

製造業の AI需要予測

統計・機械学習・現場の定性知を
つなぐ実装アーキテクチャ

*AI-Driven Demand Forecasting in Manufacturing:
An Implementation Architecture Connecting Statistics,
Machine Learning, and Frontline Tacit Knowledge*

PUBLISHED BY

taiziii Research

Manufacturing DX Desk

本書の位置付け

本書は、製造業のDX担当者（事業部のDX推進部門、生産管理・SCM部門の部長・課長クラス）を読者に想定した、需要予測の実装に関する実務レポートである。「AIで需要予測ができる」という総論ではなく、**どのデータを、どのアーキテクチャで、どの順序で組み上げれば、現場で動く需要予測になるのか**という技術的な具体に踏み込む。

需要予測は、稼働率向上と並んで製造業DXの「永遠のテーマ」である。予測が外れれば、過剰在庫として運転資金を縛り、欠品として販売機会を逃す。その上流に位置するため、需要予測の質は生産スケジュール・在庫・調達のすべてに波及する。にもかかわらず、多くの企業の予測は今なお担当者の経験と表計算ソフトに依存し、AI・機械学習の導入は道半ばにある。

本書が一貫して主張するのは、需要予測の高度化は「**より賢いアルゴリズムを買うこと**」ではなく「**3種類のデータをつなぐアーキテクチャを設計すること**」だという点である。すなわち、(1) 出荷・受注などの内部実績データ（統計の土台）、(2) 為替・天候・市況などのマクロ・外部環境データ、(3) 商談状況や顧客計画といった営業の定性情報。この3層をどう統合するかが、予測精度と現場での使い勝手を分ける。アルゴリズムは、この統合構造の上に乗る部品にすぎない。

本書は公開情報および一般的な実装知見に基づく。記載する精度水準（MAPE等）は製品特性・データ品質により大きく変動するため、絶対値ではなく相対的な傾向として参照されたい。

目次

1 なぜ需要予測がDXの起点になるのか

予測誤差が在庫・欠品・調達に波及する構造と、現状の壁

2 予測手法の地図 — 統計から因果モデルまで

手法ごとの適用領域、精度の天井、選択の判断軸

3 リファレンス・アーキテクチャと設計判断

手法ルーティング、予測階層のリコンサイル、本番で崩れる3つの落とし穴

4 3層のデータ統合という本質

内部実績×マクロ×営業定性をどうつなぐか

5 予測を計画につなぐ — FVAとS&OPの再設計

ナীব比での評価、価値を破壊する介入の発見、介入を例外に閉じ込める

6 先進事例 — 「現場で使える設計」で勝つ

キリンビール×ブレインパッド、ダイキンの実装に学ぶ

7 導入ロードマップと評価設計

スモールスタートの順序、RAE/FVA評価、つまずきポイント

需要予測DXの成否は、アルゴリズムの選定よりも、何のために予測するのかという目的の定義と、データ統合の設計、そして予測を計画行動に変換する運用の仕組みで決まる。本書の要点を5つにまとめる。

1

需要予測は、あくまでも意思決定の手段である

需要予測それ自体は目的ではない。生産・在庫・調達といった意思決定を、より良く下すための手段にすぎない。したがって評価すべきは予測精度の絶対値ではなく、その予測が在庫削減・欠品防止といった業務上の意思決定をどれだけ改善したかである。「精度の向上」を自己目的化した瞬間に、需要予測DXは迷走する。

2

どのような意思決定をするかによって、必要な精度は異なり、選択される予測モデルも自ずと異なる

数カ月先の補充・生産計画には、時系列モデルや機械学習回帰による高精度な予測が有効である（精度重視）。一方、1～2年先の設備投資や中期計画の判断には、需要のドライバを分解する因果モデルが適する（説明性・意思決定重視）。単一の万能モデルを探すのではなく、予測の時間軸と目的に応じて手法を使い分けることが現実解となる。

3

特に生産予測においては、需要発生頻度と需要量のばらつきによっても手法を変える必要がある。加えて、どの単位で需要予測を行うかも検討を要する

規則的に動く主力品目と、時々しか動かない補修部品とでは適した手法が異なる。需要の発生頻度とばらつきで品目を分類し、平滑な品目には統計・機械学習、間欠・散発な品目にはCroston／SBA等を充てる。さらに「SKU単位か、製品群×地域×週か」という予測の階層・粒度の設計が、精度と現場での使い勝手を大きく左右する。

4

需要予測のインプットは主に過去実績・マクロデータ・営業データの3層で、特に営業データの組み込みがカギ

過去実績は統計の土台、マクロデータ（為替・天候・市況）は因果モデルの説明変数、そして営業データ（商談状況・顧客の内示・受注見込み）は統計に現れない需要の先行情報を担う。とりわけ営業データを自由記述のまま放置せず、確度や数量として構造化し予測に組み込めるかが、他社が真似しにくい競争優位を生む。

需要予測モデルの運用は、人手の介入のコントロールがカギ

不都合な事実として、手間をかけた予測の多くは「前期実績をそのまま使う」ナイーブ予測にすら勝てていない。FVA（Forecast Value Added）で各工程を測ると、営業や経営の善意の介入がかえって精度を悪化させる例が頻出する。統計・MLの予測をデフォルトとし、人手の介入を価値が確認できる例外に限定することが、運用フェーズの核心である。

01 なぜ需要予測がDXの起点になるのか

予測誤差が在庫・欠品・調達に波及する構造と、現状の壁

KEY FINDINGS

- 需要予測はサプライチェーンの最上流に位置し、その誤差は在庫・欠品・調達・生産のすべてに増幅して波及する
- 多くの企業の予測は依然として担当者の経験と表計算に依存し、属人化と「当たらない前提の不在」が壁になっている
- 需要予測は「100%当てる」ものではなく、誤差を前提に在庫・調達のバッファ設計と一体で運用する対象である
- AIによる需要予測は精度向上そのものより、在庫削減・欠品防止・工数削減といった業務インパクトで評価すべき

1.1 予測誤差は下流に増幅して伝わる

需要予測は、生産計画・在庫計画・調達計画・人員計画のすべての出発点にある。予測を誤れば、その誤差は単独で収まらず、下流の各計画に増幅しながら伝播する。需要を過大に見積もれば、作りすぎた在庫が運転資金を縛り、保管コストと陳腐化リスクを生む。過小に見積もれば、欠品による販売機会の損失と、緊急増産・緊急輸送による追加コストを招く。

とりわけ製造業では、調達リードタイムの長い部材や、段取り替えに制約のある生産ラインを抱えるため、予測の誤りを後工程で吸収しにくい。需要予測の質を1段引き上げるとは、その下流すべての計画の質を底上げすることを意味する。これが、数あるDXテーマの中で需要予測が「起点」と呼ばれる理由である。

1.2 現状の壁 — 属人化と「当たらない前提」の欠如

多くの企業で、需要予測は今なお熟練担当者の経験と勘、そして表計算ソフト上の手作業に依存している。この方式には3つの構造的な壁がある。第一に**属人化**。予測の根拠が担当者の頭の中にあり、退職・異動とともに失われる。第二に**処理能力の限界**。SKU数が数千・数万に及ぶと、全品目に目を配ることは物理的に不可能で、主要品目以外は粗い一律ルールで処理される。第三に**外部要因の取り込み困難**。為替や市況、顧客の生産計画といった先行情報を、人手で網羅的に織り込むことは難しい。

需要予測は「当たらない」ことを前提に設計する

実務上の重要な前提として、需要予測は決して100%は当たらない。目指すべきは「完璧な予測」ではなく、「誤差の大きさを把握し、それを在庫の安全在庫や調達のバッファ設計に織り込んで運用する」ことである。予測精度の改善と、誤差を吸収する計画側の設計は、つねに一体で考える必要がある。本書が予測単体ではなく計画連携（第5章）まで踏み込むのはこのためである。

1.3 評価軸は精度ではなく業務インパクト

AI需要予測の導入で陥りやすい罠が、「予測精度の向上」自体を目的化してしまうことである。精度指標が数ポイント改善しても、それがビジネス上の成果に直結するとは限らない。1日1個しか売れない品目の予測を磨くより、1日1,000個動く主力品目の誤差をわずかに縮める方が、在庫・欠品へのインパクトははるかに大きい。

したがって導入の評価は、過剰在庫の削減額、欠品率の低下、発注・計画業務の工数削減といった**業務インパクト**を主指標に置き、精度指標（後述のMAPE等）はその先行指標として扱うのが適切である。どの品目の精度を上げることが最も効くのかという「投資の優先順位」こそが、最初に設計すべき論点になる。

02 予測手法の地図 — 統計から因果モデルまで

手法ごとの適用領域、精度の天井、
選択の判断軸

KEY FINDINGS

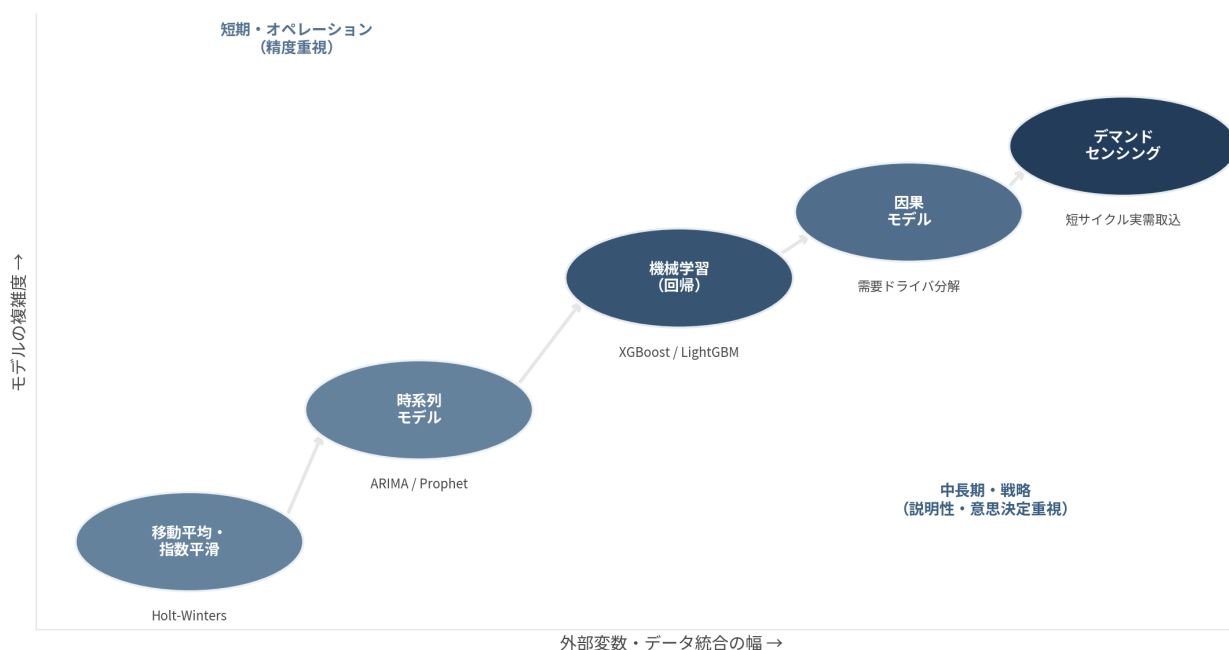
- 予測手法は移動平均・指数平滑から時系列モデル、機械学習回帰、因果モデル、デマンドセンシングへと進化してきた
- 手法は「短期・オペレーション（精度重視）」と「中長期・戦略（説明性・意思決定重視）」に大別され、用途で使い分ける
- 外部変数を統合できるかが精度の天井を決める。実績のみの予測には早期に限界が来る
- 単一の万能モデルは存在せず、複数手法のアンサンブルと、目的別の使い分けが現実解である

2.1 手法の進化 — 精度型から意思決定支援型へ

需要予測の手法は、大きく次の段階を経て進化してきた。最も単純な**移動平均・指数平滑（Holt-Winters等）**は、過去実績の傾向と季節性を捉える。次の**時系列モデル（ARIMA、Prophet等）**は、トレンド・季節性・周期をより構造的に分解する。さらに**機械学習回帰（XGBoost、LightGBM等）**は、価格・販促・天候といった多数の説明変数を同時に取り込み、非線形な関係を学習できる。

その先にあるのが**因果モデル**と**デマンドセンシング**である。因果モデルは「なぜその需要が生まれるのか」というドライバを分解し、説明性を持った予測を返す。デマンドセンシングは、直近の実需シグナル（受注・出荷・POS等）を短サイクルで取り込み、足元の変化に素早く追従する。

需要予測手法の進化：精度型から意思決定支援型へ



出所：一般的な需要予測実装知見に基づき taiziii Research 作成

2.2 用途で分ける — オペレーションと戦略

重要なのは、これらが「新しいものが古いものを置き換える」関係ではなく、**用途で使い分ける**関係にあるという点である。需要予測は、用いる目的によって2種類に分かれる。

ひとつは**オペレーション向け**。在庫補充や数カ月先の生産計画に用いるもので、環境変化が比較的小さい短期を対象とするため、時系列モデルや機械学習回帰による高精度な予測が有効になる。ここでは予測精度そのものが価値に直結する。

もうひとつは**戦略向け**。設備投資や中期計画の判断に用いるもので、1～2年先の中長期を対象とする。この領域では精度を追うことに意味は薄く、業界トレンド・競合動向・地政学リスク・規制変化といった要素をシミュレートし、需要の因果関係を踏まえて意思決定の質を上げることが目的になる。ここで効くのは因果モデルである。

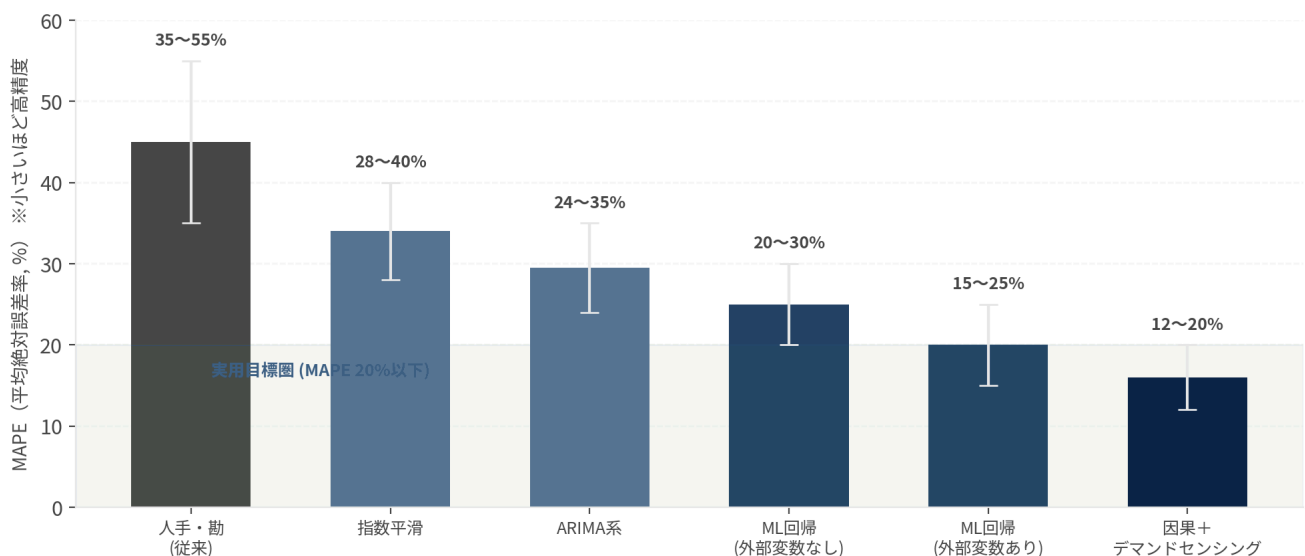
判断軸：時間軸 × 規則性 × 説明責任

手法選択の実務的な判断軸は3つ。**時間軸**（短期=精度型、中長期=意思決定型）、**需要の規則性**（規則的=統計/ML、不規則・新製品=因果・定性重視）、**説明責任**（経営会議で根拠を説明する必要があるれば、ブラックボックスなML単独より説明可能な因果モデルやドライバ分解を選ぶ）。この3軸で、品目群ごとに手法を割り当てる。

2.3 精度の天井は「外部変数の統合」で決まる

手法の高度化以上に精度を左右するのが、外部変数をどこまで取り込めるかである。実績データのみを入力とする予測は、過去の規則性が将来も続く前提に立つため、市況の転換点や需要構造の変化に弱い。為替・景気指標・天候・販促・競合動向といった外部変数を統合できて初めて、精度の天井が引き上がる。

予測手法と精度水準のイメージ：外部変数の統合が効く



※ 数値は製品特性・データ品質により大きく変動するため、相対的な傾向を示す概念図。需要の規則性が低い品目ほど絶対水準は悪化する。

出所：taiziii Research 作成（概念図／数値は相対傾向を示す例示。MAPEの実用目安20～30%は電通デジタル等の実務解説に基づく）

上図は手法別の精度水準を概念的に示したものである。注意すべきは、これらの数値が製品特性とデータ品質に大きく依存する点である。規則的に売れる主力品目では高い精度が出る一方、需要の規則性が低い品目や新製品では、どの手法を使っても絶対水準は悪化する。「自社のどの品目で、どの精度が出るのか」を実データで把握することが、手法選択の出発点になる。

過学習という落とし穴

機械学習を用いる際に最も警戒すべきは過学習である。過去データのノイズや偶然のパターンまで学習してしまうと、検証データでは高精度に見えても、未知の将来に対してまったく通用しなくなる。「過去データに100%合致するモデル」はむしろ危険信号であり、必ず学習期間と検証期間を分けた検証（バックテスト）で、未知データへの汎化性能を確認する必要がある。

03 リファレンス・アーキテクチャと設計判断

手法ルーティング、予測階層のリコンサイル、本番で崩れる落とし穴

KEY FINDINGS

- 4レイヤー構成は前提にすぎない。差がつくのは「どの品目に、どの手法を、どのレベルで当てるか」という設計判断である
- 全SKUに同じモデルを当てるのは誤り。需要パターン（平滑／変動／間欠／散発）で分類し、手法をルーティングする
- 「どの階層レベルで予測するか」が精度を左右する。中間レベルで予測し統計的にリコンサイルする手法が定石
- 特徴量リーク・新製品コールドスタート・再学習頻度の3つが、PoCでは見えず本番で崩れる典型的な落とし穴

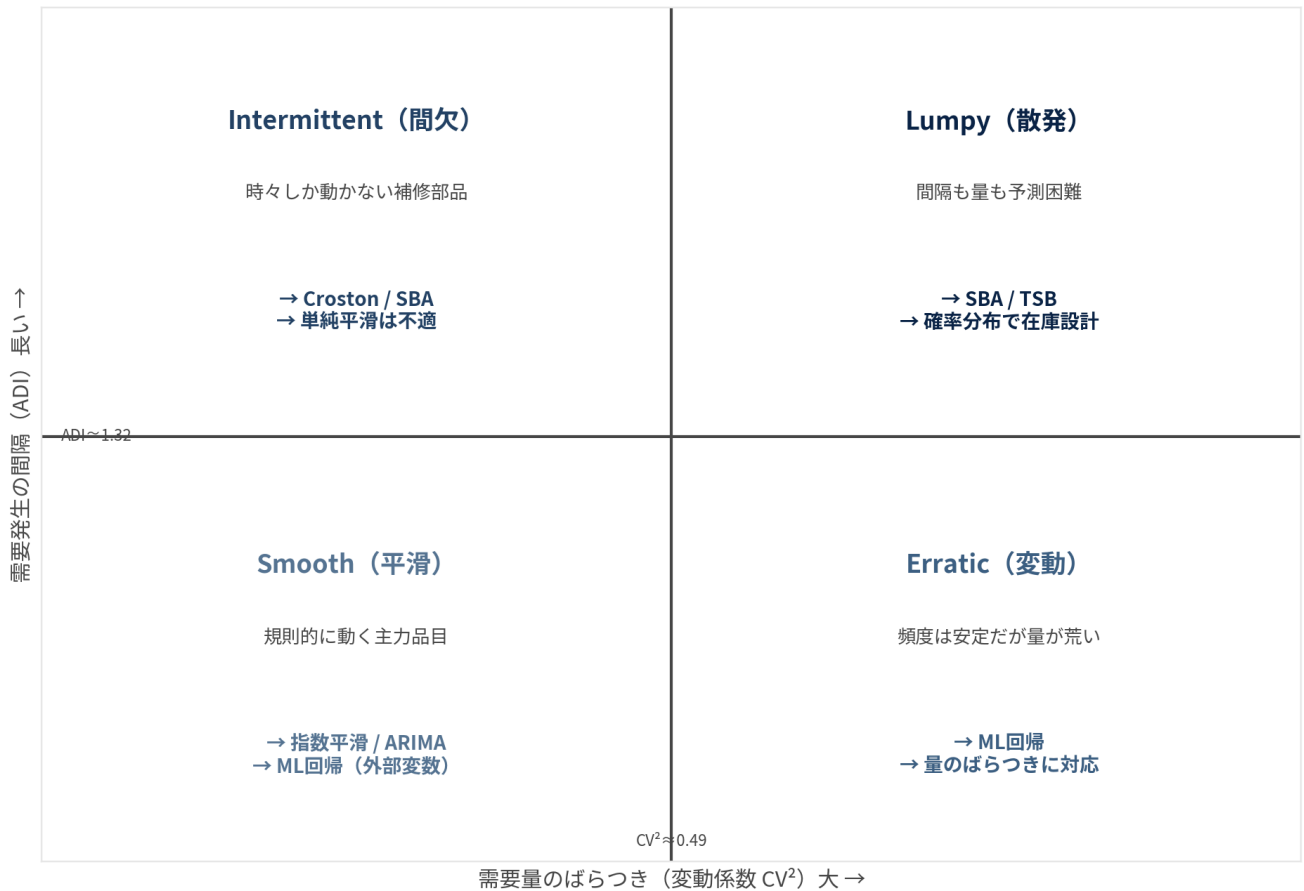
需要予測のシステムが「データソース→データ基盤→予測エンジン→計画・実行」の4レイヤーで構成されること自体は、もはや解説に値しない常識である。本章が扱うのは、その**各レイヤーで実際にどう設計判断を下すか**という、PoCと本番運用を分ける勘所である。データ基盤の選択論（データレイク／レイクハウス／データメッシュ、OT-IT統合）は本シリーズの別テーマに譲り、ここでは需要予測に固有の3つの設計判断 — **手法のルーティング、予測階層の設計、本番で崩れる落とし穴** — に絞る。

3.1 手法のルーティング — 全SKUに同じモデルを当てない

最も多い失敗が、「需要予測AIを導入する」と称して、全品目に単一のモデル（多くは高度なML）を一律適用することである。これは確実に失敗する。なぜなら、定期的に毎日売れる主力品目と、半年に一度しか動かない補修部品とでは、そもそも適した数学が異なるからだ。

実務の定石は、全品目を**需要パターンで分類し、パターンごとに手法をルーティング**することである。分類軸は2つ。需要発生の間隔（ADI：定期的に動くか、時々しか動かないか）と、需要量のばらつき（変動係数 CV^2 ：量が安定か、荒いか）。この2軸で品目は4象限に分かれる。

需要パターンの分類 → 手法のルーティング
(Syntetos-Boylan分類)



出所：Syntetos-Boylan分類 ($ADI \approx 1.32$, $CV^2 \approx 0.49$ を境界とする一般的な実務基準) に基づき taiziii Research 作成

平滑 (Smooth) な主力品目には、指数平滑・ARIMAや、外部変数を入れた機械学習回帰が効く。一方、**間欠 (Intermittent)** な補修部品に通常の指数平滑を当てると、需要ゼロの期間に予測値がずるずる下がり、突然の需要に対応できない。ここには**Croston法**やその改良版である**SBA (Syntetos-Boylan Approximation)**を使う。Crostonは需要の「大きさ」と「発生間隔」を分けて推定する手法だが、構造的にやや過大バイアスを持つことが知られており、SBAはそのバイアスを補正する。間隔も量も予測困難な**散発 (Lumpy)** 品目に至っては、点予測を諦め、需要の確率分布を推定して在庫の安全係数を設計する方が合理的である。

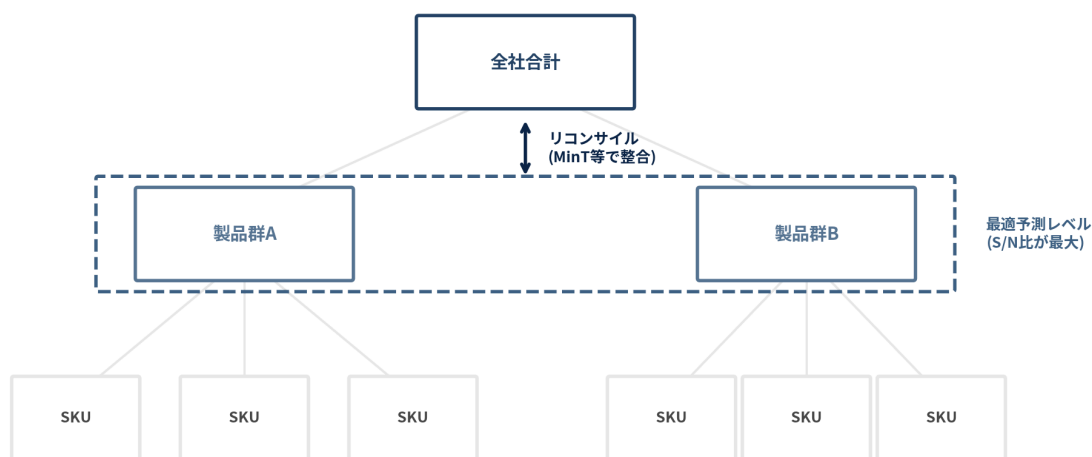
設計判断：分類を「自動化」して運用に組み込む

重要なのは、この分類を導入時に一度やって終わりにしないこと。品目は時間とともに象限を移動する（新製品が立ち上がり平滑になる、終売品が間欠化する）。ADIと CV^2 を定期的に再計算し、品目を自動で再分類して手法を切り替える仕組みを組み込む。「どのSKUにどのモデルが当たっているか」を人が管理しきれなくなる数千SKU規模では、この自動ルーティングが運用の生命線になる。

3.2 予測階層の設計 — 「どのレベルで予測するか」が精度を分ける

次の設計判断は、予測を行う粒度の階層である。SKU単位で個別に予測すると、それを足し上げた合計が、全社合計を直接予測した値と一致しない。逆に全社合計で予測すると、現場の発注・生産には粒度が粗すぎて使えない。この「下位の積み上げと上位の予測が合わない」問題は、多くの現場で軽視されているが、在庫配分の意思決定を直接歪める。

階層予測のリコンサイル：「どのレベルで予測するか」が精度を分ける



SKU単位で予測すると合計が合わない／合計で予測すると現場で使えない
→ 中間レベルで予測し、上下に統計的リコンサイルで配賦するのが定石

出所：階層時系列予測の一般理論（Hyndman et al.）に基づき taiziii Research 作成

定石は、**信号対雑音比が最も高い中間レベル**（多くの場合「製品群×地域×週」程度）で予測を行い、そこから上位（全社合計）と下位（個別SKU）へ統計的に配賦・整合させる**リコンサイル（reconciliation）**である。単純な按分（トップダウン）でも単純な積み上げ（ボトムアップ）でもなく、各レベルの予測誤差の構造を考慮して最適な重みで統合する手法（**MinT=Minimum Trace**等）を用いる。一般的な小売・SCMの事例では、適切なリコンサイルにより素朴な積み上げ／按分に対して**15～25%**程度の精度改善が報告されている。

この階層設計は、第2章で述べた手法選択や次章（第4章）のデータ統合と独立ではない。「どのレベルで予測し、どこに営業の定性情報を乗せ、どう上下に配賦するか」を一体で設計することが、アーキテクチャの骨格になる。

3.3 本番で崩れる3つの落とし穴

PoCでは高精度を示したモデルが、本番運用で崩れる。その原因はアルゴリズムではなく、次の3点にほぼ集約される。DX部門が最初から潰しておくべき設計論点である。

① 特徴量リーク（データリーケージ）

予測時点では本来知り得ない情報を、誤って学習データに含めてしまう問題。たとえば「その月の確定実績」や「事後に判明した販促結果」を特徴量に混ぜると、検証では驚異的な精度が出るが、本番では同じ

情報が予測時点に存在せず、性能が一気に劣化する。**各特徴量が「予測を立てる時点で実際に入手可能か」を1つずつ検証する**規律が要る。

② 新製品のコールドスタート

過去実績のない新製品は、時系列モデルが原理的に予測できない。ここで有効なのが第3.2節の階層構造である。SKU単体ではなく、**属する製品群のトレンドを起点（アンカー）**に、価格帯・ターゲット・販路から妥当なシェアを推定する。階層予測を組んでおけば、コールドスタートを当て推量ではなく構造的に扱える。これは新製品比率の高い業界ほど効く設計である。

③ 再学習頻度の設計

モデルは放置すれば需要構造の変化とともに劣化する。一方で、毎回フル再学習すれば計算コストとモデルの不安定化（予測が每期大きく振れる）を招く。**「いつ・何をトリガに再学習するか」**を設計する必要がある。定期再学習（月次等）に加え、予測誤差が閾値を超えたら再学習する異常検知トリガを併用し、安定性と追随性のバランスを取る。この再学習ループの設計こそが、精度を「一度きり」でなく「継続的に」維持する条件になる。

04 3層のデータ統合という本質

内部実績×マクロ×営業定性をどうつなぐか

KEY FINDINGS

- 需要予測の精度は、内部実績・マクロ外部環境・営業定性という3層のデータをどう統合するかで決まる
- 内部実績は統計の土台、マクロは因果モデルの説明変数、営業定性は統計に現れない先行情報を担う
- 営業の定性情報を「自由記述のまま」にせず、案件確度・数量レンジとして構造化することが鍵
- 3層の統合は技術課題であると同時に、営業・生産・経営をまたぐ組織課題でもある

本書の中心的主張は、需要予測の高度化が「3種類のデータをつなぐこと」に集約されるという点にある。アルゴリズムは、この統合構造の上に乗る部品にすぎない。3層それぞれの役割を見ていく。

需要予測を支える3層のデータ統合



出所：taiziii Research 作成

4.1 第1層：内部実績データ（統計の土台）

出荷・受注・在庫・生産実績といった内部データは、すべての予測の土台である。ここで設計すべきは**予測の粒度**だ。SKU単位か製品群単位か、地域別か全社か、日次か週次か月次か。粒度が細かいほど計画に

は使いやすいが、データのばらつきが大きくなり予測は難しくなる。逆に粗くまとめれば予測は安定するが、現場の打ち手には結びつきにくい。この粒度設計が、後続のすべての精度と運用性を規定する。

同時に、実績データの品質が問われる。返品・サンプル出荷・特需といったイレギュラーが実績に混入したままでは、モデルがそれを通常需要と誤認する。データのクレンジングと、異常値の扱いのルール化が、統計の土台を固める前提作業になる。

4.2 第2層：マクロ・外部環境データ

為替、景気指標、天候、市況、競合動向、規制変化といった外部環境データは、因果モデルや機械学習回帰の**説明変数**として機能する。これらは需要の転換点を先読みする手がかりになる。たとえば天候は季節商材の需要を、為替は輸出比率の高い製品の需要を、市況は素材系製品の需要を動かす。

実装上のポイントは、これらを外部API等から**継続的に自動取得し、予測の特徴量として整形する**パイプラインを組むことである。一度きりの手作業での取り込みでは運用が続かない。どの外部変数が自社の需要と相関するかは、データに基づく特徴量選択で見極める。相関が確認できない変数を増やすことは、過学習のリスクを高めるだけで精度には寄与しない。

4.3 第3層：営業の定性情報（最大の差別化要因）

3層の中で、最も価値が高く、同時に最も扱いが難しいのが営業の定性情報である。商談の進捗、顧客の生産計画や内示、大口案件の受注見込み、失注のリスク。これらは統計データには一切現れない**需要の先行情報**であり、現場の営業担当者の頭の中にしか存在しないことが多い。

この情報を予測に乘せられるかどうか、他社が容易に真似できない競争優位を生む。鍵は、定性情報を**自由記述のまま放置せず、構造化すること**にある。具体的には、案件を確度（受注確率）と数量レンジ、想定時期といった定量的な属性に分解し、統計予測の出力に重ね合わせる。たとえば「統計予測のベースライン+確度で重み付けした商談パイプライン」という形で合算すれば、現場の肌感を定量予測に統合できる。BtoBのフォーキャスト実務でも、SFA/CRM上の各案件に受注確率・受注予定時期・想定金額を付与し、それらを合算して受注見込みを算出するのが定石である。

具体例：営業由来の先行情報を予測に組み込む — 城南電機工業

自動車用照明機器などを手がける城南電機工業（静岡市、従業員約120名）は、需要予測にこの考え方を実装した好例である。同社のようなBtoB部品メーカーでは、顧客から事前に「**発注内示**」（今後の発注予定量の通知）を受け取る。これは統計実績には現れない、営業・顧客接点由来の代表的な先行情報だ。ただし**内示と最終的な実際の納入数には差異があり**、その差が余剰在庫や欠品を生んでいた。

同社は、**製品ごとの受注実績（統計の土台）と発注内示（先行情報）の両方を入力**としてLightGBMモデルに学習させ、内示と実需のギャップそのものをAIに予測させた。結果、**改善幅の大きい製品で予測の誤差率が52%から24%へと半減した**。注目すべきは、全製品を一律に自動化せず、**過去1年間で予測精度が良かったものを選別し、精度が悪化する製品にはAIを使わない**という運用にした点である。これは前章（第3章）の「需要パターンで手法をルーティングする」という設計判断の、中小企業における実装そのものといえる。

城南電機の事例は、第3層の本質を示している。すなわち、先行情報（内示や商談）は「**正解**」ではなく、**実需とのギャップを含んだ信号**である。これを統計実績と組み合わせ、ギャップの構造をモデルに学ばせることが、営業データ活用の核心になる。同様に、ケーブル等の部品を扱う石原金属化工も、出荷実績と顧客の内示情報を組み合わせて3か月先の需要を予測し、発注業務の属人性を排除している。

統合は技術課題であると同時に組織課題

3層のデータ統合は、技術だけでは完結しない。内部実績は生産・物流部門、マクロデータは経営企画、営業定性は営業部門と、データの持ち主が組織的に分断されているためである。「予測モデルを誰が持ち、各部門のデータをどう供出させ、予測結果に誰が責任を持つか」というガバナンスを設計しなければ、技術的に優れたモデルも現場で回らない。この組織論は第5章で扱う。

05 予測を計画につなぐ — FVAとS&OPの再設計

ナイーブ比での評価、価値を破壊する介入の発見と抑制

KEY FINDINGS

- 衝撃的だが、多くの企業の需要予測は「前年同月をそのまま使う」ナイーブ予測にすら勝てていない
- FVA（Forecast Value Added）で各工程の貢献を測ると、人手の介入や役員レビューがむしろ精度を悪化させる例が頻出する
- S&OPの設計問題の核心は「合意形成すること」ではなく「価値を破壊する介入をどう止めるか」にある
- 品目には予測可能性の上限（forecastability）があり、それを越えた精度追求は資源の浪費。在庫設計で吸収すべき

5.1 不都合な真実 — その予測、ナイーブ予測に勝っているか

S&OPや合意形成プロセスの解説に入る前に、多くの企業が直視できていない不都合な事実から始めたい。**手間とコストをかけた需要予測の多くは、「前期の実績をそのまま来期の予測とする」ナイーブ予測（ランダムウォーク）にすら勝てていない。**

サプライチェーン8社を対象にした著名な調査では、予測の52%がランダムウォークより精度が悪かったという結果が報告されている（S. モーリッジ／CatchBull、2013年）。つまり、高価な予測システムと多くの人員を投じた予測の半分以上が、「何もせず前期実績を使う」より劣っていた。同調査では、多くの企業はナイーブ予測を10～15%以上は上回れていないとも指摘される。需要予測DXを語る前に、自社の予測が本当に「何もしないより良い」のかを疑う必要がある。

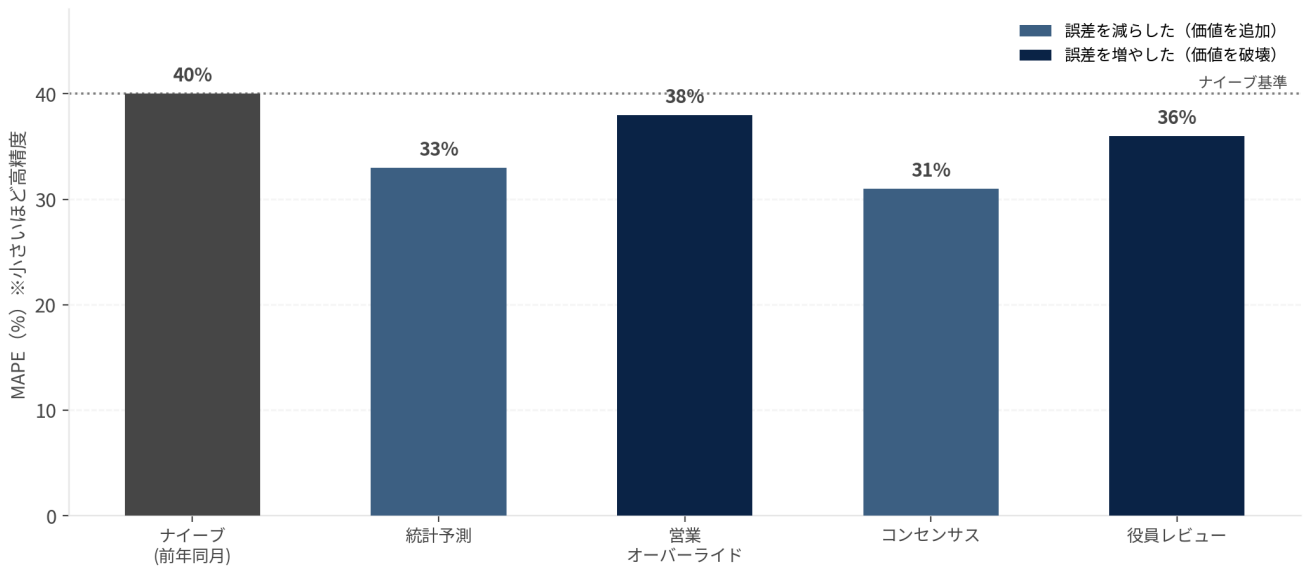
最初に測るべきは精度ではなく「ナイーブ比の優位性」

予測の良し悪しは、MAPEの絶対値では測れない。MAPE 20%は、ナイーブ予測がMAPE 40%の難しい品目なら優秀だが、ナイーブが15%で当たる簡単な品目ならむしろ価値を破壊している。評価の基準は**RAE（相対絶対誤差＝予測誤差 ÷ ナイーブ予測誤差）**に置く。RAEが1.0未満なら価値を足しており、1.0以上ならナイーブに負けている。この一指標が、後述するすべての工程評価の土台になる。

5.2 FVA — どの工程が価値を足し、どの工程が壊しているか

では、予測プロセスのどこに問題があるのか。それを可視化する手法が**FVA（Forecast Value Added）**である。FVAは、予測が作られる各工程（統計予測 → 営業オーバーライド → コンセンサス → 役員レビュー）の予測値をすべて記録し、それぞれがナイーブ予測と比べて誤差を減らしたか増やしたかを測る。製薬の治験で新薬とプラセボを比較するように、各工程が「本当に効いているか」を実証的に検証する考え方である。

FVA（Forecast Value Added）：各工程は本当に価値を足しているか



※ 例示。人手のオーバーライドや役員レビューが、かえって統計予測より精度を悪化させる現象は実務で頻出する。

出所：FVA分析の一般的フレームワーク（M. Gilliland / SAS等）に基づき taiziii Research 作成（数値は例示）

FVAを実際に回すと、多くの企業が直視したくない結果が出る。**人手のオーバーライドや役員レビューといった「偉い人が関わる工程」ほど、統計予測の精度をかえって悪化させている**ことが珍しくない。役員レビューの工程がコンセンサス予測よりMAPEを悪化させる例は実際によく見られる。営業は売上目標へのバイアスで数字を盛り、経営は願望で数字をいじる。善意の介入が、統計が見つけた信号をノイズで上書きしてしまう。

S&OPの設計問題の再定義

ここからS&OPの本質的な設計問題が見えてくる。それは「いかに全部門を集めて合意形成するか」ではない。**「価値を足している工程を残し、価値を破壊している介入をいかに止めるか」**である。FVAで各工程の貢献を定量化し、マイナス寄与の工程は廃止するか、人を介在させるルール自体を見直す。合意形成の場を増やすことが目的化したS&OPは、むしろ予測を悪くしうる。

5.3 S&OPの再設計 — 介入を「例外」に閉じ込める

この洞察に立つと、S&OPの設計指針は明快になる。第一に、**統計・MLが出すベースライン予測をデフォルトとし、人手の修正は「例外」に限定する**。全品目を会議で議論するのではなく、FVAで人の介入が明確に価値を生む品目群（新製品、大口商談が動く品目、構造変化の只中にある品目）だけに人の時間を集中させる。残りは自動で流す。

第二に、介入する場合は**必ず根拠を記録する**。「営業がなぜ統計予測を上方修正したのか」を案件情報として残し、後からその修正が当たったかをFVAで検証する。これにより、第4章で論じる営業定性情報の構造化が、単なるデータ収集ではなく「介入の質を継続的に監査する仕組み」として機能する。日本企業で頻発する、営業主導の「売れる分の先取り」による滞留在庫という“需給のねじれ”も、この監査ループで規律できる。

第三に、共通KPIを置く。在庫回転率・充足率に加え、**RAEとFVA**を部門共通の評価指標に据え、「誰の数字が、ナイーブと統計に対して、本当に価値を足したか」を全部門が同じ物差しで見る。予測の数字を「営業の願望」でも「経営の願望」でもなく、実証的に検証される共通の前提に変えること — これがS&OPの核心である。

5.4 予測可能性の上限を見極める

最後に、精度追求そのものに上限があることを設計に織り込む。すべての需要には、信号（規則的な構造）とノイズ（本質的に予測不能なゆらぎ）が混在する。どれほど高度なモデルを使っても、ノイズは原理的に予測できない。**品目ごとに「これ以上は当てられない」という予測可能性（forecastability）の限界が存在する。**

この限界は、ナイーブ予測の誤差を基準にしたRAEで近似的に把握できる。RAEが既に低い（=予測しにくさの割によく当たっている）品目に、さらにモデルを高度化する投資をしても、得られる改善は小さい。逆に、そもそもノイズ支配的で予測不能な品目を無理に当てにいくのは資源の浪費であり、そうした品目は**予測精度ではなく、在庫の安全在庫やバッファ設計で吸収する**のが正しい。第1章で「予測は当たらない前提で計画と一体設計する」と述べたことは、この予測可能性の上限という形で定量的に裏づけられる。

需要予測アーキテクチャとS&OPの接続点

第3章のアーキテクチャは、この計画・実行層でS&OPと接続する。予測エンジンが返すベースライン予測がS&OPの需要計画インプットとなり、FVAで価値が確認された人の介入だけを上乘せし、供給制約と突き合わせて実行可能な計画に落とす。技術的な予測パイプライン（第3章）と、介入を規律する合意形成プロセス（本章）が噛み合っ初めて、需要予測DXは在庫と欠品の削減という成果を生む。

06 先進事例 — 「現場で使える設計」で勝つ

キリンビール×ブレインパッド、ダイキンの実装に学ぶ

KEY FINDINGS

- 性格の異なる先進2社の事例も、勝負を分けたのは予測精度の高さではなく「現場で使える設計」だった
- キリンビールは「全自動化しない」設計を選び、人の役割を例外監査に再定義。約70%の業務時間を削減
- ダイキンは人を上回る高精度モデルを作りながら、現場の納得を得られず一度つまずいた
- 両社に共通するのは、制約条件の組み込み・業務標準化・運用を回し続ける体制という実装の工夫である

本書の主張 — 「需要予測の成否はアルゴリズムではなく、データ統合のアーキテクチャと、予測を計画に変換する仕組みで決まる」 — は、先進企業の実装にも明確に表れている。性格の異なる2社を取り上げ、どこが優れているのかを具体的に見る。

先進2社の需要予測：「精度」ではなく「現場で使える設計」で勝つ

キリンビール × ブレインパッド 「MJ（未来の需給をつくる）」プロジェクト		ダイキン工業（内製・DICT） 空調製品の需要予測の内製開発
直近1〜2週間先の製造数量算出	対象	空調製品の販売計画・需要予測
全自動化せず 人の最終判断をアラートで支援	優れた点	高精度モデルより 「現場が使える」を優先
製造要件・倉庫能力など 制約条件を組み込んで自動算出	仕組み	LightGBM採用＋業務標準化 プロトタイプ先行で巻き込み
業務時間 約70%削減 年間1,000時間超を創出	効果	人を上回る精度＋ 内製で継続改善体制
属人性の解消と 人の判断の両立	本質	精度ではなく 現場適合と運用継続

出所：各社・ブレインパッドの公表資料、講演資料に基づき taiziii Research 作成

6.1 キリンビール × ブレインパッド — 「全自動化しない」という設計判断

キリンビールは2021年にSCM部を新設し、データ活用を得意とするブレインパッドと3か年の「MJ（未来の需給をつくる）プロジェクト」を組成した。その第2弾として運用を始めたのが、直近1〜2週間先の

製造数量を算出する「製造計画作成アプリ」である。従来は、担当者が出荷実績・需要予測・在庫数量を確認し、各自がExcelで手作業で製造数量を算出していた。**工場の製造要件や倉庫の保管能力といった多数の制約条件を加味する必要があり、業務経験のある担当者しか対応できない属人的で高負荷な業務だった。**

この事例で注目すべきは、「**全自動化を狙わなかった**」設計判断である。アプリは複雑な製造数量の算出工程を自動化する一方で、**人間の目による最終チェックを効率的に行うためのアラート機能を実装した。**すなわち、機械が大量の制約計算を担い、人は機械が出した結果のうち注意すべき点だけを確認する。これは第5章で論じた「統計・MLをデフォルトとし、人手の介入を例外に閉じ込める」という設計思想の、現場での実装そのものである。結果として属人性の解消と業務効率化を両立し、当該業務で**約70%の業務時間削減、年間1,000時間以上の創出**を見込むとされる。

優れている点：制約条件の組み込みと人の役割の再定義

この事例の本質は2点。第一に、製造要件・倉庫能力という**現場の制約条件をモデルに組み込み**、予測値を「実行可能な計画」にまで落とし込んだこと（第5章のS&OP接続）。第二に、人を予測から排除するのではなく、**人の役割を「ゼロから計算する人」から「機械の出力を例外監査する人」へ再定義**したこと。需要予測DXは「人の仕事を奪う」のではなく「人の判断を高付加価値な部分に集中させる」という、導入の現実的な勝ち筋を示している。

6.2 ダイキン工業 — 「高精度モデル」が現場で止まりかけた教訓

ダイキンは2017年設立の社内育成機関「ダイキン情報技術大学（DICT）」を通じてDX人材を内製し、空調製品の需要予測を内製開発した。担当者はARIMA系・ディープラーニング系・決定木系を一通り試した上で**LightGBMを採用し、人を上回る予測精度のモデルを実現した**。技術的には成功である。しかし、この事例が示す教訓はその先にある。

第一の壁は、データサイエンティストが**当初SCM部門の業務内容を理解しておらず**、評価指標の設定やモデリングで苦労したこと。第二の壁は、高精度モデルを業務に組み込んで提案しても、現場の反応が芳しくなかったこと。「**精度は良いが、どの程度信用できるのか、どう使えばいいのか分からない**」という**現場の不安を、精度の高さだけでは払拭できなかった**のである。これは第1章で述べた「精度は目的ではない」という主張の、痛みを伴う実例である。

状況を打開したのは、技術ではなく進め方だった。**実際に動くプロトタイプを早期に作って現場に使ってもらい**、改善と試作を重ねて「これなら使える」という納得を引き出した。本番開発ではSCM部のメンバーがプロダクトオーナーを務め、現場が当事者として関与する体制を組んだ。さらに、各担当者が個別のExcelで行っていた作業を緻密に分析し、**一見バラバラな業務から共通工程を見出して標準化した上で**予測モデルを組み込んだ。

優れている点：精度より「運用に乘せ、回し続ける」設計

ダイキンの担当者は、不確実性の時代の需要予測に求められる3要素を挙げている。①**市場の動きを素早く察知して再予測できること**、②**客観的に判断できる領域をできるだけ増やすこと**、③**予測結果を振り返り知見を蓄積して改善し続けること**。これは本書が論じてきた、再学習ループ（第3章）・データ統合（第4章）・FVAによる継続改善（第5章）と正確に対応する。高精度モデルを作る能力以上に、それを**現場の業務に適合させ、内製で回し続ける体制**を築いたことが、この事例の優れた点である。

6.3 2社に共通する成功要因

2社は、業種も体制（外部パートナー活用と内製）も対象業務も異なる。しかし共通するのは、**勝負を分けたのが予測精度の高さそのものではなく、「現場で使える形に設計し、人の判断と噛み合わせ、回し続ける仕組みを作ったこと」**だという一点である。これこそ、本書が一貫して主張してきたことに他ならない。次章では、この教訓を踏まえた具体的な導入の進め方を示す。

07 導入ロードマップと評価設計

スモールスタートの順序、RAE/FVA評価、つまずきポイント

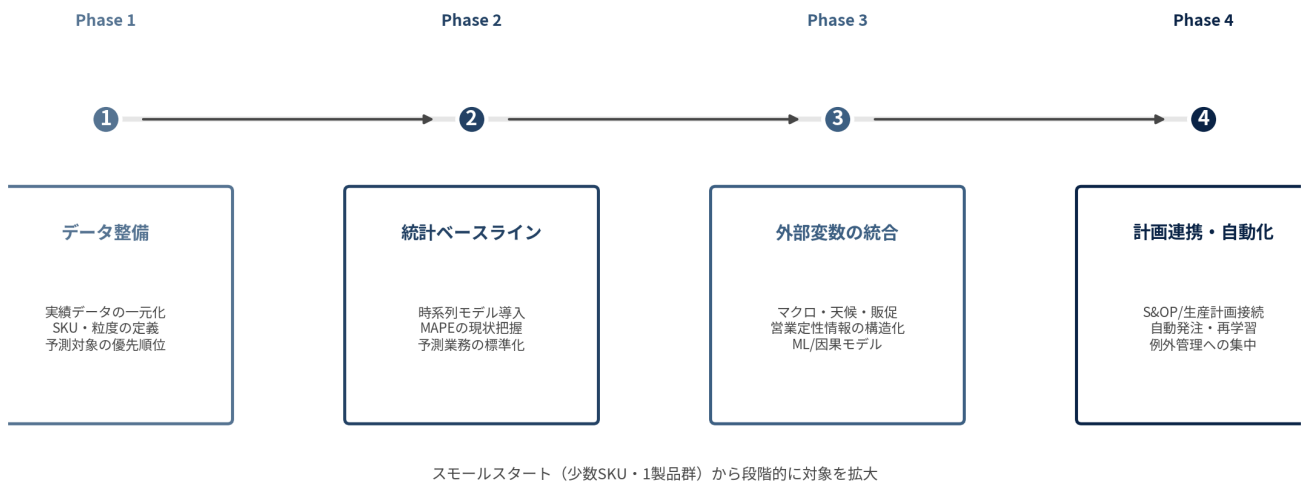
KEY FINDINGS

- 導入は「データ整備→統計ベースライン→外部変数統合→計画連携・自動化」の4段階で進める
- 最初から全社・全SKUを狙わず、少数SKU・1製品群のスモールスタートで成果を示してから横展開する
- 精度評価はMAPEだけでなくRAE（ナイーブ比）・FVA（工程別貢献）を主軸に置く
- 典型的なつまずきは、データ品質の軽視、精度の自己目的化、価値を破壊する介入の放置、再学習ループの不在

7.1 4段階のロードマップ

需要予測DXは、一足飛びに完成形を目指すのではなく、段階的に積み上げる。

需要予測DXの導入ロードマップ



出所：taiziii Research 作成

Phase 1：データ整備。実績データを一元化し、予測の粒度（SKU・期間）を定義し、どの品目群から着手するかの優先順位を決める。第1章で述べたとおり、業務インパクトの大きい主力品目を優先する。

Phase 2：統計ベースライン。まず時系列モデルなどシンプルな手法で予測を立て、現状のMAPEを把握する。ここで「今どれだけ当たっていないか」を定量化し、予測業務を標準化する。いきなり高度なMLに飛びつかず、土台を固める段階である。

Phase 3：外部変数の統合。マクロ・天候・販促といった外部変数を取り込み、営業の定性情報を構造化して重ね合わせる。機械学習回帰や因果モデルを導入し、精度の天井を引き上げる。本書が最大の差別化要因と位置づける段階である。

Phase 4：計画連携・自動化。予測をS&OP・生産計画・調達に接続し、自動発注や再学習ループを実装する。担当者は全品目を手で見のではなく、予測が外れやすい例外品目の管理に集中できるようになる。前章のキリンビールの「アラートで例外だけを人が見る」設計は、この段階の到達点を具体的に示している。

7.2 精度評価の設計

精度指標として実務で広く使われるのが**MAPE（平均絶対誤差率）**である。予測値と実績値の差を実績値で割った絶対値の平均で、パーセント表示のため品目間の比較がしやすい。ただしMAPEは需要が小さい品目で値が極端に振れる弱点があり、用途によってはMAE（在庫量を直感的に把握したい場合）やRMSE（大きな外れを重く見たい場合）を併用する。

ただし第5章で論じたとおり、MAPEの絶対値だけを追うのは危険である。評価の主軸には**RAE（ナীব予測比）**と**FVA（工程別の価値貢献）**を据える。すなわち「自社の予測はナীবに勝っているか」「どの工程が価値を足し、どの介入が壊しているか」をまず測る。その上で、精度指標は**品目の規模で重み付け**し、需要量の大きい品目の改善を重く評価して業務インパクトと整合させる。Phase 2で統計ベースラインを置く目的は、この「ナীব・統計に対する優位性」を測る基準点を作ることにある。

典型的な4つのつまずき

① **データ品質の軽視**：イレギュラーが混入した実績のまま高度なモデルを乗せても精度は出ない。土台のクレンジングを飛ばさない。② **精度の自己目的化**：ナীব比（RAE）を見ずにMAPEだけを追うと、予測可能性の上限に達した品目に資源を浪費する。③ **価値を破壊する介入の放置**：FVAを測らないまま人手のオーバーライドや会議体を増やし、統計予測よりかえって悪い数字を「合意」してしまう。④ **再学習ループの不在**：導入時は高精度でも、モデルを更新しなければ需要構造の変化とともに劣化する。

7.3 まとめ — アーキテクチャとプロセスの両輪

本書を通じた主張は一貫している。需要予測DXの成否は、最新のアルゴリズムを買うことではなく、(1) 内部実績・マクロ・営業定性という**3層データを統合するアーキテクチャ**を設計し、(2) その予測を**S&OPと再学習ループという計画・改善プロセス**に接続できるか、という両輪で決まる。技術と組織のどちらか一方が欠けても、需要予測は現場で回らない。前章の2社が示したのも、まさにこの両輪の重要性だった。

着手にあたっては、全社一斉ではなく、業務インパクトの大きい製品群でのスモールスタートから始め、Phase 1から順に積み上げることを推奨する。明日からの一步は、自社のどの品目が、どれだけ当たっていないのかを、まず実データで測ることである。そこから、本書が示したアーキテクチャとプロセスの設計が始まる。

taiziii では、事業戦略の立案からシステム設計・開発、そして運用まで、ワンストップでご支援できる体制を整えております。AI導入支援やコンサルティングをはじめ、幅広い事業領域でお客様の多様な課題解決に取り組んでいます。本書で扱った需要予測についても、「自社の予測がナীব予測に勝っているか」「どの品目にどの手法を当てるか」「営業データをどう予測に組み込むか」といった具体の検討段階から、お気軽にご相談いただけます。

お問い合わせ：<https://taiziii.com/contact/>