機械学習プラットフォーム・サービスの選び方

自己紹介

- 田上健太
- @haizine_regonn
- 2018年島根県にIターン
- フリーランスとして活動
- 機械学習コンテンツの
  ポッドキャスト配信
- VR技術やVTuber活動

機械学習(データ分析)PaaS

Amazon SageMaker

ABEJA PLATFORM

nehan

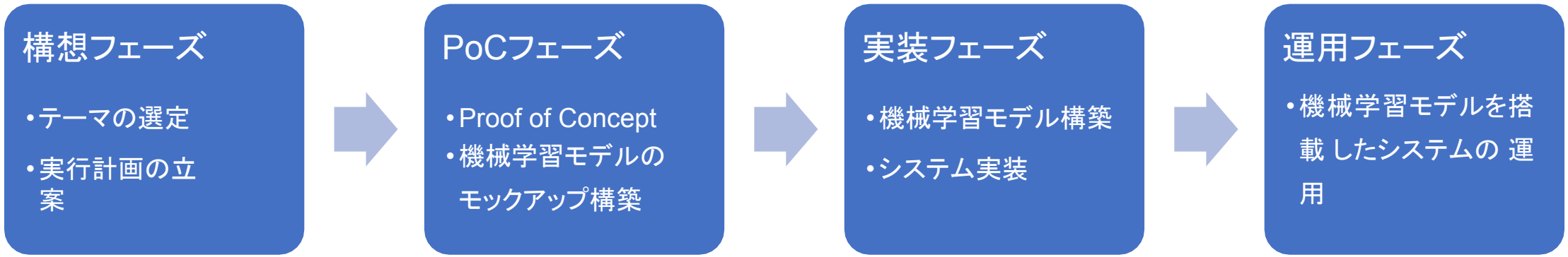
data iku

Watson Machine Learning

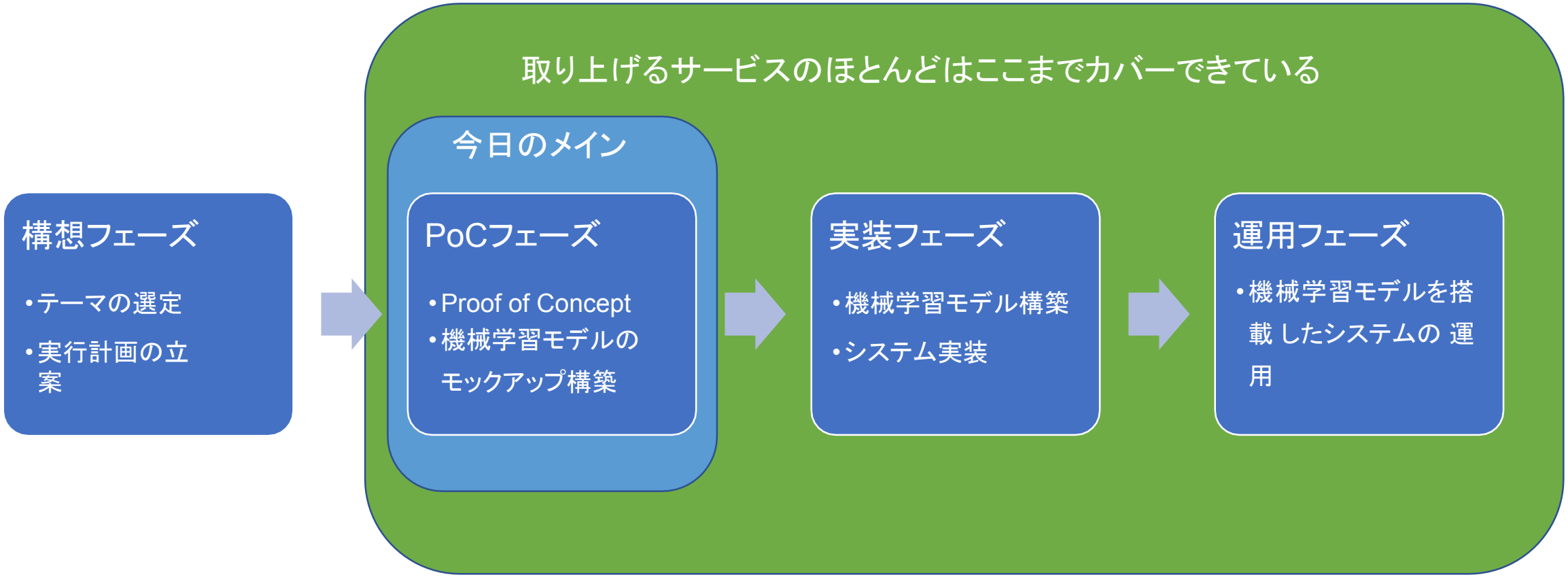
DataRobot

Neural Network Console

機械学習プロジェクトの全体像



機械学習プロジェクトの全体像



本日あまり話さないこと

- TensorFlow、PyTorch等のフレームワーク紹介
- それぞれのプラットフォームの詳しい使い方等
  - 一部サービスでのデモンストレーションあり

今回の調査対象(2019年3月調査時)

- Amazon/AWS
- Google/GCP
- IBM/Watoson
- Microsoft/Azure
- Sony
- Uber

調査以降の動向

- Microsoft
  - ドラッグ&ドロップの 機械学習ツールを発 表

日本発の分析サービスも色々出てきた

MatrixFlow

nehan

最近の機械学 習プラットフォームの傾向

GUIでの機械学習

- フローを組み立てればよい
- 非エンジニア向けというわけではなく、ある程度知 識のあるエンジニアが素早くプロトタイプを作れる

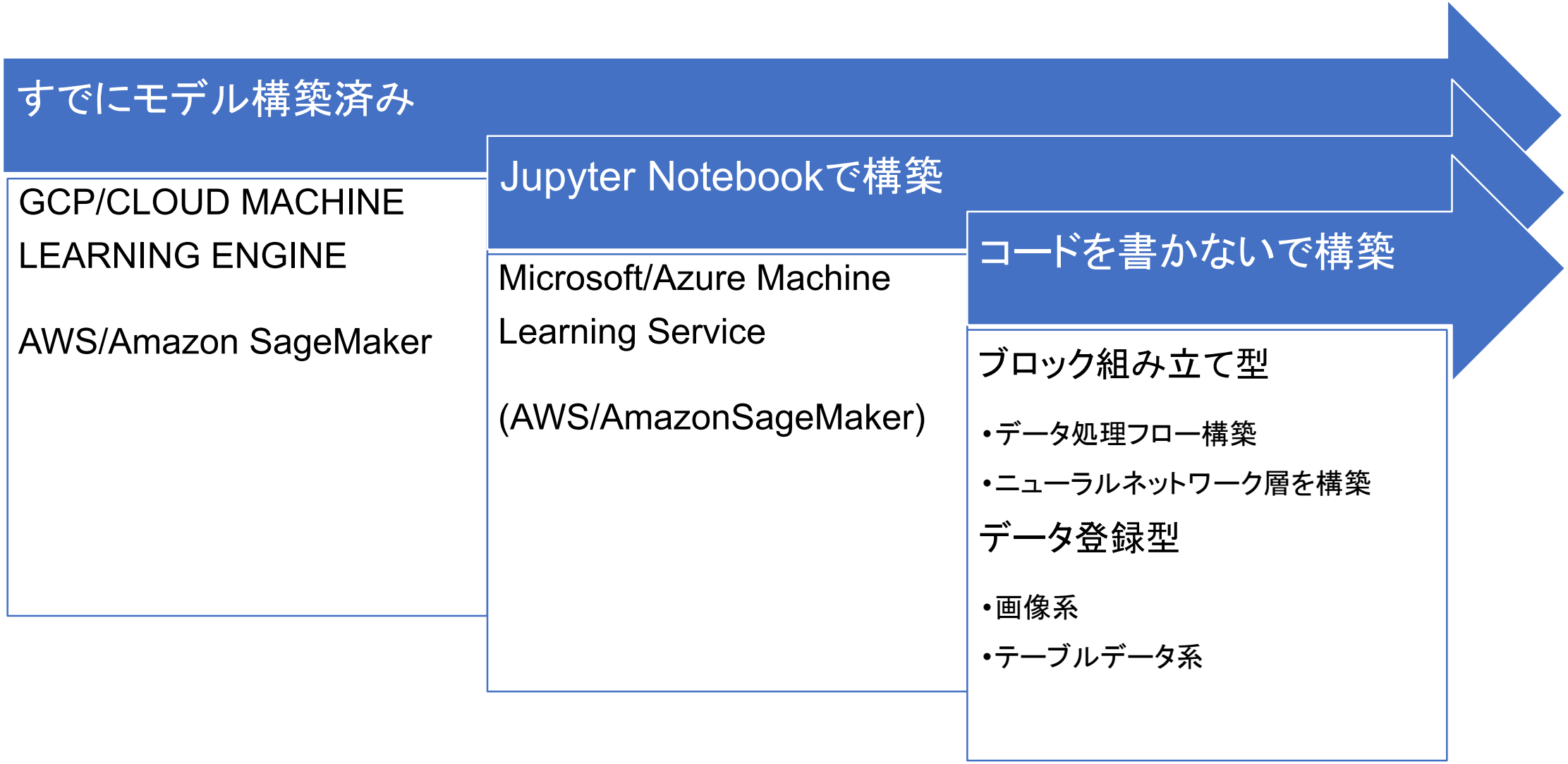
データを渡すだけで自動分析

- ある程度の精度を出すだけだったら十分
- AIの民主化

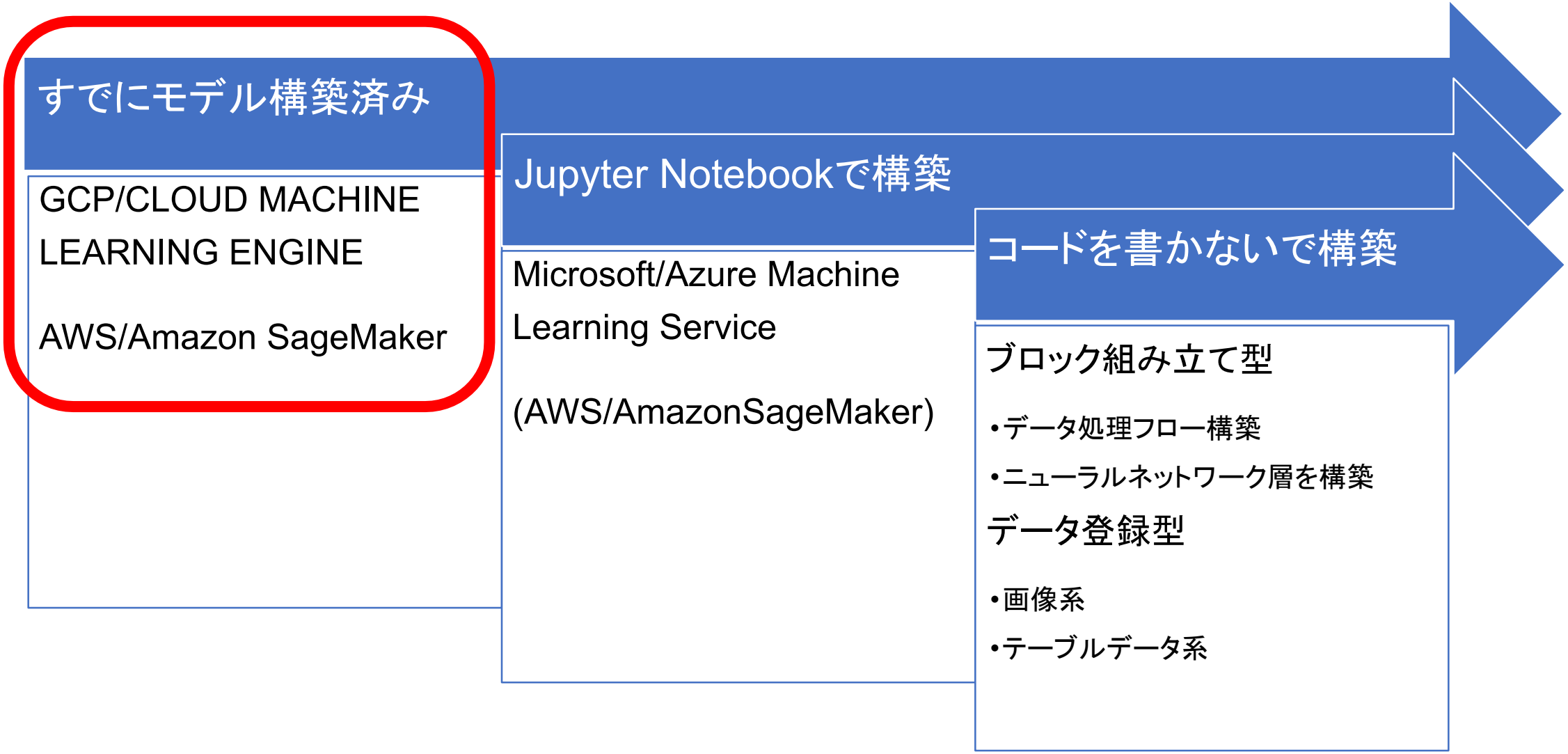
API利用等の本番運用もプラットフォームでできる

- データさえあれば、動くものを作ることができる
- パラメータ調整等もプラットフォーム上で完結する

サービス公開に向けて

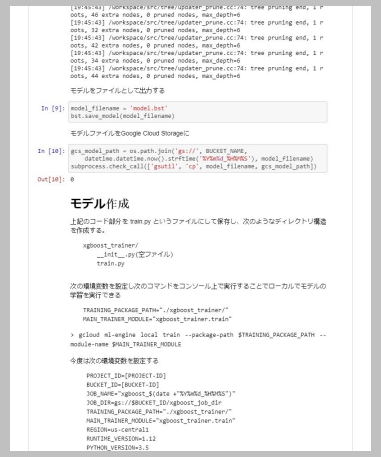

すでにモデル構築済み
GCP/CLOUD MACHINE LEARNING ENGINE
AWS/Amazon SageMaker
Jupyter Notebookで構築
Microsoft/Azure Machine Learning Service
(AWS/AmazonSageMaker)
コードを書かないで構築
ブロック組み立て型
•データ処理フロー構築
•ニューラルネットワーク層を構築
データ登録型
•画像系
•テーブルデータ系

サービス公開に向けて


すでにモデル構築済み
GCP/CLOUD MACHINE LEARNING ENGINE
AWS/Amazon SageMaker
Jupyter Notebookで構築
Microsoft/Azure Machine Learning Service
(AWS/AmazonSageMaker)
コードを書かないで構築
ブロック組み立て型
•データ処理フロー構築
•ニューラルネットワーク層を構築
データ登録型
•画像系
•テーブルデータ系

すでにモデルが構築済み

- GCP/CLOUD MACHINE LEARNING ENGINE
  - TensorFlow/scikit-learn/XGBoost等のライブラリで構築済みであれば、TPUやGPUを利用した高速なモデルトレーニング、評価、デプロイがで きるようになっている
  - ローカル環境でモデルを構築して、GCP上でジョブとして実行する
    - Docker, TensorFlow関連が強い



モデルをファイルとして出力する

```
In [9]: model_filename = 'model.bst'
        bst.save_model(model_filename)
```

モデルファイルをGoogle Cloud Storageに

```
In [10]: gcs_model_path = os.path.join('gs://', BUCKET_NAME,
             datetime.datetime.now().strftime('%Y%m%d_%H%M%S'), model_filename)
         subprocess.check_call(['gsutil', 'cp', model_filename, gcs_model_path])
Out[10]: 0
```

モデル作成

上記のコード郡を train.py というファイルにして保存し、次のようなディレクトリ構造を作成する。

```
xgboost_trainer/
    __init__.py(空ファイル)
    train.py
```

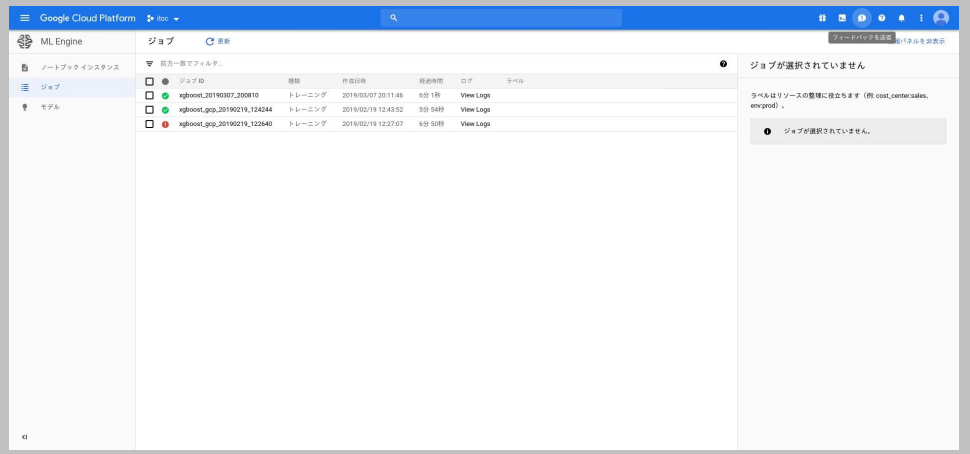
次の環境変数を設定し次のコマンドをコンソール上で実行することでローカルでモデルの学習を実行できる

```
TRAINING_PACKAGE_PATH="./xgboost_trainer/"
MAIN_TRAINER_MODULE="xgboost_trainer.train"

> gcloud ml-engine local train --package-path $TRAINING_PACKAGE_PATH --
module-name $MAIN_TRAINER_MODULE
```

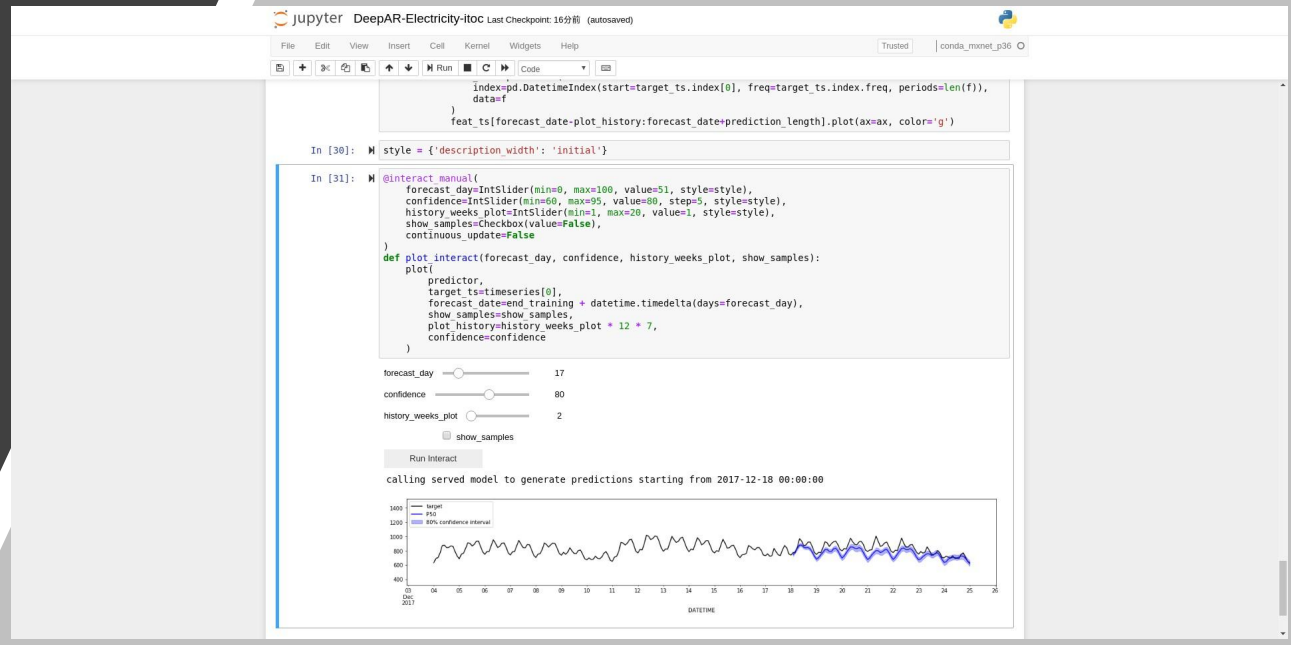
今度は次の環境変数を設定する

```
PROJECT_ID=[PROJECT-ID]
BUCKET_ID=[BUCKET-ID]
JOB_NAME="xgboost_$(date +"%Y%m%d_%H%M%S")"
JOB_DIR=gs://$BUCKET_ID/xgboost_job_dir
TRAINING_PACKAGE_PATH="./xgboost_trainer/"
MAIN_TRAINER_MODULE="xgboost_trainer.train"
REGION=us-central1
RUNTIME_VERSION=1.12
PYTHON_VERSION=3.5
```

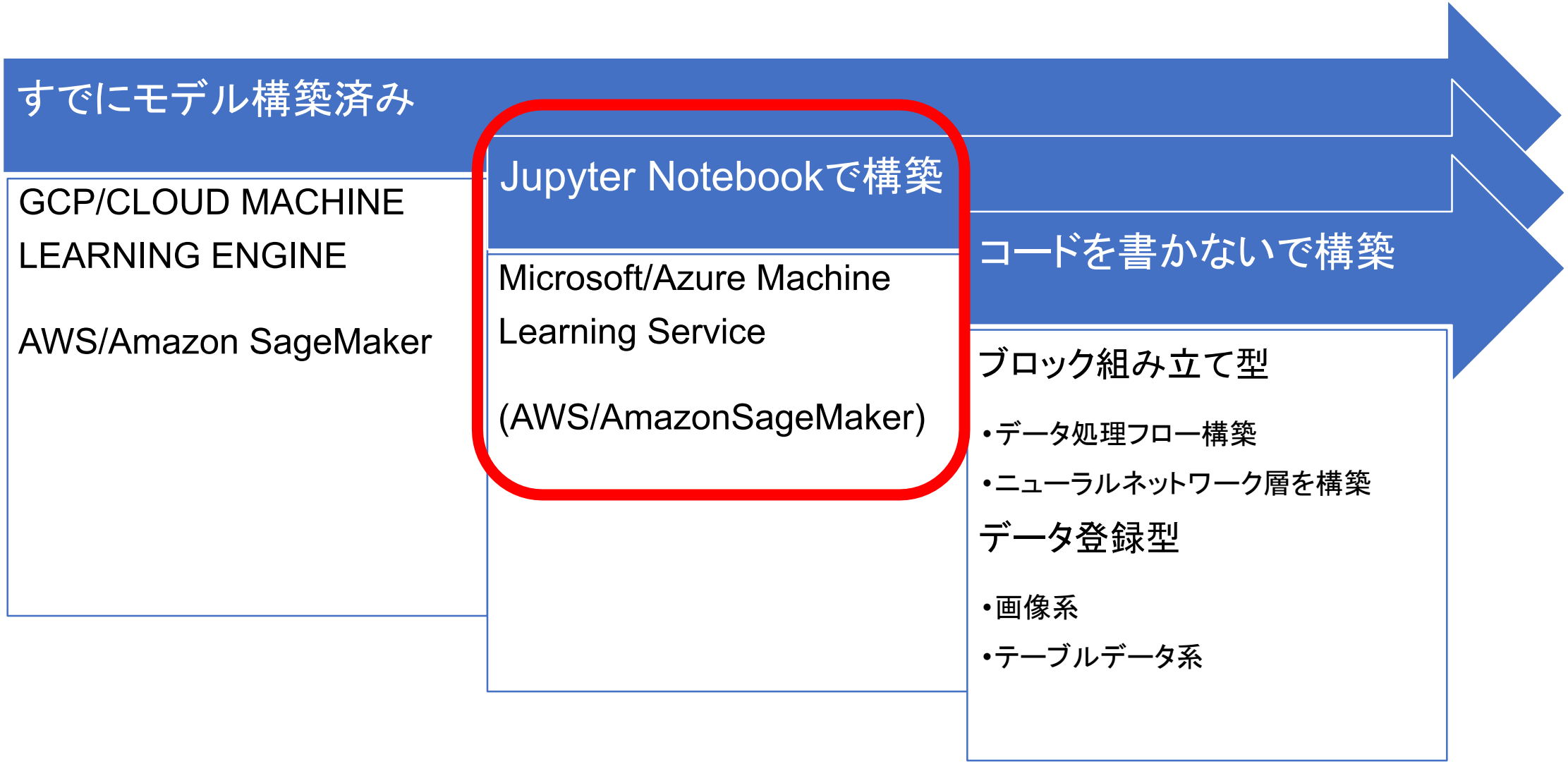


すでにモデルが構築済み

- AWS/Amazon SageMaker
  - Tensorflow/Chainer/PyTorch/MXN et といったフレームワークの SageMaker 用Wrapperが用意されて おり、それに合わせてコードを変更 すれば、SageMaker上でトレーニン グを行い、デプロイができる
  - S3上にワークスペースを作成して、AWSのインスタンス上でJupyter Notebookを動かす
  - ONNX(オニキス)という共通のディープラーニングモデルフォーマットに強い

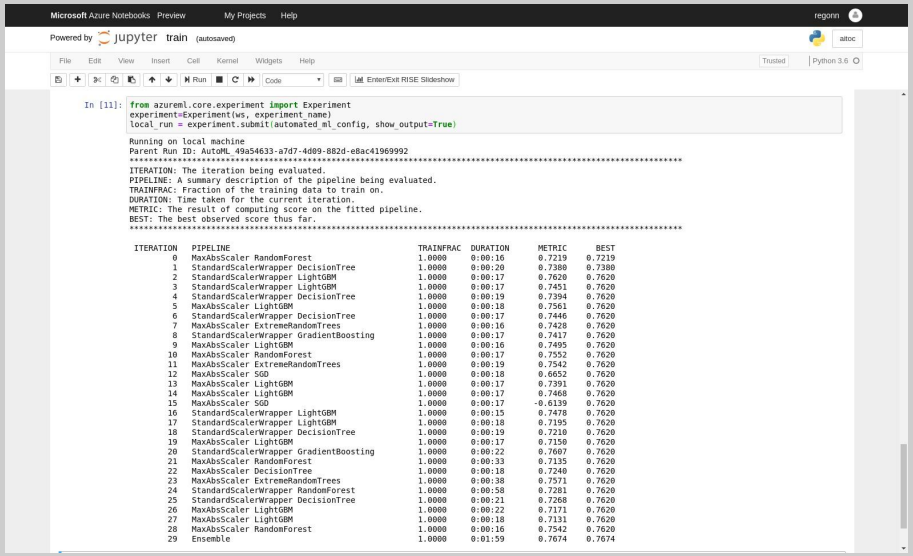


サービス公開に向けて

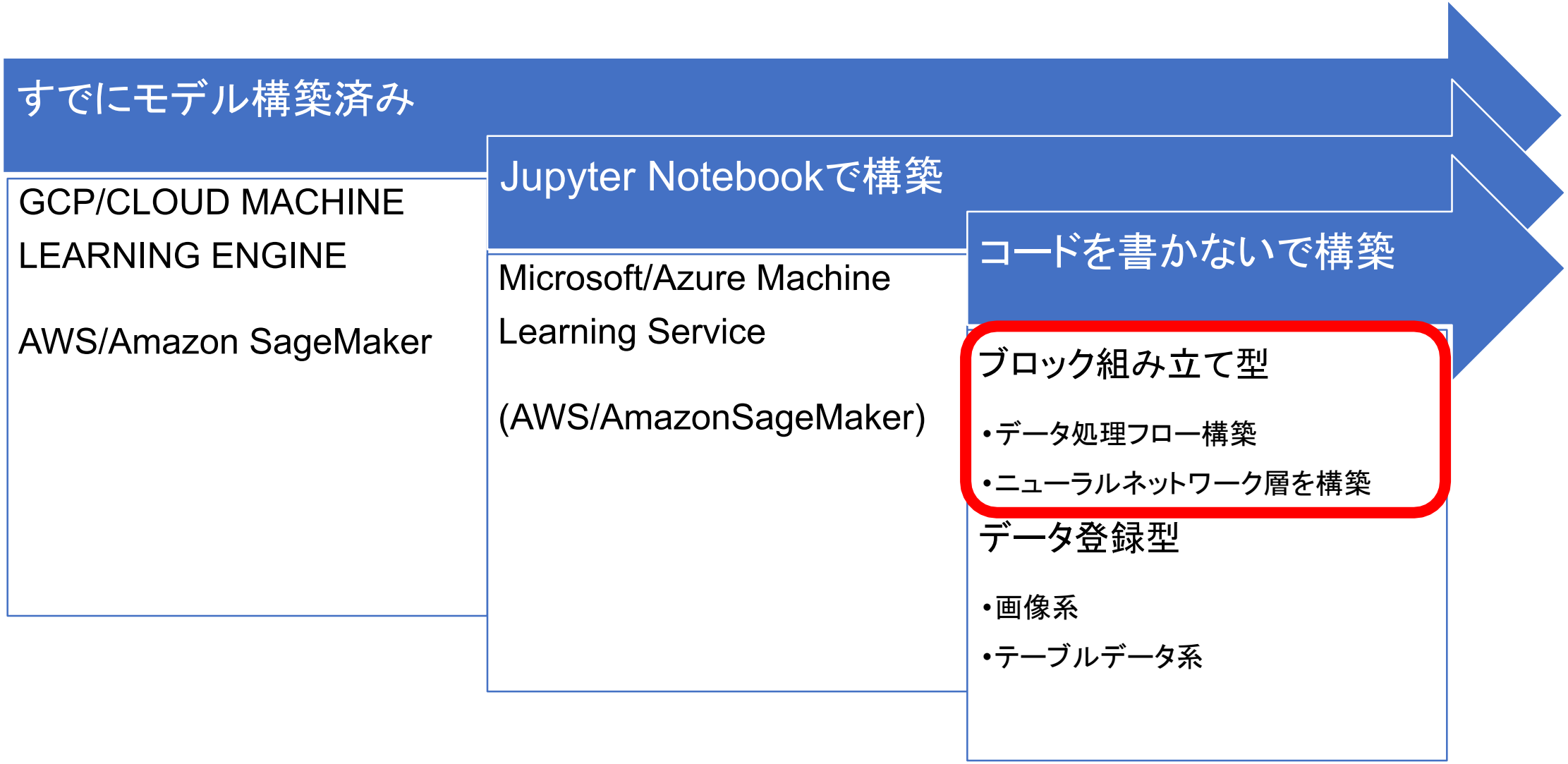

すでにモデル構築済み
GCP/CLOUD MACHINE LEARNING ENGINE
AWS/Amazon SageMaker
Jupyter Notebookで構築
Microsoft/Azure Machine Learning Service
(AWS/AmazonSageMaker)
コードを書かないで構築
ブロック組み立て型
•データ処理フロー構築
•ニューラルネットワーク層を構築
データ登録型
•画像系
•テーブルデータ系

Jupyter Notebook上でモデル構築をしたい

- Azure Machine Learning Service
  - 独自のライブラリでデータ処 理を 行 うためなれが 必　要 だけど、学習部分は色々な分類機を用いて学習を行ってくれる
  - Azure上でJupyter Notebookを動かす
  - Xboxサービスを展開しているためIoTのノウハウが強い

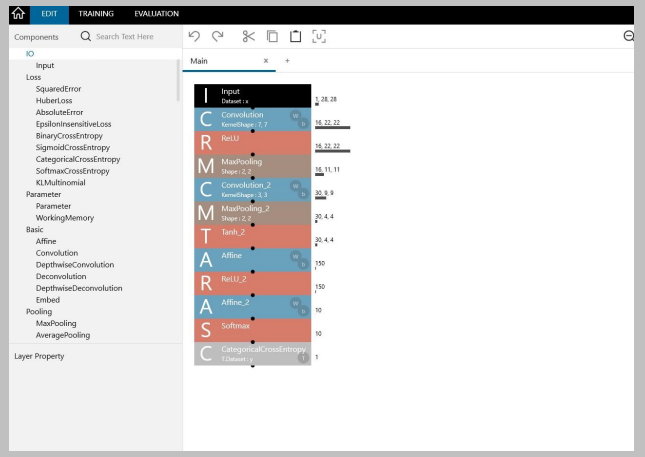


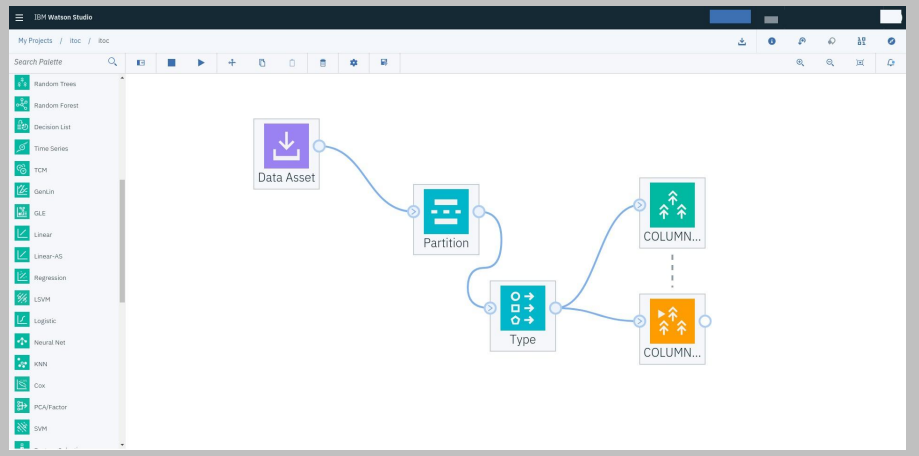
サービス公開に向けて


すでにモデル構築済み
GCP/CLOUD MACHINE LEARNING ENGINE
AWS/Amazon SageMaker
Jupyter Notebookで構築
Microsoft/Azure Machine Learning Service
(AWS/AmazonSageMaker)
コードを書かないで構築
ブロック組み立て型
•データ処理フロー構築
•ニューラルネットワーク層を構築
データ登録型
•画像系
•テーブルデータ系

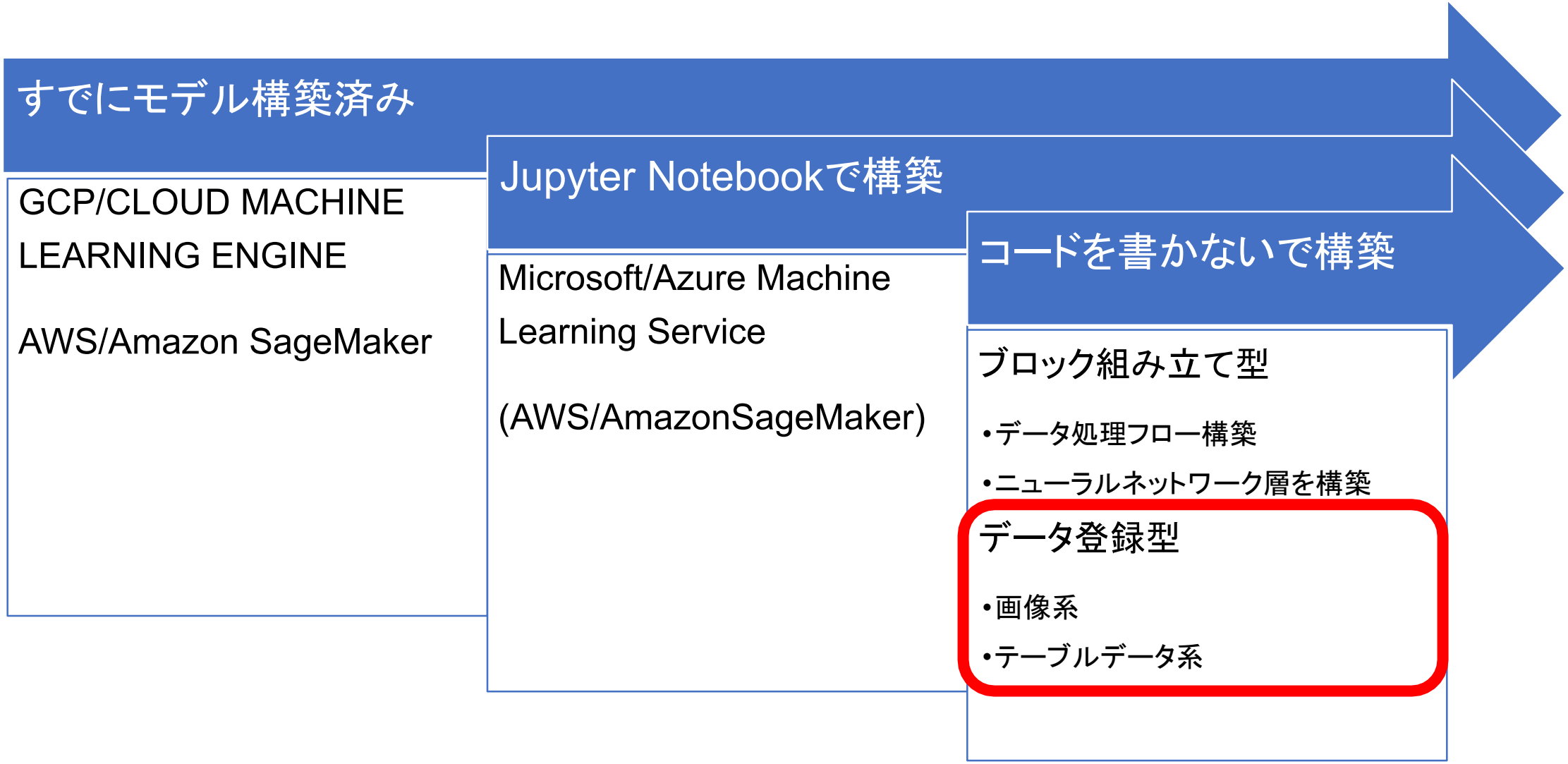
ブロックで組み立てる

- 先程も書いたが最近特にサービスが多い (MatrixFlow,nehan,Dataiku)
  - IBM/Watson Studio
  - Microsoft/Azure Machine Learning Studio
    - (後でデモンストレーション予定)
  - Sony/Neural Network Console
    - サービスまでデプロイはできな いが、画像処理ディープラーニ ングの相性が良い



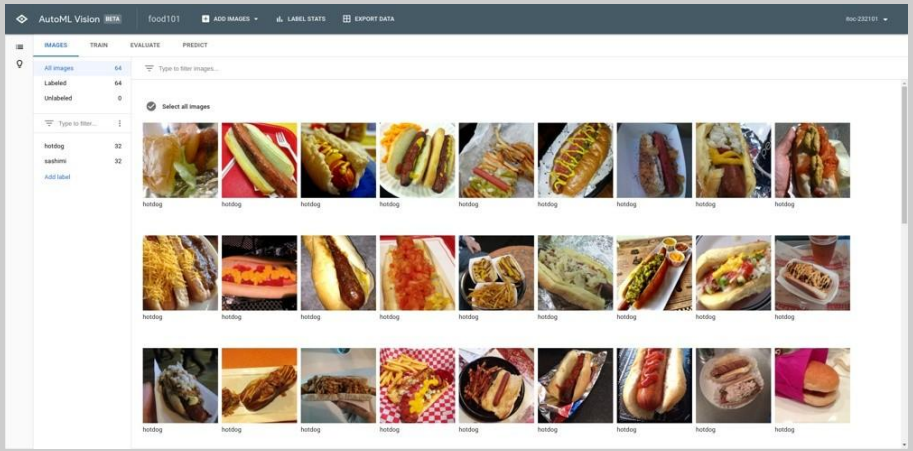


サービス公開に向けて


すでにモデル構築済み
GCP/CLOUD MACHINE LEARNING ENGINE
AWS/Amazon SageMaker
Jupyter Notebookで構築
Microsoft/Azure Machine Learning Service
(AWS/AmazonSageMaker)
コードを書かないで構築
ブロック組み立て型
•データ処理フロー構築
•ニューラルネットワーク層を構築
データ登録型
•画像系
•テーブルデータ系

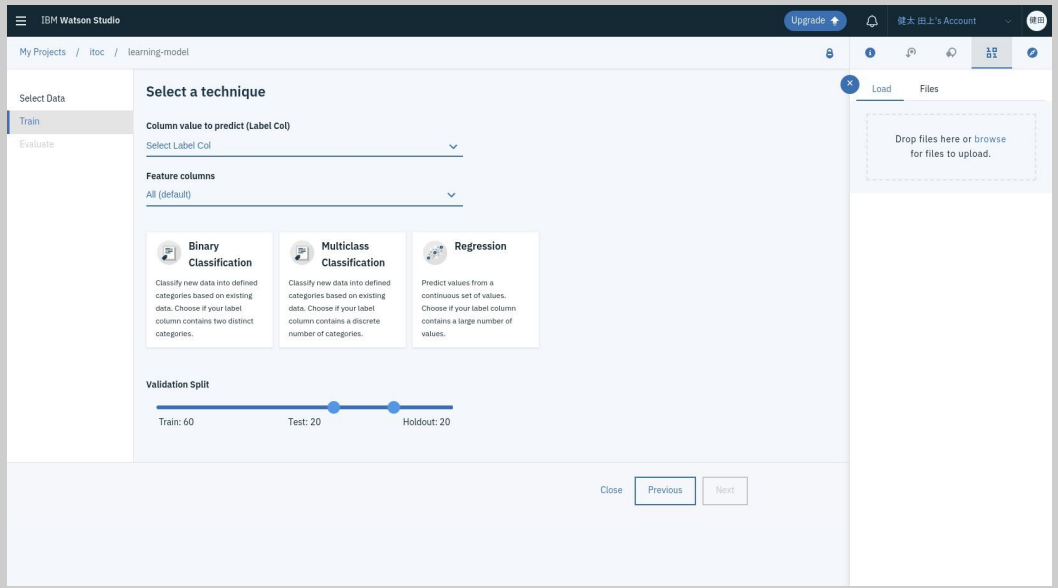
データ登録型(画像系)

- GCP/Cloud AutoML
  - 画 像を登 録してラベルを付 けるだけで学習を行うことが可 能



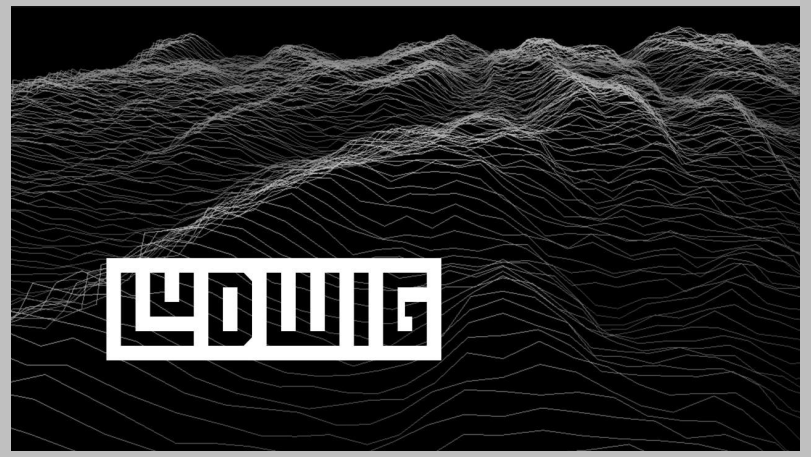
データ登録型(テーブルデータ)

- IBM/Watson Machine Learning
  - CSV等でデータを登録して、項目を設定するだけで学習が行えて、デプロイまで可能



データ登録型(テーブルデータ)

- Uber/Ludwig
  - カラム情報を記載したyamlファイル　とデータCSVを準備すれば、ローカ ルのコンソールで学習ができてしま う
  - 画像データ等もある
- Facebook/Prophet
  - 時系列に特化しており、DataFrame を渡すだけで、未来の予測をしてくれる



デモンストレーション

- 実際に新しくなったMicrosoftのAzure Machine Learning serviceの新機能 Visual interface を触ってみる
- UberのLudwigでもやってみる
- Kaggleにコードを書かずに挑む!!