

アナリティクスで継続して成果を生み出す仕組み

山田 敦^{†1}

^{†1} 日本アイ・ビー・エム（株）

ビッグデータから競争優位を構築するために、多くの企業がアナリティクスの活用を進めている。優秀なデータサイエンティストを多数獲得・育成できれば、果たして成果を生み続けられるのだろうか。これまで数多くのお客様に対してアナリティクス活用の支援をしてきた筆者は、アナリティクスで継続して成果を生み出すには、企業は組織としての仕組みを持つべきであると考え、そう考えて筆者が実践してきたことを、本稿で、企業の仕組みとしてまとめた。この仕組みは、(1) データ活用プロセス、(2) データ基盤管理、(3) 意思決定と運用管理の3つから構成される。

1. はじめに

近年、ビッグデータから競争優位を構築するために、多くの企業がアナリティクス、つまりデータ分析の活用を進めている。本稿で論じるアナリティクスの活用とは、有用なデータにアクセスし、データから有意義な洞察を導き、その洞察をビジネス上の価値を生む行動に結びつける活動のことを指す。その適応領域は、「個」への対応がより求められるようになった営業・マーケティング領域や、テクノロジーと利用環境の変化により今また注目を集めるようになってきたモノのインターネット (IoT: Internet of Things) の領域であり、それらが近年注目を集めている。しかしこれらの適用領域に限らず、たとえば文献[1]に示されているように、企業経営を支えるさまざまな業務領域で、アナリティクスの活用が進められている。

筆者はこれまで、アナリティクスの活用を専門とするコンサルタントという立場で、数多くのお客様の活動を支援してきた。その支援内容には、筆者自身が、洞察の導出と行動の定着化の推進をリードするケース、お客様のアナリティクススキルの育成を支援するケース、アナリティクスを活用する仕組み構築を支援するケースなどがある。

これまでの支援活動において、アナリティクスで成果を生めないケースに陥る場面も数多く見てきた。本稿では、お客様がこういったケースに陥らないように筆者が実践してきたことを、企業の仕組みとしてまとめた。この仕組みについて、著者は以前文献[2]で述べた。本稿では、その内容に対して、筆者が実践してきたことをさ

らに加えてまとめる。この仕組みは、優秀なデータサイエンティストを多数獲得・育成できれば備わるものではない。企業が、組織の機能として持つべきものである。

第2章では、筆者が見てきた、アナリティクスで成果を生めないケースについて述べる。第3章では、アナリティクスで継続して成果を生み出す仕組みを概説する。第4章、第5章、第6章で、その仕組みの各要素について詳しく述べ、第7章で本稿をまとめる。

2. アナリティクスで成果を生めないケース

アナリティクスで成果を生めないケースはいろいろある。筆者がこれまで見てきた主なケースを表1にまとめた。本章では、この表の各項目について説明する。

2.1 単一のプロジェクトに関すること

個々のアナリティクス・プロジェクト実施に着目したとき、それが成果を生まないケースを、A. ビジネス・バリュー、B. アナリティクス、C. データ整備の3つの活動領域に分けて説明する。これら3つの活動領域については、第4章で後述する。

A. ビジネス・バリュー

A1. 目的がない—最近では少なくなったが、「データがあるので、これを使って価値ある何かができないか」という依頼が以前は多かった。この「何か (つまり目的)」をプロジェクト開始前に明確にしないと、筆者の経験上プロジェクトは大抵うまくいかない。

A2. ゴールが主観的—「これまで知られていなかったことを発見したい」という依頼がこのケースである。部

表1 アナリティクスで成果を生めないケース

単一のプロジェクトに関すること	
A. ビジネス・バリュー	
A1.	目的がない
A2.	ゴールが主観的
A3.	期待効果が定義・評価されていない
A4.	アナリティクスのアウトプットが業務への適切なインプットとなっていない
A5.	業務でのアナリティクス活用が現場に浸透しない
B. アナリティクス	
B1.	分析開始前に解き方の方針が決まっていない
B2.	分析開始前に必要なデータ有無の調査ができていない
B3.	仮説を適切に活用していない
B4.	導いた洞察のロジックに問題がある
B5.	導いた洞察のロジックを伝えられていない
C. データ整備	
C1.	他部門からデータを提供してもらえない
C2.	データ整備の期間・工数を適正に見積もれていない
複数のプロジェクトに関すること	
D1.	有望なプロジェクトを選定するプロセスがない
D2.	複数のプロジェクトを支えるデータ基盤がゴミ箱状態になる

長にとってはこれまで知らなかったことだが、現場のエンジニアは大抵知っていることも多い。またその逆もある。このように人の主観に依存するゴールを設定すると、プロジェクトを終われない。

A3. 期待効果が定義・評価されていない—効果とは1次元の物差しであり、高い／低い、大きい／小さいなどと程度を表現することができるものでなければいけない。たとえば、苦情が減る、開発期間が短縮する、品質が向上するなどは、期待効果である。一方、暗黙知が可視化できる、悪い要因が明らかになる、新しい発見がある、などは一見メリットがありそうな表現だが、効果ではない。定量化できる効果は、試算により定義・評価し、定量化が難しい効果は、経営への貢献を構造的に示すという方法で定義・評価しておくべきである。期待効果が定義・評価されていないと、意思決定者は投資判断できない。

A4. アナリティクスのアウトプットが業務への適切なインプットとなっていない—機器の予防保全の例で説明する。保守員にとっては、市場の機器の壊れやすい順番を教えてくれれば、保守業務を効率的に実施できるとしよう。仮に、いつ壊れるのか、どの部品が壊れるのかを、アナリティクスでいかに精度良く導けたとしても、それが保守員が業務で求めるものでなければ、結局アナリティクスの結果は使われない。

A5. 業務でのアナリティクス活用が現場に浸透しない—多くの場合、人間は変化を嫌う。そのため現場は、ア

ナリティクスを使った新しい業務を受け入れようとしないうちに、チェンジマネジメントと呼ばれる変革プログラムを継続的に推進しないと、アナリティクスで導いた洞察は、使われなくなる。

B. アナリティクス

B1. 分析開始前に解き方の方針が決まっていない—A4. でアナリティクスのアウトプットが定まったとして、帰帰問題に定式化するのか、組合せ最適化として定式化するのかなど、解くための基本アプローチが定まらない状態で、やみくもにデータ分析を実施しても、アウトプットにはたどり着けない。

B2. 分析開始前に必要なデータ有無の調査ができていない—B1. で決めた基本アプローチを進めるために必要なデータは入手できそうかを、サンプルデータやテーブル定義書で確認しないで分析作業を始めると、期待するアウトプットに至れない。

B3. 仮説を適切に活用していない—ビッグデータの時代、「仮説に頼るな」と主張する論者もいるが、一概には賛成できない。1つ目の理由は、データ数は多いが変数の数が少ないケースだと、仮説がないと意味のある変数を作れないからだ。クルマのプロブデータ（定期的にクルマから集められるデータ）を使ってクルマの乗られ方の分析を実施した経験がある。変数は、緯度、経度、日時、車速など、たかだか1桁個の変数しかない。この少ない変数を機械的に演算しても意味のある変数は得られないだろう。しかし仮説を使うと、通勤利用や街乗りなどの乗られ方を識別できる変数を作れる。2つ目の理由は、分析の目的が、予測精度の向上である場合、限られた時間で結果を最大化するために、使える仮説を使わないのはナンセンスである。よって筆者のようなコンサルタントは、現場のキーパーソンにインタビューをして、最大限仮説を集め、予測精度に寄与しそうな仮説から順番にロジックを実装する。一方もし分析の目的が、これまで知られていなかった発見なのであれば、既知のことは興味の対象外なので、仮説に頼らない方がよいケースもある。つまり、状況に応じて仮説は使うべきである。仮説を考えたり収集したりせずに分析すると、期待するアウトプットに至れない可能性が高まる。

B4. 導いた洞察のロジックに問題がある—通常、検証されたロジックを組み上げて、結論としての洞察を導く。ロジックとは、「こういう前提において、こういうメッセージが言える」というものだ。ロジックの検証方法には、データで実証する、社内外を探し証拠を見つける、現場のキーパーソンへのインタビューで合意

を得る、多数決をとる、などがよく使われる方法である。結論としての洞察を構成するロジックに、検証されていないロジックが含まれると、導いた洞察の信ぴょう性がなくなる。

B5. 導いた洞察のロジックを伝えられていない—B4.で導いた結論としての洞察を、意思決定者に伝え、投資判断を仰ぐ。そのために、意思決定者が必要とする情報を、言葉と文字で簡潔に表現し伝えることが必要である。導いた洞察のロジックに問題がないとしても、意思決定者にその洞察を適切に伝えることができないと、プロジェクトを継続して推進することができない。

C. データ整備

C1. 他部門からデータを提供してもらえない—データの量と種類が多いほど、アナリティクスで有意義な洞察を導ける可能性は高まる。たとえば、販売領域の分析を行うために、製造領域のデータが有効なケースがあるとす。しかし、生産部門にとっては、データを提供するメリットがない。そのため生産部門からデータの提供を断られるというケースである。こういった問題に対処するため、近年は、部門横串でデータとその活用に責任を持つCDO (Chief Data Officer) を置く会社も始めている。

C2. データ整備の期間・工数を適正に見積もれていない—データ分析をする立場から、データの規模は、3種類に大別できる。つまり、1) PCに格納し処理できる規模、2) サーバ機にあるRDB (Relational Database) にロードし処理する規模、3) RDBでは処理性能が足りず、Hadoopや専用アプライアンスの利用が必要な規模、の3つである。データの規模によって、データ整備の期間・工数が異なる。したがって、仮にB1.で解き方の定式化ができたとしても、データ規模をプロジェクト開始前に把握しておかないと、想定外の期間・工数がデータ整備に使われ、予定した分析作業に十分な期間・工数を費やせなくなる。

2.2 複数のプロジェクトに関すること

アナリティクスの活用は、1つのテーマで終わりではない。販売・マーケティング領域を例にとっても、有望顧客層の発見、顧客離反防止、商品のリコメンデーション、販売員の最適配置などさまざまなテーマが競争優位に貢献しそうだ。こういった複数のプロジェクトが進む状況において、気を付けるべきことがある。

D1. 有望なプロジェクトを選定するプロセスがない—アナリティクス・プロジェクトの成果は、使えるデータ

に依存するため、やってみないと分からない特性がある。効果が見込めないプロジェクトは早めに中断し、より効果が見込めるプロジェクトに投資することが望ましい。しかし、プロジェクトへの投資を評価する体制、基準、プロセスがない。

D2. 複数のプロジェクトを支えるデータ基盤がゴミ箱状態になる—データの量と種類が多いほど、アナリティクスで洞察を導ける可能性は高まる。そのため複数種類のデータを十分な量入れたデータ基盤を整備し、複数の分析テーマで共用することが多い。複数の分析テーマが複数のデータソースにアクセスするというこの「たすき掛け」構造の規模は、時間の経過とともに大きくなる。各テーマから挙げられるデータへの要求に、都度答えるような対応を続けると、データ基盤はたちまちゴミ箱状態になりかねない。トーマス・C・レドマン (Thomas C. Redman) は文献[3]で、よく見られるデータ品質の課題を7つ挙げている：

- 必要なデータを見つけられない
- 不正確なデータ
- 不十分なデータ定義
- 不十分なデータプライバシー／データセキュリティ
- ソースによるデータの非一貫性
- データ過剰
- 組織的な混乱

データ基盤がこのような状態になると、分析テーマの推進スピードが著しく低下する。

3. アナリティクスで継続して成果を生み出す仕組みの概要

優秀なデータサイエンティストを多数獲得・育成できたとしても、それだけで前章で述べたケースを回避できるとは思えない。アナリティクスで継続して成果を生み出すには、企業は組織としての仕組みを持つべきである。そう考えて筆者が実践してきたことを、以下の3つの要素から成る仕組みとしてまとめた。

- (1) データ活用プロセス
- (2) データ基盤管理
- (3) 意思決定と運用管理

この仕組みを実践することで、第2章で述べた良くないケースにお客様が陥ることをこれまで回避してきた。

この仕組みの模式図を図1に示す。もしアナリティクスを使ってどういう価値を生みだせばよいかが決まっていな場合には、図1のデータ活用戦略から始めるのが

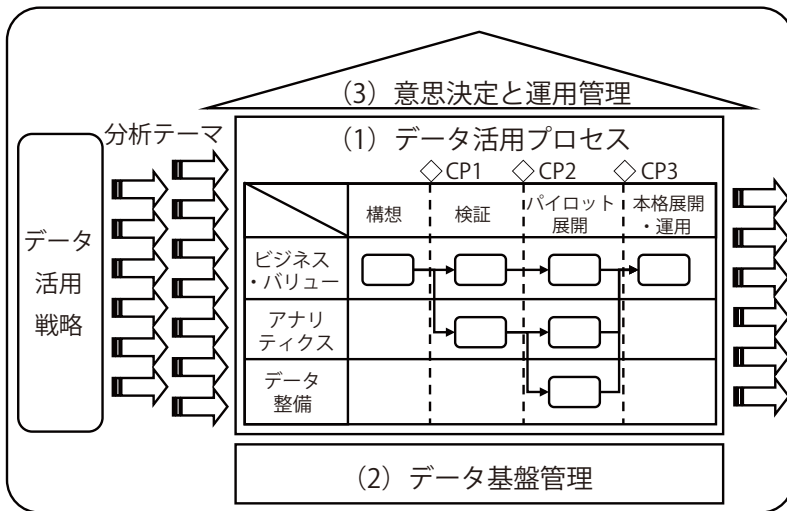


図1 「アナリティクスで継続して成果を生み出す仕組み」の模式図

よい。データ活用戦略については、本稿では詳しく述べないが、たとえば文献[4]などが利用できる。分析テーマが定まったら、それらを、(1)、(2)、(3) から成る仕組みに沿って進めていく。以下に3つの要素について概説する。

(1) データ活用プロセス

データ活用プロセスとは、分析テーマに着手してから洞察を導き行動に結びつけるまでの一連の活動の流れである。図1では、データ活用プロセスに従って複数の分析テーマが左から右に向かって推進されるイメージを模式化している。データ活用プロセスを企業の仕組みとして決めると、実務者はそれを道しるべとして活動を進められる。プロセス、作成物サンプル、注意すべきポイント等をあらかじめ準備し、これらを個別テーマにアレンジするという使い方である。この要素によって、表1に列挙したケースのA1.～A5., B1.～B5., C2. への改善が見込める。その結果、分析テーマの品質向上と工期短縮の効果を見込める。

(2) データ基盤管理

この要素では、複数の分析テーマから依頼されたデータ整備に関する要件をまとめた上で、データ基盤のデータアーキテクチャへの影響を検討する。そしてデータアーキテクチャへのマイナスの影響を小さく抑えながら、分析テーマからの依頼に回答する。そのためのプロセス、判断方針などを仕組みとして整備する。図1では、複数の分析テーマの推進を、データ基盤管理が支えている様子を模式化している。この要素によって、表1に列挙したケースのD2. への改善が見込める。その結果、分析テーマの工期短縮の効果を見込める。

(3) 意思決定と運用管理

この要素では、(1) データ活用プロセスに対してチェックポイントを決め、そのタイミングで分析テーマの進捗を評価し、経営効果最大化の観点で、見込みのないテーマの中断と見込みのあるテーマの継続を判断する。図1のCP1, CP2, CP3がそのタイミングである。併せてデータ基盤の管理に必要なハード・ソフト・要員の投資についても判断する。このための、意思決定ボードの体制、意思決定プロセスを仕組みとして整備する。この要素によって、表1に列挙したケースのC1, D1への改善が見込める。その結果、アナリティクス活用を1つの独立採算事業であると見たときの、事業の投資対効果に改善が見込める。

4. データ活用プロセス

第1章で、「本稿で論じるアナリティクスの活用とは、有用なデータにアクセスし、データから有意義な洞察を導き、その洞察をビジネス上の価値を生む行動に結びつける活動のことを指す」と述べた。本稿では、この実現プロセスを、4つのステップと3つの活動領域に整理して説明する。ステップとは、時間軸での活動の区分けであり、経営判断に適したチェックポイントを設けるために、構想、検証、パイロット展開、本格展開・運用の4つで構成する(図1)。活動領域とは、活動内容による活動の区分けであり、どのようなスキルを持つ要員でプロジェクトを構成すればよいかを検討しやすくするために、ビジネス・バリュー、アナリティクス、データ整備の3つで構成する(図1)。第1章で述べたアナリティクスの活用の定義において、洞察をビジネス上の価値を生む行動に結びつける活動が、ビジネス・バリュー、データから有意義な洞察を導く活動が、アナリティクス、有用なデータにアクセスできるようにする活動が、データ整備である。まず4つのステップから説明していこう。

4.1 4つのステップ

4.1.1 構想ステップ

このステップでは、(1) 解くべき問題の設定、(2) 解き方の方針策定、(3) データがあることの確認、主にこの3つの活動を行い、分析テーマの構想書をまとめる。

(1) 解くべき問題の設定では、現状と比べて何が変わ

るのか、見込める効果は何か、アナリティクスが何を導くのか、導いた洞察を業務でどう使うのか、に対する答えを準備する。(2) 解き方の方針策定では、アナリティクスの基本アプローチ、つまり回帰、分類、クラスタリング、テキスト分析などの中でどの手法を幹とするのかを決める。また目的変数、説明変数の候補を選ぶ。(3) データがあることの確認では、テーブル定義書やサンプルデータを確認し、データが紐づくか、データ量は足りているかを見極める。

この構想ステップでは、原則データ分析作業を行うべきではない。筆者はこの構想ステップを、通常4週間程度で実施することが多いが、データ分析作業をやり始めると到底この期間では納まらない。机上で、解く問題、解き方、データの3点を精査し、解くに値しそうだ、解けそうだ、と客観的に思われるように分析テーマを構想書としてまとめる。

机上での精査は、データ分析作業をするよりも、高いスキルが求められる。つまり、インタビュー等を通して現場のニーズを的確に把握しながら、解き方つまり問題の定式化を同時に考えて、解くに値し、かつ解くことができる問題を机上で見極めるスキルが求められる。数回のインタビューが許される場合、筆者は、前半では現場の意見を最大限引き出すための資料や質問を準備し、後半では頭の中で解き方の筋道を立てることができた問題候補を提示し、「この問題が解けたら役に立つか」を問うように進める。

作成した構想書が会社で認められると、通常すぐに次のステップに進むのだが、もしデータ量が想定する分析環境で扱える量を超えている場合は、分析環境を手配する時間が必要となる。したがって、次ステップに進む前に、データ量と分析環境を確認する。

4.1.2 検証ステップ

このステップでは、前ステップで作った構想書の実現性と有効性を、データを使って検証する。そしてゴールまでのマスタ計画を作成する。筆者は、2カ月程度でこのステップを実施することが多い。実現性の検証が、まさにデータと格闘し洞察を導く最もアナリティクスを駆使する活動である。

機械学習・データマイニングを使う分析テーマであればCRISP-DM[5]というデータマイニングの方法論に従い、データの理解、データの準備、モデリングの手順で進めるとよい。この過程で、構想書に書いた幹となるアプローチに対し肉付けを行う。たとえば、回帰を幹としたテーマであれば、前処理にクラスタリングを追加する

などが1つの肉付けの例である。

一方の有効性は、効果試算により検証する。効果試算ロジックを作り、前提を明確にして定量効果を試算する。定量的な効果試算が難しい場合は、定性効果とその経営への貢献ロジックを明確にする。

4.1.3 パイロット展開ステップ

このステップは、(1) 業務で効果が得られること、(2) 業務で機能すること、の2点を確認するステップである。本格展開・運用ステップの一部として実施することもある。

まず1点目だが、従来通り業務を行うグループと、アナリティクスを活用して業務を進めるグループを作り、一定期間後のグループの成績を比べる実験を行う方法がよく用いられる。この実験の結果は、意思決定者である上位マネジメントと現場の賛同を得るための重要なアウトプットとなる。

2点目の、業務で機能すること、を確認するためには、新業務の設計と、チェンジマネジメントと呼ばれる活動を推進する。つまり、アナリティクスを活用した新しい業務を作り、それに沿って現場が活動できることを確認する。しかし、人間は大抵変わることには抵抗するものである。チェンジマネジメントとは、この抵抗を克服し変革を現場に浸透させる方法である。トップのリーダーシップの下で、変革の必要性を社員に説き、社員を動機付けるために、繰り返しコミュニケーションを図る。時間のかかる活動である。パイロット展開ステップにおいては、パイロットに参画した一部の現場メンバを対象に、業務で機能することを確認できればよい。この確認結果を活用し、次ステップではチェンジマネジメントの活動を本格展開していく。

ときに、アナリティクスによる予測と、現場の勘と経験のどちらが勝つか勝負しようとアプローチする人がいるが、これは勧められない。文献[1]で述べられているように、アナリティクスは、人の判断力に取って代わるものではなく、むしろ判断プロセスで検討すべき新しい洞察を与えるものでなければならない。現場の方々が新しい洞察を得られるように、アナリティクスのアウトプットの見せ方や見せるタイミングなどを現場の方々の意見を聞いて調整する。

4.1.4 本格展開・運用ステップ

このステップでは主に、業務を運営するための体制・プロセス構築と、ITシステム化が必要な部分についてのシステム開発を実施し、本格展開に向けてこれらを現場に導入する。本格展開後は主に、運営上の問題と効果

を監視し、必要があれば、業務、分析モデル、ITシステムのメンテナンスを行う。また一時的な活動に終わらせないためのチェンジマネジメント活動を継続する。

上記システム開発については、BI (Business Intelligence) システム開発を手がけるベンダが持つ開発方法論に従って開発を進めることが多い。システム開発の主な活動は、分析モデルの洗練、分析結果を表示するレポート画面の構築、データマートなどのデータリポジトリの構築、データソースにあるデータを加工しデータリポジトリにロードするデータ統合の構築である。

4.2 3つの活動領域

活動領域は、前述したように、どのようなスキルを持つ要員でプロジェクトを構成すればよいかを検討しやすくするために、図1に示すビジネス・バリュー、アナリティクス、データ整備の3つで構成する。筆者は、お客様プロジェクトのチームを組むとき、各ステップにおける各活動領域で必要となるスキルを把握しているため、適切なチーム編成ができる。企業が仕組みを実装し運営するときも、必要なスキルを理解し、要員を獲得・育成することが望ましい。

4.2.1 ビジネス・バリュー

この領域は、アナリティクスが生み出す業務上の価値を定義し、上位マネジメントの支援を得ながら業務の現場にそれを定着化させる活動である。深い業務ナレッジを持ち業務上の問題の本質を見極めるスキルと、上位マネジメントおよび現場と合意形成をするために何を実施すべきかを理解し実行できるスキルとが必要である。

たとえば、ある目標を達成するための実行計画について上位マネジメントと合意を得るためには、「なぜその計画がベストなのか」、「ほかの選択肢として何を検討し、それらはなぜダメなのか」、「その計画を実施するための時間・工数・コストの試算結果はどういう値か」、「その試算で想定した前提は何か」、「どういう体制で推進するのか」、「推進リーダーはこの計画を了承しているのか」、などといった、上位マネジメントが問うであろう質問への回答を準備し、合意形成の場に臨む。この準備を行うためには、上位マネジメントからの質問を想定し、その回答に必要なアウトプットを定義し、アウトプット作成に必要な作業、工数、スキルを見積もり、作業計画を作成した上で、計画に沿った作業の進捗・リスクを管理する。世の中でコンサルタントと呼ばれる人は通常、こういった活動を行うためのスキルを、トレーニングと実践で身に付けている。

4.2.2 アナリティクス

アナリティクスにも種類がある。発生した出来事をレポートする記述的アナリティクス、統計や機械学習を使って将来起きることを洞察する予測的アナリティクス、最適化手法を使って最善の結果をもたらす行動を推奨する処方的アナリティクス、テキストデータを分析するソーシャルメディアアナリティクスが代表的である。

この「アナリティクス」領域では、洞察を生むアナリティクスモデルを試作・開発する。機械学習、テキスト分析、最適化などについてのナレッジを持ち、モデルのロジックを構成できるスキル、またモデルを実装するためにモデリングツール (IBM SPSSやRなど) を使えるスキル、導いた洞察がなぜビジネス目標に貢献するかを説明するスキルが求められる。世の中でデータサイエンティストと呼ばれる人は、必ず備えているスキルである。

4.2.3 データ整備

アナリティクスに必要なデータを調達し、それを使える状態に整備する活動である。データマートのデータモデルを設計し、データソースや共通データ基盤からマートへデータを抽出・変換・ロード (ETL) する仕組みを構築するスキルが求められる。データがRDBで扱える範囲を超える場合、Hadoop、ストリームコンピューティングなどのビッグデータ技術を活用してデータ基盤から直接データを調達するスキルが必要となる。併せて、データを調達するために、データ基盤の管理者への要求をまとめ、調達時間など含め管理者と調整・折衝するスキルも求められる。

5. データ基盤管理

4.2.3節で説明したデータ整備が、特定の分析テーマに関する活動であるのに対し、本章で述べるデータ基盤管理は、複数の分析テーマの活動を支える、分析テーマ横断の活動である。

5.1 ビッグデータアーキテクチャの管理

データ基盤には、複数種類のデータを十分な量いれ、複数の分析テーマで共有する。データ基盤管理として第1に大切な仕事は、データを入れる器であるハードウェア/ソフトウェア技術の選択である。ビッグデータ時代の代表的な技術として、大量のデータをためて処理するHadoop、流れてくるデータをメモリ内で高速処理するストリームコンピューティング、SQLを専用プロセッサで並列処理する専用アプライアンスなどがある。これ

までの汎用的なRDBと異なり、これらの技術は、特定の用途において威力を発揮する。また技術の進歩も早い。将来計画されている分析テーマを理解した上で、戦略的なデータ要件を策定し、どの技術をどのタイミングで選択/増強していくのかを計画し、管理していくことが求められる。

5.2 データガバナンス

2番目に大切な仕事は、データ自身の管理である。表1のD2.で説明したように、複数の分析テーマが複数のデータソースにアクセスするという「たすき掛け」構造の規模は、時間の経過とともに大きくなる。各テーマから挙げられるデータへの要求に、都度答えるような対応を続けると、データ基盤はたちまちゴミ箱状態になりかねない。

これらの課題に対処するために、データガバナンスとも呼ばれる、データを管理する活動の推進が重要となる。データガバナンスとは、組織が、データを企業資産として活用できるようにするために、要員、プロセスおよび技術を統合する活動である。その活動には、データアーキテクチャの管理、データ品質の評価と維持管理、ガバナンスを遂行するための組織および担当者の役割の定義、セキュリティとプライバシー、メタデータの管理などが含まれる。

データの規模と分析テーマ数が小さいときは、このデータガバナンスの活動は冗長に思えるかもしれない。しかしその規模が大きくなるにつれて、データガバナンスなしではデータ基盤は機能しなくなる。早めに備えをしておくことが大切である。

5.3 データ活用プロセスとの計画共有

第4章で説明したデータ活用プロセスと本章で説明したデータ基盤管理とを連携させることで、本稿で述べる仕組みはより有効に機能する。その連携のポイントは、両活動で計画を共有することである。データ活用プロセスから要求を受けて、データ基盤管理がデータを準備するのが、活動の基本構図だが、データ基盤管理がデータを準備するには時間がかかる。もし別の分析テーマから整合が必要な要求を受けるケースでは、それらの整合にも時間がかかる。したがって、データ活用プロセスは、どのタイミングでどのデータが必要になるのかを計画し要求する。

その要求を受けて、データ基盤管理の活動は、複数の分析テーマからの要求に不整合が生じないようにデータ

の投入・整備を計画し、実施する。時には、分析テーマからのリクエストに満額回答できないケースもあるだろう。データアーキテクチャに影響を与える重要な決定が必要な場合や、分析テーマのスケジュールに大きなマイナスインパクトを与える場合は、後に述べる意思決定での審議事項として挙げる。

このように計画を共有することで、分析テーマがデータを待つ時間が減り、かつデータ基盤がゴミ箱状態になりにくくなる。

6. 意思決定と運用管理

6.1 意思決定の仕組み

意思決定では、それぞれの分析テーマの継続と、データ基盤管理への投資を併せて判断する。分析テーマは、キャッシュを生む。一方データ基盤管理は、それ単独ではキャッシュを生まない。しかしデータ基盤が管理されなければ、分析テーマがキャッシュを生むまでの時間が著しく長くなったり、生むキャッシュの量が小さくなる。したがってこの2つの活動は、併せて投資判断されるべきである。アナリティクス活用を1つの独立採算事業であると見て、この2つの活動をセットで事業判断するのである。

どの分析テーマを継続し、どのテーマを中断するのか。文献[1]では、「最も重要なビジネス課題に対してアナリティクスを適用すべきだ」と述べている。継続する分析テーマについて、ポートフォリオを管理していくことが重要である。

6.2 データ活用プロセス上の意思決定ポイント

分析テーマの推進に対し意思決定するポイントを、第4章で述べたデータ活用プロセスに対して決める。図1の模式図で、CP1、CP2、CP3と記載したのが、その継続判断のポイントである。各ポイントで、継続の意思決定のために確認すべき事項の骨子を、表2にまとめた。これらのポイントで、見込みのないテーマは中断し、見込みのあるテーマに資源を投入していくのである。

6.3 運用管理

運用管理の第1に大切な仕事は、意思決定を円滑に行えるように、意思決定に必要な情報を準備することである。たとえば、各分析テーマの進捗状況、データ基盤の状態などの、意思決定会議における審議項目の準備である。運用管理がおろそかになると、意思決定が形骸

表2 意思決定のポイントと判断内容骨子

意思決定のポイント	確認事項の骨子
CP1	<ul style="list-style-type: none"> ● 分析テーマが定義され、その内容は解くに値する ● 分析テーマを解くための基本アプローチが示されている ● 分析着手が可能なデータ量・分析環境であることを確認済み
CP2	<ul style="list-style-type: none"> ● 分析テーマを実現することによる定量効果、定性効果が示されている ● 分析テーマを解けることが、データを使って検証できている ● 本番展開までのマスタ計画が策定されている
CP3	<ul style="list-style-type: none"> ● 業務で効果が得られることが、パイロット活動により確認できている ● 業務で機能することが、パイロットを実施した対象（地域や部署）で確認できている ● 本番展開に向けたマスタ計画が、パイロットの結果を踏まえて見直されている

化し、生み出す効果にも影響を与える。PMO（Program Management Office）と一般に呼ばれる機能を設置し運営するとよい。

第2に大切な仕事は、アナリティクス事業を活性化するための各種施策の推進である。意思決定だけを続けていけば、成果が出続けるわけではない。たとえば、実務者のスキルをどのように測り、どうやって高めるのか。実務者を動機付けるための文化をどのように醸成していくのか。こういった課題に対する施策を意思決定者に提案し、推進していくことも大切な仕事である。

7. おわりに

本稿では、企業がアナリティクスで継続して成果を生み出すために、筆者がお客様企業を支援する中で実践してきたことを、企業の仕組みとしてまとめた。筆者は、

このような仕組みを企業が持たないと、アナリティクスの活動は一過性の活動に終わると考えている。読者の皆様の組織においても、仕組みについて検討してほしい。

本稿では、仕組みの論理的な機能について述べた。役割を持って動いている既存部門がある中で、この仕組みの各要素を担う部門をどう物理的に設計するのかについては触れていない。今後機会があれば、このことについても論じてみたい。

参考文献

- 1) プレンダ・L・ディートリク 他（著）、山田 敦（監訳）：IBMを強くした「アナリティクス」ビッグデータ 31 の実践例、日経 BP 社 (2014)。
- 2) 山田 敦：ビッグデータから競争優位を構築する（第4回～第6回）アナリティクスで継続して成果を生み出す仕組み、日経 BigData (2014), <http://business.nikkeibp.co.jp/article/bigdata/20140919/271502/>
- 3) トーマス・C・レドマン（著）、栗原 潔（訳）：戦略的データマネジメント、翔泳社 (2010)。
- 4) 池田和明：ビッグデータから競争優位を構築する（第1回～第3回）ビッグデータの地政学、日経 BigData (2014), <http://business.nikkeibp.co.jp/article/bigdata/20140515/264676/>
- 5) Wikipedia：Cross Industry Standard Process for Data Mining, http://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

山田 敦（非会員） ayamada@jp.ibm.com

1995年博士課程修了（機械工学、工学博士）。同年 IBM 東京基礎研究所に入所し、ソフトウェア・アルゴリズムの研究開発に約10年間従事。この間にカーネギーメロン大学の客員研究員を務める。コンサルティング部門に異動後、2009年に新設されたアナリティクス専門部門で、先進的アナリティクスチームのリーダーを務め、現在に至る。データ分析に関するコンサルティングプロジェクトを多数実施。書籍『IBMを強くした「アナリティクス」—ビッグデータ 31 の実践例』（日経 BP 社）を監訳。

採録決定：2015年3月9日

編集担当：丸山 宏（統計数理研究所）