

ORACLE

# HeatWave AutoMLによる機械学習入門

オープンソースカンファレンス2024 Fukuoka

**山崎 由章**

MySQL Master Principal Solution Engineer

MySQL Global Business Unit

2024年12月7日

# アジェンダ

1. 機械学習概要
2. 機械学習の課題とHeatWave AutoMLによる課題の解決
3. HeatWave AutoML デモ
4. HeatWave AutoML 活用事例



# 機械學習概要

# 機械学習とは？

Arthur Samuel の定義

明示的にプログラミングすることなく  
コンピュータが能力を獲得する研究分野

Tom Mitchell の定義

コンピュータプログラムが、タスクのクラス $T$ と  
その性能指標 $P$ に関して「経験 $E$ から学習する」とは、  
「 $T$ 内のタスクに対する $P$ が $E$ により改善する」ことを言う

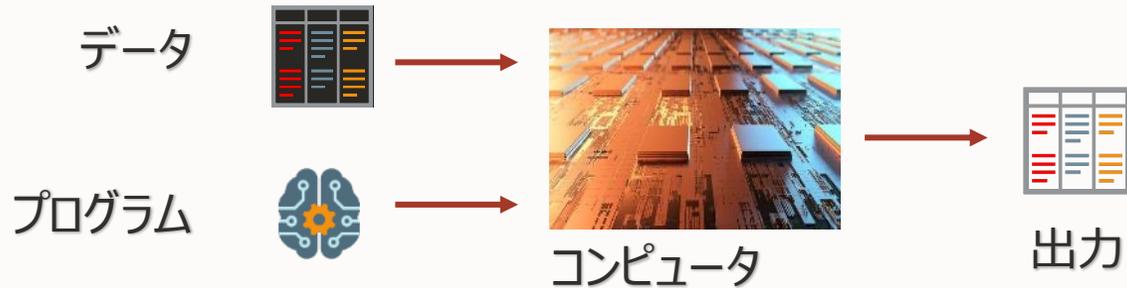


# 機械学習とは？

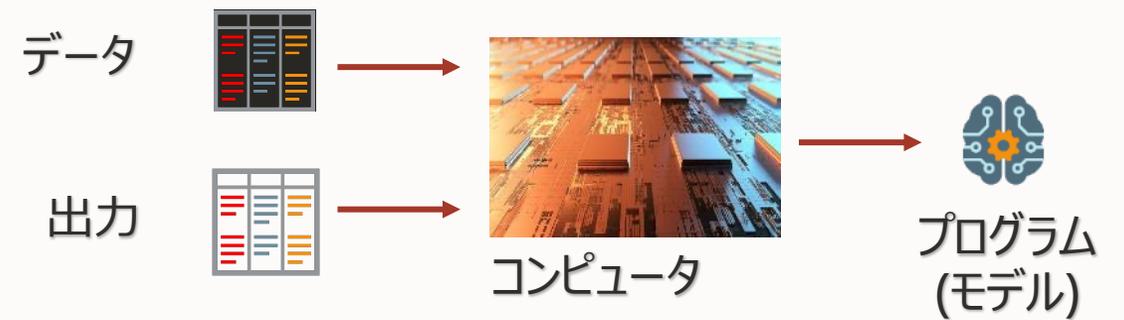
Arthur Samuel の定義

明示的にプログラミングすることなくコンピュータが能力を獲得する研究分野

## プログラミングの場合



## 機械学習の場合



# 機械学習とは？

Tom Mitchell の定義

コンピュータプログラムが、タスクのクラスTとその性能指標Pに関して「経験Eから学習する」とは、「T内のタスクに対するPがEにより改善する」ことを言う

具体例で考えると・・・

- タスク：明日の売上を予測する
  - 経験：過去の売上データ
  - 性能指標：実際の売上との誤差
- ⇒ 過去の売上データから明日の売上データを予測し、その誤差を少なくしていく

# 機械学習の種類

<h2>教師あり学習</h2> <p>正解を含むデータを提供し、 そのデータから学習してモデルを作成する</p>	<h2>教師なし学習</h2> <p>正解データなしで入力したデータから 特徴やルールを学習してモデルを作成する</p>
<h2>半教師あり学習</h2> <p>少数の正解を含むデータと 大量の正解を含まないデータを提供し、 それらのデータから学習して モデルを作成する</p>	<h2>強化学習</h2> <p>「エージェント」が「環境」の中で 最も多く報酬が得られるような 振る舞いを試行錯誤しながら学習し、 モデルを作成する</p>



# 機械学習の種類

<p style="text-align: center;"><b>教師あり学習</b></p> <ul style="list-style-type: none"><li>• 分類：いくつかのクラスに識別する</li><li>• 回帰：連続する値を予測する</li><li>• 時系列予測： 過去の出来事から将来の値を予測する</li></ul>	<p style="text-align: center;"><b>教師なし学習</b></p> <ul style="list-style-type: none"><li>• クラスタリング： 似た特徴を持つものを識別する</li><li>• 異常検知： 異常なパターンを発見する</li></ul>
<p style="text-align: center;"><b>半教師あり学習</b></p>	<p style="text-align: center;"><b>強化学習</b></p>



# 機械学習の活用例

**分類 (Classification) :** いくつかのクラスに識別する

活用例)

- Eメールがスパムか否かを判定する
- ローンを承認していい顧客かどうか判定する
- 顧客の解約率の予測する (解約率を高・中・低に分類する)

**回帰 (Regression) :** 連続する値を予測する

活用例)

- 住宅価格を予測する
- 将来の売上を予測する
- 顧客の待ち時間を予測する



# 機械学習の活用例

**時系列予測 (Timeseries Forecasting) :** 過去の出来事から将来の値を予測する

活用例)

- 来月の需要を予測する(発注量を予測する)
- 冬場の必要発電量を予測する

**異常検知 (Anomaly detection) :** 異常なパターンを発見する

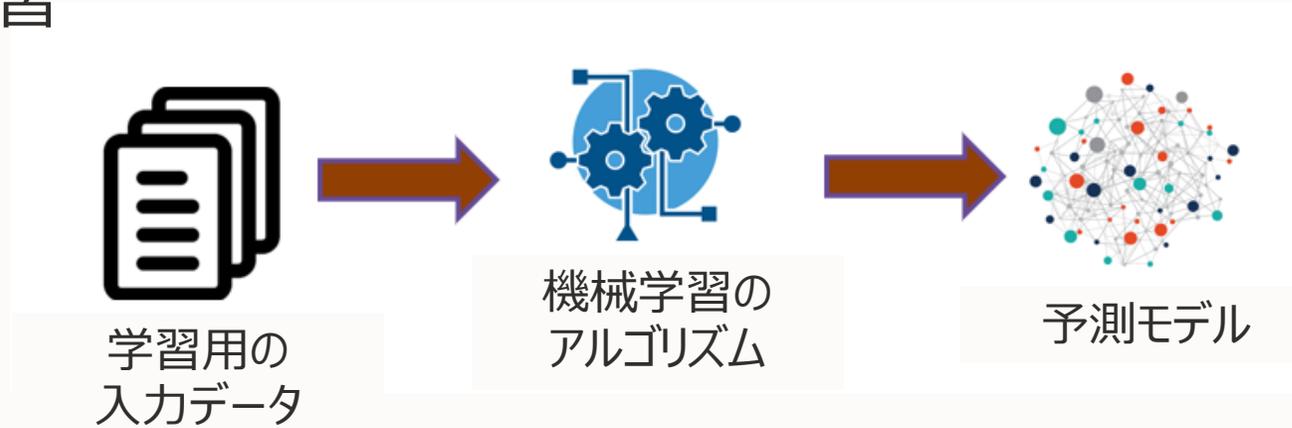
活用例)

- 製造データから、欠陥品を発見する
- システムのログデータから不正を検出する

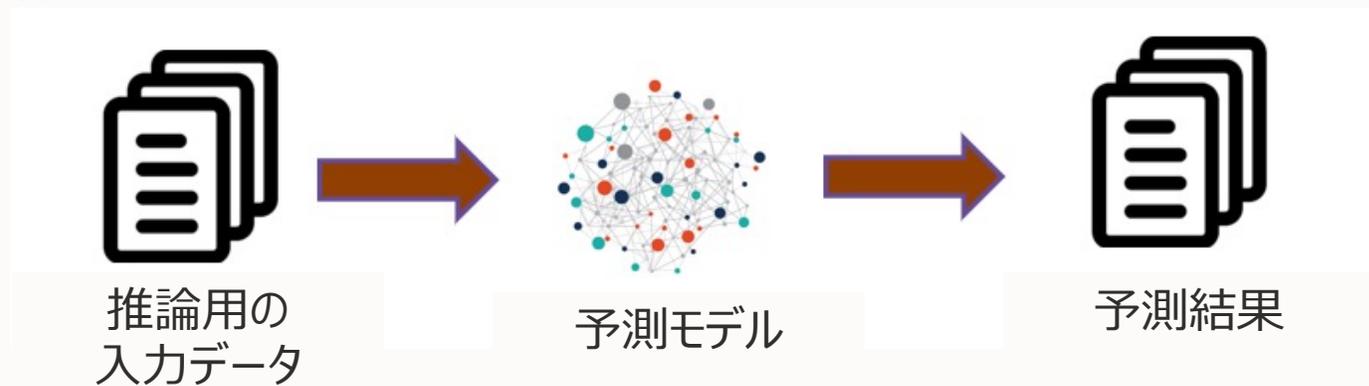


# 学習(Training)と推論(Inference)

## 学習



## 推論



### 学習と推論の流れ

1. 学習用のデータから機械学習のアルゴリズムを使って学習し、モデルを作成する
2. そのモデルを使って推論用のデータから結果を予測する



# 機械学習アルゴリズムの例

## 分類

- デシジョンツリー
- ロジスティック回帰
- ランダムフォレスト
- ニューラルネットワーク
- サポートベクターマシーン

- 
- 
- 

## 回帰

- 線形回帰
- ベイズ線形回帰
- ランダムフォレスト
- ブーストデシジョンツリー
- ニューラルネットワーク

- 
- 
- 

様々な機械学習アルゴリズムが存在する



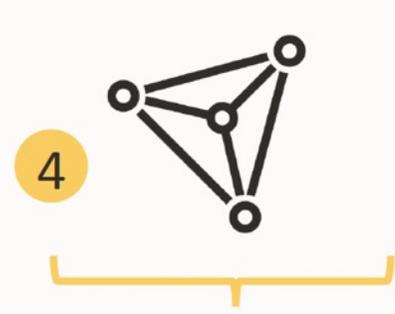
# 機械学習を現場で活用する流れ (モデルの作成まで)

ビジネス上の目的を設定する



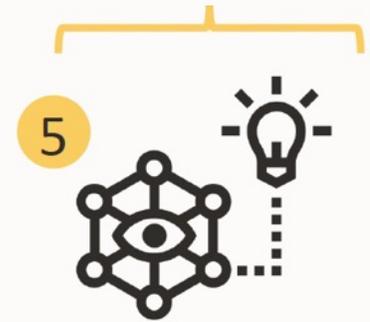
必要なデータを集める、  
データの形式を整える

データの中から特徴量(※)を抽出し、  
最適なアルゴリズムを選択する  
※予測の手掛かりとなる変数



データを使って学習する、  
ハイパーパラメータ(※)を  
チューニングする

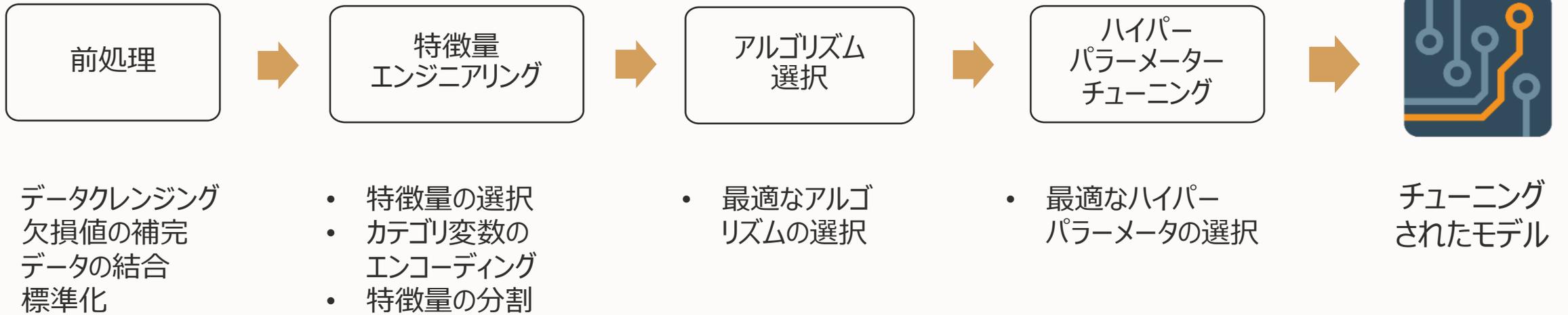
モデルが完成する



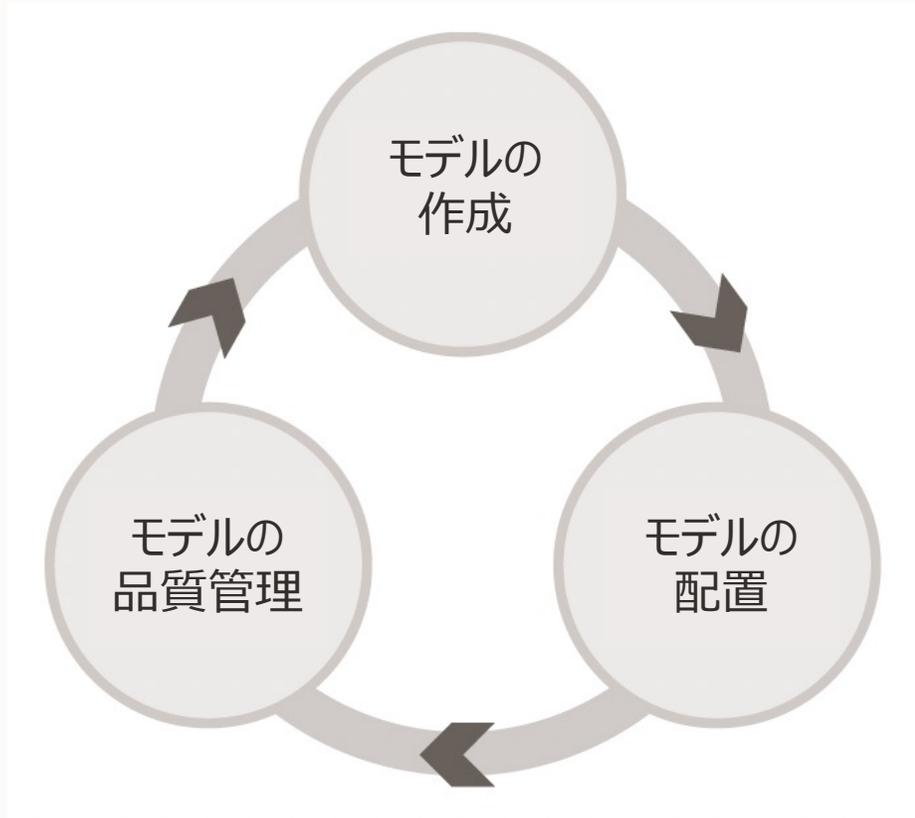
※機械学習アルゴリズムの挙動を設定するパラメータ  
(モデル学習する前段階で設定するパラメータ)



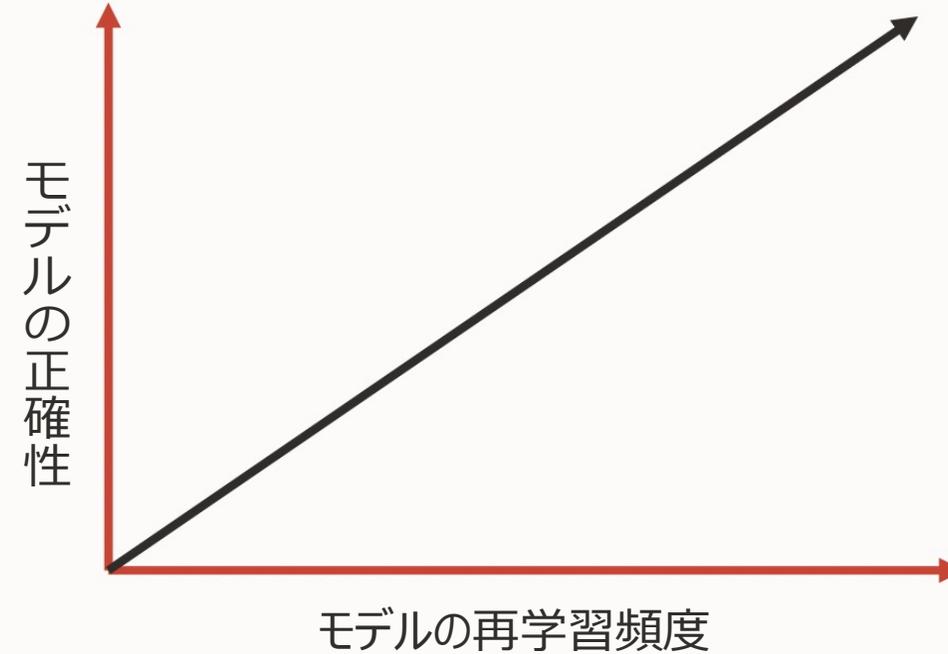
# モデルのチューニング過程



# 機械学習のライフサイクル



より最新のデータ、より関連性の高いデータを使って再学習することでモデルの正確性を向上できる



# 機械学習の課題と HeatWave AutoMLによる 課題の解決

機械学習は  
まるで **魔法** のよう

なのになぜみんな使わないの？

機械学習

は

難しい &

お金がかかる

# 機械学習の工程と各工程での課題

工程

大量のデータを  
収集

結論を導く  
アルゴリズムを選択

パラメーター  
を選択

統計的な演算を行ってモデルを  
作成するためのアルゴリズムを  
「トレーニング」

実際のデータに対して  
推論を実行



データの移動と加工は手間  
とコスト、時間がかかる面倒  
な作業



正しいアルゴリズムと  
パラメーターを選択できる  
か？



推論や予測の精度に大きな影響があるため  
重要ではあるが、トレーニングに非常に  
長い時間がかかる (※)



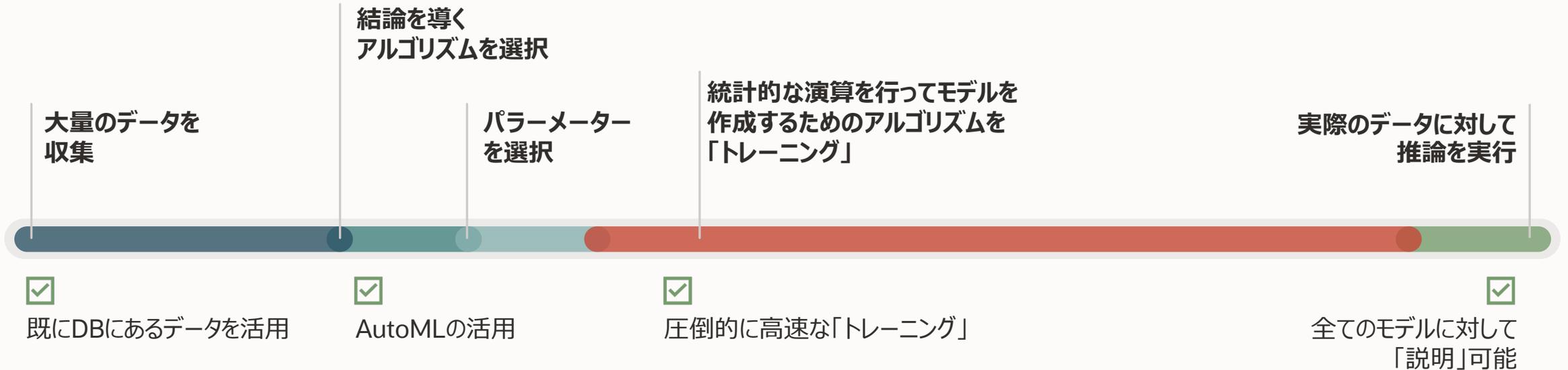
ブラックボックス化、  
なぜそのモデルがその  
選択をしたのか分からない

※トレーニングは、推論に比べて非常に時間がかかる



# 機械学習の工程と各工程での課題

工程



## 機械学習の悩みをHeatWave AutoMLが解決



HeatWave AutoMLを使うと

**SQLだけで簡単にMLを活用できる**

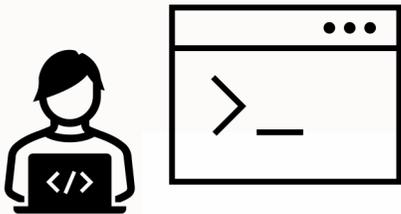
# HeatWave AutoMLのインデータベース機械学習

機械学習特有の複雑さを排除し、SQLだけで簡単に機械学習を実現

MySQL Workbench



mysqlsh, mysql client



MySQL Database



SQL

クエリプッシュダウン

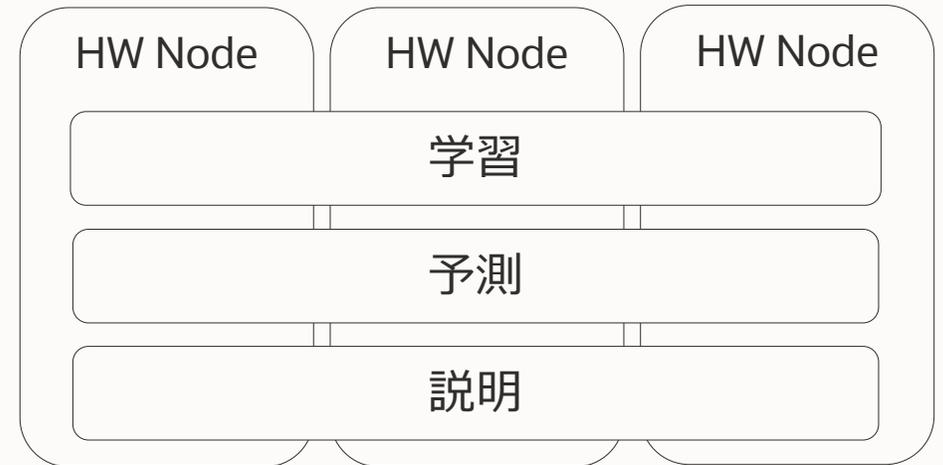
学習データのロード

モデル

テストデータのロード

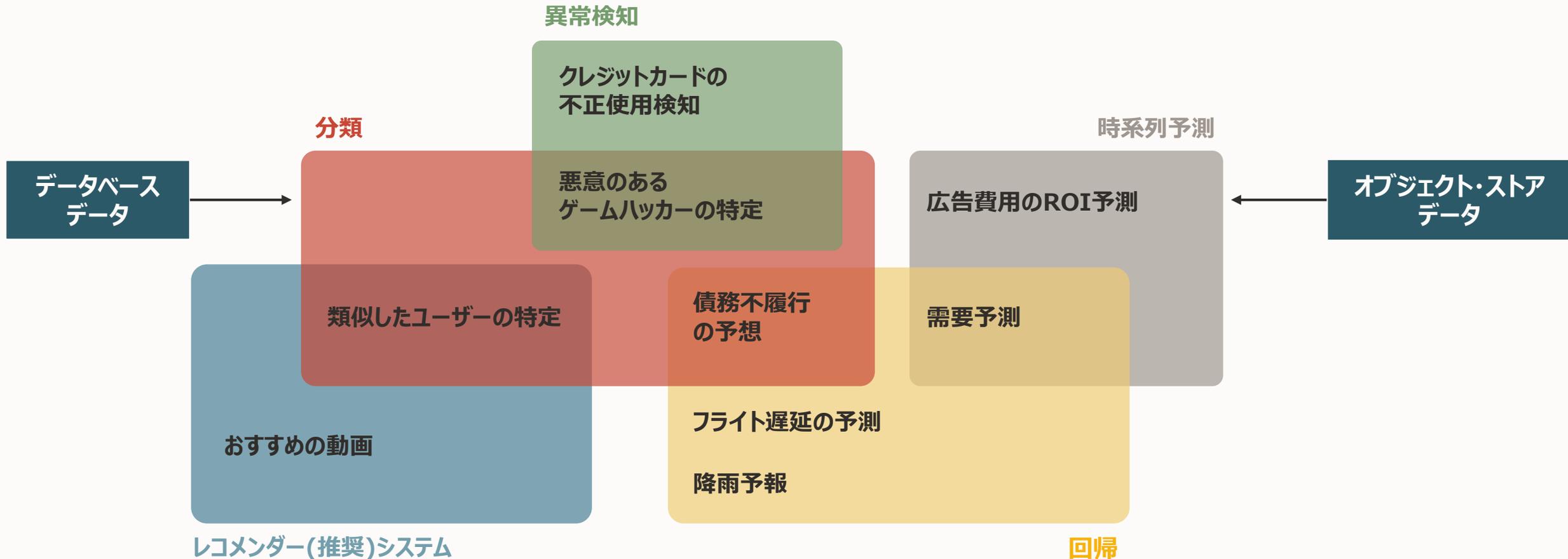
予測、説明

HeatWave Cluster



# 今あるデータを活用して機械学習を活用したアプリケーションを簡単に開発可能

- MySQL HeatWaveに追加されたAuto ML用のプロシーチャーを実行するだけでMLを実現
- 分類、回帰に加え、時系列予測、異常検知、レコメンダー（推奨）システムを実現可能



# HeatWave AutoMLのデモ

# HeatWave AutoML活用事例

# Aiwifi様によるHeatWave AutoML導入事例

- 機械学習の経験がないエンジニアがたった1ヶ月で機械学習を活用した処理を実装
- 実装済みのこと
  - WiFiに接続している人の行動に基づいて、非顧客、従業員を検出
  - WiFiに接続している人が興味のあることを予測
- 今後実装を検討していること：優れた顧客体験分析のための感情分析、行動予測、WiFiネットワークの障害検出

## Machine Learning

### Our (DB) tech journey

We were always keeping ML implementation in mind, but kept thinking about our roadblock - the database.

**The challenge: Develop and implement ML models that worked for our customers.**

Our development team took 1 month to learn and implement Heatwave AutoML - with no previous Machine Learning experience.

First class teams working together.

ORACLE + 

## Machine Learning

### What we've done

- Non-customers and employee detection based on behavior. (BETA)
- Interests prediction (WIP)

### How the future looks.

- Sentiment analysis for superior CX analytics.
- Behavior prediction
- WiFi network fault detection (Internet, Infrastructure, overall connection experience).



[www.aiwifi.io](http://www.aiwifi.io)

  
MySQL™  
HeatWave



# 某製造業大手によるHeatWave AutoMLを活用した生産性向上事例

- 機械学習モデルの「説明」機能を生産性向上に活用
  1. 製品不良に大きな影響を与えている要因を判断し、製品不良品率の改善
  2. 消費電力に大きな影響を与えている要因を判断し、消費電力削減
- IoTにより収集したデータをMySQL HeatWaveに格納し、HeatWave AutoMLを使って機械学習モデルを作成
- HeatWave AutoMLでは「説明」により、機械学習モデルを使った推論に大きな影響を与えている要因を判断可能
  - 要因毎の影響度を数値で確認可能

# 某製造業大手によるHeatWave AutoMLを活用した生産性向上事例

- 製品不良率改善の流れ
  1. 不良品判定時の各種データをIoTにより収集し、MySQL HeatWaveに格納
    - 例) 製品の容量、製品の温度、冷却装置の温度、ガス圧、など
  2. 収集したデータと不良品判定の結果を元に、HeatWave AutoMLで「**分類**」のデータモデルを作成(教師あり学習により、正常な製品と不良品の傾向を学習)
  3. 作成されたモデルに対する「説明」を確認し、どの要素が製品の良/不良の判断に大きな影響を与えているかを確認  
=> 大きな影響を与えている要因について、優先的に改善対策を取ることで、製品の不良品率を改善
- 消費電力削減の流れ
  1. 工場のある機械が稼働している時の各種パラメーターのデータをIoTにより収集し、MySQL HeatWaveに格納
    - 例) プレス圧、プレススピード、回転数、コンベアのスピード、など
  2. 収集したデータと消費電力のデータを元に、HeatWave AutoMLで「**回帰**」のデータモデルを作成(教師あり学習により、消費電力と工場稼働時の各種データの相関関係をを学習)
  3. 作成されたモデルに対する「説明」を確認し、どの要素が消費電力に大きな影響を与えているかを確認  
=> 大きな影響を与えている要因について、優先的に改善対策を取ることで、消費電力削減を改善



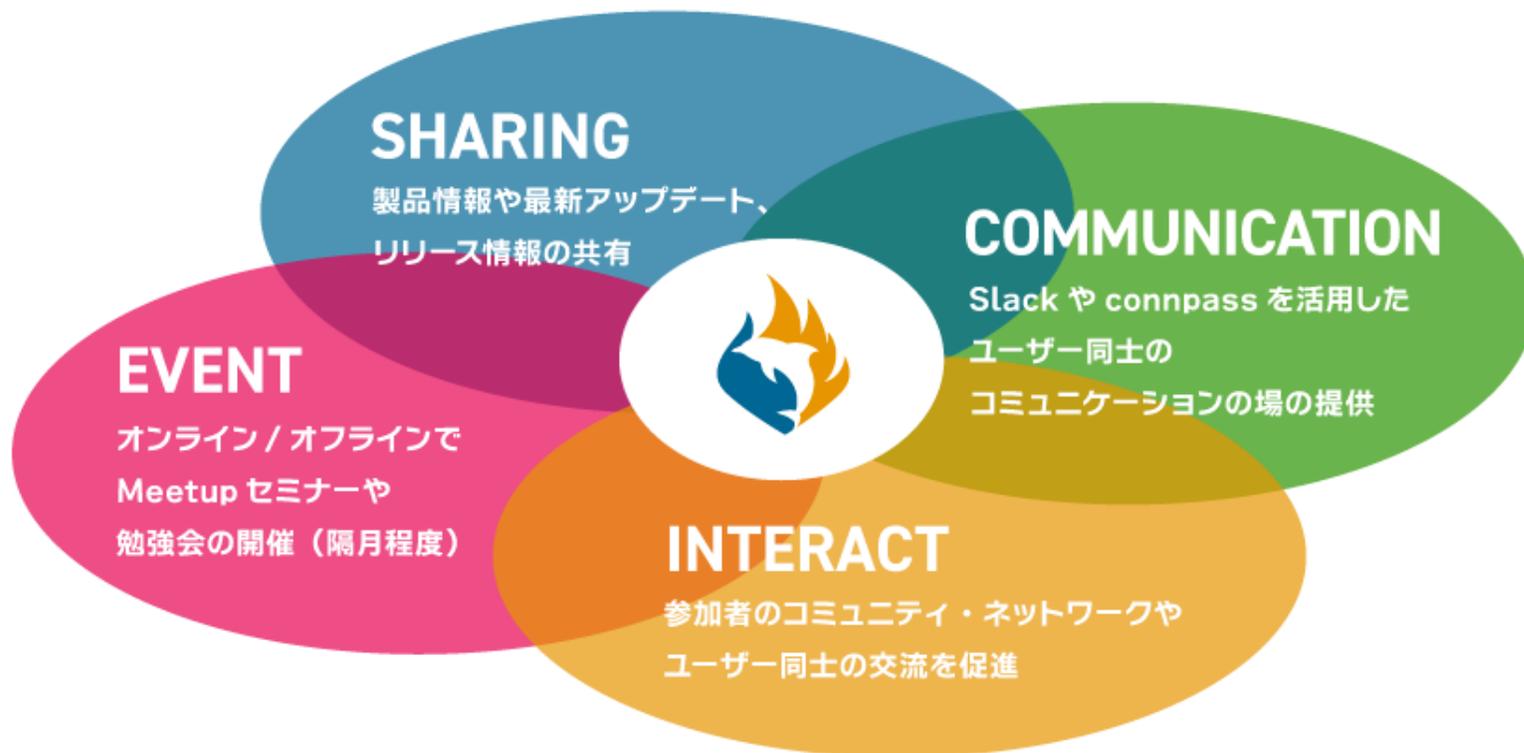
# Appendix



# HeatWavejp

MySQL HeatWave Japan User Group

HeatWavejp (MySQL HeatWave Japan User Group) は、MySQL HeatWave Database Service の日本での普及と発展を目的としたユーザーグループです。  
ユーザー同士でノウハウやナレッジを共有できるコミュニティ活動を行っていきます。  
MySQL HeatWave をより良く知り、学びあい、盛り上げていきましょう。



## ユーザーグループへの参加



[https://join.slack.com/t/heatwavejp/  
shared\\_invite/zt-1n3vprmlY-  
N3Si6Y7luGVXx9KNTmjdZg](https://join.slack.com/t/heatwavejp/shared_invite/zt-1n3vprmlY-N3Si6Y7luGVXx9KNTmjdZg)



## イベント情報 & イベント参加



<https://heatwavejp.connpass.com/>



## SNSフォロー & 情報収集



@HeatWavejp  
<https://twitter.com/Heatwavejp>



# HeatWave MySQL / HeatWave AutoML のドキュメント、参考情報

- HeatWave User Guide  
<https://dev.mysql.com/doc/heatwave/en/>
- HeatWave User Guide / HeatWave AutoML  
<https://dev.mysql.com/doc/heatwave/en/mys-hwaml-machine-learning.html>
- HeatWave User Guide / HeatWave Quickstarts / Iris Data Set Machine Learning Quickstart  
<https://dev.mysql.com/doc/heatwave/en/mys-hwaml-iris-quickstart.html>
- Oracle LiveLabs: Get started with MySQL HeatWave Machine Learning  
<https://apexapps.oracle.com/pls/apex/r/dbpm/livelabs/view-workshop?wid=3306&clear=RR,180&session=112536597636012>
- 機械学習でJupyter NotebookからHeatWave MLを使ってみよう  
<https://thinkit.co.jp/article/19676>



## MySQL問合せ窓口

TEL: 0120-065-556

Email: [mysql-sales\\_jp\\_grp@oracle.com](mailto:mysql-sales_jp_grp@oracle.com)

ORACLE