データマイニングを企業で成功させる方法

大内　雅晴

1. はじめに

近年のデータマイニングに対する注目度の高まりと業務の中で蓄積された膨大なデータを何とか活用したいというニーズから，データマイニング・プロジェクトが数多くの企業で実施されています．しかし，ビジネス課題を正しく把握せずにデータマイニングを実施した結果，効果的な成果を得られず，データマイニング試行の段階から抜け出せない状況が散見されます．そのような企業に正しい方法論を導入し，効果的なデータマイニング・プロジェクトを遂行できるように導く必要があります．本論文は，筆者が日本アイ・ビー・エム㈱で実施した多くのコンサルテーションの成功例に基づいて，データマイニング・プロジェクトに関する進め方を標準化したものです．

2. データマイニング・プロジェクトの目的

データマイニング・プロジェクトの目的は，企業の抱える業務上の課題をデータ分析に基づいて解決することです．そのためには担当者が正しいデータマイニング・プロジェクトの進め方を理解し実践することが重要になります．

3. データマイニング・プロジェクトの前提条件

データマイニング・プロジェクトは，分析の要件定義から始まります．プロジェクトのトリガーはデータ分析課題です．ところで，このデータ分析課題は，どのようにして出てきたのでしょうか．企業でデータマイニングのプロジェクトをスタートするにあたって色々なケースがあります．例えば，

- 既にデータ分析の方針を決めている，
- 蓄積されたデータを活用したい，
- 最新のデータ分析手法が何らかの結果を導いてくれると期待している，

など．

この中には，担当者自身がデータ分析課題を明確にしてないケースもあり，その場合，そのままデータマイニング・プロジェクトに入ると，混乱を生じる危険性があります．業務上の課題を解決するという命題に対して最初に行うべきことはビジネス分析です．データ分析課題は，ビジネス分析の結果の一つとして明確化されます．

図1にあるようにビジネス分析を行うことにより，本来あるビジネス課題において，データマイニングによって解決できる部分の割合や効果目標が設定できます．また，包括的な捉え方をすると，ビジネス課題を

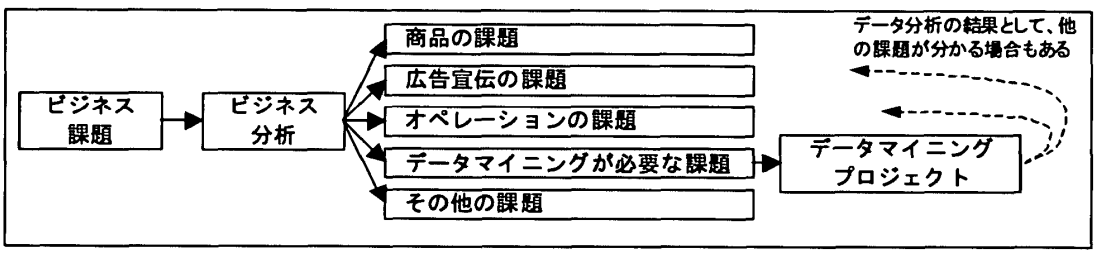

図1　ビジネス分析とデータマイニング・プロジェクトとの関係

おおうち　まさはる
日本アイ・ビー・エム㈱
〒103-8510 東京都中央区日本橋箱崎町 19-21

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

解決するためには，データマイニングだけではなく，あらゆる角度から可能な対応策を施す必要があるといえます．したがって，ビジネス分析に立ち戻って対象となる課題についてどのような方法によって解決できるかを確認する必要があります．もしかしたら，データマイニングの必要はなく，全く異なった解決アプローチが有効であるかもしれません．また，データマイニングはデータ分析の一手段であるので，場合によってはデータマイニングではなく単純統計でよい場合もあります．

4. データマイニング・プロジェクトの形態

データマイニング・プロジェクトには，

(1) 分析だけで業務として完結，

(2) 分析の後に施策や他の業務に連結

の2種類の形態があります．分析だけで完結するものは，主に経営者やデシジョンメーカの観点から見ることを目的として実施するものです．経営者が洞察を深めるため，また自らの意思決定を確信するための分析です．他方，業務に連結するデータ分析は，投資対効果（ROI）に直結する分析であり，業務プロセス再構築を伴います．この形態のプロジェクトは，前者に比べて規模が大きく，後工程としてデータウェアハウスやアプリケーション構築のプロジェクトが実施されます．

運用の観点から見たプロジェクト形態では，社内にシステムを導入し社内の要員にデータ分析のノウハウを蓄積していくケースと，外部にアウトソーシングするケースに分かれます．

5. データマイニング・プロジェクトの位置付け

データマイニング・プロジェクトは，企業内におけるデータマイニング活用の浸透の度合いにより次の3段階に分けられます．

(1) Proof of Concept,

(2) パイロット・プロジェクト，

(3) 本格運用プロジェクト．

それぞれの段階によって，目的，期待する成果，プロジェクトの運営方法などが異なります．これらを表1にまとめました．

最初のProof of Conceptは，データマイニングの有効性を判断する段階です．この段階では，企業のサンプル・データを入力として，データマイニングを疑似体験します．ここで完全な結果を求めることは行わず，あくまでデータマイニングが使えそうか，ということを目的とします．

Proof of Conceptの段階では，早期に結果を出すためにスキルのあるソリューション提供会社のデータ分析担当者が作業を行うことが多く，導入を検討する企業側の担当者は，データ提供およびユーザーの観点から分析結果の解釈を行います．また，データ分析を実施する環境はソリューション提供会社が提供する環境となります．

さて，Proof of Conceptが完了すると，実際のプロジェクトがスタートします．実際のプロジェクトでは最初から全面展開するには，分析という業務の性質からいってリスクがありますので，最初にパイロット・プロジェクトを実施します．パイロット・プロジェクトの成果は，実際の分析結果です．企業の担当者はパイロット・プロジェクトを通して自分達の費用と責任の下で，データ分析ソリューションを全面展開す

表1 データマイニング・プロジェクトの段階

プロジェクト	Proof of Concept	パイロット・プロジェクト	本格運用プロジェクト
状況	利用可能性の検討	社内での初期導入・試行	社内での全面活用
分析環境	ソリューション提供会社の環境または社内の環境	社内の環境（アウトソーシングの場合はソリューション提供会社の環境）	
分析の主体	ソリューション提供会社の分析担当者	社内の担当者とソリューション提供会社の分析担当者の共同作業	社内の担当者（アウトソーシングの場合は、社内の担当者とソリューション提供会社の共同作業）
成果	データマイニング有効性の判断材料	経営判断に有効な分析結果または業務に活用可能な分析結果	経営判断に有効な分析結果または業務に活用可能な分析結果かつ、利用目的に合った出力形態
利用データ	最低限のサンプリング・データ	目的に沿って絞り込んだデータ	運用目的に合った分析を行う上で必要な全てのデータ

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

るか否かを決定します．

パイロット・プロジェクトでは，まだ企業内部にデータ分析スキルが蓄積されていないので，プロジェクト全体をソリューション提供会社のプロジェクト・マネージャーがコントロールします．しかしながら，データ分析の本質的な洞察は，業務知識の中にありますので，データマイニングの主体は企業側の担当者です．企業側の担当者とソリューション提供会社のチーム・メンバーは一致協力して，目的に沿った分析を遂行します．パイロット・プロジェクトでは試行錯誤が多く，確立した手順で進めにくいものです．逆にいうと，パイロット・プロジェクトの結果としてデータ分析手順が確立します．そのような理由で，データ加工や分析結果の整理などは手作業となり，これらの作業を行うためのデータ加工エンジニアが必要となります．このデータ加工エンジニアには，スプレット・シート，SQL，さらにはデータ加工用のプログラムを作成するスキルが求められます．

パイロット・プロジェクトは，お客様の投資を最小に抑えて実施します．また，投資の内訳として固定資産をなるべく抱えないようにするためにパイロット・プロジェクトはアウトソーシングし，本格運用を自社で行うことも一つの選択枝となります．

また，マシン・リソースを考慮してデータを絞り込むことも時には必要となります．前段階のProof of Conceptでもデータを絞り込みますが，あえて違いをいうならば，Proof of Conceptではランダム・サンプリングによる絞り込み，パイロット・プロジェクトでは分析シナリオ上で意味のあるデータ抽出による絞り込み，といえます．

さて，パイロット・プロジェクトも無事完了し，データマイニング・ソリューション全体として有効性が実証されると，本格運用に向かいます．この段階では，企業の課題に対するデータ分析シナリオが明確になっています．そしてデータ加工や分析シナリオの実施にかかるワークロードの省力化が必要となります．そのため，本格運用プロジェクトの前段階としてデータウェアハウス構築およびアプリケーション開発プロジェクトを実施します．この開発プロジェクトは，一般的な情報システムの開発プロジェクトです．しかしながら，そのプロジェクトによって構築されるデータウェアハウスには，パイロット・プロジェクトの結果から導き出されたデータ分析シナリオに利用されるエンティティーが配置されています．すなわち，使えるデータウェアハウスが構築されることになります．

本格運用プロジェクトになると，データマイニングの形態の違いが明確に現れます．分析だけで完結するプロジェクトでは作業の主体は営業企画部など，企業の中でデータ分析を実施している担当者です．パイロット・プロジェクトでの経験を経て，担当者はデータマイニング・ツールなどのデータ分析ソフトウェアを使いこなせるようになっています．担当者が本格運用プロジェクトを実施するにあたって最低限必要となるものは，データウェアハウスです．自分達の分析に用いるデータをタイミング良く提供してくれるデータウェアハウスがあれば，決められた分析シナリオに沿った分析を自力で実施可能となります．

一方，施策や他の業務と連結する形態ですと，データ分析者の手を離れて一般業務の担当者が利用することが十分考えられます．それらのユーザーがデータ分析ソフトウェアを使いこなすことはありません．データ分析を目の見えない部分に封じ込めたアプリケーションが効果を発揮します．これはデータ分析ロジックをハード・コーディングしたものではなく，データマイニングのモデリングによって生成されたモデルを組込んだ柔らかいアプリケーションとなります．

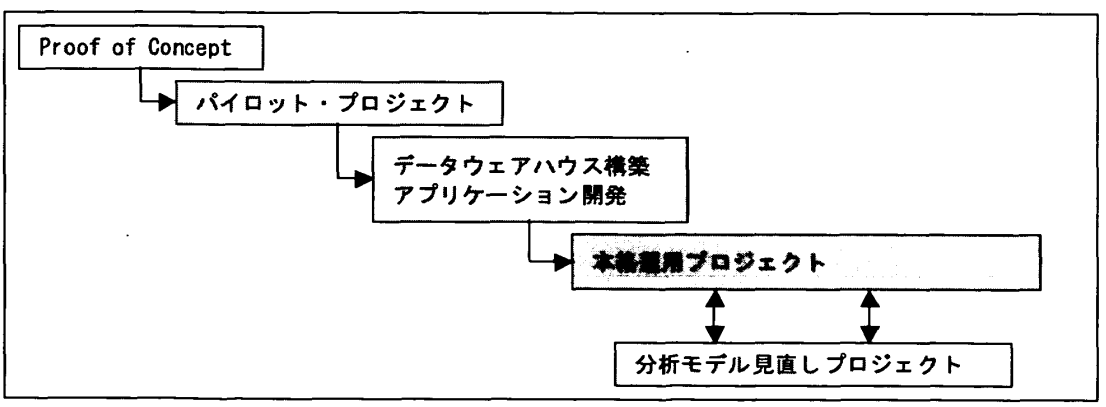

図2　データマイニング・プロジェクトの位置付け

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

どちらのプロジェクト形態でも，そこに使われるデータ分析シナリオやデータ分析モデルを使い続けることは，リスクがあります．市場の状況変化に沿った分析モデルを使用する必要があり，そのために本格運用プロジェクトに使われるシナリオや分析モデルを見直すためのプロジェクトを適切なタイミングで実施することとなります．

6. データマイニング・プロジェクトの局面

データマイニング・プロジェクトにおける局面について説明します．データマイニング・プロジェクトも一般のシステム開発プロジェクトと同様に，局面化することによって手戻りが最小限になり，進捗管理が容易になります．データマイニング・プロジェクトの各局面は，次の通りです．

(1) ビジネス課題確認，
(2) 分析要件定義，
(3) 分析シナリオ設計，
(4) データ準備計画，
(5) データ準備，
(6) データ分析実施，
(7) データ分析結果評価．

それぞれの局面間の関係について，図3に示します．

データマイニング・プロジェクトの各局面の他に「データ基礎分析」という作業があります．これは分析シナリオ設計以後の局面で頻繁に行うものであり，データそのものがプロジェクト全体へ大きな影響を与えることを示唆しています．それぞれの局面において生データ，加工データ，そして分析結果データを見ることによって，データに潜む間違いを発見し，更に対象課題に対する仮説の醸成に貢献します．

データマイニング・プロジェクトは，システム構築プロジェクトに比べて工程管理は格段に難しくなります．その理由は，成果物が分析を行ってみないとハッキリしないからです．システム構築プロジェクトでは，成果物の達成度合いが時間とともに確実に増加します．ところが，データマイニング・プロジェクトでは，分析結果が出るまでは実質的な成果物はなく，また分析が済んだとしても有益な分析結果を得られない場合もあり，先行きが不透明です．システム構築プロジェクトが成果物を作り上げていくプロジェクトであるとすると，データマイニング・プロジェクトは成果物を発見するプロジェクトであるともいえます．このような特性を持つデータマイニング・プロジェクトではリスク管理が重要となります．

7. スキルと体制

データマイニング・プロジェクトを実施する前提として，データマイニング・プロジェクトは業務プロセス再構築プロジェクトであることを認識する必要があります．そこでは，次の役割を担うメンバーが必要となります．

(1) ビジネスインテリジェンス（BI）コンサルタント
(2) データマイニング・アナリスト

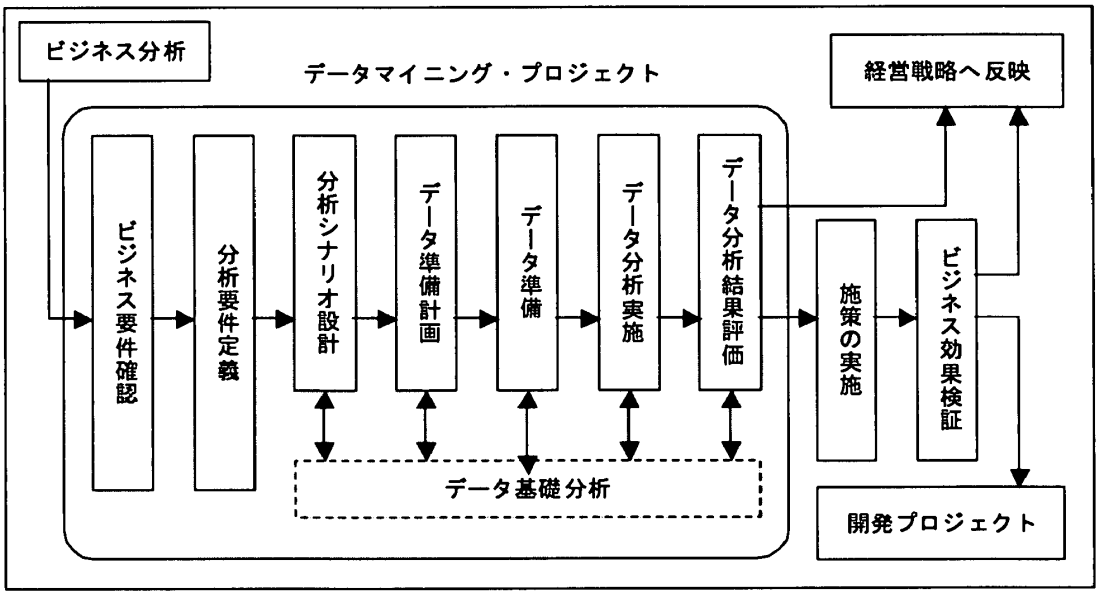

図3 データマイニング・プロジェクトの各局面の関係

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

(3) プロジェクト・マネージャー

(4) データ加工エンジニア

プロジェクト・マネージャーに関しては，システム構築とは異なるプロジェクト管理能力が要求されるので，アナリスト，コンサルタントとしての経験を経た人が望まれます．プロジェクトの規模によっては，コンサルタントが兼任する場合もあります．

データマイニング・プロジェクトではデータ分析という科学的な手法を用いてビジネス課題を解決するという特質から広範囲のスキルが要求されます．データマイニングが世間一般に紹介された当初，ニューラル・ネットワークなどのデータマイニング・アルゴリズムと適正なモデルを作るための技術ばかりが注目されました．しかしながら，技術に偏重した人材によるプロジェクトでは局所最適解しか求めることが出来ないという現象が起こります．すなわち，最高の技術によって精度の高いデータマイニング・モデルを作成したにもかかわらず，業務上の活用が図れないという状況です．このような経験から，図3に示したデータマイニング・プロジェクトの局面定義において，ビジネス分析の成果を受けて実施するビジネス要件確認の局面を重要視するようになりました．

データマイニング・プロジェクトの局面ごとに見ると必要とされるスキルに特徴があります．まず，ビジネス要件確認と分析要件定義においては，業界および業務についての知識が大切になります．また経営者の意識の奥にある真の課題を引出すためにコミュニケーション能力が要求されます．また，この段階で考慮することとして，全体的なゴールを設定することです．部分最適に陥らず，企業として全体最適を得るにはファイナンス的な判断が必要となります．それは投資対効果を見極めることです．単に一施策のコストとその効果ではなく，体制の維持や要員の育成なども考慮した上で，数年後までに得られる利益の累積が現在価値で判断したときにプラスとなることを確認するスキルが必要となります．

次の分析シナリオ設計がデータマイニングの専門家としての力量を発揮する局面です．データ準備計画においては，データウェアハウスとデータマートのスキルが必要となります．さらに，プロジェクト全体をコントロールするためには当然のことながらプロジェクト管理スキルが不可欠となります．これらのスキル要件をまとめたのが図4です．この図の中央に位置するBusiness Insightsはデータマイニング・プロジェクトを数多く経験する中で体得する洞察力です．これは単なるスキルではなく，経営者の気持ちになることによって身に付くものと考えられます．スキルを極めても乗越えられないものが意識の問題です．仕事としての割り切りを超えたモチベーションが人を育てることは，この領域にも当てはまります．

8. コンサルタントの役割

データマイニングの導入期においてBIコンサルタントの役割は重要です．データマイニングがこれだけ普及した理由は，その適用対象をビジネス領域としたことです．そのビジネスにデータマイニングという部品を組込むための方法論をBIコンサルタントが提供します．

かつてデータマイニングに必要なのはニューラルネットワークや多変量解析のスキルだと思われていました．しかし，優秀な技術者を擁し与えられたデータ分析課題に対して最適な解を求めるが，ビジネスとしては満足が行かない例がよく見られました．例えば，需要予測の課題ですが，その需要予測により数ヶ月先の生産計画を調整する場合があります．ここでデータ分析者は極めて精度の高い予測モデルを構築することに全力を尽くします．しかし，BIコンサルタントは，ビジネス要件確認の局面において本来の目的である最適在庫を実現するための要件を見直し，ビジネスとして最も効果のある課題の再定義を行います．その結果

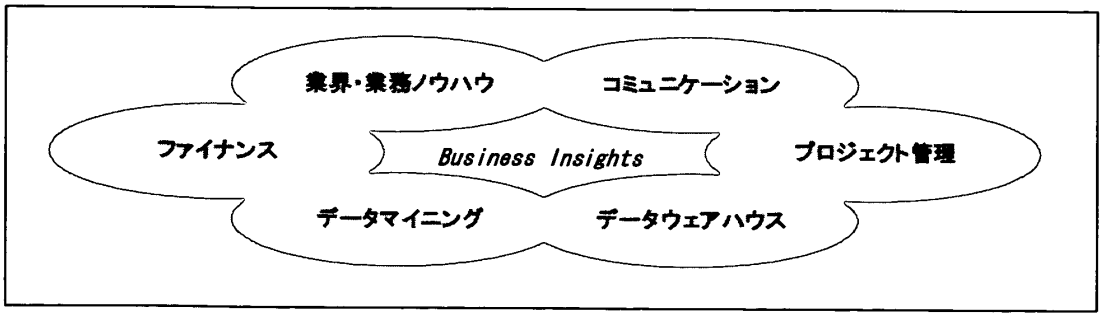

図4 データマイニング・プロジェクトのスキル

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

としてサプライチェーンの見直しの方が重要であるならば，そのようにプロジェクトの方向修正を行います．データマイニングに拘らず本来のビジネス課題を解決するためにBIコンサルタントは適切な指針を与え，その方向がデータ分析による課題であるならば最適な分析シナリオへと導く役割を果たします．

分析要件定義では，データマイニングの入出力を定義します．アプリケーション開発業務における外部設計と同様に，前後のプロセスとの間のデータのやり取りを決定します．この局面までは，業務の担当者が重要な役割を果たします．したがいまして，業務担当者の参画なくしては本当に有効なデータマイニングは行えません．

この局面におけるプロジェクト成功要因は，代替案を用意しておくことです．まず，最適と思われる業務シナリオの他に達成可能性の高い代替案も用意します．また，業務シナリオの中でのデータ分析も幾つかの代案を用意しておきます．これによって，行き詰まったときの速やかなリスク回避が可能となります．

ここまでの作業でBIコンサルタントは，下記のことを注意してプロジェクトをコントロールします．

(1) 企業が抱える課題の本質を引出す（担当者の思いが間違っている場合もあります）．
(2) データマイニング・プロジェクトは業務プロセスの改善プロジェクトであることを，企業の担当者に理解していただき，担当者自身に主体的に参画していただく．
(3) ビジネスに対する効果を優先させる．
(4) 実現可能性を見極める．

業務シナリオを定義するためには，社内の各部門の担当者と多くのセッションを行うことになります．このセッションを円滑に実現することもプロジェクト推進において大切な成功要因となります．

9. おわりに

筆者は，日本アイ・ビー・エム㈱でデータマイニング活用のコンサルテーションを企業へ提供しておりますが，その成果の多くは公になることはありません．その理由は，その成果自身が企業のノウハウであり競争他社との差別化要因となるからです．時折，新聞などでデータマイニング導入事例を発表する企業がありますが，ほとんどは，その企業の宣伝と思われます．その多くは，分かりやすいバスケット分析の成果を披露しています．しかし，ミルクと紙オムツとの相関のように，競合他社に知れても大丈夫な，ビジネスゲインのない結果しか公表しないようです．

企業がデータマイニングを業務に組り込むには，多額の費用がかかります．もちろん，成功した場合には，その投資を上回るリターンと競争優位条件を手にすることになります．しかし，データマイニングの安易な導入は失敗を招きます．データマイニングの導入は，ビジネス戦略からくるビジネス・モデルの変革です．そのことを企業の経営者が理解して取り組むこと，すなわち，トップのリーダシップが最も重要な成功要因といえます．

参考文献

[1] Kohonen, T., 1995, Self-Organizing Maps, Springer-Verlag Berlin Heidelberg New Yourk.
[2] Bigus, Hoseph P., 1996, Data mining with nural networks, McGraw-Hill.
[3]『IBM DB 2 インテリジェント・マイナーインテリジェント・マイナー使用　バージョン6リリース1』SH 88-7360-00.

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．