、セマンティックセグメンテーション
  - 自然言語の分類
  - 音声分類
- 最適化問題

コンペのカバー範囲：業務プロセス

- 実務の一部。しかしスキルが問われる重要なタスク。



３．得られたこと

コンペで得られたこと

HITACHI
Inspire the Next

① 分析設計・ベースライン作成の高速化

② 技術の使いこなしスキル

③ 新しい技術を知るチャネル

① 分析設計・ベースライン作成の高速化

HITACHI
Inspire the Next

- 多くの経験からコードへの落とし込みが早くなる
  - ビジネス課題 ⇒ 分析設計 ⇒ ベースライン
  - データ利活用で解ける課題へ誘導

- 検証の仕組みの重要性を体感
  - 手元のスコアと、リーダーボードのスコアのギャップ
  - シェイクダウンなどの失敗経験を実務へ活かす

② 技術の使いこなしスキル

HITACHI
Inspire the Next

- 分析にはライブラリや技術がたくさんある
  - LightGBM, tensorflow, torch, BERT,,, etc.
  - チューニングスキルが必要、未経験だと学習コストが必要
  - 多くのタスクで多くのライブラリ利用経験を持つべき

- 知っている << 使ったことがある
  - 理論は大事だが、極端に言うと実践は理論を上回る

- 経験に裏付けされた技術の取捨選択
  - 技術の有用性・限界を知っているからこその適切な分析設計へ

③ 新しい技術を知るチャネル

- 新しい技術を知る／試すチャンス
  - コンペ駆動型の勉強法
  - コンペで使ってみて本当に役立つかを判断できる

- コンペから得た人脈
  - チームを組めるので会社や国の枠を超えた人脈が出来る
  - そこから新しい技術やノウハウを知ることができる

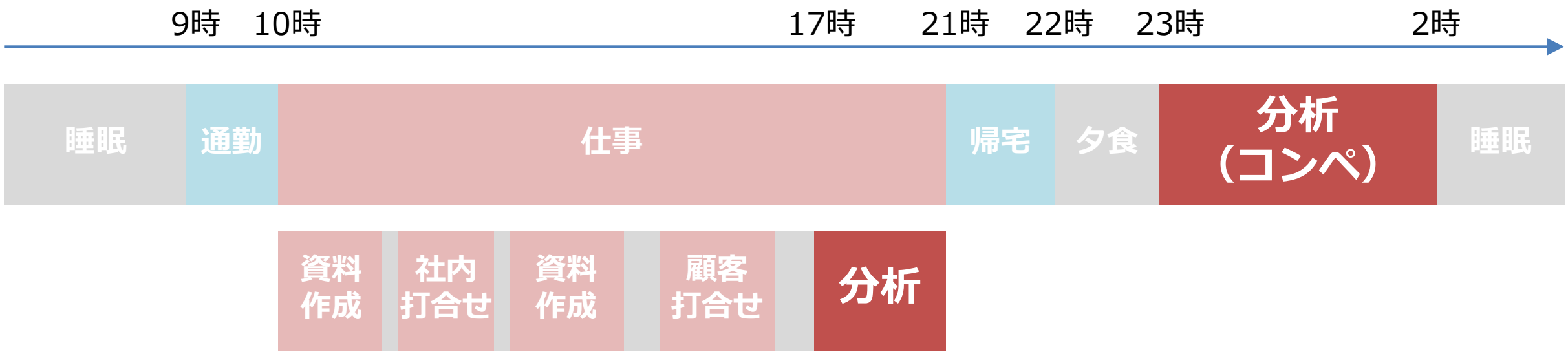
４．私の活用の仕方

こんな感じで取り組んでます

- まず学習目的を定める
- 振り返りは大事
- 仕事ではないので
  楽しいと思うコンペを選ぶ
- メダルや順位は気にしない

とある平日

- 仕事で分析しながら、
  早く家に帰って分析したいと思う日々。。。



コンペの魅力とは

HITACHI
Inspire the Next

楽しい

興味持った方は是非試しにやってみてください！
嵌りすぎて寝不足になってしまったらすみませんｗ

HITACHI
Inspire the Next

Kaggleで磨く 機械学習の実践力
――実務xコンペが鍛えたプロの手順

- 出版社: リックテレコム
- B5変型判　376ページ
- 定価：3,300円（税込）
- ISBN：978-4-86594-326-9
- 2022年5月23日刊行
- 諸橋政幸 著

■本書の主な内容

HITACHI

Inspire the Next