

## 第1章

# データサイエンス(機械学習のアルゴリズム)による データ解析が社会にもたらす変化



この章のゴール

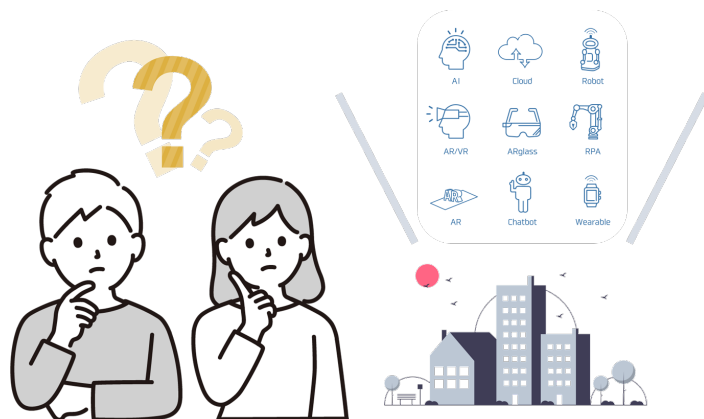
AI、テクノロジーをはじめとする最新技術の活用における社会的な変化を知り、「データサイエンス(機械学習のアルゴリズム)によるデータ解析」がどのように役立てられているかを理解する。

## はじめに(社会における変化)

### 社会における技術の進化とデータサイエンスの活用

人工知能(AI)、ロボティクスなどをはじめ、技術がめざましい進歩を遂げ、私たちの生活に大きな変化を与えています。

第1章では、「データ×AI 時代」と言われる現代において、社会でどのような変化が起きているのかを知り、自分たちが置かれている状況や現在地を把握します。AIをはじめとする最新テクノロジーが私たちの身近でどのように活用されているのかを理解しましょう。



### 新たな移動通信システム 超高速・超大容量のデータ通信 – 5G

新型コロナウイルス感染症の感染拡大が世界中で深刻な中、ひそかに 2020 年 3 月には、新たな超大容量通信システムとして通信キャリアによる「5G」のサービス提供開始が話題となりました。

引き続き、感染症拡大予防策の一環で人の移動が制約されている 2021 年現在、通信技術を活用したビジネスは拡大し、人々の生活においてもライフラインを支える重要技術としてより一層、発展が見込まれます。

人の動きが減少した一方で、モノの移動ニーズは拡大をしています。5G は、超大容量・超高速・超低遅延・超大量接続を実現するとされています。2020 年に急発展を遂げた非対面・非接触の技術と 5G の通信によって、まずは物流の世界から劇的な革新がもたらされています。5G の始まりによって、データを活用する様々なサービスに変革が起きようとしています。

## IoT が叶える、モノによるコミュニケーション

1999 年にイギリスの技術者によって提唱され、日本では 2016 年頃より、社会やビジネスにおける最新技術として注目されていた、「IoT:Internet of Things(モノのインターネット)」の技術を応用したサービスはこの数年で飛躍的に進化し、モノ・機械・人間それぞれの行動を情報収集し、つなぐことができるようになりました。

これらは一例ですが、こうした技術の根幹に関わり、今後の発展の鍵を握っているのがデータサイエンス、機械学習などと深いかわりのあるビッグデータ解析技術です。これらの分野の学問は限られた人にとってのものではなく、未来の社会を形成する人材が必須で持つべきリテラシーとも言われています。

最近では、専門知識がなくとも機械学習モデルを構築することができる技術も浸透してきており、AI を構成する機械学習の技術を「誰もが使える時代」が始まっているとも言えます。本書は、進化し続ける世の中の動きを捉えながら、データサイエンスの概念を理解し、修得できる仕組みになっています。



# 機械学習とデータ×AI の関係性

## データサイエンスとは

データサイエンスとは「アルゴリズムや統計などといった情報科学系の理論を活用してデータを分析し、有益な知見を見出すことを追究する新しいアプローチ」（「データサイエンティストのためのスキルチェックリスト/タスクリスト概説」独立行政法人情報処理推進機構（IPA）・データサイエンティスト協会（著）より）です。プログラミング・数学などの知識を要する側面が多くありますが、データサイエンスの実践において必要な能力や知識は、それらだけにとどまりません。

取り組む課題そのものについての理解力や想像力、解決を目指す上で何をどのように実行すべきかを計画立てる力、データから導き出された結果が有益な価値に繋がるかどうか判断するための論理的思考も欠かせません。データサイエンスのプロセスでは、課題に関する情報を適切に手に入れるためのコミュニケーション力が求められますし、導き出した解決策を提案して当事者に納得してもらうためのプレゼンテーション力も大切な要素です。

また、データサイエンスの工程をグループで行う場合には、チームワークや議論する力、課題を止めることなく推進させるためにはリーダーシップが必要になる場面も出てきます。データサイエンスの実行は、このように幅広い能力が求められる、学問としての分野の垣根を超えた技術です。

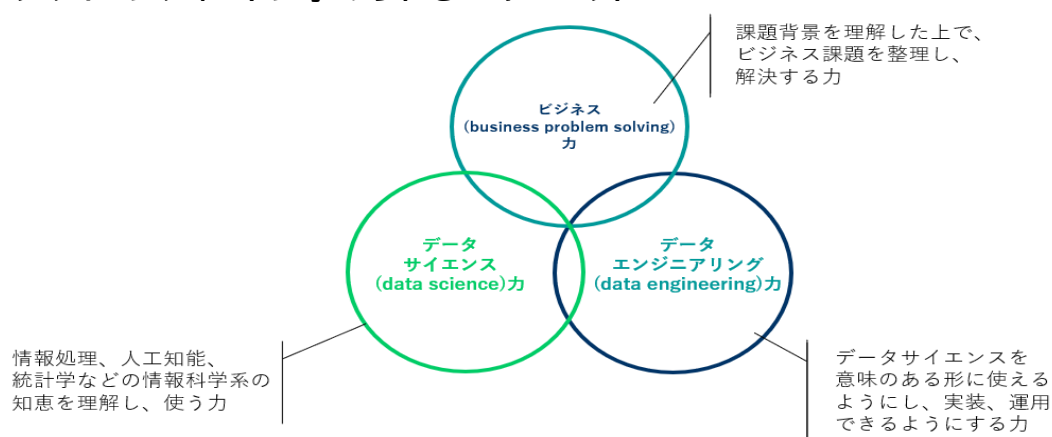
## 注目される職業「データサイエンティスト」

近年、注目の職業として話題となっているデータサイエンティストは、データサイエンス力、データエンジニアリング力、ビジネス力の 3 つのスキルをもとにデータから価値を創出し、ビジネス課題に答えを出すプロフェッショナル\*のことです。

\*ここで「ビジネス」とは社会に役に立つ意味のある活動全般を指します

様々な意思決定の局面において、データにもとづいて合理的な判断を行えるようにサポートします。ビッグデータ、機械学習などの言葉とともに、ビジネスにおけるデータサイエンスの重要性が高まり、同様にそれらを扱う職業も注目を浴びるようになりました。

## データサイエンティストに求められるスキルセット



出典：一般社団法人データサイエンティスト協会「3つのスキルセット」

URL: [http://www.datascientist.or.jp/symp/2019/pdf/1115-1155\\_skill.pdf](http://www.datascientist.or.jp/symp/2019/pdf/1115-1155_skill.pdf)

## 機械学習の基本概念

機械学習とは、AI(人工知能)を実現するための方法のうちの1つで、マシンラーニング(Machine Learning)とも呼ばれています。

コンピュータがデータを読み込むことで自動的にルールやパターンを見出し、その結果を活かして分類や予測を行う仕組み全体のことをいいます。従来、人間は過去の経験や認識能力に基づいて様々なものの分類や判断を行ってきましたが、その基準をコンピュータで再現することにより分類、判断を行うことが機械学習です。

人間が経験により得た特徴に加えて、人間が見つけることができない特徴を学習することができるこの仕組みは、様々な分野の課題を解決し技術の進化に寄与しています。

## 機械学習の歴史

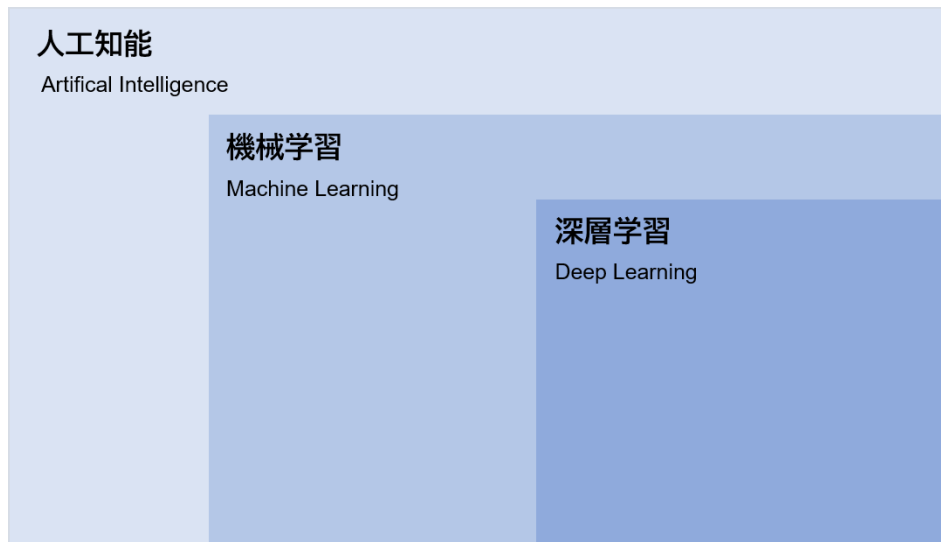
機械学習は、AI(人工知能)の研究の中で誕生したひとつの技術です。

AI が最初に注目されたのは 1950 年代と言われています。1956 年に、人工知能の学術分野をはじめて確立されたといわれるダートマス会議で、はじめてこの言葉が使われました。当時は、「コンピュータに人間の知識を搭載し、その知識を使って推論する」ことに論点をおいて研究が進められていました。現代のように、搭載された AI に「学習能力を持たせる」という発想はまだ行われていませんでした。

その後、1980 年代になると医療分野において人間の経験や知識をシステムに記録し、病気の診断に活用する取り組みが行われ、この時にはじめて、「人間の経験や知識を機械に持たせる」という視点での AI 研究が行われ始めました。一方で、この頃の AI は実用に十分な知識量のデータを取り込むという点において、例外処理や矛盾したデータへの対応などに課題を抱えていました。

それまでも何度かブームを引き起こし、その都度研究開発が活発に行われていた AI ですが、このような課題に幾度もぶつかってはブームが終息に向かうということを繰り返していました(「AI 冬の時代」と呼ばれています)。

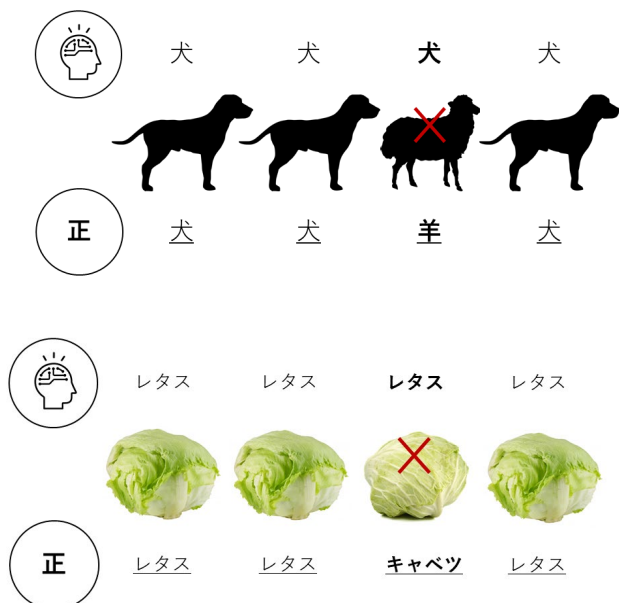
2000 年代になり、コンピュータ技術やデータ取得技術などの進歩によって、ディープラーニングや機械学習の技術を用いたビッグデータ解析に注目が集まりました。大量データを処理し、分類や予測などを実行できるようになったことが AI の進化を加速させ、現在の研究開発においても大きな原動力となっています。



## 機械学習の限界

多くの期待が寄せられる機械学習ですが、まだ完全なものとはいえません。場合によっては、機械学習よりも人間が行ったほうがよいとされる作業なども存在します。例えば、過去のパターンから全く割り出すことができないような未知のデータに出会った場合に、過去のデータで学習した機械学習システムにより予測を行うと、誤った判断結果を出す可能性もあります。

コンピュータである以上、あらかじめインプットされている「ルール」以上のことは実行できません。データの不足や偏りに対して、複雑すぎるモデルをあてはめることで誤った予測結果を導き出してしまう「過学習」や「過剰適合」など、機械学習にはいくつかの「限界」があります。



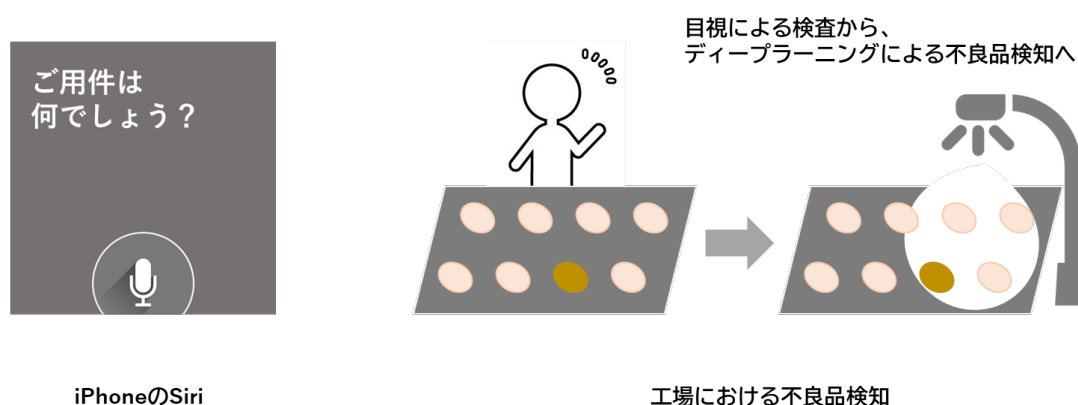
AI を使った画像データの学習における失敗例

# AI

AIとは **Artificial Intelligence** の略称で、日本語では人工知能と呼ばれます。

コンピュータにデータを与えることによってデータの特徴や法則性を見出すことのできるコンピュータのことを表しています。

近年、「AI が人間の仕事を奪う」という言葉が様々なメディアで話題となりましたが、AI は特徴や法則性をモデル化することにより、単純作業をこなしたりルールに則った作業をすることができるばかりでなく、人間のよようにその場の状況をみて判断を下したり、未来を予測することも可能です。人間の代わりに情報を調査、集約したり、人の手によって起こるうっかりミスのようなものを正確な技術でカバーすることもできます。



AI は大きく2つに分けると「汎用型(強い AI)\*」と「特化型(弱い AI)」の2つに分けられます。汎用型とは、適切にプログラムされた人間のような意識を持ち、総合判断を行える AI です。一方で特化型は、一定領域の業務に特化したタスク処理を行う AI を指します。

弱いAI	強いAI
人間のような知的処理の実現 人間の知能の一部を代替 機械的 自意識を備えていない	人間と同等の知能の実現 人間のようにものごとを認識 人のように仕事を行う 人間のような自意識を備えている
特化型	汎用型
特定の決まった作業を遂行 自動運転技術 画像認識、将棋・チェス チャットボット	特定の作業やタスクに限定せず人間と同様の、あるいは人間以上の汎化能力 特定の機能以外にも、応用して対応できるとされる

\*「強い AI」とは、人間のように思考して行動することができ、想定外の状況においても、過去の経験に基づいて問題を処理することが可能になると期待されています。ハリウッド映画などに出てくるヒーローロボットなどをイメージしていただくとわかりやすいです。ただし、これらを実現するには技術的な課題がまだ多くあり、もう少し先の未来になると言われています。

発展当初は、AI の根幹技術である機械学習の技法を使いこなすためには、ある程度のプログラミングをはじめとするデータエンジニアリング力や機械学習の数理的な理論への深い理解が必要とされてきました。しかし、現在は深い専門知識を持たなくても UI(ユーザーインターフェース)上から操作して機械学習を実行できるツールが数多く登場しています。

クラウドサービスには、様々な機械学習・AI のライブラリといわれる汎用性のあるプログラムが公開されています。多少のプログラミングスキルを身につければ、簡単な画像認識や音声認識は行うことができます。

このように、AI を構成する統計・機械学習を使うことのハードルは年々下がっており、「AI の民主化」が進んでいるとも言われています。

### ミニコラム:近年注目される、AI×倫理

一方、「AI による非公平性や差別」はしばしば大きな問題として報じられています。

人材採用においての AI 活用や、クラウドサービスを利用した写真の分類は、事例として大変有名ですが、データに偏りがあるかどうかを考慮して意思決定ができるか、AI にインプットするデータを様々な視点から偏りの無い状態で用意することができるかという点においては、まだ課題が残ります。特に倫理の面では、学習データにおける偏りが影響し、意図せず AI が差別的な判断を下してしまう事象が問題視されています。AI を扱う上でのリテラシーは技術面の理解だけでなく、倫理観も伴うものであることを忘れてはなりません。



# ビッグデータ

## ビッグデータとは

インターネット、通信技術、センサーなど IoT 技術などの発展により、自動的に大量のデータを収集、蓄積することが可能になってきました。このような技術の発展で集められたビッグデータは、価値を創出し未知の発見が可能になったことを意味しています。

ビッグデータの特徴は以下の「5 つの V」で説明されることがあります。価値創出の源泉となるその特徴は次の通りです。

Volume（容量の大きさ）	IoT の発達によって現代では膨大な量のデータを取得することが可能になりました。収集したデータは今後継続して集めていくことを想定し、活用しやすい状態にしておく必要があります。
Variety（多様性、種類）	数字だけではなく、テキスト、音声、画像、位置情報、センサー、SNS など様々な形態のデータを様々な場所から得ることができます。データが示す内容の種類も豊富で多岐にわたります。
Velocity（スピード、頻度）	世の中の移り変わりの速い現代では、時代に合ったデータ活用をするためにはリアルタイムに近い新しいデータを活用することが重要です。新しく更新頻度が高いデータは、時代に合ったサービスの提供を可能にします。
Veracity（正確さ）	正しいデータ活用のためには正確なデータが必要不可欠です。正確なデータを集めるためには、手元のデータが事実に沿ったものであるか見極める必要があります。
Value（価値）	データの活用は、社会に価値をもたらしたり、課題を解決することにより価値を創造します。

これらの特徴には、データを扱う上で注意すべき点や有意義なデータ活用のためにおさえておきたいポイントが数多く含まれています。データを扱う際には心得ておきたい事項です。



# IoT

## 取得されるデータはより細かく ― IoT はさらなる先へ

IoT:Internet of Things(モノのインターネット)を応用した技術はこの数年で飛躍的に進化しました。

今や様々な「モノ」に通信機能が搭載され、リアルタイムにデータが取得されています。

とりわけ各種センサーなどを用いて取得される情報はデータの粒度は、より細かくなっており、そのため蓄積されるデータの量も増えています。

さらには、IoT デバイスはウェアラブルデバイスとして「身につける」時代から、マイクロチップのようなものを使ってヒトそのものを認証媒体にするなど「体内に組み込む」時代へ変化を遂げようとしているとも言われています。

インプラントブルデバイスと呼ばれるこれらのデバイスから得られたデータは、機械学習・AI の技術と組み合わせることで、適切な健康管理や疾病への罹患リスクを未然に防ぐことなどが期待されています。



図:IoT を活用したモノの操作

# ロボティクス

## ロボティクスとは

ロボティクスとは、ロボットの設計、製作、制御の研究をする学問、およびビジネス活用の研究ですが、近年ではロボットに関する産業全般がロボティクスと呼ばれています。「ロボットテクノロジー」、もしくは「ロボット工学」と称されることもあります。ロボットは 1980 年代頃に産業用ロボットが現れた頃から普及し始め、現在ではお掃除ロボットやドローンなど、いわゆるアニメや映画で見られるような形態以外のものやソフトウェア型など、様々な形や用途に応じたものが社会で活躍しています。

## RPA とは

ロボティクスに関わりの深い言葉として「RPA」が挙げられます。RPA は“Robotic Process Automation”の略で、ロボットによる作業の自動化を指しています。事務作業など既に活用されている領域もあり、今後もその発展は期待されています。

## ロボティクスへの期待

人々がロボティクスに寄せる主な期待には、人手不足の解消や人間の負担の軽減などが挙げられます。また、人間では行うことが難しくてもロボットを用いれば容易にできることは数多く存在するため、物流、消防、撮影、医療、介護分野など様々な領域での活用が期待されています。

## ロボティクスとの共存

AI が搭載されたロボットのニュースとともに「AI に仕事を奪われる」ということが話題になることがあります。今まで人間がしてきたことを AI 搭載ロボットが行うことにより、人間はより「考える」ことを中心とした仕事に集中できる、と言い換えることもできます。また、未来には今はまだ存在しない仕事生まれる可能性は非常に大きいとされています。いずれにせよ、人口が減少傾向にある日本はこれから労働力の不足が懸念されますから、労働の一部を担ってくれるロボティクスに寄せる期待は大きく、救世主とも言えるでしょう。



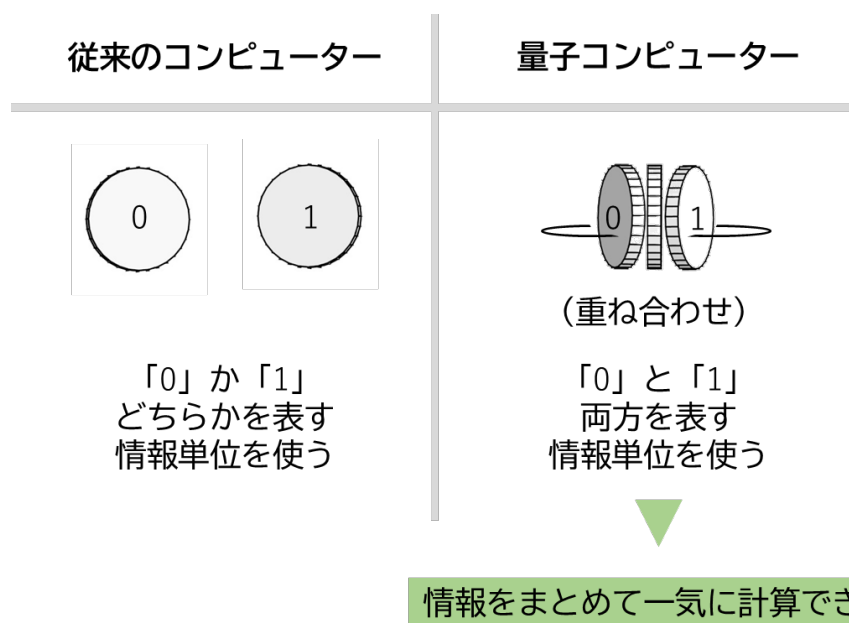
# 量子コンピュータ

## 量子コンピュータとは

量子とは物質を作る小さな単位のこと、主に物質を形作る原子、電子・中性子・陽子などのことを指し、光子やニュートリノやクォークなどといった素粒子も量子に含まれます。ナノサイズ(1メートルの10億分の1)あるいはそれよりも小さなサイズである量子の動きにはまだ解明されていないものが多く存在しますが、私たちの知る物理法則(ニュートン力学や電磁気学)には従わない独特の動きをするものが多いとされています。この量子の動きを研究する学問は「量子力学」と呼ばれ、量子コンピュータは量子力学の特徴を生かして計算するコンピュータの総称です。量子コンピュータに対して、私たちが普段使っているパソコンは古典コンピュータと呼ばれるものです。量子コンピュータは、古典コンピュータよりはるかに速い計算が可能とされています。

## 量子コンピュータの特徴

古典コンピュータの世界では情報の基本単位を「ビット」といい、その値は0か1のどちらかの状態で表現され演算が行われています。一方、量子コンピュータでの情報の基本単位は「量子ビット」といいます。量子ビットは重ね合わせという量子力学的な性質を持ち、0か1の2種類、もしくはそれ以上の状態を同時に合わせ持つことが可能です。これはたくさんの演算を並行して行うことを可能とします。



## 量子コンピュータの種類

量子コンピュータには大きく分類して量子アニーリング方式とゲート方式の2つの種類が存在します。量子アニーリング方式は、量子の動きを統計的に扱い、ミクロからマクロへの性質を導き出すことに長けています。これは「巡回セールスマン問題\*」に代表されるような組み合わせ最適化問題の演算に適しています。対してゲート方式の量子コンピュータは暗号の解読などいくつかのアルゴリズムの高速演算が可能であり、量子アニーリング方式よりも汎用的なコンピュータと言えます。どちらも古典コンピュータでは演算に時間がかかりすぎ現実的な実用に耐えられなかった演算を実用的な速度で行うことを実現しています。

\* 巡回セールスマン問題: 移動時間や交通費の最適を導き出すアルゴリズム

## 量子コンピュータへの期待

現代では量子コンピュータにしか解けないとされる問題が存在することが分かっており、その解決のためにも大きな期待が寄せられています。また、AI との組み合わせで自動運転、金融、医療など様々な分野でのロボットの開発も期待されています。

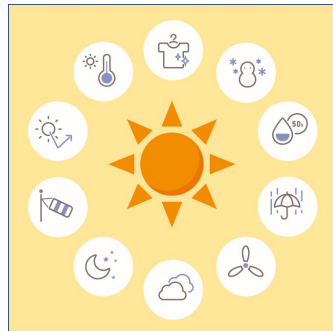
現在進行中の技術でもあります。クラウドを介して利用できるサービスも登場しており、これからより身近なものになっていくことでしょう。

# データとはなにか

機械学習を実行するにあたりデータは必要不可欠です。一口に「データ」と言っても様々な種類があり、その性質により扱い方も異なってきます。ここではデータの種類、扱い方について学んでいきます。

## 身近なデータ

私たちの身近には様々なデータがあります。天気(気温・降水量)、選挙の得票数や投票率などもデータのひとつです。



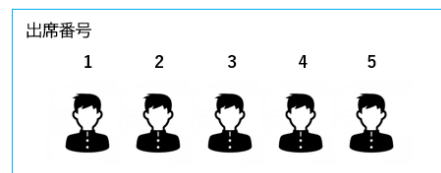
## データとは

データには、下記の 4 つの尺度があります。

### ①名義尺度…“分類”を目的とした尺度

例: 出席番号、映画館や劇場での座席番号など

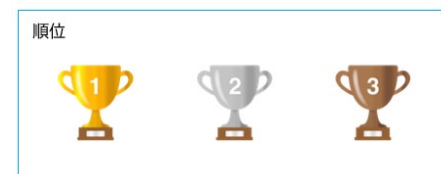
順序に意味がないもの。



### ②順序尺度…“順序”にも意味がある尺度

例: 「学年」「1 位・2 位・3 位」

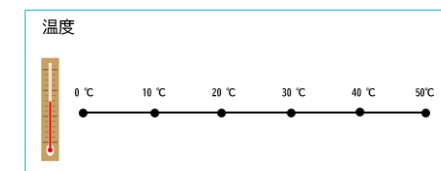
間隔は定かではないもの。



### ③間隔尺度…順序尺度で、かつ“間隔”にも意味がある尺度

例: 「温度」「時刻」

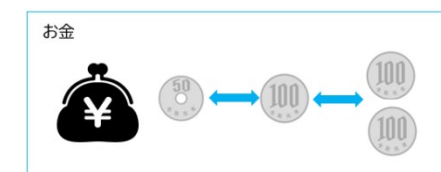
0 に意味がないもの。



### ④比尺度…間隔尺度で、差や比率に意味がある尺度。

例: 「身長」「金額」

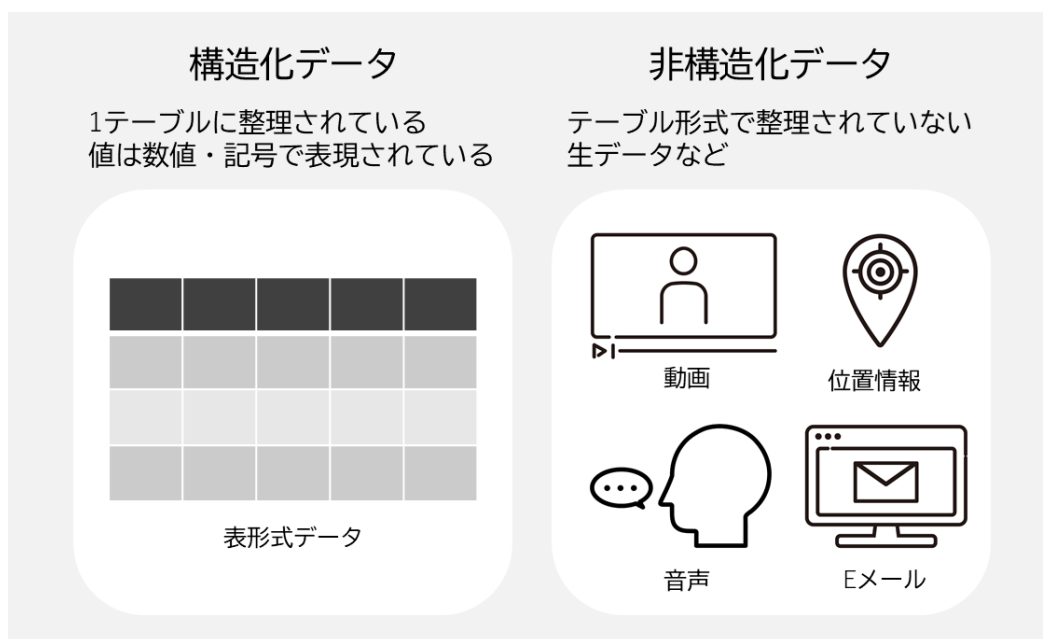
0 に意味があるもの。



## 構造化データと非構造化データの種類

### 構造化データと非構造化データ

データは、データベース上での取り扱いの観点などから「構造化データ」と「非構造化データ」の2種類に分けられ、どちらも様々な課題解決のために活かすことができます。各々の特徴はまた後述しますが、どちらもデータを取得してから AI に読み込ませる状態にするまでには複数の工程を経てデータの加工を行う必要があります。また、データをより使いやすく運用するためにはいくつか留意することがあります。

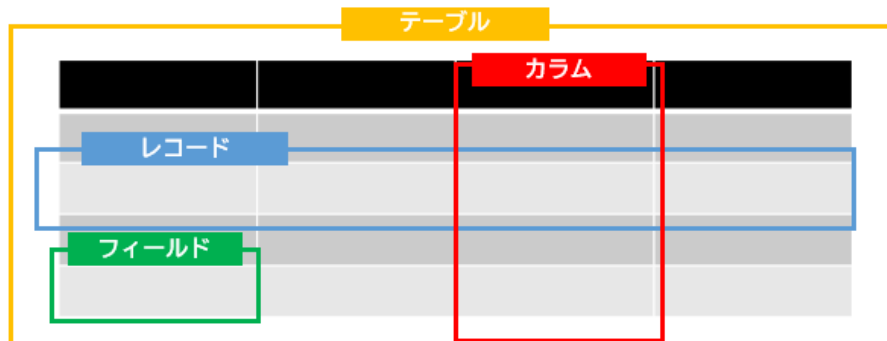
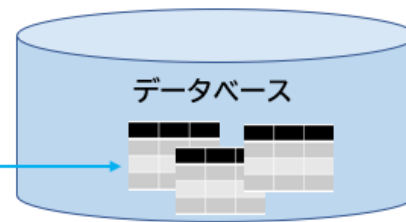


### 構造化データとは

構造化データとは、行と列の二次元、表形式に変換可能なデータで、格納される値は主に数値や文字で表現されています。ほとんどの場合、行(レコード)ごとに人や商品などを管理し、列(カラム)ごとに同じ意味合いの項目が格納され、一目瞭然で分かりやすい構造となっています。

ある事柄に関しての一つの構造化データのかたまりをテーブルといいます。また、複数のテーブルで同じカラムを持たせると、テーブルを結合させて別のテーブルで管理している情報を一元化して見ることができます。複数のテーブルを関係(リレーション)させているため、このような仕組みを RDB(リレーショナルデータベース)と呼ばれています。

リレーショナルデータベースは、データをテーブルで管理します。データベース内には、複数のテーブルを格納することができます。





## 非構造化データとは

非構造化データとは、音声・画像・テキストなど、そのままでは表形式に変換できないデータのことを言います。非構造化データを機械学習データとして活用するためには、かつては構造化データとして使えるように音声データを数ミリ秒ごとの波形(周波数)データに変換したり、画像データをピクセル単位の行列データに置き換えるという処理を行う必要がありました。

現在は、この構造化処理のほとんどは人間が行わずに、ダイレクトに処理する仕組みやクラウドサービスなどを用いることにより、容易に行えるようになっています。

さらには、構造化した先の処理、例えば音声から感情の起伏を読み取って感情別にスコアを算出したり、別のアクションに至る確率を予測する、といった仕組みも多く存在しています。



動画データ



音声データ



テキストデータ



位置情報データ

etc

## テキストデータ

テキストデータはアンケートの自由記述欄や SNS で書かれた文章などから取得することが多く、文字で構成されているため構造化データと思われることもありますが、表形式で表されているものではないため非構造化データに分類されます。テキストデータの中には、口語体や方言が混ざっていたり、一般的ではない専門用語や固有名詞など辞書には登録されていない表現が含まれていることもあります。また、一件のテキストデータには多くの情報が含まれているためそのまま構造化データとして扱うことはできず、構造化データにするための処理を行う必要があります。特に日本語は他の言語と比べて文字の種類も表現の手段も多く、この処理は複雑で大変なものです。

主な処理としては、形態素解析が挙げられます。形態素解析とは、文章を最小の単位である形態素に分解して、意味合いや品詞などを判別する自然言語処理の一つの手法です。形態素解析を行うとテキストは品詞ごとに分けられ、分析を行うことが可能な状態にすることができます。また、品詞ごとに感情極性値を与えた対応表なども存在していて、これは品詞をスコア化してネガティブなのかポジティブなのかを測ることができます。

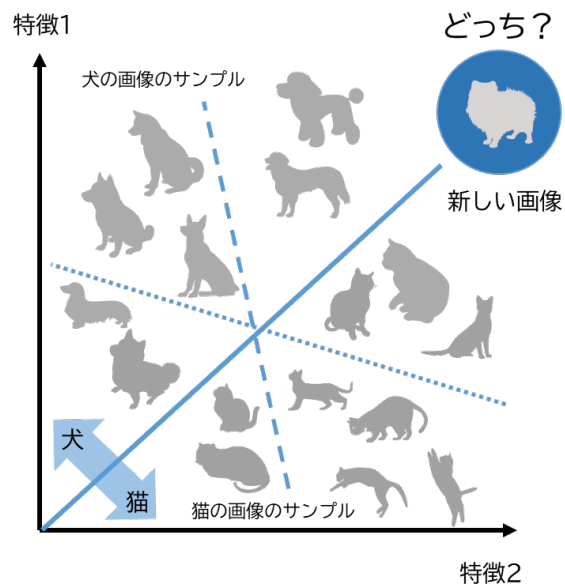
私は高校で機械学習の勉強をします



## 画像・動画データ

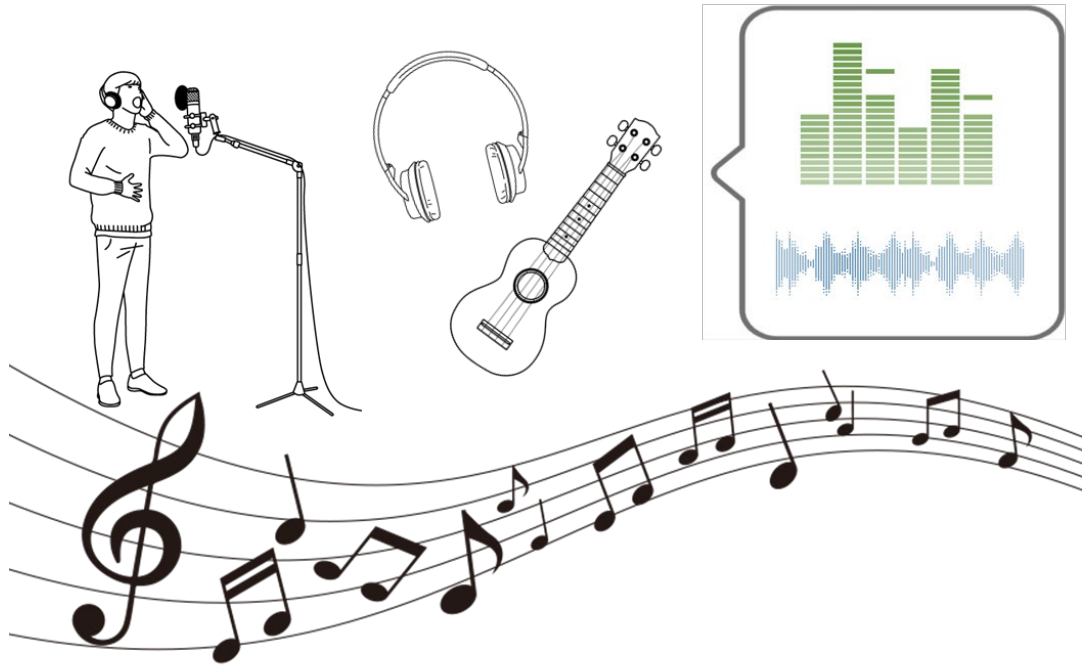
画像データもテキストデータと同様、そのままでは AI に読み込ませるデータとして使うことはできませんが、構造化処理をすることによりインプットデータとして用いられています。具体的には画像に映っている人物や場所の特定、動植物や商品などの種類の判別などです。また、手書きの文字からテキストデータを生成することも可能です。これらの構造化処理は今日ではいくつかのクラウドサービスで実現が可能となっています。

例えば、動画は 1 秒間に 30 回の画像を表示させることによって、連続した画像に見えるように構成されています。この 1 秒間に何回画像を表示させるかは、その動画のフレームレートによって決まっています。



## 音声・音楽データ

音楽および音声データも、構造化することにより AI に読み込ませるデータとして用いることができます。数ミリ秒ごとの周波数データに変換することにより話している内容をテキストデータとして生成することができますし、声の特徴から個人を特定することも可能です。



音楽データも構造化データへ変換させることが可能です。音楽データの中の音階やコードを判別したり、楽曲名を判別するなどして構造化データを生成することができます。

# AI の根幹となる技術・機械学習とは

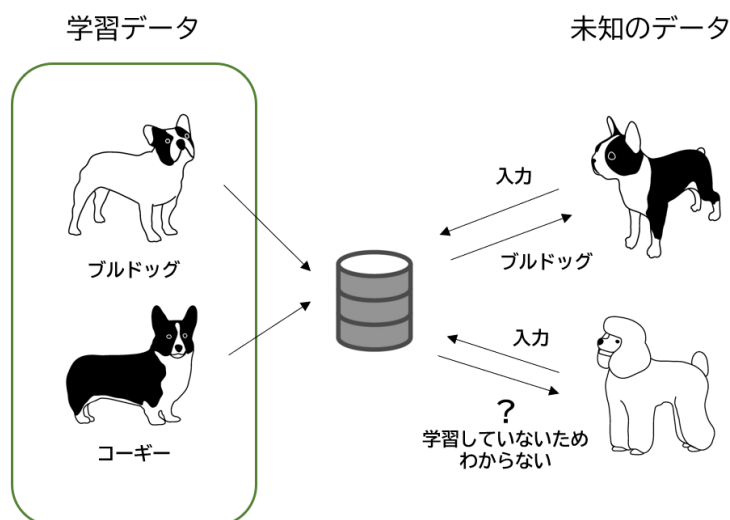
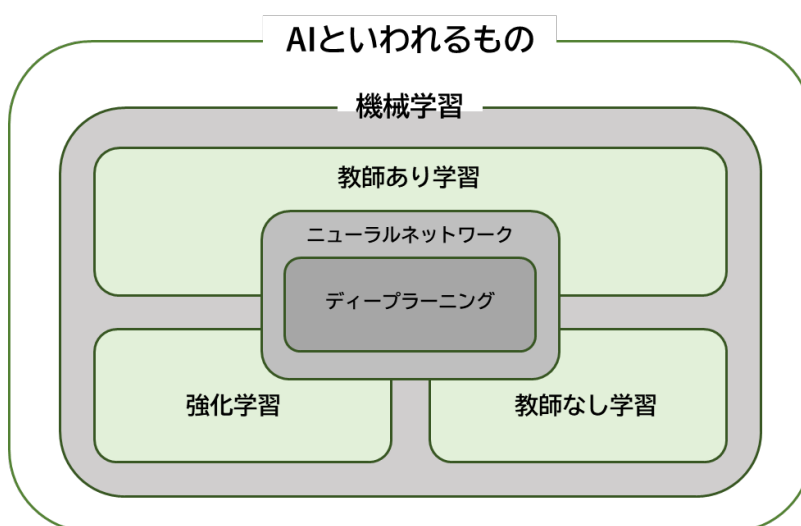
これから学んでいく機械学習とはどのようなものなのか。どんな特徴があって、何ができるものなのか？を学習します。機械学習を使って、将来どんなものが作れそうか、どんなものができたら嬉しいかなど、イメージを膨らませながら読んでみましょう。

## 機械学習とは

機械学習とは AI(人工知能)を実現するための 1 つの手段です。機械がデータを読み込んで学習することにより自動的にルールやパターンを見出されます。そのルールパターンを活かして分類や予測を行うことにより、金融・医療・物流など様々な分野での効率化や課題解決に役立てられており、AI(人工知能)など最新テクノロジーの開発や、ビジネスのデータを分析し活用するために必要不可欠な技術です。

機械学習には様々なアルゴリズムがあり、機械学習の対象に適したアルゴリズムを用いる必要があります。代表的なアルゴリズムの1つがディープラーニングです。ディープラーニングは画像認識の分野で画期的な成果を上げ注目されてから、今日ではそれ以外の分野でも成果を上げています。

機械学習には、データや解決したい課題の性質により大きく「教師あり学習」と「教師なし学習」の2つに分けることができます。



## 教師あり学習

教師あり学習とは、特徴を表す情報と正解を表す情報がセットになった状態のデータを用いてコンピュータに学習させる手法です。コンピュータに読み込ませる学習データは「特徴を表すデータ」と「正解を表すデータ」がセットであることが前提となります。特徴を表す情報は説明変数、正解を表す情報は目的変数と呼ばれています。

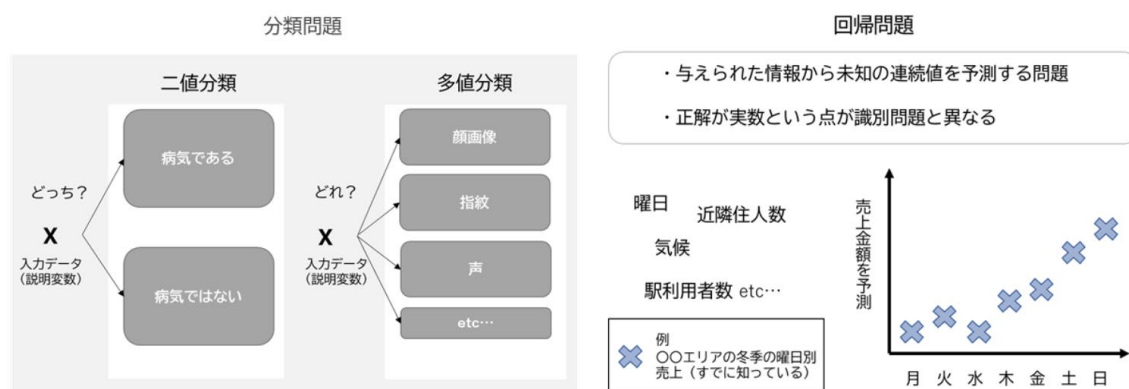


この手法で解く代表的なものが**分類問題**と**回帰問題**です。

分類問題は、説明変数の情報から「どのクラスに分類されるかという目的変数」を予測するものです。例えばレントゲンの画像から特定の病気にかかっているか否かを分類したり、顔画像、指紋、声の情報から個人を分類したりします。前者は目的変数が「病気である」「病気ではない」の2種類に分類されるため「二値分類」「二項分類」と呼ばれ、目的変数の値が3種類以上存在する場合は「多値分類」「多項分類」と呼ばれます。

分類問題の目的変数は数値ではなく、数値に置き換えたとしてもその値の大小には意味がないことが特徴で「ラベル」と呼ばれることもあります。一方、回帰問題とは説明変数の情報から目的変数に当てはまる数値を予測するものです。

テストの点数から志望校へ合格できる確率を予測したり、曜日や気候、近隣の住人数、駅の利用者数などの情報から店舗の売上金額を予測したりします。回帰問題の目的変数は、連続した数値です。



## 教師なし学習

教師あり学習では、説明変数と目的変数がセットになった状態のデータを学習データとして用いましたが、教師なし学習では説明変数のみを学習データとして扱いコンピュータに学習させる手法です。

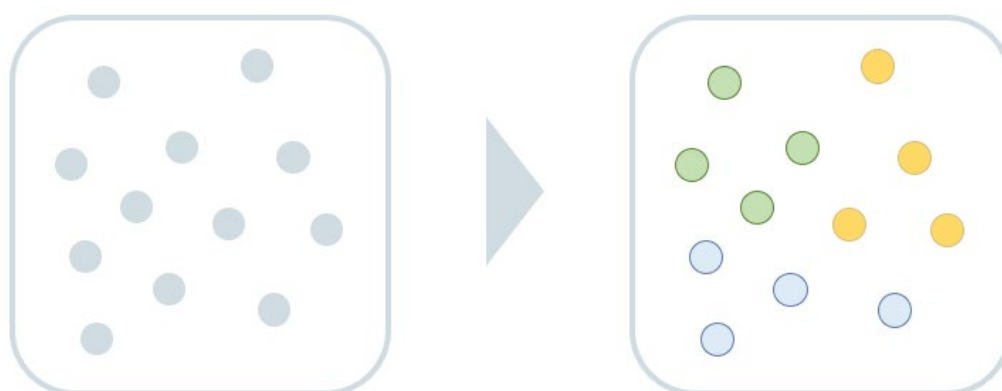
教師なし学習の主な手法には次元圧縮とクラスタリングがあります。次元圧縮は次元削減とも呼ばれ、数多ある説明変数をより少ない情報で学習データの意味合いを損なうことなく表現する手法です。説明変数の数が多いと学習データを読み解くことは大変ですが、次元圧縮を行った結果の学習データは情報量が少ないため把握しやすくなります。クラスタリングは学習データを似ているもの同士に分類する手法です。

### 次元圧縮のイメージ

氏名	国語	英語	社会	算数	理科		氏名	文系能力	理系能力
佐藤	42	45	43	48	91	➡	佐藤	55	92
鈴木	46	70	68	89	76		鈴木	77	87
田中	33	97	59	97	90		田中	62	91

### クラスタリングのイメージ

- 教育の話ばかりする人
- 料理の話ばかりする人
- アニメの話ばかりする人



教師なし学習で用いる学習データには目的変数が存在しないため、分析結果を見て人間が意味合いを解釈する必要があります。次元圧縮では圧縮された情報の意味合いを人間が考え、クラスタリングではクラスターごとの特徴を見てクラスターにラベルをつけたりします。意味合いやラベリングに妥当性があるか、納得のいく分析結果なのかといったことも人間が判断する必要があるため、学習データに関する知識が必要です。

## AI 最新技術の活用例

AI はどのように現代の社会に役立ち、私たちの生活に関わっているのでしょうか。業種別の活用例を見て、現代における AI の活躍を知り、AI の未来はどのようなものになるのか想像を膨らませましょう。

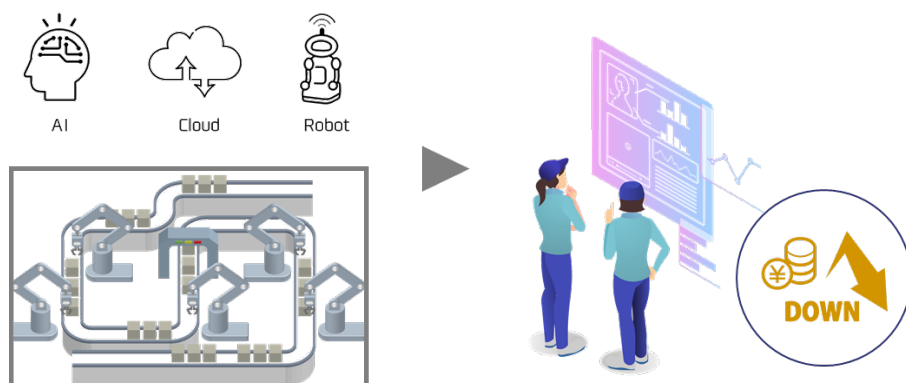
### 製造業

#### 原材料の在庫管理を AI で行い、仕入れコスト削減

原材料を仕入れて製品を生産する製造業では、原材料をなるべく安価に仕入れるとともに、多く売れ残ってしまったり、在庫がないために売る機会を逃してしまったりすることがないように、常に在庫を適正に保つことが課題です。

これまで原材料や在庫数の調整は、経験豊富な人による過去の経験や情報から判断して行われてきました。昨今では、ニュースや株価など原材料の情報から製品の需要や原料の価格を予測して、なるべく原材料を安く仕入れ、かつ在庫を抱えすぎないように製造量を調整する AI が導入され始めており、製造にまつわるコストの削減に成功しています。また、この予測結果を見ることにより、キャリアの浅い人の情報収集にも活かされています。

また、検品前に異常検知機能を用いて規格外の製品を除外して品質の低下を防ぐ、製造機器のチューニングの組み合わせをモデリングして品質の向上を図るなど、様々な面で AI が活躍しています。



### 医療

#### 患者のモニタリング結果から異常を検知。重症化を未然に防ぐ技術

医療分野では、レントゲン画像や超音波など従来のモニタリング結果から異常を検知して疾病の重症化や発病を未然に防ぐ機能を持つ AI が、身体の各所に対して稼働しています。海外では、バイオセンサーを使って慢性疾患患者の体調に対して常時モニタリングを行い、どのようなパターンになったら悪化する可能性が高いかを予測することにより、発症や入院を未然に防ぐ遠隔医療サービスが行われている国もあります。

医療技術の発展だけではなく、医療現場で働く人の事務処理の負担軽減、医療施設の効率的な活用支援など、様々な場面で患者や医療従事者の役に立っています。

AI による医療への貢献は、医療技術の発展はもちろんですが、地域による医療レベルの差の解消や人手不足が叫ばれている医療従事者への負担を軽減、医療施設を有効に活用する手段として、これからさらなる発展が期待されています。



## 道路、橋、送電線。私たちの生活を支える、街中にある AI

橋や送電線、道路などは、従来は人の目で点検して損傷具合を確かめ、修理をするかどうか判断をしていました。今では、ドローンに搭載されたカメラで撮影した画像をもとに損傷具合を AI が診断して修理が必要な部分を判断することができます。大きい橋は点検する箇所が多くたくさんの画像が必要なため、5G 通信が用いられています。

橋や送電線の点検は、これまで人員や時間、予算の確保が難しいとされていましたが、これらの問題を解消するとともに、今まで人により差があった診断結果が統一されるようになりました。また、人が立ち入ることが難しい場所や遠隔地の点検も今までより容易に行うことができる、天候に左右されず点検を行うことができるなど、AI の導入によるメリットは大きいです。

## 顧客への重要かつ複雑な説明を AI にお任せ！接客の簡略化へ

金融業界でも様々な場面で AI の活用が進んでいます。例えば銀行などでは、金融商品を販売する時に顧客に必ず行わなければならない説明がいくつかあり、そのチェックを今までは人間が行っていました。これを AI や RPA が代わりに行うことにより作業時間の大幅な減少に繋がりました。

また、債権回収時に延滞しそうな人を事前に予測する機能やクレジットカード会社の顧客情報や購入履歴から不正利用を目的とした詐欺や悪徳な顧客の特徴を抽出する AI を導入し、加入時の審査でこれらを検知して被害を未然に防ぐ取り組みなども生まれています。

これらは今まで人間の手で行っていたものですが、導入することにより業務時間や人間によるミスを減らすことで人間の作業時間の減少を生むだけでなく、行う必要のない業務を洗い出してやるべきことに注力したり、詐欺による被害を未然に防止するなど業務への効果は大きいものとなっています。

## 世の中のニーズに合わせて花の販売価格を自動調整

花は売ることができる期間が大変短く、たくさん仕入れすぎて売れ残りが発生したり、逆に在庫が不足して売る機会を逃してしまうことが多く、適正な数を仕入れるには長年の勘が必要でした。しかし最近では販売数の予測してくれる AI のおかげで、勘に頼らなくても適正な数の花を仕入れることが可能になっています。また、仕入先からの花の画像で花のランク付けが自動的にできるようになり、仕入れの効率化にも大きく寄与しています。



花以外のものについても、販売数の予測をすることにより大きな倉庫を借りずに常に適正な在庫を保つ、在庫が一定量より少なくなったら欠品になる前に自動的に発注を行う、混雑状況や天候などの状況により価格を変動させて需要や売上を安定させる「ダイナミックプライシング」など、AI が流通に関わる部分で貢献していることは多岐にわたります。