

データ活用における阻害要因と成功の鍵

—なぜAI・アナリティクス技術のビジネス適用は難しいのか—

AI・アナリティクス技術をビジネスに適用したいと考える企業は着実に増えている。しかし、実際には適用がなかなか進まず、難しさを感じている企業も多いのではないだろうか。本稿では、AI・アナリティクス技術のビジネス適用を難しくしている阻害要因と、解決するために気を付けるべきポイントについて考察する。

野村総合研究所 システムコンサルティング事業本部
システムデザインコンサルティング部 データサイエンティスト

ひろた そういちろう
廣田 壮一郎

専門は業務改革、データ活用・分析支援、先端技術の調査・適用検討など



ビジネス適用を阻害する原因と その対策

これまで人にしかできなかった作業を、AIやアナリティクス技術が代替し始めており、場合によっては人よりも良い結果を出すことができるようになった。世界中の企業は現行の業務を大きく改善したり、新しいサービスを創造したりするために、これらの技術のビジネス適用検証を開始している。初めての人でも扱いやすいプログラミング言語や近年話題となっているディープラーニング向けのフレームワークも充実してきている。各種技術の検証や実装も、比較的容易に行うことが可能となってきている。

一方、実際にビジネス適用検証を開始すると、さまざまな問題に直面することも少なくない。「データがあると思って案件をスタートさせたが、実際にはデータがなかった」「データアナリストの検証結果が理解できず、投資の判断できない」「AI活用というテーマは決まったが、何をすればいいかわからない」といった声をよく聞く。モデル構築自動

化ツールの登場や開発支援ツールが充実してきている中で、何がビジネス適用を難しくしているのだろうか。

これまでの経験の中で見えてきたのは、「従前のシステム開発ではできていた、課題や要件について整理しつつ順番に案件を進めていくという基本的なことが、AI・アナリティクス技術に関するプロジェクトに代わった途端にうまくいかなくなる」という問題が至る所で発生しているということである。これはAI・アナリティクス技術の検証においては、難解な数式やモデルの構築など、これまで検討したことのない要素が増えるのが原因の1つだと考えられる。検証結果について高い評価を得ることを念頭に置くあまり、検証を終える条件が曖昧になってしまったり、やり遂げたとしてもビジネス適用にまで至らなかったりすることも多い。このため事前に解決できる課題は取り除いてから、検証プロジェクトを開始するべきである。

表1に、案件が開始する前に整理しておいた方が良い項目の例を挙げた。実際の案件ではこれらの項目に優先順位を付けて整理し、

案件の特徴に合わせて検討項目を追加する。例えば画像認識技術について検証するのであれば、使用するフレームワークやウェブ上で配布されているモデルの活用方針などが考えられる。当然、全ての項目について案件開始前から整理しきれないわけではないので、案件を推進していく中で順次整理していく。

表1 整理しておいた方が良い項目の例

分類	チェック項目
データの質と量の確保	<ul style="list-style-type: none"> 十分な期間・量のデータが用意できるか 取得可能なデータの粒度は適切か データの活用時に環境などの制約がないか
推進チームの体制確保	<ul style="list-style-type: none"> 必要なスキルを兼ね備えた体制が構築できるか ユーザーを含めた役割分担が明確か 意思決定者・業務担当者を巻き込んでいるか
目的とゴールの明確化	<ul style="list-style-type: none"> 背景・課題・目的が確認できているか 評価指標・達成目標の基準が合意できているか 検証の進め方が合意できているか
ユーザーサイドの巻き込み	<ul style="list-style-type: none"> 意思決定者は誰か 業務担当者は誰か 現行業務の大まかな流れが確認できているか
モデルの考え方	<ul style="list-style-type: none"> インプットはどのシステムから取得するか アウトプットはどのシステム/誰が使うのか どんなモデルを作成するのか

成功のための3つのポイント

ここまではビジネス適用を阻害する原因と対策について紹介した。ここではさらに、現場が当然行っているだろうと思われがちだが、実はできていないことが特に多い3つのポイントについて、重点をおいて詳しく解説する。

(1) データの質と量の確保

AI・アナリティクス技術活用の成否には、必要なデータの質と量を明確にし、確保できるかが大きなウエートを占めている。データの質には、「正しく入力されているか」「入力基準に一貫性があるか」「複数のデータが正しくつながるか」などが挙げられる。一方、データの量とは、「十分な数や期間のデータがあるか」「適切な粒度でデータを保持しているか」「意図しない偏りが生じていないか」などを指す。

このようなデータの問題は、実際のデータの中身を見なくては分からないことが多い。多くの企業ではデータ保有については把握しているものの、その質と量に関してはつかんでいないことが多い。保有するデータの内容

確認を行うためにはそれなりのコストがかかるからである。自社内に十分なデータが存在すると考えて検証プロジェクトをスタートさせたものの、価値あるデータが十分に用意できずに、想定していた検証を行うことができないといったことがよく起きる。

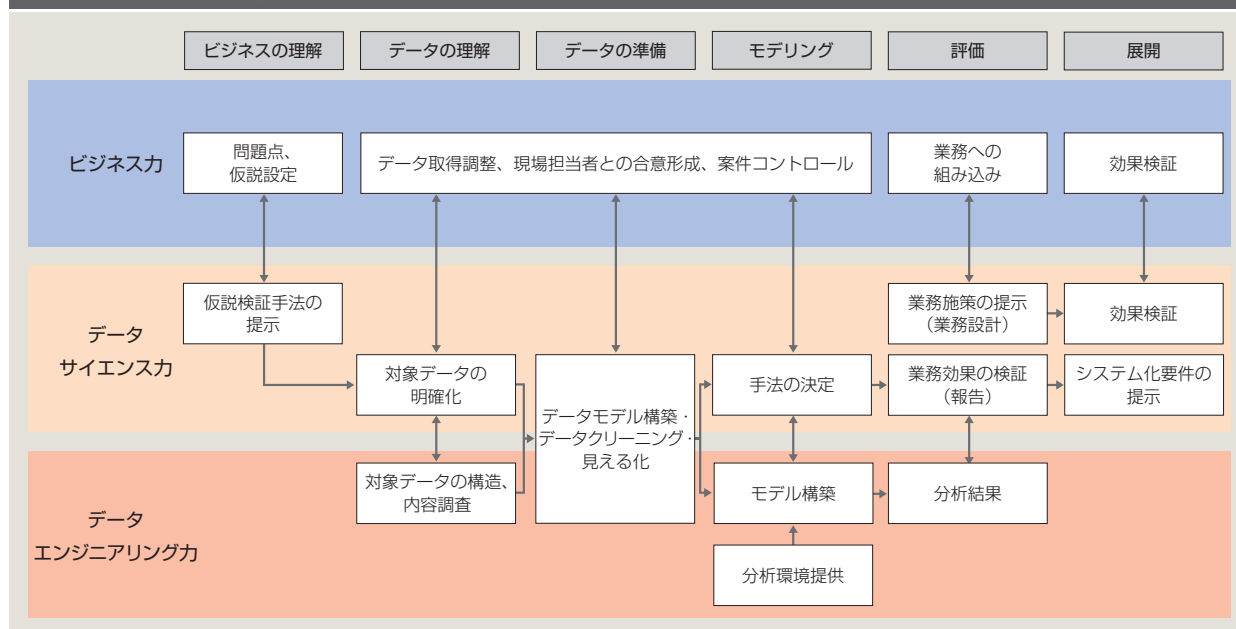
本格的に検証プロジェクトを開始する前には、データの質と量を確認しておくべきである。データの確認を主目的とした棚卸しを行うほか、データの可視化案件の一貫として確認するのも良いだろう。

データの質については、業務の核となる主要プロセスから生成された一連のデータがつながるかどうかを確認する。つなげたデータを用いて必要な情報を抽出できれば、それは質の良いデータである。

データの量に関しては、期間と絶対的な数の2つの側面から確認する。テーマにもよるが、最低でも1年～2年分のデータが数百件以上は存在してほしい。

自社保有のデータのみでのモデル構築が難しい場合、外部から調達する方法と、社内で新しくデータを取得するように業務を変更す

図1 アナリティクス系プロジェクトのプロセスモデルとスキル、タスクの関係



する方法がある。外部から調達する場合は、ユーザーの感覚を手がかりに、モデルに寄与しそうなデータをそろえると良いだろう。社内に新しくデータを取得する場合は、業務内容やデータの入手手順の変更が必要となる場合もある。業務に変更が発生することで生じる現場からの抵抗を減らすために、変更によるメリット、またデータ取得による分析の必要性を提示し、社内の協力を得る必要がある。

(2) 推進チームの体制構築

推進チームが備えておくべきスキルとして、「データエンジニアリング力」「データサイエンス力」「ビジネス力」の3つがある。CRISP-DMと呼ばれるアナリティクス系プロジェクトのプロセスモデルに、各スキルを保有するメンバーと主に担当するタスクについて加筆したものを図1に示す。

「データエンジニアリング力」を専門とするメンバーは、対象データの定義と実際に格納されている数値を照らし合わせ、発生して

いる事象を理解しながら手法を実装する。「データサイエンス力」を専門とするメンバーは、課題に対して必要なデータや手法、さらには得られるであろう示唆を想定した上で、仮説検証の手順を設計する。「ビジネス力」を専門とするメンバーは、ユーザーを巻き込みつつ解決すべき課題を明確にし、仮説設定や業務への組み込み方について整理する。

AI・アナリティクス技術と言っても分野は広く、全てのスキルを兼ね備えたメンバーを確保するのは非常に難しい。一人一人のスキルを見極めながらアサインし、検証プロジェクトを推進する必要がある。

検証プロジェクト推進時に注意する点の1つが、専門外のスキルを各メンバーにどこまで保有させるかである。自身が専門としていないタスクについて、その要点を意識しないで進めてしまったために、後工程のタスクで無理が生じてしまうことがよく起きる。それぞれが専門とするタスクのみを担当するので

はなく、互いの作業内容をレビューすることで、こうした事態を防げる。特にデータの準備と評価のタスクに関しては、専任の担当者が1人で実施した場合、ミスが生じることが多い。手戻りを減らすためにも、他のメンバーの積極的な関与が求められる。

構築すべき体制が見えてきた一方で、統計解析や機械学習の手法を実装できるデータサイエンティストが社内には存在しないといった相談を受けることがある。その場合はアナリティクスサービスを提供する企業と提携する、特定の問題に関して自動で解決してくれるサービスを使用するといったことも検討すべきである。これらのサービスを活用しながら徐々にスキルを社内に取り込み、データサイエンティストを育成していく。データサイエンティストの候補には、一定のITリテラシーを保有するメンバーをアサインすべきである。データサイエンティストには、システム構成の理解や実装力が求められる。これらを後から習得するのは大変である。

(3) 目的やゴールの明確化

AI・アナリティクス技術の適用を検討するプロジェクトの特徴として、目的やゴールを明確化しないままスタートしていることも多い。業務に利用できないものができてしまう、結果を突き詰めることに終始しプロジェクト終了のタイミングが分からないといったことが起きてしまうことがある。

プロジェクトを開始する際には目的を明確にし、目的を達成するために必要なゴールを設定すべきである。評価指標の算定方法や達成目標となる基準だけでなく、予定していた検証を全て終え、示唆をまとめることまで

含めて、ゴールと設定すべきだろう。

達成目標を超えたタイミングで検証を終えられればベストだが、達成目標を必ずしも超えられるとは限らないのがこの類いのプロジェクトの難しいところである。一通り検証を進めた結果、基準に達しなかったとしてもそこで検証を終了する可能性があることをユーザーに事前に理解してもらい、その場合、その結果をもってその後どうするかも議論しておくとうまいだろう。

今後のデータ活用に向けて

近年はデータサイエンティストが「21世紀最もセクシーな職業」と評されるなど、華々しい側面ばかりがフォーカスされている。しかし、統計解析手法を適用したり、機械学習のモデルを構築したりするモデリングのタスクは、データサイエンティストの仕事の中のほんの一部でしかない。ユーザーとの調整やデータの前処理といった地道な作業も多いのが現実である。企業のAI・アナリティクス技術活用も同様で、データの整備や人材育成、標準化活動などに取り組めたかが、今後の成否を分けることになるだろう。

AI・アナリティクス技術の活用に向けては、本稿で解説したものの以外にもさまざまなポイントがある。実際に経験して気付く点も多く、まだまだトライ&エラーが必要な分野である。いきなり投資対効果を得ることは難しいかもしれないが、地道な取り組みを進めていくことで今後、AI・アナリティクス技術のビジネス適用に向けた成功確度を上げていくことができると考えている。 ■