

機械学習を促進する: InterSystems IntegratedML

技術概要



開発・配備が容易な機械学習ソリューションで 人工知能の人材争奪戦に勝利する

技術概要について

Forrester Research によると、「98%の企業が、収集したデータから洞察を得る上での課題に直面しており、その主な原因は社内の専門知識の不足によるものである」とのことです¹。お客様の組織では、現在はこのような洞察を得られなくても競争優位性を維持できているかもしれませんが、デジタル変革に向けた急速な変化はすぐにお客様のビジネスに影響を及ぼすことでしょう。

そのため、現在ではあらゆる業界の組織が、データを使用し、人工知能(AI)と機械学習(ML)を統合して、ビジネスを向上させることに注力しています。MLによって、業務を改善、自動化し、イベントと行動を予測して、これらの予測に基づいて処方的なプログラムによる処理をプロアクティブに実行できるようになります。

たとえば、お客様の組織で、MLと予測分析を使用して消費者の行動と好みを的確に把握することで、新たな購入者を特定してターゲット設定したり、セールを実施するのに最適なタイミングを特定したりすることができます。医療機関の場合は、コード化された診断結果や、患者の入院、転院、および退院データを分析して、再入院率を減らすことができます。簡単に言うと、MLは時間とリソースを節約し、予測の精度を向上させ、より適切な意思決定によってより良い成果を生み出すことができます。

これは素晴らしいことだと思われるかもしれませんが、1つ問題があります。MLモデルの開発は難しく、専門知識を必要としますが、そうした専門知識を持つ人材は不足しており、人件費もかさみます。

幸いにも、業界の専門家は新しいツールを開発してMLをより活用しやすいものにと取り組んでいます。AutoMLとIntegratedMLもその一例です。

この技術概要をお読みになることで、以下のことをご理解いただけるようになります。

- MLがお客様のビジネスの成功に不可欠であるのはなぜか
- 組織でMLを活用する上で人材不足がいかに障壁となっているか
- AutoMLとは何か、そしてAIの人材争奪戦を勝ち抜くためにいかに役立つか
- InterSystems IntegratedML™とは何か、そしてどのようなことを意図して設計されているか
 - 既存のソフトウェア開発者がMLモデルとML対応アプリケーションを開発できるようにする
 - トレーニングされたデータサイエンティストの生産性を向上させる
 - 業務プロセスと分析プロセスを合理化して、カスタマーエクスペリエンス、業務効率、生産性を向上させる
 - 予測の精度を高め、より優れたビジネスの成果を生み出して、競合他社との差別化を図る

¹ Forrester Opportunity Snapshot(2019)。Data Insights Are Key to Differentiated Customer Experience: A Unified Data Analytics Platform Enables Timely and Contextually Relevant CX(カスタマーエクスペリエンスを差別化するための鍵はデータの洞察: 統合データ分析プラットフォームによりタイムリーでコンテキストの関連性が高いCXを実現)

この技術概要は、ML チームの生産性を向上させたいと考えているか、ML 専門家を雇うことなく ML を活用しようとしている、基幹業務のエグゼクティブ、マネージャー、IT 専門家を対象としています。

機械学習: その価値

AI を応用した ML では、経験と推論に基づいてデータに関する学習を行うようマシンをトレーニングします。特にそのようにプログラムされていなくても、継続的に結果が改善されていきます。

ML は広範なデータを分析して、分析や業務に関するさまざまな要件を満たすために使用されるモデルを開発することができます。オフラインでは、ML モデルはビジネスユーザによる顧客の行動の把握と、効率性の問題への対処を支援します。これらは応用の一例に過ぎません。ML は、オンラインで、または業務フローで配備される場合には、目に見える際立った成果を生み出すことができます。たとえば、顧客が Web サイトを閲覧しているときにお勧めの製品やサービスを表示したり、サプライヤーが納入できないおそれが大きい場合、セール前にあらかじめアラートを送信したり、取引を承認する前にその取引が不正である可能性があるかどうかを特定したりすることができます。販売とマーケティング、研究開発、法務、人事、カスタマーサポート、製品開発、さらに財務にいたるまで、組織のあらゆる部門の業務が ML の恩恵を受けることができます。実際、ML はほぼすべての業種に価値をもたらしており、より多くの組織が ML を導入するにつれて、必ずユビキタス化していきます。

あなたも、日々の生活で既に ML を経験していることでしょう。Amazon Alexa や Apple の Siri のような仮想パーソナルアシスタントから、スパムメールフィルターやマルウェア検出プログラム、Facebook での新しい友だちやグループの提案、オンラインカスタマーサポートを提供するチャットボット、自動運転を行うスマートカーにいたるまで、あらゆる場面で。

機械学習: その課題

機械学習には多くのメリットがあるので、もっと多くの企業で採用されてもいいはずです。それを阻んでいるものは何でしょうか。主な原因は、ML は使いこなすのが難しく、高度な専門知識が必要であるということです。

ML は、その理論、技術、手法、ツールについて理解している専門家を必要とします。今日、このような専門家は極めて少なく、引く手あまたです。アメリカ合衆国労働省労働統計局の最新のデータによると、米国のデータサイエンティストの数は合計 32,000 人未満です²。AI 専門家とデータサイエンティストの不足にさらに追い打ちをかけるように、そうした人材の多くは既に、Amazon、Facebook、Google、Microsoft といったデジタルジャイアントに高給で雇われています。そのため、一般の組織が既に乏しいリソースを巡って張り合うことは難しくなっています。

² <https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-research-scientists.htm>

今日の AutoML ツールの多くには
限界があります。それらは
ML モデルを作成できても、
リアルタイムのビジネスプロセス内で
モデルを実行する機能は
提供していません。
それが、競合製品と InterSystems
IntegratedML とで
大きく異なっている点の 1 つです。

AutoML: AI の人材争奪戦に打ち勝つ

AutoML (Automated Machine Learning: 自動化された機械学習) は、現在の ML に関する人材の活動範囲を拡大しようとしている組織や、ML の取り組みを開始したばかりの組織のための、成長著しい新たなテクノロジーです。

AutoML は、ML モデルの作成を自動化および簡素化するという、データサイエンスの比較的新しいアプローチです。特徴量エンジニアリングを実行し、生データを ML モデルに適した形式に変換するプロセスを自動化します。モデルの選択、トレーニング、結果の分析を自動化し、さまざまなパラメータを使用して異なる ML アルゴリズムをテストして、特定の問題に対する最も正確なモデルを作成します。データサイエンティストのチームが存在する組織の場合、これによって ML モデルの構築に使用される手動および試行錯誤のプロセスの多くが自動化され、データサイエンティストの生産性が向上し、時間と労力を節約することができます。

スタッフに ML の専門家がまだいない場合は、特徴量エンジニアリングを実行したりモデルの作成やトレーニングを行ったりするのは難しい場合があります。しかし今後は、AutoML によって、組織ではデータサイエンティストが必ずしもいなくても、有用な ML モデルを作成できるようになります。まずはシンプルなユースケースと AutoML を使用することから始め、それと同時に、分析と ML 開発プロセスにおけるより多くの作業をこなせるよう開発者をトレーニングすることができます。

ただし、今日の AutoML ツールの多くには限界があります。それらは ML モデルを作成できても、リアルタイムのビジネスプロセス内でモデルを実行する機能は提供していません。それが、競合製品と InterSystems IntegratedML とで大きく異なっている点の 1 つです。

InterSystems IntegratedML: AutoML をより強力に

InterSystems IntegratedML は、完全なデータ管理ソフトウェア環境である、InterSystems IRIS® Data Platform に組み込まれた機能です。IntegratedML は、従来の AutoML のすべての機能と特長を備えています。また、InterSystems IRIS に組み込まれているため、モデルやデータを抽出または移動することなく、リアルタイムのイベントやトランザクションに応じてこれらのモデルを動的かつシームレスに実行する高度なアプリケーションを開発し、配備することができます。

たとえば、ある銀行ではクレジットカードを発行しており、カードでの個々の取引を承認する前に、不正使用のおそれがないかどうかを確認する必要があります。そのため、InterSystems IRIS を使用して開発した、リアルタイムでの処理が可能な高パフォーマンスのクレジットカードアプリケーションを実行しています。このプラットフォームには、すべての顧客とクレジットカード取引の人口統計データと財務データがすべて格納されています。このアプリケーションには、各クレジットカード取引について、それらの取引が不正であるか有効であるかをはじめとする、数百のデータ要素を保存できます。

IntegratedML を使用すれば、銀行の既存のアプリケーション開発者は、ML モデルを自動的に作成し、過去の取引に基づいて高リスクの取引を識別することができます。開発者が目的のフィールド(「is_fraudulent」など)を選択するだけで、IntegratedML が最適なモデルとパラメータを作成してくれます。

ただし、従来の AutoML とは異なり、InterSystems IntegratedML ベースのモデルは、クレジットカードアプリケーションにシームレスに統合し、各入金取引に対してリアルタイムで実行することができます。そして、モデルによって不正取引のおそれ大きいことが確認された場合、アプリケーションはプログラムによる適切な操作を実行することができます（取引を中止してカードの所有者に電話をかけたりメッセージを送信したりするなど）。

また、IntegratedML によって、アプリケーションが実稼働環境で実行されて新たなデータが生成されても、モデルをより容易に最新の状態に保てるようになります。クレジットカード詐欺の場合、アプリケーションが詐欺の 1 つの手法を検出して防いでも、犯罪者は必ず新たな手法を編み出そうとするでしょう。最新のデータを含むすべてのデータがデータプラットフォーム内に保存されているため、手動でデータを抽出して異なる環境に移動する必要はありません。最新のデータを使用し、後れを取ることなく新しい攻撃パターンを検出して防止することで、継続的にモデルを改良していくことができます。

InterSystems IRIS と IntegratedML を使用することで、リアルタイムのイベントに対してインテリジェントで処方的なプログラムによる処理を実行するアプリケーションを開発し、重要な競争優位性とビジネス上のメリットを得ることができます。それによって、新しい製品やサービスを市場にいち早く投入し、新しいイニシアチブに率先して取り組んで、顧客の行動における変化に迅速に対応することができるようになります。

IntegratedML: 人件費を削減して生産性を向上させる

IntegratedML によって、開発者は ML の知識がほとんどまたはまったくなくても、SQL を使用して高度な ML モデルを開発できるようになります。

これは、お客様がデータサイエンティストを一切雇わないよう提案するものではありません。データサイエンティストのチームを擁する大規模な企業の場合は、IntegratedML によってデータエンジニアとデータサイエンティストの時間を大幅に削減することができます。たとえば、2018 年に [Kaggle ML and Data Science](https://www.kaggle.com/headsortails/what-we-do-in-the-kernels-a-kaggle-survey-story) が実施した調査³によると、データサイエンティストはほぼ 40%の時間をデータの収集とクリーニングに費やしていました⁴（下の図 1 を参照）。データの準備と特徴量エンジニアリングに IntegratedML を使用することで、こうした作業からデータサイエンティストを解放して、モデルの最適化などのより重要で価値の高い作業に集中してもらえるようになります。

IntegratedML によって、開発者は ML の知識がほとんどまたはまったくなくても、SQL を使用して高度な ML モデルを開発できるようになります。お客様の組織に既にデータサイエンティストのチームが存在する場合は、IntegratedML によりチームの時間と労力を大幅に節減できます。

³ <https://www.kaggle.com/headsortails/what-we-do-in-the-kernels-a-kaggle-survey-story>

⁴ <https://businessoverbroadway.com/2019/02/19/how-do-data-professionals-spend-their-time-on-data-science-projects/>

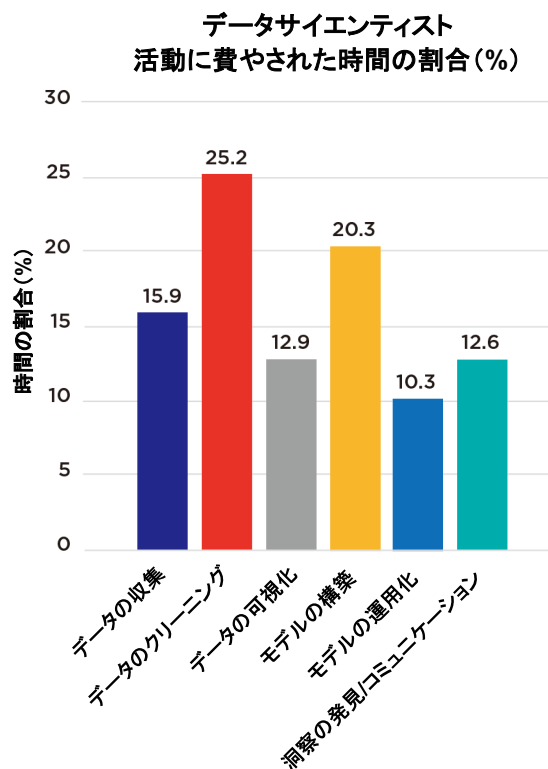


図 1: データサイエンティストが費やした時間の割合³

ML を使用し始めたばかりの組織の場合も、普段ビジネスアプリケーションを構築してデータのことを熟知しているソフトウェア開発者やアナリストであれば、InterSystems IntegratedML を使用して ML を自由に活用することができます。IntegratedML は、最適なモデルの特定、パラメータの設定、モデルの構築やトレーニングといった基本的な作業を自動化します。また、ML モデルの実稼働アプリケーションへの統合プロセスも短縮します。開発者は、IntegratedML により精通してプロセスや達成される成果について理解できるようになれば、オプションのパラメータを変更して、値を自分で設定することができます。また、IntegratedML によって、データ収集や特徴量エンジニアリング/選別ではなく、実際のモデル最適化に時間を費やすことができるようになるため、データサイエンティストの生産性も向上します。

InterSystems Integrated ML: しゅみ

IntegratedML によって、モデルのトレーニング(ソースデータからの適切な入力特徴の識別など)、モデルのパラメータの調整、実行は、ごく少数の SQL コマンドを使ってすべて完了できます。

```
CREATE MODEL WillSurvive PREDICTING (Survived) FROM Titanic
```

CREATE MODEL コマンドは、機械学習モデルのメタデータをセットアップします。開発者はモデルの名前 (WillSurvive)、予測されるターゲットフィールド (Survived)、ターゲットフィールドとすべてのモデルの入力フィールドの入力元となるデータセット (Titanic) を指定します。FROM 構文はごく一般的に使用されており、任意のサブクエリの式を指定できます。このデータセットに関連付けられたメタデータを使用して、ターゲットと入力フィールドのデータ型を推測して、解決するモデルの問題を完全に定義することもできます。

```
TRAIN MODEL WillSurvive FROM Titanic
```

TRAIN MODEL コマンドは、トレーニングに使用されるデータを指定して AutoML エンジンを実行します。このエンジンは一連のリレーショナルデータを入力として受け取ります。FROM 構文は一般的に使用されているため、異なるデータセットを使用して同じモデルを複数回トレーニングできます。たとえば、異なる顧客セグメントでマーケティングキャンペーンモデルをトレーニングしたり、新しいトレーニングデータが使用可能になったときにモデルを定期的に再トレーニングしたりすることができます。

AutoML エンジンは、必要な機械学習のタスクをすべて自動的に処理します。選択したデータから関連する候補となる特徴を特定し、データと問題の定義に基づいて適切なモデルタイプを検討するほか、ハイパーパラメータをチューニングして 1 つ以上の実行可能なモデルを作成します。

開発者は、InterSystems AutoML、H₂O、DataRobot Enterprise AI Platform などのさまざまな AutoML エンジンから選択できます。すべての AutoML エンジンオプションは InterSystems IRIS にシームレスに統合されているため、開発者はそれらのオプションのことを意識せずに済みます。

```
SELECT PREDICT(WillSurvive) As Predicted FROM Titanic  
SELECT PROBABILITY(WillSurvive FOR 1) FROM Titanic
```

開発者はさまざまな
AutoML エンジンから選択できます。

トレーニングが完了すると、モデルは 2 つのスカラー関数 PREDICT()と PROBABILITY()のうちどちらかを使って結果を返します。PREDICT()は、トレーニング済みモデルで指定されているとおり、特定の列に対して、最も可能性の高い値または推測される値を返します。分類の問題の場合、PROBABILITY()はトレーニング済みモデルで計算された確率を返し、モデルのターゲットフィールドがユーザ定義の値と同じになります。これらのシンプルなスカラー関数は、クエリ内の任意の場所で使用できるほか、他のフィールドや関数と自由に組み合わせて使用することもできます。IntegratedML が提供する革新的な主要機能の 1 つは、特定のクエリコンテキスト内の使用可能なフィールドを、モデルを実行するために必要な入力フィールドに透過的にマッピングできることです。

IntegratedML は、モデルの作成またはトレーニングに使用される特定のテーブルまたはクエリ以外のデータソースへのマッピングなどの機能により、開発者にさらなる柔軟性をもたらします。以下にその例を示します。

```
SELECT Name, PREDICT(WillSurvive WITH Sex = Geschlecht,  
Age = DATEDIFF(year, NOW(), Geburtsdatum), Fare =  
TicketPreise, Cabin = Kabine) FROM Hindenburg
```

ほとんどの AutoML ソリューションは、外部のデータプラットフォームやアプリケーションとの連携が緊密ではない低スループットのスタンドアロン環境で動作しますが、IntegratedML はそうではありません。InterSystems IRIS Data Platform 内でシームレスに動作し、ML モデルのトレーニングと実行を短縮・簡素化します。また、データやモデルを移動することなく、ML モデルを IRIS アプリケーション内でシームレスに統合することができます。ML モデルがこうして運用可能になることで、ビジネスアプリケーションでの ML の迅速な導入を妨げている最大の障害の 1 つが取り除かれることになります。

InterSystems IRIS Data Platform

インターシステムズのテクノロジーは、世界中のさまざまな業種で実装されており、その数は 15 万件に達しています。InterSystems IRIS Data Platform は、リアルタイムのデータ駆動型アプリケーションの開発を迅速化および簡素化するために構築された、包括的なデータ管理ソフトウェアプラットフォームです。InterSystems IRIS によって、開発者はビジネスインテリジェンス、AI、ML、自然言語処理、予測分析といった高度な分析機能を、リアルタイムのミッションクリティカルなビジネスプロセスに組み込むことができます。組み込みの高パフォーマンスのトランザクショナル分析データベースエンジンが、非常に大規模な運用ワークロードと分析ワークロードの両方を同時にサポートします。

InterSystems IRIS は、組み込みの ML 開発およびランタイム機能のほか、以下の機能も備えています。

データおよびアプリケーション統合 - InterSystems IRIS には、データをクリーニング、変換、正規化し、高度な統合をサポートするための、統合・相互運用機能が揃っています。広範なパッケージアプリケーション、データベース、業界標準、プロトコル、テクノロジーを使用し、即座に接続してデータを変換することができるので、データの統合・分析と予測的・処方的モデルの構築が容易になります。

さらに、SQL クエリ、予測分析、ML、自然言語処理(NLP)などの分析処理を、分散したデータソースとアプリケーションを結び付ける複合ビジネスプロセスに組み込むことができます。このような複合プロセスによって、アプリケーションのパフォーマンスに影響を与えることなく業務を合理化してアラートをトリガすることができます。

拡張性 - InterSystems IRIS は垂直方向と水平方向に拡張可能で、リソースの効率が高いため、極めて高いデータ取り込み率、大量の分析ワークロードの処理、多数の業務プロセスの同時実行、そして非常に大規模なデータセットをコスト効率の高い方法で処理、保存、分析する機能をもつアプリケーションに理想的です。

レポートと追跡可能性 - すべてのデータ(移動中のデータ、メタデータ、長時間実行されている非同期トランザクションに関連付けられたデータなど)は組み込みのデータベースに自動的に保存され、リアルタイムのレポートと分析に使用できます。視覚的な追跡機能により、統合とプロセスの動作をより容易に可視化して診断することができますようになります。

グラフィカルな開発 - グラフィカルなローコードのツールにより、開発者はプロセス、変換、ルール、ワークフローを視覚的に図表化できるため、コーディングではなくシステム間の論理的なやり取りに集中することができます。グラフィカルなモデルにより、基幹業務と IT とのコラボレーションが推進され、組織で新しいソリューションを開発したり、既存のアプリケーションをより迅速に変更したりできるようになります。

配備 - InterSystems IRIS は、パブリッククラウド、プライベートクラウド、オンプレミス、ハイブリッドの主要な配備オプションをすべて含む、広範な配備オプションを提供します。

InterSystems IRIS と IntegratedML は、リアルタイムのパーソナライズされたエクスペリエンスによる顧客満足度の向上、患者の臨床転帰の改善、障害発生前の保守ニーズのプロアクティブな予測、またはリアルタイムでの詐欺の検出と防止といった多くの目標をお客様が達成できるよう支援します。

まとめ

機械学習はこれからのトレンドであり、企業が競争を勝ち抜くには、今後機械学習を活用する必要があります。残念ながら、データサイエンティストは不足しており、支払われる給料も高騰しているため、大規模な組織が ML の活用範囲を広げたり、小規模な組織が ML を取り入れたりすることが困難になっています。AutoML のようなイノベーションは有用ですが、AutoML だけでは不十分です。

InterSystems IntegratedML は、直感的な SQL インタフェースを通じて使用できる、包括的なデータプラットフォームに完全に統合された、高度な AutoML 機能を提供します。IntegratedML によって、データやモデルを移動させることなく、データサイエンティストのスタッフも必要とせずに、リアルタイムのミッションクリティカルなアプリケーションで ML モデルを容易に配備できるようになります。InterSystems IRIS と IntegratedML の統合により、お客様は改善の好循環を生み出し、最新の生産データに後れを取ることなく対処して継続的に ML モデルを改良できるようになります。

お客様の組織にデータサイエンティストのチームが存在する場合は、IntegratedML によってチームの生産性を向上させることができます。

AI への取り組みを始めたばかりのお客様も、IntegratedML によって、人件費をかけて ML 専門家を雇うことなく今すぐ ML を活用できるようになります。

いずれの場合も、IntegratedML は以下を実現します。

- ML モデルの作成を迅速化および簡素化する
- インテリジェントなプログラムによる処理をリアルタイムで実行する
- プロセスを合理化して、カスタマーエクスペリエンス、業務効率、生産性を向上させる
- 予測の精度を向上させ、短期間でより優れたビジネスの成果を達成し、競争に打ち勝つ
- より少ないリソースでよりスマートなアプリをより迅速かつ容易に開発する
- AI の人材争奪戦を勝ち抜く

InterSystems IRIS および IntegratedML の詳細については、[InterSystems.com/jp/IRIS](https://inter-systems.com/jp/IRIS) にアクセスしてください。

インターシステムズについて

インターシステムズは、世界で最も大切なアプリケーションを支えるエンジンを提供しています。人々の生命や暮らしにかかわる医療、金融、官公庁、その他の分野で、大切なことを支援する力となっています。インターシステムズは、1978 年に設立された戦略的テクノロジープロバイダです。米国マサチューセッツ州ケンブリッジに本社を置き、世界中に拠点を持つ株式非公開企業です。その製品は世界 80 ヶ国以上の数百万ユーザによって、日々利用されています。詳細については、InterSystems.com/jp/をご覧ください。

The power behind what matters.

大切なことを、支える力に

