

最新の予測分析の ビジネス・ガイド



IBM

目次

なぜこのガイドが重要か	<u>03</u>
全体像	<u>04</u>
予測分析とAIが重要な理由	<u>05</u>
AI導入の転換期	<u>06</u>
AIはどのようにビジネスを強化することができるか	<u>08</u>
AIのはしごを登る	<u>13</u>
どんなソリューションがあるか	<u>14</u>
次のステップを踏み出す	<u>16</u>
重要なポイント	<u>17</u>
なぜ意思決定最適化を適用すべきなのか	<u>18</u>
用語集	<u>19</u>



なぜこのガイドが重要か

ビジネスにおいては、先見性がすべてです。次に何が起こるかを予測できれば、次のようなことができるようになります。

- よりスマートな意思決定を行う
- 市場投入までの時間を短縮
- 競合他社を圧倒する

最新の予測分析は、ヒストリカル・データをリアル・タイムの洞察で補強し、それを活用して未来を予測し、形づくることを可能にします。

予測分析は、分析の道程における重要なマイルストーンであり、古典的な統計分析技術が人工知能(AI)の新しい世界と出会う合流点です。

[Forrester Research](#)によると、企業は機械学習とナレッジ・エンジニアリングを組み合わせるようになる時期にきています。データを人間の知恵で補強することにより、AIアプリケーションの開発は劇的に加速します。

このガイドは、ビジネスで以下のことを行うのに役立ちます。

- 最新の予測分析の状況の把握
- AIを一層活用し・強化する機会を特定
- データ・サイエンス・チームと業務部門のステークホルダーの双方を支援し、迅速に価値を実現

最新の予測分析は、機械で生成された予測と人間の洞察を用いてビジネスを前進させることを目的としています。



全体像

AI革命が根付くにつれ、企業はデータ・サイエンス・チームに大きな問題への取り組みを求められるようになってきています。

その結果、データ・サイエンティストには、単発の研究プロジェクトに携わるだけでなく、より多くの仕事が求められるようになりました。日々の意思決定のためのリアル・タイムな洞察を提供する、反復可能で自動的な方法を見つける必要があります。

このような期待に応えるために、データ・サイエンスのリーダーは、最新の予測分析テクノロジーの可能性を業務部門のステークホルダーに説明できるだけでなく、結果を出すことも必要です。

成功するデータ・サイエンス戦略を定義し、実行する能力は、今後数年間、リーダーとそのフォロワーの間の重要な差別化要因の1つとなるでしょう。

これは簡単なことではありません。データ・サイエンス能力を高めるには、以下のようなことを行います。

- 熟練したスペシャリストで構成されるバラバラのチームの継続維持
- シームレスな協力を促進
- 予測モデルが常に業務部門から信頼されるよう、健全なガバナンス構造を導入

とりわけ、データ・サイエンス・チームと業務部門は、新しいやり方で効果的に協力する必要があります。こうした方法には、予測分析で何ができるかを理解し、AIが企業の優位性をもたらす領域を特定することが含まれます。

- 1 お客様はどのように行動するのでしょうか？
- 2 なぜ市場が変化しているのでしょうか？
- 3 ビジネス戦略の成否の要因は何でしょうか？
- 4 この先どうなるのでしょうか？
- 5 資金はどのように調達しているのでしょうか？
- 6 購買決定者はどこにいるのでしょうか？



予測分析とAIが重要な理由

予測分析は新しい概念ではありません。統計学者は長年にわたり、決定木や線形回帰、ロジスティック回帰を用いて、業務部門でデータの相関や分類を行い、予測を立てることを支援してきました。

新しい点は、予測分析の範囲が広がったことです。機械学習とディープ・ラーニングの飛躍的な進歩により、これまでビジネス投資が現実的でなかった領域で予測モデルを使用する機会が広がりました。

企業では、直感的なツール、新しい予測技術、ハイブリッドクラウドの導入モデルがかつてないほど融合し、予測分析が以前より身近なものになっています。

このような状況が、転換期を招いています。あらゆる規模の組織が初めて以下のことを行うことができるようになりました。

- 予測分析のビジネス・プロセスへの組み込み
- 規模を拡大してAIを活用
- テキストから位置情報まで、これまで未開拓だった「ダーク・データ」からの価値の抽出

部門ごとの少人数のAIプロジェクトから、企業向けのデータ・サイエンス・プラットフォームへと進化させることができれば、組織は大きな競争力を獲得することができるのです。このチャンスを逃すと、時代に取り残されるリスクがあります。

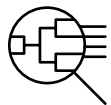
776億ドル

2022年までのコグニティブとAIシステムへの投資

(出典: [IDC](#))



AI導入の転換期



どのようなデータを解析することができるのですか？

以前：

主に規模が大きいのはリレーショナル・データで、その他の種類のデータは特別な研究プロジェクトに必要です。

現在：

リレーショナル・データ、半構造化文書、テキスト、センサー・データなど、ヒストリカルとリアルタイムの両方の大規模な分析が可能です。



どのようなツールが利用できますか？

以前：

専門知識が異なるチーム間で何度も引き継ぎが必要な、バラバラで互換性のないツール。

現在：

ドラッグ&ドロップのインターフェースとオープン・ソースのノートブックを融合させ、チーム間の協業をさらにに行いやすくします。



どのような分析技術が使えるのでしょうか？

以前：

ロジスティック回帰、線形回帰などの基本的な統計手法。

現在：

統計手法を、最先端の機械学習とディープ・ラーニングのアルゴリズムで補強。



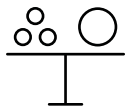
企業はどのように分析アプリケーションを導入しているのでしょうか？

以前：

アプリケーションと分析は、オンプレミスのサーバーやデータ・ウェアハウスと連携しているため、場所を問わず、常時分析できる機会に限られています。

現在：

ハイブリッドなマルチクラウドの導入は、オンプレミスのセキュリティと柔軟性・拡張性を兼ね備えながら、データが存在する場所での分析を進めるのに役立ちます。



企業はどのようにして分析をビジネス・プロセスに組み込むことができるのでしょうか？

以前：

業務の専門家が手動で分析するための静的なレポートを生成します。

現在：

新しいアプリケーションや企業アプリケーションに予測モデルをシームレスに組み込むことができます。



企業はどのようにガバナンスを導入しているのでしょうか？

以前：

部門レベルのポリシーにその都度従い、可視性やトレーサビリティがほとんどない。

現在：

一貫性のあるガバナンスとセキュリティのフレームワークにより、大規模に企業全体のポリシーを実行することができます。



企業はどのようにして人工知能を最新のアプリケーションに投入しているのでしょうか？

以前：

アプリケーション開発チームとデータ・サイエンス・チームが完全に分断されているため、それぞれの導入がカスタム・プロセスになっています。

現在：

データ・サイエンス・ライフサイクルは、AIを統合するための標準化された反復可能なプロセスを作成するために設計されています。



企業はどのように分析を推進していけばよいのでしょうか？

以前：

記述的分析から予測的分析、処方的分析に至る各ステップでは、個別のツール、スキル、投資が必要です。

現在：

統合されたプラットフォームは、分析作業の進行に対応し、新人研修を簡素化し、ニーズの変化やスキルの発達に合わせて成長させることができます。

AIはどのようにビジネスを強化することができるのでしょうか？

理論的には、予測分析に最新のアプローチを採用することは簡単です。テクノロジーはもはや障害ではなく、優れたツールが参入障壁を大幅に下げています。

しかし、実際は、価値の実現にはまだ課題があります。特に業務部門のステークホルダーは、AIをめぐる誇大広告に捕らわれ、データ・サイエンスが何を達成できるのかに非現実的な期待を抱きがちです。

ユースケースの定義

データ・サイエンスと業務部門のリーダーの最初の作業は、最新の予測分析が価値を実現できる具体的かつ実用的なユースケースを特定することです。

ユースケースの中には、以下のように、ほとんどの産業で一般的に適用できるものもあります。

- 営業・マーケティング・チーム向けの製品レコメンデーションと「ネクスト・ベスト・アクション」のモデル
- カスタマー・サポート・チーム向けのコール・センターの自動化

また、特定の業界、部門、あるいは業務チームに特化したユースケースも考えられます。これらは、実行がより困難になる傾向がありますが、独自の競争優位性を引き出す可能性も高くなります。

AIシステムへのビジネス投資をリードしているのは、どのような業務機能なのでしょうか？

46% 営業およびマーケティング

40% カスタマー・サポート

(出典: [Forrester Research](#))



一般的なユースケース

業務で新しいテクノロジーに投資する際、最も手の届きやすい果実を最初に選ぶことは、理にかなっていることが多いのです。

予測分析も例外ではありません。いくつかのユースケースは業界を超えて広く適用可能であり、ベンダーはすでに汎用的なパッケージ・モデルやサービスを開発しています。

これらのサービスは、データ・サイエンスを研究機能から日常の運用の一部に組み込もうとする業務部門にとって、優れた出発点となり得ます。導入が容易で、カスタム開発を最小限に抑え、迅速に価値を実現することができます。

コール・センターの最適化

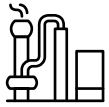
予測不可能な量のお客様からの電話、Eメール、SMS、チャット・メッセージの処理は、多くのカスタマー・サービス・チームにとって課題となっています。

インテリジェントなチャット・ボットは、従業員のプレッシャーを軽減し、お客様の待ち時間を短縮する強力かつ費用対効果の高い方法です。これらのチャット・ボットは、お客様の問い合わせを把握するために、以下のような機能を使用しています。

- AIによる音声認識
- 自然言語処理
- コンテンツ分析により、人手を介さずに社内のナレッジベースで有用な回答を探索

最新の予測分析が提供できる最も一般的な業界横断的なユースケースには、以下のようなものがあります。

- パーソナライズされたリアルタイム・レコメンデーションとオファーによる**クロス・セルとアップ・セルの増加**
- 顧客の解約を予測し、その防止のために**ロイヤルティを高める**
- お客様の声を聞き、ニーズを先取りして**提案を最適化する**
- ターゲットを絞ったパーソナライズされたキャンペーンで**マーケティングを強化する**
- 正確な予測で在庫コストを最小化し、**リソース管理を向上させる**
- 適材適所の人員配置と正確な労働力予測による **生産性の向上**
- 故障の未然防止による **メンテナンス・コストの削減**
- 正確な顧客信用スコアリングによる**リスク軽減**
- 不審な行動パターンの特定による**不正の検出**
- 未開拓の需要に対応し、予測を最新のアプリケーションに統合することにより、**新たなビジネス・モデルを切り拓く**



業界特有のユースケース

多くの業界の革新的な組織は、すでに特定のビジネス課題を解決するために、独自の予測モデルの構築に投資しています。次の2ページでは、複数の主要産業におけるAIと予測分析アプリケーションの可能性のいくつかを紹介します。



保険

保険会社は、以下のような業務に予測分析を使用しています。

- 不正な請求の検出
- 申請者ごとに関連するリスクを評価し、見積もりと保険料を最適化
- 危険な気象現象を予測し、自動車保険の請求額を削減



商業銀行

商業銀行は、以下のような業務に予測分析を使用しています。

- 取引における市場リスクおよびカウンター・パーティー・リスクの評価
- ローン申請に対する信用リスク評価
- 不正取引のリアルタイムでの検出
- 予測モデルを活用した融資承認プロセスの迅速化



エネルギー/公益

公益事業では、以下のような業務に予測分析を使用しています。

- 膨大な物理的資産のネットワーク管理
- 生産と需要パターンの予測
- 停電の事前予測
- 需要と供給の計画



政府

政府は、以下のような予測分析のユースケースなど、多くの分野での政策決定に正確な統計情報を利用しています。

- 給付金詐欺の検出
- 公共サービスの利用パターン予測
- 廃棄物処理と交通の流れの最適化



製造

製造では、以下のような業務に予測分析を使用しています。

- 製品品質のモデル化と欠陥の検出により、生産ラインのスムーズな運営を維持
- 倉庫管理・物流の最適化
- 機械学習モデルによる自律走行車用センサーの開発



小売

小売では、以下のような業務に予測分析を使用しています。

- カスタマー・ロイヤルティ・プログラムの管理
- 顧客プロフィールと高度な傾向モデルに基づき、ターゲットを絞ったレコメンデーションを行うことにより、クロス・セルやアップ・セルを促進
- 正確な需要予測が可能



食品

食品では、以下のような業務に予測分析を使用しています。

- 食品の安全性に関するデータ収集・分析の自動化
- 健康被害の発生を予測・警告し、迅速な介入が可能
- 企業の機密データを保護し、競合他社との安全な協業を実現



医療

医療機関では、以下のような業務に統計モデル技術を使用しています。

- 心電図やその他の医療機器からのデータ・ストリームの監視
- 患者の状態がいつ変化するかを予測
- 医学研究
- 患者データ・ストリームをリアルタイムで分析



リテール・バンキング

リテール・バンキングでは、以下のような業務に予測分析を使用しています。

- 迅速なクレジット・スコアリングによる顧客満足度の向上
- ハイブリッドクラウド・インフラストラクチャーによる柔軟性と堅牢性・安全性の両立
- 革新的なアーキテクチャーによるコスト削減と開発スピードの向上



運輸

運輸・物流では、以下のような業務に予測分析を使用しています。

- ルート計画の最適化
- 車両の予知保全が可能
- サプライチェーンの運用を最適化

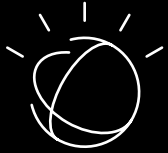


教育

教育機関では、以下のような業務に予測分析を使用しています。

- 学生の成績と定着率の予測
- 目標達成のために特別な対応が必要な生徒の特定
- 資金提供者との関係強化
- 学生の行動の追跡により、欠席率を低減

AIのはしごを登る



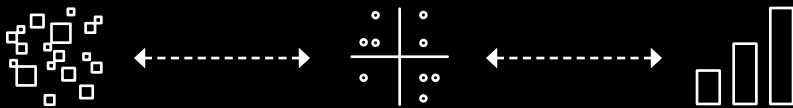
組み込み-信頼性と透明性を備えたAIの運用

分析-あらゆる場所でAIを用いて洞察を深める

編成-信頼できる分析基盤の構築

収集-データをシンプルに、アクセスを容易に

あらゆる場所、あらゆる種類のデータにアクセス



最新の予測分析で成功をおさめることは旅路のようなものです。技術的な成熟度と組織的な成熟度の両方を考慮し、ビジネスに適したレベルでAI戦略を調整することが重要です。データ・サイエンスと業務部門のリーダーは、ビジネス価値を実現するための最善かつ最速の方法を定義するために協力する必要があります。

技術的な観点から、AIの成熟度をはしごのように可視化することができます。はしごの最初のステップはデータ収集です。データがないと、分析もモデル化もできないからです。次のステップは、データ編成です。ガバナンスと見つけやすい機能を提供するためにメタ・データを追加し、必要なデータ・サイエンティストが常に正しいデータを利用できるようにします。

データの収集と編成は重要なテーマですが、このガイドでカバーする範囲を超えています。その代わりに、以下のようなトップ2の階層を登る支援をすることに焦点を当てましょう。

- 予測モデルの構築、トレーニング、テストによるデータ分析
- アプリケーションの一部としてこれらのモデルを本番環境に導入することで、AIを運用に組み込むことができます。

どんなソリューションがあるか



IBMのAIポートフォリオには、AIの階層のトップに登るために必要なものがすべて揃っています。

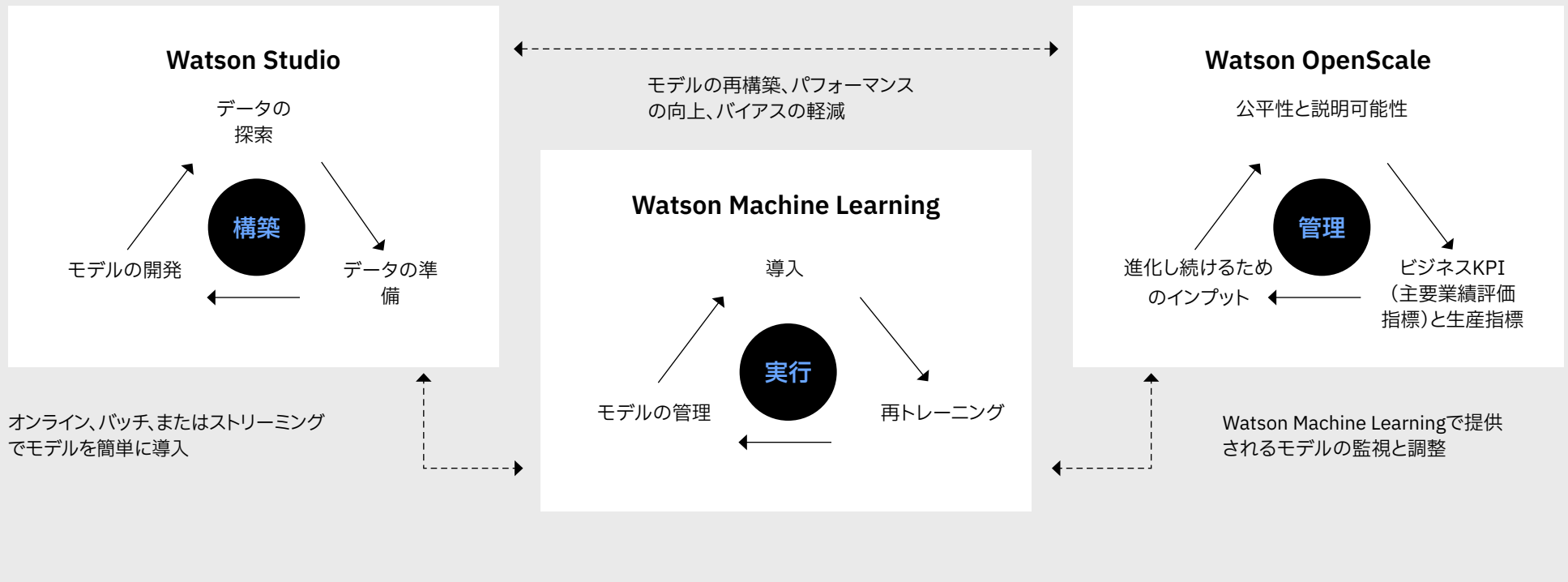
[Watson Assistant](#)などの構築済みのAIサービスは、一般的なユースケースに迅速かつ効率的に対処し、迅速に価値を実現するのに役立ちます。

独自のAIソリューションの開発を始める準備ができれば、[Watson Studio](#)とWatson Machine Learningは、予測モデルの構築、トレーニング、導入を行うためのシームレスなワークフローを提供します。これらのソリューションは、最先端のIBMツールと最高のオープン・ソースAIフレームワークの両方を活用することで、お客様を支援します。

[Watson Knowledge Catalog](#)は、モデルとデータの堅牢なデータ・ガバナンスと見つけやすい機能を提供し、[Watson OpenScale](#)は、リアルタイムでモデルを監視および管理し、精度の向上、説明可能性の増大、バイアスの軽減に役立ちます。

[IBM Cloud™ Pak for Data](#)は、これらすべての機能へのアクセスを一元化し、強力なマルチクラウド・データ・プラットフォームを提供します。

IBM Cloud Pak for Dataは、SPSS ModelerやDecision Optimizationなどのデータ・サイエンスの生産性を向上させる機能を追加し、価値を生み出すための時間を短縮し、AI/MLプロジェクトの成功の可能性を高めます。



AIのトップ2のステップは、「分析」と「組み込み」です。これらのステップに到達するために、組織はデータ・サイエンスのライフサイクルの各段階で、データ・サイエンティストとビジネス・ステークホルダーが効果的に連携できるようにしなければなりません。

ライフサイクルの全体像は、以下の3つのサブ・サイクルが相互作用することにより可視化されます。

構築

データ・サイエンティストは、ビジネス・データを探索して興味深い特徴を特定し、予測モデルを設計するために使用する、適切に構造化されたデータ・セットを準備します。

実行

運用チームは、モデルのトレーニング、テスト、導入、管理を行い、必要に応じて再トレーニングを行います。

管理

業務部門のエキスパートは、モデルの実行パフォーマンスを監視し、バイアスや説明の必要性の兆候を探し、フィードバックを提供し、再トレーニングが必要な場合はデータ・サイエンス・チームに通知します。

次のステップを踏み出す

AIの進歩の度合いに応じて、ビジネスには組織全体の予測分析の採用レベルに基づくさまざまな要件があります。

開始する

ビジネスでデータ・サイエンス機能を構築する際は多くの場合、特定の質問に答えるためのモデルの開発や、研究プロジェクトへの対応など、特別なプロジェクトから始めます。データ・サイエンティストは自分のコンピューターやラップトップでいつでも業務を行い、必要に応じて広範なチームと同期させることができます。

成長する

データ・サイエンスが広く採用されると、さまざまな部門がモデルを導入し、データ・ソースに接続して本番のアプリケーションに組み込む必要がでてきます。Watson StudioとWatson Machine Learningでは、各部門のデータ・サイエンス・チームとITチームが、このライフサイクルを通じた協力を容易に実現できます。

企業規模への移行

AIがビジネスに不可欠なプロセスに組み込まれた後、モデルやデータを管理・運営するためには、統合基盤の構築が不可欠です。IBM Cloud Pak for Dataは、シングル・ポイント・オブ・コントロールとして機能する包括的なマルチクラウド・プラットフォームに必要なインフラストラクチャーとツールを提供することができます。

実用化に向けて

データ・サイエンティストでも業務部門のリーダーでも、IBMの最新の予測分析ポートフォリオがどのようにビジネスを変革できるかを知る最良の方法は、実際に体験してみることです。以下のチュートリアル of いずれかをお試しください。

機械学習による演習を行う

IBM Watson StudioでApache SystemMLを使った演習を行い、機械学習の世界に飛び込んでください。[さらに詳しくは](#)

心不全を予測するスコアリング・モデルの作成

IBM Watson Studioを使用して、IBM Watson Machine Learningで予測モデルを構築します。[さらに詳しくは](#)

IoTセンサー・データによる機器故障の予測

IBM Watson Studioで多変量のInternet of Things (IoT) センサー・データを分析し、機器の故障を予測する方法をご覧ください。[さらに詳しくは](#)

オープンな医療データセットを分析し、洞察を得る

IBM Watson Studioを使用して、機械学習分類器を実行し、出力を評価尺度で比較します。[さらに詳しくは](#)

生データの成形と精製

IBM Data Refineryと連携し、予測分析のための大規模なデータセットを準備します。[さらに詳しくは](#)

重要なポイント

IBMの最新の予測分析ポートフォリオは、データ・サイエンスと業務部門のリーダーがAI時代の競争優位をつかむために使用できる以下のようなメリットを提供します。

スケール

- データ・サイエンスとデータ・エンジニアリングの作業を自動化することで、運用の作業量とコストを削減
- 複数の企業アプリケーションでシームレスにモデルのトレーニング、テスト、導入が可能
- ハイブリッド、マルチクラウド環境に共通のデータ・サイエンス機能を拡張する

スピード

- 事前に構築されたアプリケーションと学習済みモデルを活用し、開発を加速化
- データ・サイエンス・チームと業務チームの協力を支援し、より速く価値を実現する
- 最先端のIBMとオープン・ソース・ソフトウェアの組み合わせによるモデル構築の効率化

シンプル

- データ・サイエンスのライフサイクル全体を管理する統合基盤の活用
- 開発・導入プロセスの標準化
- 組織全体におけるデータ・ガバナンスとセキュリティの単一フレームワークの作成

Watson Studioは、業務部門が問題解決とビジネス機会の特定に集中できるよう支援します。

さらに詳しくは

Watson Machine Learningは、業務部門が必要とする結果を迅速に提供するためのモデルの導入と管理を支援します。

さらに詳しくは

なぜ意思決定の最適化と予測分析を組み合わせるのでしょうか？

IBM Decision Optimizationは、複雑な最適化問題を解決することで、データ集約度の高い業界がより良い意思決定を行い、ビジネス目標を達成することを可能にする処方的な分析ソリューションです。業務部門のリーダーは、このツールを使って以下のような活動など、資源の効率的な使用を強化しています。

- サプライチェーンにおける在庫の流れ
- 労働力のスケジューリング
- 輸送のルート

このソリューションは、機械学習アプリケーションの予測結果を利用して最適化された結果を提供するため、予測分析との相性が良いのです。機械学習は、ユーザーから与えられた観察結果に基づいて、未来に関する洞察を提供します。機械学習では、答えを知っている人が、その答えを見つけるための方法を機械に学習させます。

意思決定の最適化は、次のステップに進み、その情報をもとに行動することを可能にします。意思決定の最適化では、答えは分かりませんが、何が良い答えで何が悪い答えなのかはよく知っています。機械学習の出力を受けて、意思決定の最適化で行うべき行為を指定し、ビジネス目標を達成するための最適化の規則や制約を盛り込むことができます。

その後、意思決定の最適化は、実行可能な項目や変革のための推奨など、業務で価値を実現するための答えを返します。この活動を行うことで、意思決定の最適化では、予測分析が提供できる内容を強化します。

このソリューションにより、チームは最適化および機械学習技術と、モデル管理、導入、その他のデータ・サイエンス機能を組み合わせて、運用効率を向上させる最適なソリューションを開発することができます。

さらに詳しくは



用語集

アルゴリズムとは、特定の問題を解決するためにデータに適用できる一連のオペレーションを定義した規則のことです。データ・サイエンスの文脈では、以下のような膨大な技術を含んでいます。

- 決定木と回帰モデル
- 自己回帰・移動平均法 (ARMA)、自己回帰統合移動平均法 (ARIMA)、指数平滑化法
- 予測因子と外れ値検出を伴う伝達関数
- アンサンブル・モデルと階層モデル
- ベクトル・マシンと時間的因果モデリング
- 時空間予測のための時系列と空間AR (拡張現実)
- 敵対的生成ネットワーク (GAN) と強化

データ・サイエンス・プラットフォームでは、これらの強力なアルゴリズムに簡単にアクセスできるようにする必要があります。

人工知能 (AI)とは、コンピューター・システムがデータを解釈し、学習する能力のことです。機械学習やディープ・ラーニングのモデルを用いて構築されたシステムを指す言葉として最も一般的に使われています。AIテクノロジーを使用することで、これまで困難とされてきたさまざまな問題をコンピューターが解決できるようになります。

バイアスは、モデルの設計、トレーニング、テストを行う際に一般的な問題で、不正確な予測につながる可能性があります。実行中のモデルを監視・監査してバイアスを軽減することは、業務部門がAIをより広く採用しようとする中で、ますます重要なテーマとなっています。

分類モデルは、既に分類されているデータ点と比較することで、データ点をカテゴリーに分類することを目的としています。結果はスコアではなく、限られた選択肢の中の一つを意味する離散値です。例えば分類モデルは、お客様が購入する可能性が高いか、信用リスクが低いかどうかについて、イエスかノーかの答えを出すことができます。分類モデルは、決定木やロジスティック回帰など、さまざまな方法で構築することができます。

コンテンツ分析はテキスト、画像、音声、動画ファイルなど、さまざまなフォーマットのドキュメントに含まれる非構造化データを分析するものです。機械学習技術は、これまで人が何百、何千時間もかけてレビューし、分類していた大規模なコンテンツ・リポジトリの分析を著しく加速させることができます。

データ・サイエンスは統計学、データ分析、機械学習など全般を統合し、データを活用してビジネスの問題を解決するための幅広い学問分野です。

ディープ・ラーニングとは機械学習の一分野であり、多数の隠れ層を持つニューラル・ネットワークを使用するものです。これらの高度に洗練されたネットワークは、コンピューター・ビジョン、機械翻訳、音声認識など、ディープ・ラーニングの最先端分野で使用されています。

ディープ・ニューラル・ネットワークのトレーニングは非常に計算量が多く、通常、高性能プロセッサを搭載したマシンのクラスターが必要です。IBM Watson StudioやIBM Cloud Pak for Dataなどのハイブリッドクラウド・プラットフォームは、あらゆる規模の企業にとって、この種のインフラストラクチャーをより身近で手頃なものにすることができます。

導入とはモデルをビジネス・アプリケーションに統合し、そのモデルを実際のデータに対して実行するプロセスです。テスト環境、ステージング環境、本番環境を通してモデルを作成し、移動させるには、データ・サイエンス・チーム、アプリケーション開発者、IT運用チームが協力する必要があります。

オープン・ソースのデータ・サイエンス・ツールを、組織の既存の継続的インテグレーションおよび導入パイプラインと統合することが困難な場合があります。チーム間で何度も引き継ぎを行いながら手作業で導入することを避けるには、自動導入機能を備えた首尾一貫したデータ・サイエンス・プラットフォームが大きな利点となります。

予測モデルの開発には、従来の統計方法や機械学習アルゴリズムを使用し、お客様のデータ・セットに対してトレーニングとテストを行うことでモデルを作成し、洗練させていきます。

開発プロセスは非常に反復的で、要求される精度を達成するためには、何十、何百ものモデルでトレーニングする必要があるかもしれません。そのため、モデルの開発やトレーニングに関わるワークフローを自動化することで、大きな価値を実現することができるのです。

説明可能性は、予測モデルを使用して推奨し、ビジネスの意思決定を支援するシステムの重要な属性です。複雑で不可解な予測モデルでは、その出力を信頼するように業務部門のステークホルダー、規制当局、お客様を説得するのが難しい場合があります。Watson OpenScaleの高度なランタイム監視とロギング機能により、各決定に関する状況が提供され、AIモデルの透明性と監査可能性が高まります。

モデル構築のプロセスにおいて、データの探索は重要な要素です。この活動は、与えられたデータ・セットの興味深い特徴を明らかにし、隠れた関係を明示し、予測モデリングがビジネス価値を実現するユースケースを強調することを目的としています。

探索の段階では、データ・サイエンスのスキルと業務の知識を駆使して、答えてほしい質問と予測したい結果を定義することが重要です。その結果、領域を完全に探索し、データを適切な形に整えて次に進むまで、準備と探索のサイクルを繰り返すことになるかもしれません。

地理空間分析とは、緯度と経度、郵便番号、住所などの地理的なデータを分析することです。この分析は、実用的なデータ・サイエンスのさまざまな問題を解決するために非常に役立ちます。最新のデータ・サイエンス・プラットフォームは、地理空間情報の検出、解析、計算を容易にし、結果を可視化するためのマッピングツールとのインテグレーションを容易にするものであるべきです。

人工知能における推論は、ナレッジ・ベースに論理的な規則を適用し、不確実性がある中で結論を導き出すものです。推論では、ユーザーは単純化され、圧縮され、ランタイム・パフォーマンスが最適化された予測を得ることができます。

線形回帰は、ある値やスコアを説明したり予測したりするために、1つの独立変数を使用する統計のプロセスです。例えば、ある週に売れた製品の在庫保管単位数や、お客様がアカウントを閉鎖するリスクの割合などです。

ロジスティック回帰は、結果を予測する際に用いられる統計プロセスです。このプロセスは線形回帰とは異なり、1つの独立変数が無限の可能性ではなく、限られた数の可能な値しか持ちません。ユーザーはロジスティック回帰を、回答が1位、2位、3位などの数値順のようなカテゴリーで分類される場合に使用します。

機械学習は、統計的方法を使用し、大規模なデータ・セットから洗練された予測モデルやアルゴリズムを導き出すもので、明示的なプログラミングを必要としません。

通常、この反復プロセスは、データ・セットをトレーニング用とテスト用の2つのサブ・セットに分割することから始まります。トレーニング・セットに対してモデルをトレーニングし、テスト・セットに対して数十から数百のバリエーションでテストを行い、その予測精度を評価するのです。このプロセスを実行し、各反復のベスト・パフォーマンスを基に次世代のバリエーションを作ることによって、モデルは徐々に学習し、パフォーマンスが向上します。

機械学習の主なアプローチは、以下の2つに分類されます。
教師あり学習と教師なし学習です。

モデルの管理は、長期間にわたって精度を維持するために不可欠です。新しいデータを考慮して定期的にモデルを再トレーニングすることが重要であるため、モデルの開発、実装、導入、管理が継続的なサイクルを形成する必要があります。

この管理は、バラバラのオープン・ソース・ツールでは難しい場合があります。エンド・ツー・エンドのデータ・サイエンス・プラットフォームを使用することで、プロセスが途切れないようにすることができます。また、このプラットフォームでは、モデルのパフォーマンスが低下し始めると直ちに適切なチームに通知され、迅速な対応を取ることができます。

自然言語処理 (NLP)は、構造化されていないテキスト・データをコンピューターが分析できるようにすることに焦点を当てたAIの一分野です。一般的なユースケースとしては、音声認識、自然言語理解、感情分析などが含まれます。

ニューラル・ネットワークは、多くの機械学習アルゴリズムの間で複雑な相互作用を可能にするトレーニング・モデルのフレームワークを提供し、最適なモデルの特定に役立ちます。

人間や他の動物の脳にある神経細胞が相互に接続された構造から、人工的なニューラルネットワークの構造を着想しました。人工ニューロンを層で接続します。データは入力層から一つ以上の隠れ層を経て出力層まで構造を横断します。この横断の間、数学関数がデータを予測に変換し、その精度を評価することができます。

オープン・ソース・ソフトウェアは統計モデリングや機械学習の多くの分野で、ますます支配的なパラダイムになりつつあります。R、Python、Scalaなどの言語、Apache HadoopやSparkなどのビッグ・データ・アーキテクチャー、TensorFlowやSpark MLlibなどの機械学習フレームワークは、予測分析やデータ・サイエンスの世界では重要な存在です。

オープン・ソースのフレームワークは、モデル開発やトレーニングなど、データ・サイエンス・プロセスの特定の部分を対象とした高品質なツールの開発に重点を置いていることが多いです。その結果、すべてのツールを統合し、一貫したワークフローを実現する責任をエンド・ユーザーに負わせることが多くなっています。この作業は予測分析を企業全体に拡大し、ビジネス・プロセスにAIを組み込もうとするときに問題となることがあります。

予測分析は、ヒストリカル・データを使って特定の領域や問題をモデル化し、過去に特定の結果をもたらした重要な要因を切り分けます。このプロセスで構築されたモデルは、新しいデータから将来起こる可能性のある結果を予測します。

予測分析は、古典的な統計モデリングから機械学習アルゴリズムまで、幅広い技術を包含しています。

予測モデルは、データベースの記録、テキスト・サンプル、画像などのデータの一部を示す入力を、出力または予測に割り当てるアルゴリズムです。出力は通常、数値や割合などの連続変数か、「はい」や「いいえ」などの離散的なカテゴリーです。予測モデルには大きく分けて2つのタイプがあります。回帰モデルと分類モデルです。

データの**準備**は、データ・サイエンス・プロセスの最初のステップの1つです。ほとんどのプロジェクトでは、まずデータ・セットを洗練させ、詳細な分析に耐える品質を確保することから始めます。

多くの場合、ソース・データをクリーニングし、モデル化や分析がしやすいフォーマットに変換する必要があるかもしれません。機械学習モデルを構築する場合、教師あり学習で使用するためにデータを手動でラベル付けするための投資も必要になる場合があります。

回帰モデルは、複数の変数を含むデータ・セットがあり、それらの間の関係を分析したい場合に役立ちます。具体的には、ある特定の変数が、他の変数が変更されたときに、どのように変化するかを回帰モデルで明らかにすることができます。

線形回帰は、値やスコアを予測するために使用することができます。例えば、ある週に売れた製品の在庫保管単位数や、お客様がアカウントを閉鎖するリスクの割合などです。

統計的モデリングは、データ・セットに関する確率的仮定に基づくモデルの作成を含む数学の領域です。業務部門では、統計モデルを用いてデータ・セットの重要な特徴を分析し、データの分類や予測の生成に使用できる相関を特定してきました。

教師あり学習は、すでに正しくラベル付けされたデータ・セットを用いて、機械学習モデルをトレーニングする方法です。モデルは出力変数(通常はカテゴリーまたは値)を生成するため、出力とラベル付けされた入力を比較することでその精度を容易に評価することができます。線形回帰、ランダム・フォレスト、サポート・ベクター・マシンはいずれも教師あり学習アルゴリズムの代表例であり、ほとんどの予測モデルはこれらの技術で構築されています。

予測モデルのテストはトレーニングとともに、AI処理におけるデータの精度を判断するために不可欠です。予測モデルは、精度を高めるために継続的にテストする必要があります。モデルが失敗だった場合、アナリストは根本的な原因を特定し、モデルを改善するために再トレーニングとテストを行う必要があります。

テキスト分析は、言語規則、自然言語処理、機械学習を用いて非構造化コンテンツを測定します。これは、人間の脳が行っているのと同様ですが、より高速にデータを確認するプロセスです。テキスト分析では、全データの約90%を占める非構造化コンテンツから、より多くの洞察と発見が可能になります。

予測 モデルのトレーニングは、機械学習、ディープ・ラーニング、その他のAIプロセスにおいて、どのデータが役立つかを判断するための重要な要素です。正確な予測を行うようにトレーニングされたモデルは、リアルタイムのデータをスコアリングするために使用することができます。モデルは、変化する行動パターンを調整するために定期的に再トレーニングする必要があります。

教師なし学習は、ラベル付けされていないデータを用いて機械学習モデルをトレーニングする方法です。その目的は通常、データ内の興味深いパターンや構造をモデル化し、強調することです。クラスタリングやアソシエーション問題は、非構造化学習の一般的な領域です。例えば、お客様をセグメント化したり、お客様間の類似性を特定するための興味深い新しい方法を見つけることです。

可視化とは、データをグラフィカルに表現するプロセスで、多くの場合、チャートや図表を使用します。人間はデータを理解するために視覚化する必要があります。このプロセスは、結果を業務部門のステークホルダーに提示するときにも、プロジェクトの初期段階で新しいデータセットを探索するときにも重要です。

予測分析プラットフォームでは、可視化ツールを備えた直感的なグラフィカル・インターフェースを提供する必要があります。これらの機能により、大規模なデータ・セットでも数分で意味を理解することができますようになります。



© Copyright IBM Corporation 2019

日本アイ・ビー・エム株式会社
〒103-8510 東京都中央区日本橋箱崎町19-21

米国で制作
2019年3月

IBM、IBMロゴ、ibm.com、IBM Cloud、IBM SPSS ModelerおよびIBM Watsonは、世界中の多くの国々で登録された International Business Machines Corp.の商標です。その他の製品名およびサービス名は、IBMまたは他社の商標である可能性があります。その他の製品名およびサービス名は、IBMまたは他社の商標である可能性があります。IBMの登録商標の現在のリストは、Webページ「著作権および登録商標情報」www.ibm.com/legal/copytrade.shtmlでご確認いただけます。

本書は最初の発行日時点における最新情報を記載しており、IBMにより予告なしに変更される場合があります。IBMが事業を展開しているすべての国で、すべての製品が利用できるわけではありません。

本ドキュメントで説明するパフォーマンス・データは、特定の稼働条件下で得られたものです。実際の結果は異なる場合があります。本書の情報は"現状のまま"で提供されるものとし、明示または黙示を問わず、商品性、特定目的への適合性、および非侵害の保証または条件を含むいかなる保証もしないものとします。IBM製品は、IBM所定の契約書の条項に基づき保証されます。

お客様は自己の責任で関連法規を順守しなければならないものとします。IBMは法律上の助言を提供することはなく、また、IBMのサービスまたは製品が、お客様においていかなる法を順守していることの裏付けとなることを表明し、保証するものでもありません。

Statement of Good Security Practices:

IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.