注目しています。その技術！

# 食品分野における機械学習：基礎知識と応用事例

宮川　弥生，橋田　泰彦

株式会社知能情報システム

## 1.　緒　　言

機械学習は，複雑なデータを解析することで新たな関係性を発見し，科学研究においても価値ある洞察をもたらす手段として認められている．一般に，統計的検定や単回帰分析といった古典的な解析手法によりデータの関係性を予測することは可能だが，機械学習アルゴリズムを活用することで，より複雑なデータに対して高い予測精度を実現できる場合がある．

本稿では，機械学習が従来の解析手法とどのように異なるのかを例を交えて平易に解説する．また，必要な基礎知識を併せて紹介する．なお，本稿においては機械学習という語を，昨今話題となっているAI，生成AI，深層学習といった関連技術を含む広義の意味で用いるものとする．

近年では，インターネット上の情報や生成AIの支援を活用することで，機械学習の実践方法に関する知識を容易に取得し，あるいは簡単な機械学習プログラムを実行させることができるようになった．しかしながら，実際に機械学習に取り組む段階になると，思わぬところで手詰まりを感じたり，得られた予測結果の解釈に難しさを覚えることも少なくない．

株式会社知能情報システムでは研究用ソフトウェアの開発やソフトウェア技術に関するコンサルティングを行っており，機械学習に関する問い合わせは近年増加している．このような背景を踏まえ，本稿では，機械学習を実践する際に留意すべき点や，つまずきやすい実践上の課題についても取り上げ，より実用的な観点から理解を深めることを目指す．

## 2.　機械学習は何がすごいのか

機械学習はデータ解析において非常に強力な手法であると広く認識されている．しかし，機械学習がどのような点で優れているのかを具体的に理解している読者は必ずしも多くない．本節では，具体例を用いて機械学習の優位性を平易に解説する．

初学者が機械学習を学ぶ際は，プログラミング言語としてPythonを利用するとともに，Pythonの機械学習ライブラリであるscikit-learnを利用するのが一般的である．ライブラリとは，プログラミングに必要な機能や便利な処理をまとめて提供するツール群を指す．scikit-learnは，機械学習アルゴリズムだけでなく，学習用の模擬データセットも提供している．本節では，模擬データセットの1つである「ワイン データセット」を用いる．

ワイン データセットは，ワインの化学的特性と品種情報を含む178個のサンプルから構成されている．化学的特性としては，アルコール濃度，リンゴ酸含有量，灰分，灰分のアルカリ度，マグネシウム含有量，総フェノール含有量，フラバノイド含有量，非フラバノイドフェノール類含有量，プロアントシアニン含有量，色の強さ（彩度），色相，吸光度比（280 nm/315 nm），プロリン含有量の13項目が測定されている．このデータセットを用いてどの特性を予測対象とするかは解析目的によって異なる．例えば，ワインの産地偽装を防ぎたい場合には化学特性をもとに品種を推定する課題となる．一方，成分分析を簡略化したい場合には比較的容易に測定できる特性（例えば色相）をもとに他の成分濃度を予測する課題となる．課題に応じて解析は柔軟に設計される．なお，対象データの特性を定量的な数値として表した機械学習モデルに入力するための情報を特徴量といい，予測対象の特徴量を目的変数（または従属変数）という．また，目的変数を説明するために用いられる特徴量を統計解析の文脈では説明変数（または独立変数）ということが多いが，機械学習の文脈では実務的にこれを特徴量とよぶこともある．特徴量の意味するところは文脈によって異なるが，本稿では後者の意味で使用する．

ここでは例として，アルコール濃度と色の強さを特徴量とし，古典的な統計手法と機械学習を使った場合のワインの品種を予測する場合を取り上げる．

線形判別分析（Linear Discriminant Analysis，LDA）

〒600-8815　京都市下京区中堂寺栗田町91 KRP9号館609号室

† E-mail: miyagawa.yayoi.public@chino-js.com



© 2025 Japan Society for Food Engineering

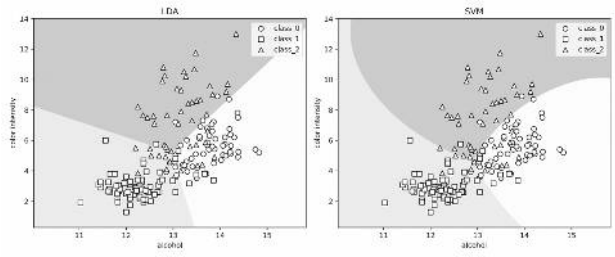



図 1　LDA と非線形 SVM によるワインの品種の分類結果
左図 : LDA による分類結果，右図 : 非線形 SVM による分類結果．シンボルはデータ点を表し，シンボルの違いはワインの品種（class_0 ～ class_2）の違いを表す．また，プロット エリアの背景色はモデルによる予測領域を示し，白の背景色は class_0，薄いグレーの背景色は class_1，濃いグレーの背景色は class_2 の予測領域を示す．

は，もともと統計学の分野で開発された古典的な手法である．現在では LDA も広義には機械学習モデルに分類されることがあるが，線や平面によってデータを分割する単純な仕組みに基づくため，複雑なデータ分布を正確に分類するには限界がある．一方，サポート ベクター マシン（Support Vector Machine, SVM）は，データの分布に応じて境界を自動的に学習し，柔軟に曲げることができる．このように，古典的な線形手法と現代的な機械学習モデルとの間には，表現力と柔軟性に大きな違いがある．これら 2 つのモデルで実際に解析した結果を視覚化した（図 1）．図 1 はデータ点を品種ごとに区別して表示した散布図と，それに基づく予測領域（各品種に分類されると予測される領域）とを示している．LDA を用いた場合には予測境界は直線で分離され，SVM を用いた場合には曲線的な境界が形成されているのが確認できる．また，この場合のモデルの性能を確認すると，LDA の場合の正解率（全体のデータのうち，正しく予測できた割合）は 0.80±0.06 で，SVM の場合の正解率は 0.84±0.05 であり，SVM の方がデータの分布をより正確に捉えられていることがわかる．なお，詳細は割愛するが，ここでは 5 分割クロス バリデーションという手法によって正解率を評価している。

このように，機械学習アルゴリズムは，従来の線形モデルでは表現できなかった複雑な関係性を学習し，高精度な予測を実現できる点で優れている．本節の例からも，機械学習の有用性が具体的に理解できるであろう．

### 3.　機械学習はどれほど簡便か

前節では，機械学習モデルが複雑な境界を表現できることを示した．本節では，そのような高性能なモデルの学習手続きが実際にはどれほど簡便であるかについて示す．まず，以下に機械学習モデルの学習と，その結果の可視化を行うための Python コードを示す．このコードには，必要なライブラリの読み込み，データの前処理，モデルの学習，予測結果の可視化に至るまでの一連の手続きが含まれている．

このコードからわかるように学習に必要な手続きは非常に少ない．実際，データの前処理とモデル学習に必要なコードはわずか 3 行であり，それ以外は主に可視化に関する部分である．コード中に付されたコメントを参照しても，学習に関連する処理は簡潔に記述されていることが確認できる．このように，ライブラリが整備されている近年は機械学習モデルの学習そのものは極めて簡単に実施できる．すなわち，ライブラリが提供する標準的な手続きに従うことで，高度なモデルを短時間で構築し，解析に応用することが可能となっている．

```
# 必要なライブラリを読み込む.
import matplotlib.pyplot as plt
from sklearn.datasets import load_wine
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.inspection import DecisionBoundaryDisplay
import pandas as pd
from matplotlib.colors import ListedColormap

# ワインデータセットを読み込む.
wine = load_wine()
wine_df = pd.DataFrame(wine.data, columns=wine.feature_
names)
X_plot = wine_df[["alcohol", "color_intensity"]]
y = wine.target

# データの前処理を行う.
scaler = StandardScaler()
X_plot_scaled = scaler.fit_transform(X_plot)

# 機械学習モデル (SVM) の学習をする.
svc = SVC(kernel="rbf", C=1, gamma="scale").fit(X_plot_
scaled, y)

# 学習結果を可視化する.
colors = ["#FFCC99", "#99CCFF", "#CCFFCC"]
cmap = ListedColormap(colors)
fig, ax = plt.subplots(figsize=(6, 5))
disp = DecisionBoundaryDisplay.from_estimator(
    svc,
    X_plot_scaled,
    response_method="predict",
    alpha=0.4,
    ax=ax,
    cmap=cmap,
)
ax.scatter(X_plot_scaled[:, 0], X_plot_scaled[:, 1], c=y,
cmap=cmap, edgecolor="k", s=40)
ax.set_title("SVM")
ax.set_xlabel("Alcohol")
ax.set_ylabel("Color Intensity")
plt.tight_layout()
plt.show()
```



© 2025 Japan Society for Food Engineering

## 4. 機械学習はなぜ難しいと感じるのか

機械学習の学習手続き自体は簡単であると前節で述べたが，実際に取り組み始めると多くの初学者が困難を感じる場面に直面する．本節では，機械学習が難しいと感じる主な要因について整理し，それぞれについて解説する．

### 4.1 データの量と質に関する課題

古典的な統計手法に比べ，機械学習を適用する場合はより多くのデータ数を必要とする．とくに，特徴量の数が多い場合にはそれに比例して必要なデータ数も増加する．一般に，特徴量が数十ある場合には数百件以上のデータが望ましいとされ，深層学習に至ってはさらに大規模なデータセットが必要とされる．たとえば，ChatGPT で用いられている大規模言語モデル（GPT-3）においては，数百億語規模のデータが学習に用いられたことが知られている [1].

データ数が一見多いように見えても，内容が類似したデータばかりである場合，実質的な情報量は限定される．また，目的変数の分布にも注意が必要であり，特定のカテゴリ（クラス）にデータが偏っている場合，バランスの取れた学習が困難となる．たとえば，ワインの品種予測モデルを構築する場合，複数の品種のデータが満遍なく揃っていることが望ましい．近年では，少数クラスのデータを人工的に増加させる手法（データ拡張）が発展しており，偏りが避けられない場合にはこれらを活用することも可能である．しかし，基本的にはバランスの良いデータセットを確保することが重要である．

さらに，データの欠損にも注意が必要である．欠損値を含むデータは原則としてそのままでは学習に利用できない．大量にデータを取得したつもりでも，欠損データを除去した結果，使用可能なデータが大幅に減少してしまうこともある．実験データにおいて完全なデータを得ることはむしろ稀であり，可能であれば適切な補完処理を施して活用することが望ましい．補完方法は，データの平均値等による単純な補完や，仮定した確率分布を利用した補完データのパターンを利用した機械学習による補完，など多数存在するため，データの特性に応じた手法を選択する必要がある．

また，外れ値にも注意が必要である．外れ値は学習モデルの性能を低下させる原因となるため，適切に検出し除去することが求められる．外れ値の検出手法としては，四分位範囲（IQR）による方法や，標準偏差に基づく方法などがあり，いずれもデータの特性に応じた選択が重要である．

このように，データの量，質，分布，欠損，外れ値といった問題は，機械学習に取り組む際に避けて通れない重要な課題である．

### 4.2 特徴量に関する課題

十分なデータが揃ったとしても，必ずしも高性能なモデルが得られるとは限らない．次に重要となるのが，特徴量の適切な扱いである．機械学習において特徴量に関する作業は，大きく特徴量選択と特徴量エンジニアリングに分けられる．

まず，特徴量選択（Feature Selection）とは，学習に寄与しない，あるいは冗長な特徴量を除去する作業を指す．たとえば，ある特徴量 A が，他の特徴量 B および C の単純な計算結果であり，新たな情報を持たない場合，特徴量 A はモデル性能の向上に寄与しないと判断され，削除されることがある．さらに，主成分分析（Principal Component Analysis, PCA）などにより各特徴量の寄与度を定量的に評価し，寄与度の低い特徴量を除去するアプローチも有効である．近年では，これらの作業を自動化するツールやライブラリも整備されている．

一方，特徴量エンジニアリング（Feature Engineering）とは，既存の特徴量を加工し，より有用な新しい特徴量を作成する作業を指す．たとえば，単独では意味をもちにくい複数の測定値を組み合わせることで，新たなパターンを捉える指標を構築することができる．適切な特徴量エンジニアリングにより，モデルの性能を飛躍的に向上させることが可能となる．

このように，特徴量を適切に選択し，必要に応じて新たな特徴量を設計することは，機械学習の性能向上において不可欠な工程である．

### 4.3 予測データの適切な準備

モデル構築後に留意すべき点として，予測に用いるデータの分布が挙げられる．学習時に使用したデータと，実際に予測に用いるデータの分布が大きく異なる場合，予測精度が大幅に低下することがある．とくに製薬や材料設計の分野では，この問題は重要視されており，「モデルの適用範囲（Applicability Domain）」とよばれている．適用範囲の判定には，k 近傍法（k Nearest Neibours, kNN）に基づく手法が一般的に用いられる．これは，予測対象データが学習データの近傍に位置しているかどうかを判定するものであり，適用範囲外のデータに対しては予測結果の信頼性が低下する可能性がある．

## 5. 機械学習に必要な知識

これまでに述べたように，適切なライブラリを利用するための基本的なプログラミング知識があれば，誰でも機械学習の実践を始めることができる．本節では，機械学習を実践するにあたって最低限必要となる知識



© 2025 Japan Society for Food Engineering

について整理する．具体的には，(1) ライブラリとプログラミング言語について，(2) 機械学習の基本ステップ，(3) 評価指標，(4) 数学的基礎知識，(5) 古典的手法との関係，の順に説明する．

### 5.1 ライブラリとプログラミング言語

実践的な観点からは，機械学習にはライブラリを使用するのが便利であり，以下のライブラリの使用方法を理解することが望ましい．これらのライブラリはPythonで提供されているため，Pythonの習得が推奨される．まずはこれらの基本的なライブラリを用いて機械学習を体験することで，機械学習がどのようなものかを直感的に理解することができる．

表1　機械学習で利用される代表的なライブラリ

| 用途 | ライブラリ名 |
|---|---|
| データの操作や修正 | NumPy，pandas |
| データの可視化（プロット等） | matplotlib，seaborn |
| 機械学習 | scikit-learn |

### 5.2 機械学習の基本ステップ

機械学習の一般的なステップは以下のとおりである．

(1) データの収集
(2) データの前処理
(3) モデル構築に使用する特徴量の選択
(4) モデルの選択
(5) 学習と予測
(6) 評価
(7) モデルの改善（パラメーター調整，再学習）

理想的には1回のサイクルで満足できるモデルが得られることが望ましいが，実際のプロジェクトでは (1) ～ (7) のサイクルを複数回繰り返すことが一般的である．

前述したワイン データセットの例においては，データの収集および前処理はすでに実施されており，アルコール濃度と色相を特徴量に選択した段階が (3) に該当する．また，SVMを用いて学習，予測を行ったことが (4) および (5) に対応する．次に，構築したモデルの性能を評価し，必要に応じて改善する段階が (6)，(7) にあたる．

### 5.3 モデルの評価指標

機械学習では，予測したい対象によって2つの問題に分けられる．一般に，予測対象がカテゴリ（例：品種，ラベル）である場合には「分類問題」，連続値（例：濃度，数値）である場合には「回帰問題」とよばれる．問題に応じてモデルの性能を評価する指標も異なるため，それぞれに適した指標を理解しておく必要がある．以下に分類問題と回帰問題それぞれの代表的な評価指標を紹介する．なお，本稿では簡易な説明にとどめ，詳細な定義や理論的背景については専門書やインターネット上の資料を参照されたい．

分類問題では，予測対象が複数のクラスに分かれる．とくにクラスごとのデータ数に偏りがある場合には，単純な正解率だけではモデルの性能を評価できない．例えば，100件のデータのうち90件がクラスA，10件がクラスBである場合，いかなる場合もクラスAと予測するだけで正解率は90%になる．しかしながら，このモデルはクラスBは予測できておらず，実用上は不十分である．このような場合でもモデルの性能を評価するためには，表に示した他の指標が重要となる．

回帰問題では，予測値と実測値の差を測る指標が用いられる．これらの指標を用いることで，モデルがどれだけ正確に連続値を予測できているかを客観的に評価することができる．一般に，外れ値の影響を抑えたい場合にはMAE（平均絶対誤差）を，大きな誤差を重視したい場合にはMSE（平均二乗誤差）を用いることが多い．通常はどちらか一方と，モデルを総合的に評価する指標である$R^2$（決定係数）によって評価を行う．

表2　機械学習における性能指標

| 問題の種類 | 性能指標 | 説明 |
|---|---|---|
| 分類 | 正解率 | 全体のデータのうち，正しく予測できた割合 |
| | 適合率 | 正解と予測したデータのうち，実際に正解であった割合 |
| | 再現率 | 実際に正解であったデータのうち，正解と予測した割合 |
| | F1スコア | 適合率と再現率を総合的に評価する指標 |
| 回帰 | 平均絶対誤差（Mean Absolute Error，MAE） | 予測値と実測値の絶対誤差の平均 |
| | 平均二乗誤差（Mean Squared Error，MSE） | 予測値と実測値の誤差の二乗平均 |
| | 決定係数（$R^2$） | 予測値が実測値をどの程度説明できているかを表す（1に近いほど良好） |

### 5.4 機械学習に関連する数学的基礎知識

機械学習はライブラリを利用するだけでも実践可能である．しかしながら，より高い性能のモデル構築を主体的に目指すためには，一定の数学的素養が欠かせない．とくに，機械学習に関連する数学は，アルゴリズムの学習原理やモデルの性能向上の仕組みを理解するだけでなく，データの前処理や特徴量選択，モデルの選択といった実践的な工程にも直結する重要な役割



© 2025 Japan Society for Food Engineering

表３　機械学習に必要な基礎数学のキーワードと内容

| キーワード | 内容 | 利用場面など |
|---|---|---|
| 確率統計 | 期待値，分散，相関係数，決定係数，確率分布，尤度，推定 | 特徴量の選択，モデルのもっともらしさの理解 |
| 線形代数 | ベクトル，行列，内積，固有値，テンソル | モデルの中身や学習の原理の理解 |
| 微積分 | 偏微分，連鎖律，勾配，勾配降下法 | モデルを最適化するための仕組み（学習の原理） |

を果たす．

以下に，主要な数学分野のキーワードとそれぞれが活用される場面を示す．これらの知識は一度にすべてを網羅する必要はないが，興味や必要性に応じて段階的に学び進めることが，より深い理解と実践力の向上につながるであろう．

### 5.5　古典的手法との関係

これまで機械学習の優れた点を紹介してきたが，古典的手法が不要になるわけではない．実際には，機械学習の実践においても古典的統計手法の理解は不可欠である．

古典的な手法では，あらかじめデータの予測式を明確に設定した上で解析を行う必要があるためモデルの解釈がしやすく，十分なデータ量が確保できない場合には，機械学習よりも古典的な回帰分析や判別分析のほうが効果的な場合もある．一方，第２節でも述べたように，機械学習モデルではデータに応じて柔軟に最適な境界やパターンを自動的に学習することができるため，より複雑なデータ構造にも高い精度で対応できる．例えるなら，古典的な手法では，近似に用いる多項式の次数や項数をあらかじめ人が決めてから学習を行う必要がある．それに対して機械学習モデルでは，どのような形状の関数が最適かをデータに応じて柔軟に学習し，必要に応じて近似の複雑さ（次数や項数に相当するもの）も自動的に調整することができる．

実際，データの前処理段階では特徴量間の相関関係を確認する作業が不可欠であり，PCA などの古典的手法による次元削減も頻繁に行われる．このように，機械学習と古典的手法は競合するものではなく，むしろ補完的な関係にある．両者を適切に理解し使い分けることが実践的なデータ解析において重要である．

## 6.　食品分野の研究紹介

本節では食品分野の機械学習の適用例を紹介する．

### 6.1　代替肉のテクスチャー予測

代替肉は，環境問題，健康，動物福祉の観点から近年注目を集めている非動物性タンパク質である．実際の肉に類似した見た目，テクスチャー，味を指向して開発されており，これまで試行錯誤的な手法によって多大な資源と時間を要してきた [2]．この課題に対して Kircali Ata らは，機械学習を用いて，代替肉にとって重要な特性である硬さ（hardness）と噛み応え（chewiness）を代替肉の組成情報から初めて予測することに成功し，効率的な代替肉開発のためのフレームワークを提案した [2]．また，構築したモデルにより硬さおよび噛み応えに寄与する特徴量を明らかにした．この研究では，まず実際に代替肉を作成し，テクスチャー アナライザーを用いて圧縮時の力およびエネルギーから硬さと噛み応えを計測した後，得られたデータに基づいてモデル構築を行った．

使用された特徴量は，灰分，炭水化物，脂質，食物繊維，水分，タンパク質，目標水分量の７項目である．機械学習モデルとしては，Ridge 回帰，ランダムフォレスト，XGBoost が用いられた．モデルの構築には，Python の scikit-learn および XGBoost ライブラリが利用された．

複数のモデルの性能比較にあたっては「Leave-one-out 交差検証」が適用された．これは，データ量が少ない場合に貴重なサンプルを最大限活用するための手法である．データから１つのサンプルを除外して残りのサンプルでモデルを学習した後，除外していたサンプルを用いて予測を行い，その性能を評価する．このプロセスをすべてのサンプルに対して繰り返し，最終的に各性能指標の平均値によってモデルの性能を評価する．

交差検証の結果，Ridge 回帰モデルが最も高い性能を示したと報告されている．また，噛み応えに関する特徴量の重要度分析においては，すべてのモデルで脂質および炭水化物が高い重要度を示し，これらの成分が噛み応えに大きく寄与していることが示唆された．さらに，特徴量選択を実施した結果，過学習が抑制される効果も確認された．加えて，特徴量の重要度には使用される植物タンパク質の種類が大きく影響することが指摘されており，異なる代替肉製品間での特徴量重要度評価については慎重に検討する必要があると報告している．

この研究は，従来は試行錯誤に頼っていた開発に，機械学習による合理化，効率化の道を示した点で意義深い．



© 2025 Japan Society for Food Engineering

### 6.2 ビールの複雑な風味の改善

Schreurs らは，250 種類の異なるビールを分析し，化合物の濃度から味，香り，口当たり，ならびに消費者の評価を予測する機械学習モデルを構築した [3]．学習に用いたデータセットは，1 種類のビールあたり 226 の化学特性（アルコール濃度, pH, 糖質など）を特定し，これらを訓練されたパネルによる官能試験結果および 180,000 人以上の一般消費者によるレビュー評価と結びつけた大規模なものである．このデータセットを用いて 10 種類の機械学習モデルを訓練した結果，Gradient Boosting が最も良好な性能を示した．さらに，この最も性能の高かったモデルを詳細に解析することで，ビールの風味や好みに寄与する化合物が特定された．

注目すべき点として，モデルによって重要なパラメーターとして特定された化合物の中には，これまでビールの風味要素として確立されていなかったものや，従来はビール品質に悪影響を及ぼすと考えられていたものも含まれていた．とくに，ビールの劣化に関連するエステルであるエチルフェニルアセテートが，消費者の好みに寄与する重要な要素として特定されたことは，興味深い発見である．

さらに，モデルから求めた重要度の高い上位 15 個の特徴量と，統計手法で求めた重要らしいと判定された上位 15 個の特徴量との相関を調べたところ，高い相関を示したのは 6 項目にとどまった．一部の化学成分では相関が非常に低く，従来の相関分析のみに依存していた場合には重要性が見逃されていた可能性が示唆された．この結果から，ビールの風味のような複雑な現象をより的確に捉えるためには，従来の単純な相関分析に代わるモデリング手法の有効性が示された．

加えて，重要度の高い化合物を複数添加したビールでは，消費者の評価が向上することが確認された．本研究は，飲料の香りの予測が極めて難しいとされる分野において，ビッグ データと機械学習を用いて化学特性，風味，消費者評価の関係性を明らかにし，優れた風味を持つ新たなカスタマイズ食品開発の基礎を築いたと著者らは結論づけている．

### 6.3 人工電子鼻技術（e-Nose）によるポテトチップスの酸化評価

（フライド）ポテトチップスは世界で最も人気のある塩味スナック製品であるが，脂質酸化が進みやすく保存性が低いため，脂質酸化の発生と進行を迅速かつ正確に検出する手法を開発することが食品安全と健康上の観点から求められている [4]．Bose らは，酸化の指標であるヘキサナールを e-Nose を用いて検出することで，ポテトチップスの酸化の検出を行う SMART（Specific, Measurable, Achievable, Realistic, Timely）手法の開発を試みた [4]．e-Nose（人工電子鼻）とは，人の鼻のように複雑な多種多様な匂いをリアル タイムで識別するセンシング技術のことである．酸化初期のヘキサナール濃度は低濃度であるため生化学的アッセイや水素炎イオン化検出器を用いたガスクロマトグラフィー（GC-FID）分析でも定量が不可能とされるが，e-Nose は極めて低濃度のヘキサナールを迅速かつ正確に定量できるとして着目を浴びている [4]．

まず e-Nose に搭載されたセンサー群の中から，酸敗マーカーであるヘキサナールに対して良好に応答するものを SVM を用いてスクリーニングし，選択したセンサーについて標準ヘキサナールを用いてキャリブレーションを行った．次に，包装後のポテトチップスを 一定条件で保存し，本研究で新たに定義した腐敗指数（spoilage indices）を保存期間中に測定した．この腐敗指数とヘキサナール濃度との回帰モデルを構築し，ポテトチップスの安全性（ヘキサナールの検出有無）と消費者の総合的な受容性（酸化に伴う風味変化をどこまで受容できるか）を同時に評価することが可能となった．従来必要とされていた GC-FID 分析や侵襲的な化学分析を省略し，フライ食品の状態を効率的に評価できる点で新たな手法と言える．この研究においてはセンサーのスクリーニングにおいて機械学習が利用されている．

## 7. 結　　言

本稿では，機械学習の優位性と実践に必要な基礎知識，ならびに食品分野における具体的な適用例を紹介した．機械学習はツールの発展により手軽に試せる一方で，高精度なモデル構築にはデータの質，特徴量の選択，適切な評価，古典的手法への理解が不可欠である．本稿が，機械学習に初めて取り組む読者にとって実践への指針となり，今後の発展に関心を持つきっかけとなることを願う．

## 引　用　文　献


1) T. Brown, B. Mann, N. Ryder, et al; Language models are few-shot learners. Adv. Neural Inf. Process. Syst., **33**, 1877-1901 (2020).
2) S. Kircali Ata, J. K. Shi, X. Yao, X. Y. Hua, S. Haldar, J. H. Chiang, M. Wu; Predicting the textural properties of plant-based meat Analogs with machine learning. Foods, **12** (2), 344 (2023).
3) M. Schreurs, S. Piampongsant, M. Roncoroni, et al; Predicting and improving complex beer flavor through machine learning. Nat. Commun., **15** (1), 2368 (2024).
4) A. Bose, N. Bhattacharyya, P. Bhattacharjee; A SMART methodology for assessment of hexanal in potato crisps using electronic nose technology: sensor screening by scalar machine learning classifier method. J. Food Sci. Technol., **61** (1), 150-160 (2024).




© 2025 Japan Society for Food Engineering