
実務環境での機械学習の 可能性を最大限に引き出す

Cloudera Machine Learningプラットフォームにおける
MLOps機能とその効果

CLOUDERA

FEATURING
FORRESTER®

機械学習における 拡張性の問題を克服

機械学習モデルを大規模に導入し、それを維持していくことは、多くの企業にとって大きな課題になっています。

機械学習モデルの導入プロセスは、本質的に実験的なものになります。しかしどんな企業の場合でも、機械学習の目的は、ビジネスに新しい価値を生み出すことにあります。このため、機械学習モデルは最終的に実務環境に移行する必要があり、実験は方程式の半分にすぎません。

ほとんどの企業の場合、ラボの環境から決して十分整った状況とは言えない実務環境に一足飛びに移行する形になります。しかし、実務環境に移行するモデルが少数であっても、その管理、監視、および管理に関わる労力は通常と変わりません。このため、数百あるいは数千規模の機械学習モデルへの拡張は、気が遠くなるような作業を伴うことになります。

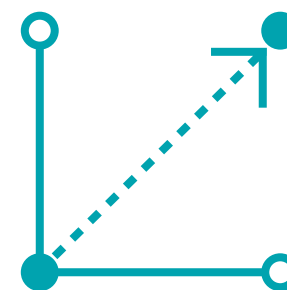
そこで、機械学習運用 (MLOps) の登場となります。MLOpsには、定義済みのプロセスやベストプラクティス、機械学習チームが実験から実務移行するための各種機能などが含まれ、大規模なモデルの管理、導入、監視、維持、統制を支援します。

これらのMLOps機能をカプセル化して取り込んだCloudera Machine Learning (CML) with MLOpsによって、導入企業は、成功裏に機械学習のライフサイクル全体を運用できるようになります。担当チームは、1つの包括的なプラットフォーム上で、データパイプラインをソースから直接シームレスに利用したり、あらゆる言語やライブラリを使ったモデルの実験やトレーニング、さらにモデルを実務に導入するなど、モデルや基礎となるデータを、透過的にコントロールしながら管理することができます。

Forrester ResearchのアナリストであるDr. Kjell Carlsson氏は、「MLOps in Enterprise Production Machine Learning」の中で、MLOpsに対して企業が共通に抱く質問について答えています。本書では、持続的で拡張性に優れたエンタープライズ機械学習の実務環境において、MLOpsが果たす重要な役割と、CML with MLOpsによって、いかに企業向け機械学習の状況を変革することができるかという点について、詳細にご説明します。

出典

Artificial Intelligence: The Next Digital Frontier, McKinsey Global Institute, 2017年6月





88%のエンタープライズ
機械学習プロジェクトは、
未だ実験段階にあります。

MLOpsが不可欠となる理由

MLOpsは、データサイエンティストや機械学習エンジニアのシームレスなコラボレーションや、迅速な反復、さらに実務環境に対する機械学習モデルの効率的な移行を支援します。

実務環境に対する機械学習モデルの迅速な移行だけでなく、対象モデルの容易な管理や保守を可能にするなど、MLOpsによって正確で拡張性に優れた機械学習モデルから、確実にビジネス価値を引き出せるようになります。

CML with MLOpsは、正にこのような役割を果たします。CML with MLOpsは、機械学習モデルの構築、トレーニング、導入、運用、拡張を可能にする共通のプラットフォームを提供します。さらに、機械学習チームと業務ユーザーのコラボレーションを可能にします。



「多くの場合、業務側の関係者は、実際に導入が終わるまで、何が本当に必要かを正確には理解していません。MLOpsによって業務効率の向上に貢献するソリューションを迅速かつ繰り返し提供することで、イノベーションを加速し、さらに付加価値を高めていくことができます」

MLOps in Enterprise Production Machine Learning,
Forrester, 2020年5月





可視化およびコントロール

CML with MLOpsは、Cloudera Enterprise Data Cloudを利用するチーム全体に対して、反復可能なライフサイクル全体の監視とリネージトラッキング機能を提供します。このインフラストラクチャーによって、厳格なセキュリティやガバナンス対応、コントロール管理を行い、IT環境に対するデータのソーシングから実務環境に至るまで、同一のデータを共有しながら、様々な分析機能間の連携を図ることができます。

機械学習のライフサイクル全体を可視化することで、チーム間で発生するシステムのサイロ化を防ぎ、効率的なワークフローを実現して、機械学習が「ブラックボックス化」する状況を排除することができます。担当チームは、可視化とコントロールの強化によって、モデルのガバナンスルールを設定し、業務環境におけるミッションクリティカルな変更に対応したり、包括的なモデル運用を実施することで、確実な成果を持続的に提供できるようになります。



パワフルで簡単

データサイエンスチームは、CML with MLOpsを使用することで、必要なリソースやIDE (統合開発環境)、ライブラリへのアクセスにおいて、これまで障壁となっていた、IT環境やサイロ化したポイントソリューションの影響を緩和することができます。

CML with MLOpsは、ハイブリッドな形でのシステムの導入をサポートし、Microsoft AzureやAmazon Web Servicesといった、複数のクラウドサービスを横断する形で利用できるようにします。一旦モデルを導入したら、効率的なツールを使用して、実務環境にあるこれらのモデルを適切に維持し、業務全体のユースケースに対応できるよう運用を拡大していくことができます。またCML with MLOpsによって、モデルのメタデータや用語、さらに予測のトラッキングといった共通の基準を導入し、企業全体でコスト効果に優れた、反復可能で透過的な導入、監視、ガバナンス対応を実現することができます。



成果および信頼性

大規模なワークフローを実務に組み込み、機械学習プロジェクトを完全な透明性を持って管理することによって、業務プロセスを自動化し、予測分析を活用して、新しい製品やサービスを生み出すことができます。

企業のリーダーは、透過的かつ説明可能で、統制の取れたCML with MLOpsによって、業務上の意思決定に向け、確信をもって機械学習モデルや予測機能を活用することができます。CML with MLOpsは、業務チームに対して、基礎データをより良く理解するためのビジュアルインターフェースを提供し、データサイエンティストとの間でのコミュニケーションやスキルのギャップを埋めます。

出展: MLOps in Enterprise Production Machine Learning, Forrester, 2020年5月

実務環境で機械学習を大規模に展開するための主要な機能

機械学習のライフサイクルを実務環境に移すためには、オープンスタンダード主導の正確な監視、モデルのカタログ化、さらにデータソースからビジネスユースケースに至る継続的なガバナンス機能を提供できる、MLOpsを備えた包括的なMLプラットフォームが不可欠となります。

実務環境に対して機械学習を大規模に展開するためには、多くの機能が必要となりますが、Forresterでは、最も重要となる機能を、導入と提供、監視、そしてコラボレーションとガバナンスという、実務における機械学習の3つの主要カテゴリーに分類しています。



3X

「急成長している企業は、継続的にモデルを実務環境へ導入し、監視および維持できる可能性が3倍高くなっています」

MLOps in Enterprise Production Machine Learning, Forrester, 2020年5月



導入および提供

CML with MLOpsによって、モデルの導入を効率化し、ストレスのない予測可能な対応を可能にします。また、企業における機械学習のワークロードに特化した設計となっているため、IT管理者やデータサイエンティストは、ダウンタイムのない高可用性に優れた機械学習モデルを、大規模に提供することができます。

データサイエンティストは、ノートブックや、オープンソースツール、フレームワークを含め、IDEを自由に選択して機械学習モデルを構築し、パッケージングの上、展開することができます。つまり、ユーザーは、使い慣れていないツールやフレームを使用せざるを得ないという状況に陥ることがないのです。CML with MLOpsは、Python、R、Scaleといったエコシステムに加え、RStudio、PyCharm、Jupyter NotebookなどのIDEをサポートしています。

機械学習チームは、オンプレミスやマルチクラウド、ハイブリッドなど、モデルの稼動にあたって最も合理的な環境を選択して導入することができます。



監視

データサイエンティストや機械学習エンジニアは、CML with MLOpsを使用することで、アップタイムやスループットといった技術的な指標に加え、モデルの正確性や差異といった予測指標をトラッキングすることができるようになります。運用担当者は、運用や正確性、業務上のKPIに対するモデルのパフォーマンスを、トラッキングできます。業務環境でのミッションクリティカルな問題に対するアラートによって、業務環境にあるモデルの使用を中断することなく、潜在的な問題に対応することができます。

さらに、CPUとGPUにおけるコンピューティングリソースの自動拡張に関して、ガードレールとなる設定を行うことが可能なため、使用量を許容範囲内に収めることができ、予算の容易な管理を実現できます。



コラボレーションおよびガバナンス

CML with MLOpsは、データパイプラインからデータサイエンスのワークロード、さらに本番環境のモデルに至るまで、包括的なガバナンスを確保できるように設計されています。担当チームは、反復可能なワークフローを導入し、モデルの運用を効率化して、セキュリティおよび包括的な権限管理機能を備えたモデルを、業務環境に提供することができます。

また大規模な機械学習モデルのガバナンスをサポートするCML with MLOpsによって、対象モデルをカタログ化することも可能です。データサイエンティストや機械学習エンジニアは、優れたモデルカタログや、オープンソースのガバナンスおよびメタデータのフレームワークであるApache Atlas上に構築されたリネージ機能を、ライフサイクル全体で活用できるようになります。

MLOps - エンタープライズ機械学習における将来的な基盤

「MLOps機能を開発しなければ、間違いなくビジネスの競争力が低下するでしょう」¹

MLOps in Enterprise Production Machine Learning, Forrester, 2020年5月

Forresterは、包括的なMLOpsプラットフォームへの投資を行う企業は、機械学習の取り組みからより大きなビジネス価値を生み出すことができるようになるだろうと予測しています。該当企業は、「モデルで使用されるデータの分布、モデルが行う予測、そして結果として生じる成果をビジネス反映できるため、市場の変化をより迅速に検出して適応することができるようになります」¹。

逆にForresterは、MLOps戦略なしに機械学習を進めようとする企業は、実務へのモデルの移行に苦戦し続けるだろうと予測しています。このような企業は、機械学習の運用に必要な周回の手順を踏まないことによって、機械学習の取り組みから十分な業務成果を引き出せなくなる可能性があります。

大規模な実務用機械学習の実現

MLOpsが実験と実務の間のギャップを埋めるものだとなれば、CML with MLOpsは、革新的な実務用機械学習の大規模な実現に向け、その足元を固める役目を果たします。CML with MLOpsは、分析やモデリングに必要なデータに対するデータサイエンティストの自由なアクセスを可能にし、ワークフローとスキルのギャップを埋め、さらに実験からフルスケールの実務環境へモデルを容易に移行できるようにします。

データサイエンティストは、規制対応を確実に遵守しながら必要なリソースにアクセスすることができ、また、機械学習チームは業務ユーザーと協力しながら、企業のニーズや機械学習モデルが業務に与える影響を把握することができます。

15%

業務に対して、機械学習を成功裏に導入できた企業の収益率は、業界平均よりも15%高いことが報告されています²。

出典:

1 MLOps in Enterprise Production Machine Learning, Forrester, 2020年5月

2 Artificial Intelligence: The Next Digital Frontier, McKinsey Global Institute, 2017年6月

詳細情報

本書のトピックに関する詳細については、ウェビナー「Enabling Production MLOps at Scale—Hands on with Cloudera Machine Learning」（実務機械学習の大規模な実現 - Cloudera Machine Learning ハンズオン）でご覧頂くことができます。

Cloudera について

Clouderaは、データによって、現在は不可能なことも将来は可能になると信じています。Clouderaによって、複雑なデータを明解で実践的なインサイトに変えることができます。Clouderaは、エッジからAIに至るまで、あらゆる場所のあらゆるデータに対応するエンタープライズデータクラウドを提供しています。Clouderaはオープンソースコミュニティの絶え間ない革新に支えられ、世界最大規模の企業のデジタル・トランスフォーメーションの促進に貢献しています。

© 2020 Cloudera, Inc. All rights reserved. ClouderaおよびClouderaのロゴはCloudera Inc.の商標として米国その他の国に適用されています。その他の商標については、その商標が表す企業に帰属します。このCorporate Guideの情報は事前告知なしに変更される場合があります。

Cloudera株式会社

〒104-0031 東京都中央区京橋2-7-14 ビュレックス京橋

<https://jp.cloudera.com/>

お問い合わせ：info-jp@cloudera.com

CLouDERA