

# 機械学習を「社会実装」すること

～社会実装を阻む罣とその解決方法の考察～

Moe Uchiike

# 自己紹介 - 内池 もえ -

## 誰？

- 株式会社ブレインパッド アナリティクス本部 アナリティクスサービス部所属 リードデータサイエンティスト
- 機械学習による需要予測プロジェクトのプロジェクトマネージャーとして日々奮闘
- 手掛けたアルゴリズムは現在グローバルでリリースされており、今この瞬間も現場で使われ続けている
- 松尾研究室主宰 DL4USの2期生。最終課題は「おでんの需要予測」



ということで、

本日は「**機械学習プロジェクトをきちんと本番運用まで持っていった実務家の立場**」からお話しさせていただきます！

# 目次

---

アイスブレイク      パンデミック時の需要予測、どうする？

---

第 1 章              機械学習プロジェクトの現実

---

第 2 章              社会実装までのプロセスと「罨」のマッピング

---

第 3 章              社会実装を阻む「罨」と、その解決策

---

まとめ              本日も伝えなかったこと

---

アイスブレイク：パンデミック時の需要予測、どうする？

# アイスブレイク：パンデミック時の需要予測、どうする？

まずはアイスブレイクということで、1 つ課題を出します。以下の状況を踏まえ、自分なりの答えを出してみてください。

## 状況

- 皆さんは、世界中に店舗を展開している和食チェーンにおいて、機械学習による需要予測プロジェクトのPMを務めています。
- このプロジェクトでは、翌月に必要になる食材の需要量をモデルで予測し、その予測結果をもとに食材が発注・納品されることを目指します。
- 既に皆さんは数々の試練を乗り越え、いよいよ本番稼働というタイミングになりました。
- ところが、このタイミングでパンデミックが起ってしまった。このパンデミックはいつ収拾がつくかわかりません。
- 学習データは過去 3 年分しかなく、かつ過去 3 年間に類似のパンデミックは起こっていません。

皆さんならPMとして、この問題にどう立ち向かいますか？

3 分間で考えてみてください

# アイスブレイク：パンデミック時の需要予測、どうする？

いかがでしたでしょうか。この問題に明確な答えがあるわけではありません。

ですが、**ありとあらゆることを考えなければならない**ことがわかります。この事実に気づくことが機械学習の「社会実装」に直結します。

## 恐らく皆さんが考えたこと

- そんな時期に予測が当たるはずがないじゃないか！
- そもそも店舗は開いているのか？
- 需要予測が「大外れ」した場合の経済的損失やフードロスは？
- その責任は一体誰が取るのか？
- そもそもこの期間にモデルを稼働させるのか？
- 仮に稼働させるとして、予測値は「後処理」するべきではないか？
- その「後処理」は何が適切か？ルールベースなのか？
- ローンチを遅らせてみるのはどうか？
- ローンチを遅らせたとして、パンデミックの期間のデータはモデルに学習させていいのか？

我々が立ち向かわなければならないのは、まさにこのような問題の数々！

## 第 1 章：機械学習プロジェクトの現実

# 第1章：機械学習プロジェクトの現実

## ほとんどのプロジェクトが社会実装されずに終わる

まず認識しなければならないのが、ほとんどの機械学習プロジェクトが社会実装されずに終わるという事実です。



✓ PoCはそう簡単に上手くいかない

✓ PoCが上手くいったとしても、その後のフェーズも一筋縄ではいかない

✓ 遡って、「上流フェーズでの問題設定が適切ではなかった」ということもある

PoC : Proof of Conceptの略語で、実証実験（実際に上手くいくか確かめる活動）のこと

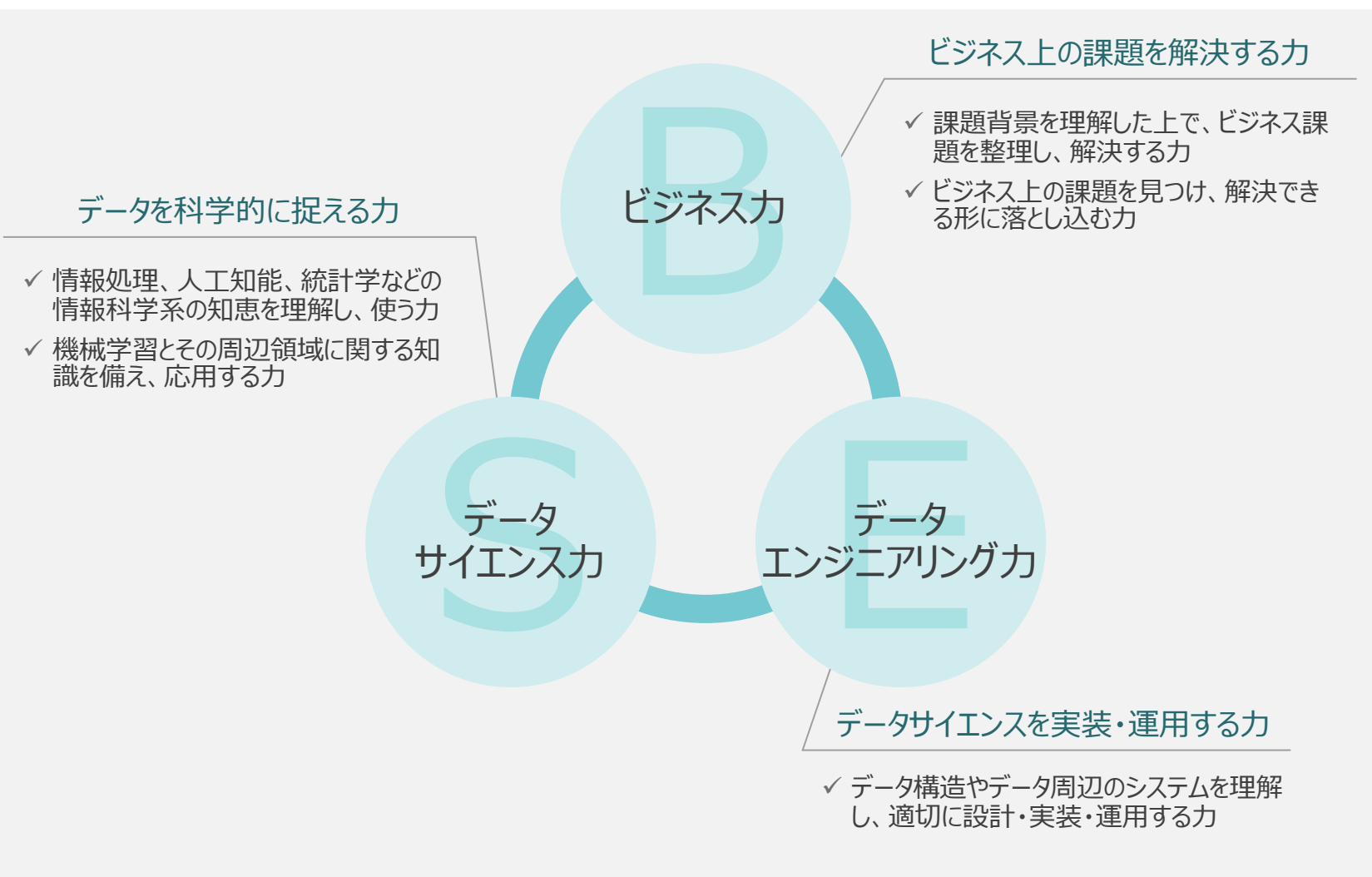


機械学習・データサイエンスに真摯に取り組む立場として、これほど悲しいことはない……

なぜなのか？

# 第1章：機械学習プロジェクトの現実 なぜ社会実装できないのか

機械学習プロジェクトが社会実装されずに終わってしまう理由は、一言でいうと「理想と現実のギャップ」です。



- 技術側とビジネス側の「食い違い」の時代はさすがに終わり、徐々に左記のモデルが共通認識になりつつある
- 左記のモデルは決して間違っていない
- しかし、これはあくまでも「理想」。現実的には多くの「罫」が待ち構えている

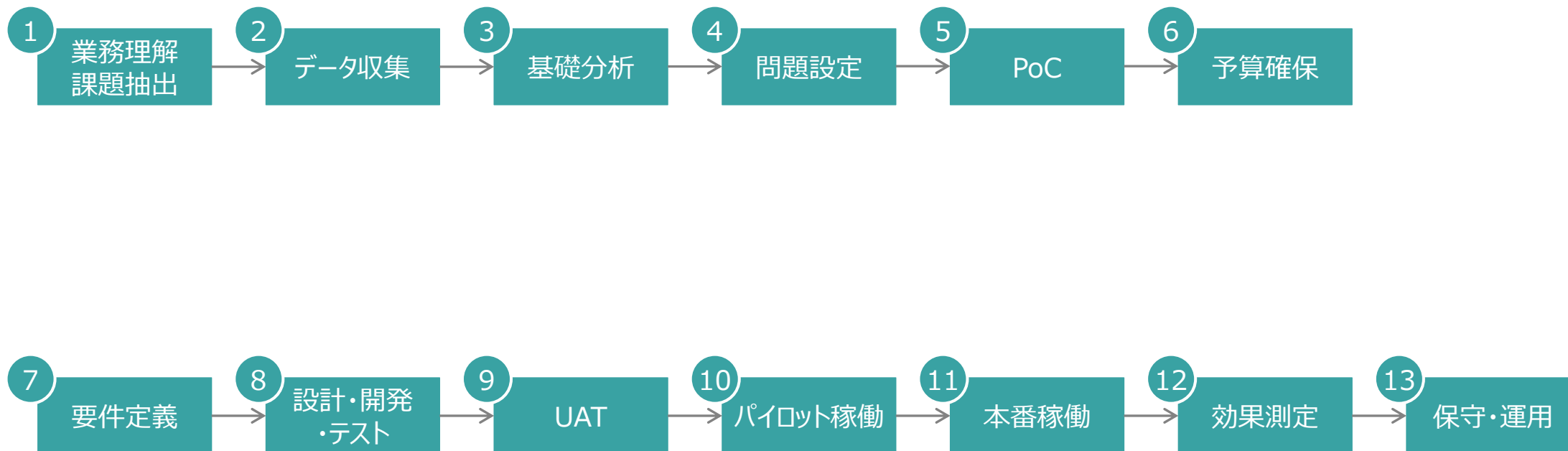


- 「視野を広く」「視界の解像度を上げて」「時には泥臭く」取り組んでいく必要がある
- **データサイエンスとはそういうジャンル**

## 第2章：社会実装までのプロセスと「罨」のマッピング

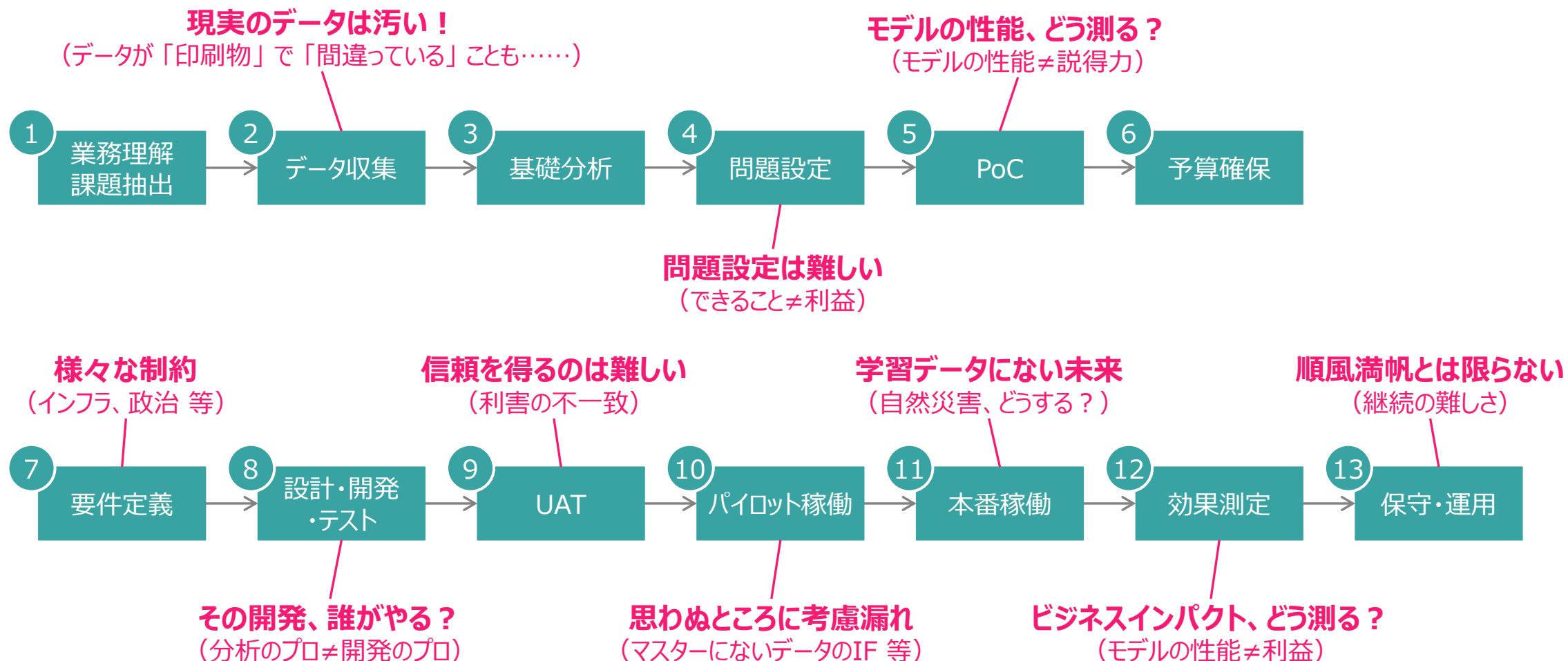
## 第2章：社会実装までのプロセスと「罫」のマッピング 一般的な機械学習プロジェクトのプロセス

一般的な機械学習プロジェクトでは、Kaggleのように初めから「綺麗な問題」が用意されているわけではありません。もはや、機械学習プロジェクトのメインはEDAやモデル構築ではないと言っても過言ではありません。



## 第2章：社会実装までのプロセスと「罣」のマッピング 現実

前スライドで機械学習プロジェクトの膨大なタスクについてお話しさせていただきました。  
しかし、それだけではありません。これらのプロセスには、たくさんの「罣」が待ち構えています。



### 第3章：社会実装を阻む「罣」と、その解決策

# 第3章：社会実装を阻む「罣」と、その解決策

## 【データ収集】現実のデータは汚い！

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

構造化されたデータがクラウド上の列指向DBに格納されており、容易に収集可能

### 罣

- 「利用できるデータがたくさんある」と聞いていたが、**そのデータは実は印刷物**であった
- データはあるが、同じデータなのに**データ取得断面によって数字が違う**
- データはあるが、データ保有部門との関係が悪く、**データを渡してもらえない**
- データはあるが、**常に上書き処理されており、蓄積されていない**

### 解決策

- ✓ 現状のデータの品質について、関係者一同で事前に認識を合わせる
- ✓ 知恵を絞ってデータクレンジングを行い、なんとか利用できる形にする
- ✓ 役職者からトップダウンで業務命令が下るように関係者との調整に奔走する
- ✓ データを定期的に蓄積するスキームを作り、すぐにデータの蓄積を開始する

### 一言コメント

構造化されたデータがあればラッキー！ぐらいに思っておきましょう。適切にデータを管理している組織は少ないです。



## 第3章：社会実装を阻む「罫」と、その解決策

### 【問題設定】問題設定は難しい

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

#### 理想

解くべき問いと解ける問いが一致しており、特段疑問を抱かずに機械学習の問題に落とし込める

#### 罫

- （例）賞味期限の長い「かんぴょう」の需要量を完璧に予測できるモデルができたが、**ビジネス的に意味があるとは到底思えない**
- （例）**実は需要量をニアピンで当てることよりも、「大外し」によるロスをなくすることが重要**であることを後から指摘された

#### 解決策

- ✓ 機械学習の問題としての解きやすさと、ビジネス的な効果の双方を考慮して施策をデザインする
- ✓ 本当に解くべき問いは何なのか？を必要なステークホルダーを巻き込んで議論する
- ✓ 「解くべき問い」と「解ける問い」が一致しない場合、振り出しに戻って再度検討する、あるいは機械学習以外の最適化問題に落とし込む

#### 一言コメント

「解ける問い」と「解くべき問い」は往々にして一致しません。  
時には機械学習以外の方法を検討する必要もあります。（数理最適化等も課題解決のための強力な手段です）

考えてみよう：モデル性能、どう測る？

## 考えてみよう：モデルの性能、どう測る？

以下の状況を踏まえ、自分なりの答えを出してみてください。

### 状況

- 皆さんは、世界中に店舗を展開している和食チェーンにおいて、機械学習による需要予測プロジェクトのPMを務めています。
- 今まさに、解くべき問題を「来月必要な食材の需要量」と設定し、PoCを回そうとしています。
- 食材は米、野菜、肉、かんぴょうなど様々です。
- A国では「かんぴょう巻」が絶大な人気を誇っていますが、B国ではあまり人気がありません。

皆さんなら、どのような指標・考え方でモデルを評価しますか？

3 分間で考えてみてください

# 考えてみよう：モデルの性能、どう測る？

いかがでしたでしょうか。この問題に明確な答えがあるわけではありません。

ですが、**ありとあらゆることを考えなければならない**ことがわかります。この事実に気づくことが機械学習の「社会実装」に直結します。

## 恐らく皆さんが考えたこと

- 予測値と実績値の誤差を最小化すればよいのだから、素直にMAEで評価すればいいのではないかな？
- いや、「大外れ」は賞味期限の問題で修正がきかないのだからRMSEで評価すべきなのではないかな？
- 「当てるべきもの」と「当てなくていいもの」が存在するのではないかな？（例えば「かんぴょう」の需要量を当てても意味がない）

## 私ならこういうことも考える

- 「過剰予測」と「過小予測」にどのように重みづけをするかな？（過剰在庫と販売機会損失の重みを天秤にかける）
- 国別、あるいは地域別に必要とするモデルの振る舞いは異なるのではないかな？
- 最終的に「良いモデル」であることをどう定義し、どう証明するか？（絶対評価とするか？何かと比べて相対評価とするか？それぞれの場合の効果試算をどのように行うか？）

我々が立ち向かわなければならないのは、まさにこのような問題の数々！

# 第3章：社会実装を阻む「罣」と、その解決策

## 【PoC】モデルの性能、どう測る？

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

どこからどう見てもモデルの性能、およびビジネス上の効果に疑いの余地がない

### 罣

- 機械学習に過度の期待をされており、意思決定者の**期待を意図せず裏切ってしまう**
- 意思決定者が**性能指標を理解できない**、あるいはモデルの性能がビジネス上の利益に結びつく実感が湧かず**投資判断ができない**
- モデルの解釈性が低く、その**モデルを信頼する根拠として不足がある**

### 解決策

- ✓ 「良いモデル」を緻密に定義し、事前に役職者を含めて合意を取っておく
- ✓ 誰にでもわかりやすく、かつ本質を損なわない指標を定義した上でバックテストする
- ✓ 無理に深層学習に寄せず、回帰木などの解釈性の高いモデルも候補に入れた上で、総合的に判断できるようにコーディネートする

### 一言コメント

基本的に、「交差検証しました。はいOK！」とはならないと考えておくべきです。  
PoC時の性能評価は、まさに「社会実装」できるかどうかを左右する重要事項です。

# 第3章：社会実装を阻む「罣」と、その解決策

## 【要件定義】様々な制約

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

クラウド上に仮想環境を立ち上げ、ML Opsが効率よく回るように要件を定義できる

### 罣

- 既存のデータ基盤との兼ね合いやセキュリティ上の制約により、**クラウド上の仮想環境へのスムーズな移行が難しい**
- **既存の仕組みを変えたくない部署**の発言権が社内で強い
- **機械学習の運用に強いチームが存在しない**

### 解決策

- ✓ 既存の基盤をある程度活かした仕組みを構築し、段階的に移行していくプランを立てる
- ✓ 役職者を巻き込み、定期的にディスカッションの場を設けて意思決定を促す
- ✓ 運用負荷が極力低くなるように要件を定義する、あるいは新たに運用チームを編成し、開発と運用のサイクルがスムーズに回る体制を築く

### 一言コメント

本番環境に乗せるとなると、PoCと比べて何倍も強い制約が課される場合が多いです。「理論の理解」や「実装力」では太刀打ちできない領域もあります。得意な人に任せてしまうのも手です。

# 第3章：社会実装を阻む「罣」と、その解決策

## 【設計・開発・テスト】その開発、誰がやる？

### 理想

データサイエンスとエンジニアリングの両方に長けた人材をアサインする

### 罣

- PoCが終わり、いよいよ開発フェーズとなったものの、いざ本番環境で開発するとなると**どのように開発していけばいいかわからない**
- 代々あらゆる開発を担っているベンダーが常駐しているが、データサイエンスに関する知見は乏しく、**本番環境での開発を担える見込みがない**
- データサイエンスに長けたメンバーと本番環境での開発に長けたメンバーがそれぞれいるが、**コミュニケーションに難があり両車輪が動かない**

### 解決策

- ✓ データサイエンスの担当者の他に、エンジニアリングの担当者をアサインする（可能であれば初期段階からアサインしておき、スムーズに本番環境の開発に入れるように準備しておく）
- ✓ 上記の解決策を講じた上で、エンジニアリングの担当者が間に入り、開発に必要な知見の共有や、メンバー間のコミュニケーション促進の役割を担う

### 一言コメント

データサイエンスとエンジニアリングの両方に長けた人材のアサインが理想ですが、そう上手くはいきません。ある程度のエンジニアリングの知識は必要ですが、基本的には分業を前提に現実的な解を見つけていきましょう。

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

# 第3章：社会実装を阻む「罣」と、その解決策

## 【UAT】信頼を得るのは難しい

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

モデルの性能が良く、現場からの評判も上々。スムーズに次のフェーズに移行できる

### 罣

- そもそも性能の良いモデルを提供しても**現場担当者には旨味がなく**、既存のオペレーションを変えたくない層から**ネガティブな意見が出る**
- 予測が外れたごく一部について現場担当者に固執されてしまい、**モデルを信頼してもらえない**
- 確かにモデルの性能は良いが、実際に現場のオペレーションに組み込んでみたところ、**使いにくい部分があることがわかった**

### 解決策

- ✓ 予測が当たった場合のメリットについて、経営目線だけでなく、現場目線で整理する
- ✓ 全体としての評価や、ポジティブと考えられる要素について丁寧に説明する
- ✓ 予測が外れた原因を可能な範囲で分析し、説明して納得してもらう
- ✓ ユーザーからの意見を漏れなく吸い上げ、改善すべき点については改善を試みる

### 一言コメント

「機械学習だから……」という主張は、社会実装のフェーズでは通用しないことが多々あります。きちんと「使う側のメリット」を提示し、Win-Winの関係でプロジェクト進めていくのが正解です。



# 第3章：社会実装を阻む「罨」と、その解決策

## 【パイロット稼働】思わぬところに考慮漏れ

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

PoCと同様の性能が出ており、目立ったバグもなく順調に稼働している

### 罨

- **PoC時と同等の性能が示せず、本番稼働に踏み切れない**
- マスターの内容が刻々と変化しており、**存在しないカテゴリカル変数が特徴量として投入されてしまう**
- 定められた時刻までに必要なデータがIFされてこず、**必要な特徴量がNULLのまま予測処理が行われてしまう**

### 解決策

- ✓ PoC時と同じ品質のデータが使えるとは限らないことを認識する
- ✓ 可能な限り本番に近い形でバックテストを実施し、ある程度の性能が出ることを担保しておく
- ✓ 機械学習モデルによる予測値を過信せず、異常値を回避するための仕組みを複数用意し、セーフティネットを張り巡らせておく

### 一言コメント

たった一度の失敗で、機械学習モデルのような「わかりにくいもの」に対する信頼は崩れ去ります。そうならないために、「事故が起きない仕組みづくり」を徹底しましょう。

考えてみよう：オリンピックヤーの需要予測、どうする？

# 考えてみよう：オリンピックイヤーの需要予測、どうする？

以下の状況を踏まえ、自分なりの答えを出してみてください。

## 状況

- 皆さんは、世界中に店舗を展開している和食チェーンにおいて、機械学習による需要予測プロジェクトのPMを務めています。
- 様々な困難を乗り越え、ようやくシステムローンチすることができました。
- ひとまず問題なく動いており、現場からの評判も上々です。
- しかし、100店舗を展開しているヨーロッパのA国で、来年オリンピックが開催されることに気づきました。
- 学習データは3年分しかなく、オリンピック期間の需要量については見当が付きません。  
(ここではオリンピックに準ずる規模のイベントもなかったと仮定します)

皆さんならPMとして、この問題にどう立ち向かいますか？

3分間で考えてみてください

# 考えてみよう：オリンピックヤーの需要予測、どうする？

いかがでしたでしょうか。この問題に明確な答えがあるわけではありません。

ですが、**ありとあらゆることを考えなければならない**ことがわかります。この事実気づくことが機械学習の「社会実装」に直結します。

## 恐らく皆さんが考えたこと

- オリンピック開催国は、インバウンド需要の増加により売上が大幅増になるはず
- 過去のオリンピック、あるいはそれに準ずるイベント時のデータで学習しているモデルを構築するのがベター
- 一方、オリンピック、あるいはそれに準ずるイベント時のデータを持っていないため、事実上それは不可能
- 通常通りにモデルが予測をすると、売上の大幅増を過小評価してしまう可能性がある。どうすべきか？
- 需要の大幅増が見込まれる場合、何らかの特徴量として投入できるモデルに改良するべきではないか？
- そこまでしなくても対応できる手段は何かないか？

我々が立ち向かわなければならないのは、まさにこのような問題の数々！

# 第3章：社会実装を阻む「罫」と、その解決策

## 【本番稼働】学習データにない未来

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

ローンチ後、システムは順調に稼働。社会情勢に大きな混乱はなく、モデルは質の高い予測をし続けている

### 罫

- **学習データの期間に存在しないイベント**が行われることとなった
- **観測史上最大の台風**が日本列島に上陸し、猛威を振るう見込みである
- **突然のパンデミック**。モデルはパンデミック時の需要量の傾向を学習しておらず、妥当な水準の予測ができる保証がない

### 解決策

- ✓ 解釈性の高いモデルやルールベースのアルゴリズムとの2段構えの仕組みにしておき、必要に応じてスイッチできるようにしておく
- ✓ 緊急時に運用回避できるよう、緊急時用のオペレーションを組み、日頃から周知しておく
- ✓ 各拠点の状況を丁寧にヒアリングしつつ、モデルの利用可否や再開タイミングについて1つ1つ判断する

### 一言コメント

今まさに起きている状況で、様々な機械学習プロジェクトが苦戦を強いられていることは想像に難くありません。

# 第3章：社会実装を阻む「罣」と、その解決策

## 【効果測定】ビジネスインパクト、どう測る？

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

**理想** ローンチ後、期待を上回る性能を示し続けており、もはやモデルのビジネスへの貢献には疑いの余地がない

### 罣

- モデルの性能は高い値を示しているが、それが具体的にどのように**ビジネスに貢献できているかがわからない**
- ビジネスへの貢献度合いを測れていないため、中長期的に改善していくための**予算が出ない**
- ビジネスへの貢献度合いを可視化した**が、モデルの性能に反してビジネスインパクトが小さい**

### 解決策

- ✓ モデルの性能が必ずしもビジネス上の効果に結びつかないことを認識し、ビジネス上のインパクトを定量的に可視化する指標を新たに作成する
- ✓ モデルのパフォーマンスとビジネス上のインパクトの相関関係や因果関係を分析し、モデルがビジネスへの貢献を果たしていることを裏付ける
- ✓ モデルの性能がビジネスへの貢献に結びつかない原因を探り、対策を講じる

### 一言コメント

ビジネス上の効果が見えなければ、次の投資判断にダイレクトに影響してきます。  
そうならないために手を打つことが、機械学習の「社会実装」の拡大に繋がります。

# 第3章：社会実装を阻む「罣」と、その解決策

## 【保守・運用】順風満帆とは限らない

- 1 業務理解  
課題抽出
- 2 データ収集
- 3 基礎分析
- 4 問題設定
- 5 PoC
- 6 予算確保
- 7 要件定義
- 8 設計・開発  
・テスト
- 9 UAT
- 10 パイロット稼働
- 11 本番稼働
- 12 効果測定
- 13 保守・運用

### 理想

ローンチ後、特に問題なくモデルを運用できており、システム改修の必要性も特段ない

### 罣

- 手運用が必要な場面が多く、**運用に多くの工数がかかる**
- データ取得元のテーブル仕様に認識していない変更があり、**予測前処理時にエラーを吐いてしまう**
- 特徴量として使っていた**データが諸般の事情で使えなくなる**（データ提供元の方針変更、組織の意向等）

### 解決策

- ✓ 手運用の必要が少なく、尚且つ他のシステムと疎結合な設計にする
- ✓ 変更情報をキャッチできないことがないよう、ステークホルダーとの情報共有を密に行う。あるいは、情報共有のスキームを作っておく
- ✓ それでも解決しない場合は、関係者とコミュニケーションを取り、工数を確保の上、解決に向けて奔走する

### 一言コメント

機械学習プロジェクトでも、“負の遺産”を残さないようにシステム・運用を設計することが必要不可欠です。また、保守・運用フェーズにおいては関係者とのコミュニケーションが肝になることが多々あります。

まとめ：本曰お伝えしたかったこと



## まとめ 本日お伝えしたかったこと

1

✓ 「第3次AIブーム」が叫ばれる昨今だが、「社会実装」するとなると甘くはない

2

✓ 「社会実装を阻む罫」に対してアンテナを張り巡らせる必要がある

3

✓ 「視野を広く」「視界の解像度を上げて」「時には泥臭く」プロジェクトに取り組む

## 関連情報

### 『いちばんやさしい機械学習プロジェクトの教本』

(株) ブレインパッド 荻原 祐介 著、インプレス、2018年  
<https://www.amazon.co.jp/dp/B07BXSC9XT>

### BrainPad Inc. SlideShare

機械学習プロジェクトのあれこれについて発信しています  
<https://www.slideshare.net/BrainPad>

### 『ブレインパッドにおける機械学習プロジェクトの進め方』

(株) ブレインパッド 太田 満久 作、2019年  
<https://www.slideshare.net/BrainPad/ss-149214163>

### OpenBrainPad Project

(株) ブレインパッド社内にある技術資料の公開等を行っています  
<https://brainpad.github.io/OpenBrainPad/>

ご清聴ありがとうございました！

*Thank you for your time and attention!*