

Social Tech Talk #01

OSSを用いた、これからの社会課題解決

株式会社 日立製作所
Software CoE
OSSソリューションセンタ
伊藤 雅博

1. 自己紹介
2. 事業の現場でAIを活用するには
3. MLOpsによる機械学習システムの継続的な運用と改善
4. まとめ

- 伊藤 雅博 (いとう まさひろ)
 - 所属: 株式会社 日立製作所 Software CoE OSSソリューションセンタ
 - Qiitaアカウント: [@sigmalist](#)
- 国内外の案件におけるビッグデータ・AI分野のOSS活用を推進
 - Spark / Kafka などを活用した大規模データ処理システムの開発
 - Kubeflowなどを活用した機械学習システムの構築
- OSSの活用事例や検証結果の情報発信
 - Qiita / Think IT でのWeb記事投稿
 - オープンソースカンファレンス / ApacheCon / Open Source Summit など、国内外のイベントにおける情報発信



- **ビッグデータ/AIなどデジタル分野のソフトウェアはOSSが主流**
 - 先進企業が社内で開発したソフトウェアをOSSとして公開
 - 最先端の論文に基づいた技術を世界中の開発者が素早く実装
- **ビッグデータ分野:**
 - 大量データの分散処理を行うOSSがエコシステムを構成し、これらを組み合わせてデータパイプライン/データレイクなどを構成
 - ✓ Hadoop / Spark / Kafka など
- **AI分野:**
 - データの加工・可視化・分析を行うPythonのライブラリが次々と登場し、これらを活用してデータ分析を実施
 - ✓ NumPy / pandas / matplotlib / TensorFlow / PyTorch / scikit-learn など

事業の現場でAIを活用するには

1. 提案 (ビジネス検討)

- 顧客業務・要望の理解
- 改善方法の具体化
- 事業価値の評価

2. 実現性の検証 (PoC)

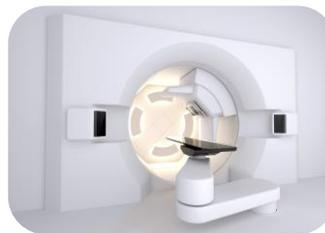
- 仮説設定
- データ分析・モデル化
- モデルの有用性を確認

3. 本番適用 (システム化)

- 業務システムとの統合
- 分析モデルの現場導入
- 分析モデルの継続的運用

今回紹介するトピック

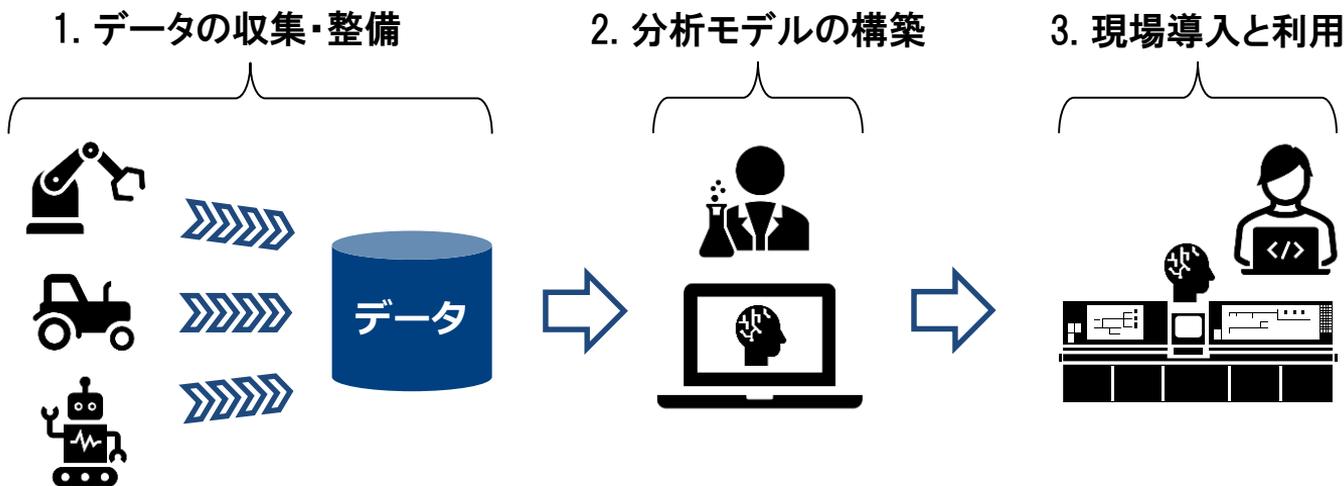
- 日立はITだけでなく幅広い社会インフラ事業を推進
 - AI技術を活用して、様々な事業における社会課題の解決に取り組み



- 事業の現場でAIを活用するには？
 - ① AIの現場導入： 既存システムとAIを統合し、現場で利用できるようにする
 - ② AIの運用監視： 導入したAIを長期間・安定的に運用できるようにする

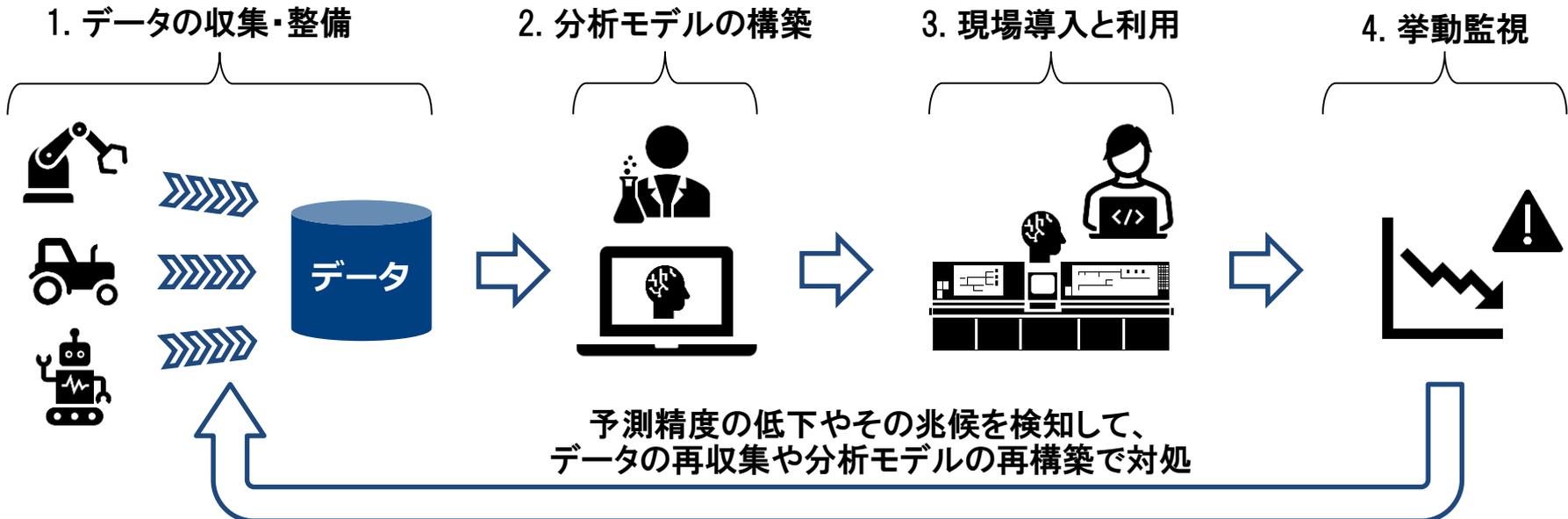
① AIの現場導入： 既存システムとAIを統合し、現場で利用する

- 例: 工場における設備故障の予兆検知
 - 既存システムからのデータ収集、分析モデルの構築、構築した分析モデルの現場導入と利用の仕組みをシステム化

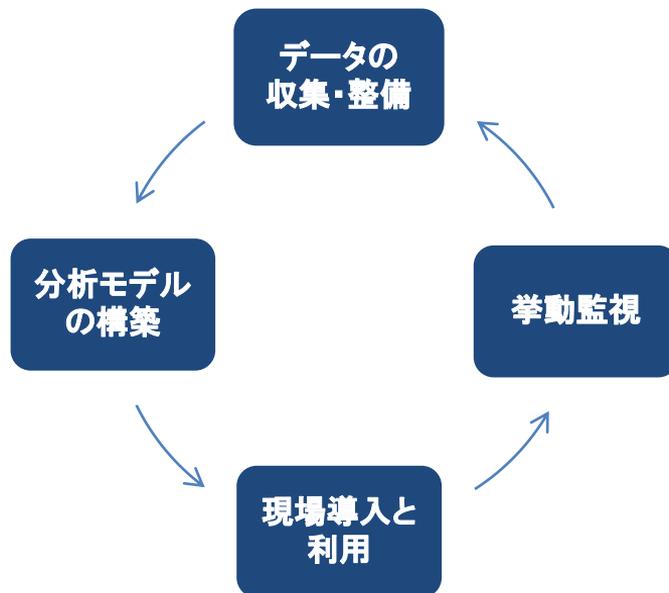


② AIの運用監視： 導入したAIを長期間・安定的に運用する

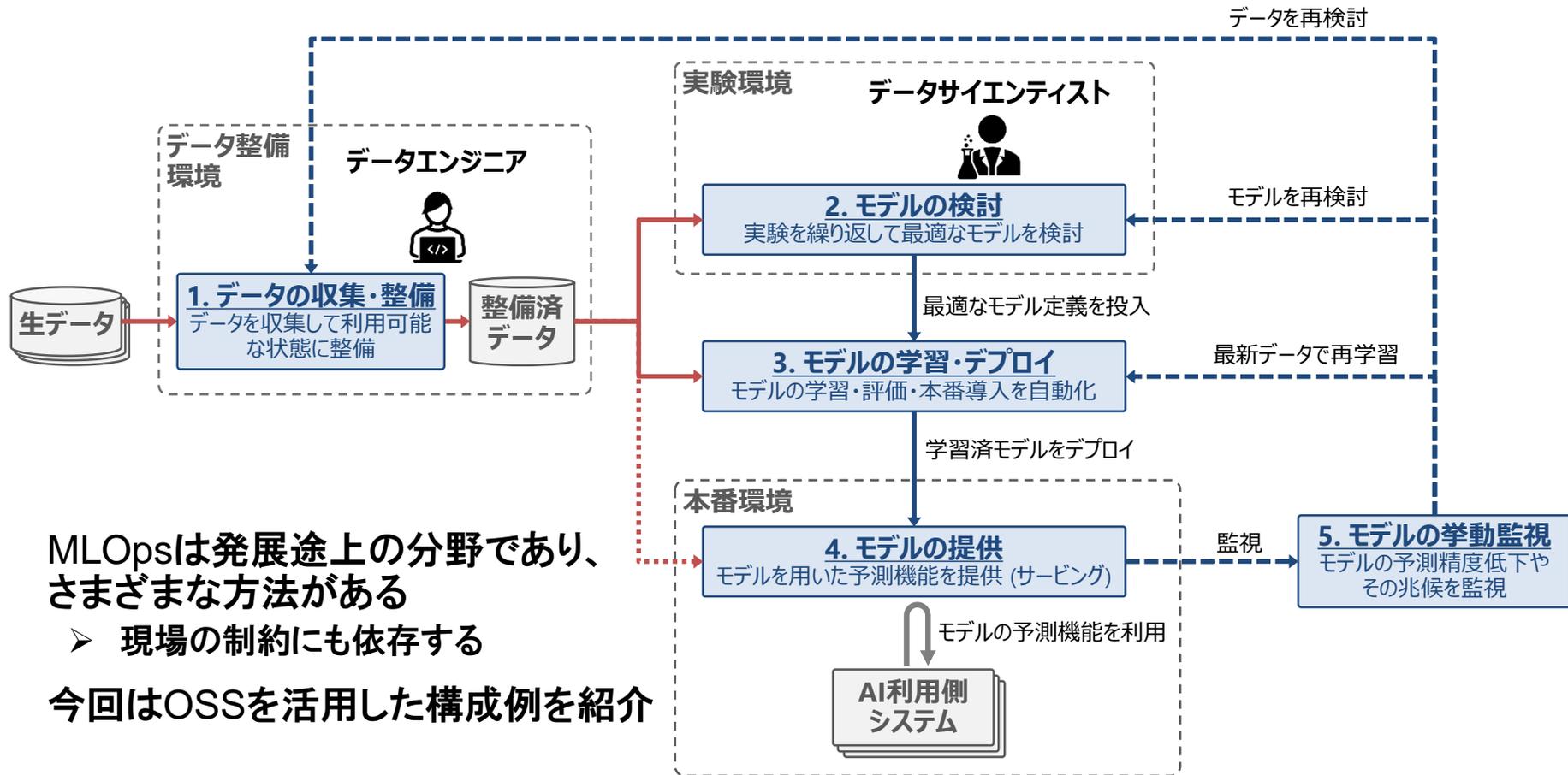
- 分析モデルの寿命： 現場導入した分析モデルの予測精度は時間とともに低下する
 - データの傾向変化、データに対する解釈の変化など、様々な要因がある
- 運用開始後も継続的な監視と問題への対処が必要



- MLOps (Machine Learning Operations) とは
 - 機械学習システムを継続的に運用するための取り組み
 - DevOps の考え方を採用し、一連の運用サイクルを体系的に自動化



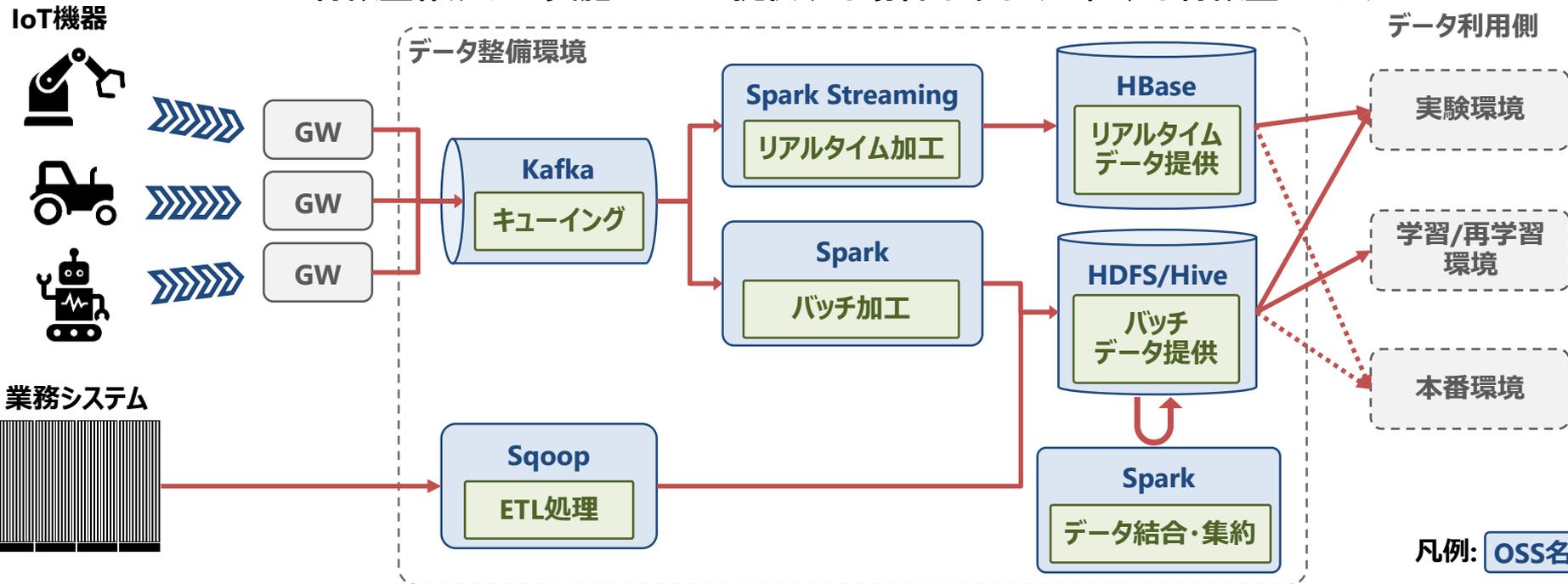
MLOpsによる機械学習システムの継続的な運用と改善



- MLOpsは発展途上の分野であり、さまざまな方法がある
 - 現場の制約にも依存する
- 今回はOSSを活用した構成例を紹介

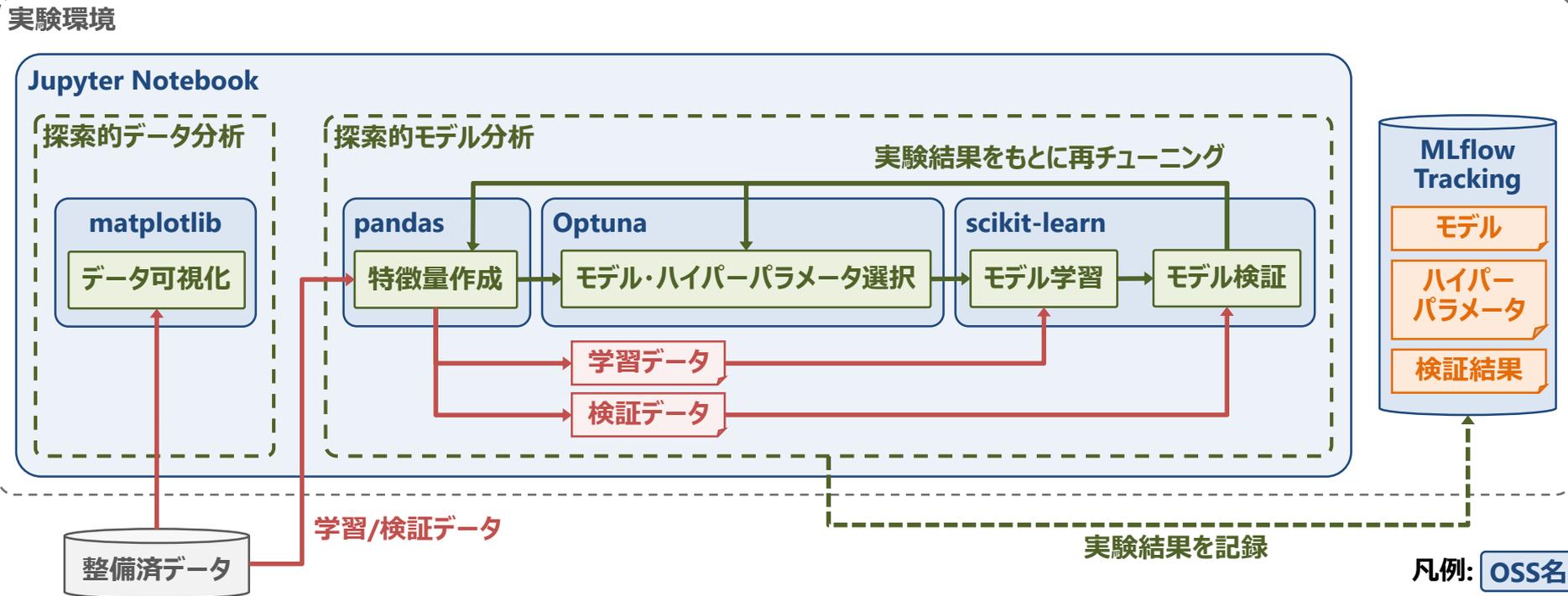
(1) データの収集・整備

- データエンジニアが分析モデルの構築・利用に必要なデータを収集・整備
 - データの収集・加工・結合・集約などを行うデータパイプラインを構築
 - データの蓄積・提供を行うデータストアを整備
 - ✓ ここで特徴量作成まで実施してから提供する場合もある（いわゆる特徴量ストア）



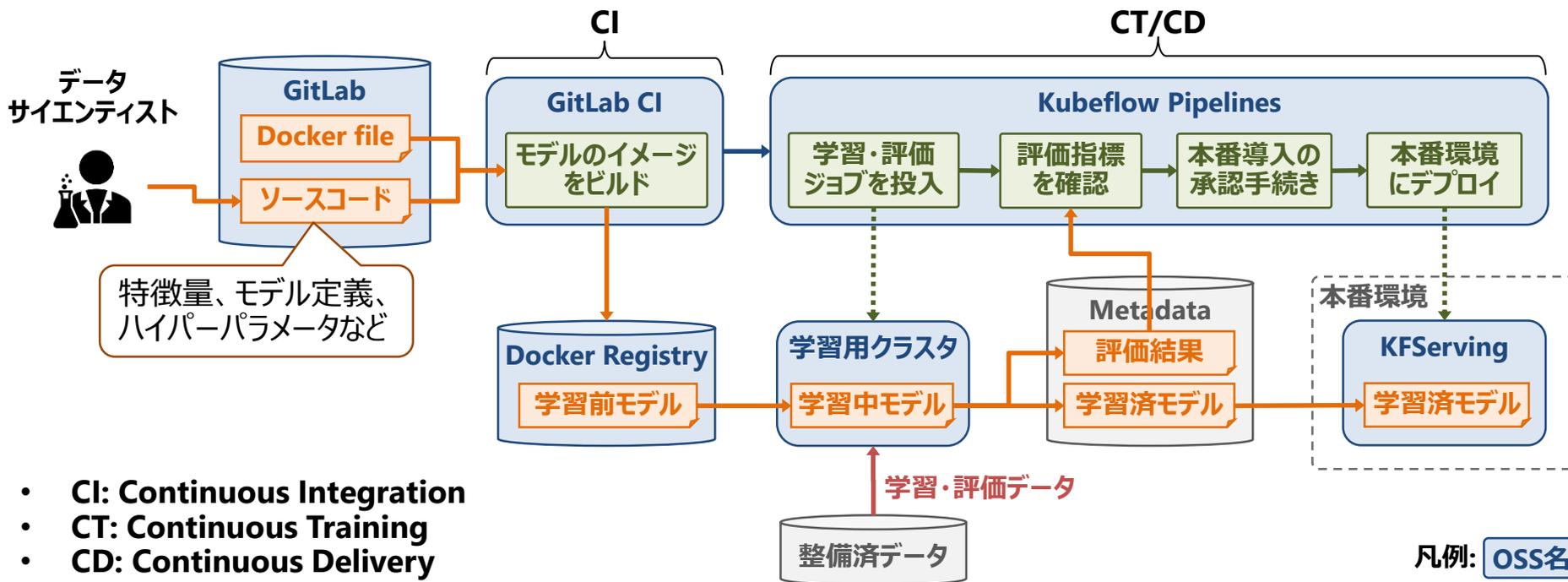
(2) モデルの検討

- データサイエンティストが試行錯誤を繰り返して機械学習モデルを構築
 - モデルの学習と検証を繰り返して、良いスコアのモデルを探索



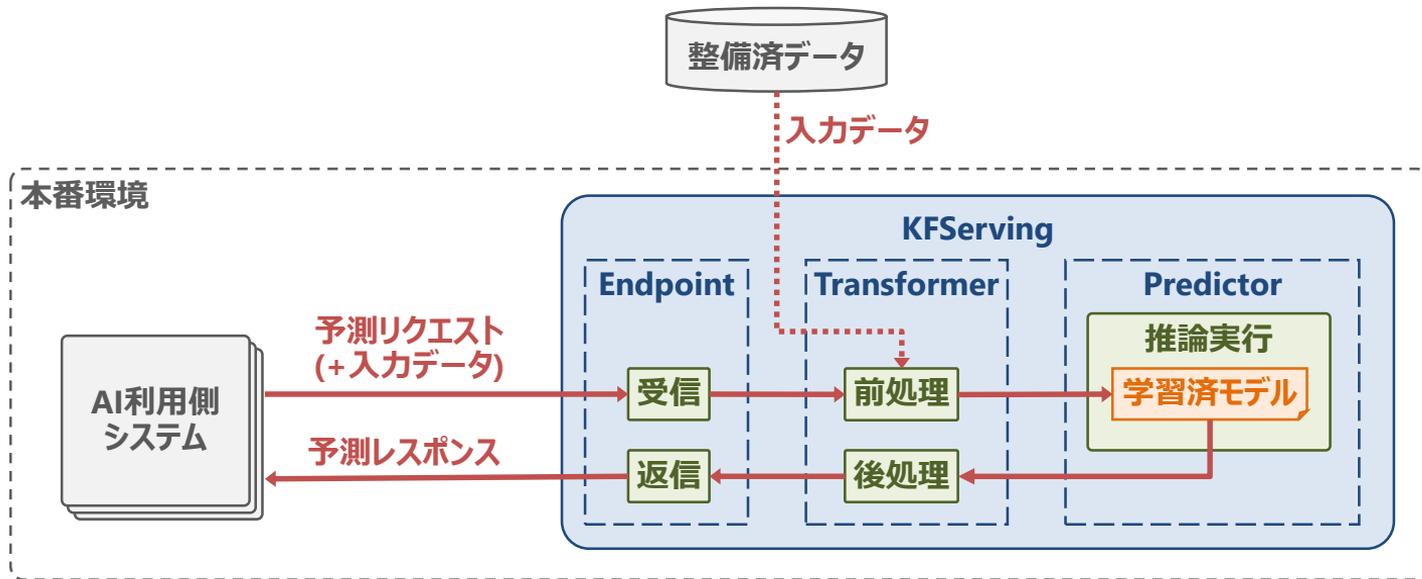
(3) モデルの学習・デプロイ

- モデルの最終的な学習・評価・承認を行い本番環境へデプロイ
 - 一連のデプロイ作業はCI/CT/CDパイプラインとして自動化



(4) モデルの提供

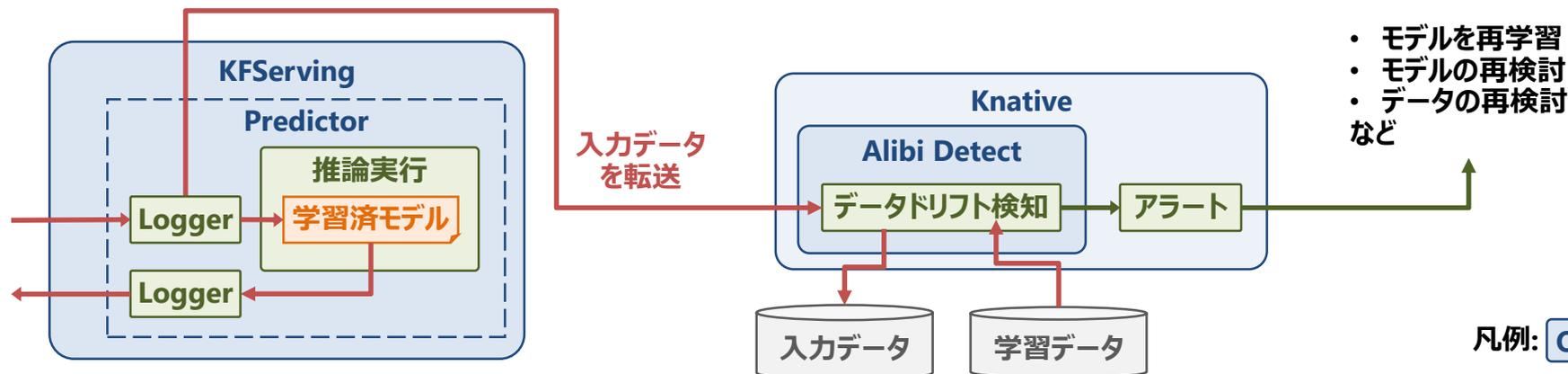
- 機械学習を利用したいシステムに対して、モデルの予測機能を提供(サービング)
 - 予測をリアルタイムに行うか、バッチ処理でまとめて行うかにより提供方式は異なる
 - リアルタイム推論の例:
 - ✓ 利用側システムから低レイテンシのWeb API(RESTやgRPCなど)経由で予測機能呼び出し



凡例: OSS名

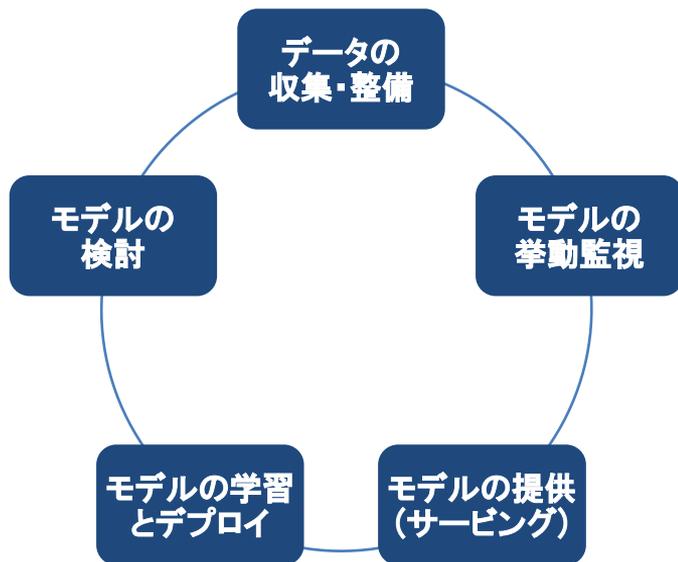
(5) モデルの挙動監視

- 本番環境にデプロイしたモデルの予測精度低下やその兆候を監視
 - 予測精度低下の兆候の例: データドリフト (入力データの分布が学習時から変化)
 - ✓ 物体Aと物体Bの判別モデルを、AとBが50%ずつのデータで学習していたが、推論時の入力データは物体Aが90%を占めるようになった
- 予測精度低下やその兆候を検知したらアラートで通知して対処
 - 最新データによるモデル再学習、モデル自体の再検討、データの再検討などで対処



凡例: OSS名

事業の現場でAIを活用するには、AIシステムの継続的な運用・改善が必要
⇒ MLOpsによりAIの継続的な運用・改善サイクルを実現可能



- MLOpsの各ステップでは様々な分野の技術が活用されている
 - ビッグデータ / AI / DevOps など
- デジタル分野のソフトウェアはOSSが主流でありMLOpsの実現にはOSS活用が不可欠といえる

日立のOSS・クラウド活用をより詳しく知りたい方におススメのイベントです。
ぜひお気軽にご参加ください。

The banner features a night cityscape with blue digital network lines overlaid. On the left, the Hitachi logo and event details are displayed. On the right, a large QR code is provided for registration.

HITACHI
Inspire the Next

大規模開発
× OSS・クラウド活用

2021.4.17(Sat.)
13:00～18:00



- Apache Hadoop, Apache Spark, Apache HBase, Apache Kafka, Apache Sqoop は、Apache Software Foundation の米国およびその他の国における登録商標または商標です。
- Python は、Python Software Foundation の商標または登録商標です。
- NumPy, pandas, matplotlib, scikit-learn, Jupyter は、NumFOCUS Foundation の商標または登録商標です。
- TensorFlow は、Google LLC の商標または登録商標です。
- PyTorch は、Facebook, Inc. の商標または登録商標です。
- MLflow は、LF Projects, LLC の商標または登録商標です。
- Optuna は、株式会社Preferred Networksの日本国およびその他の国における商標または登録商標です。
- GitLab は、GitLab B.V. の米国およびその他の国における登録商標もしくは商標です。
- Docker は、Docker Inc. の米国およびその他の国における登録商標もしくは商標です。
- ALIBI は、Seldon Technologies Limited の商標または登録商標です。
- Qiita は、Increments株式会社の商標または登録商標です。
- HITACHI は、株式会社日立製作所の商標または登録商標です。
- その他記載の会社名、製品名などは、それぞれの会社の商標もしくは登録商標です。

HITACHI
Inspire the Next 