

2020年12月4日

## コンサルティングレポート

# AIによる商品購買予測とアソシエーション分析

## シリーズ「事業経営へのAI活用」④

経営コンサルティング部[大阪] チーフコンサルタント 山村 一夫

AI（人工知能）で何ができるのだろうか。AIは事業経営に積極的に活用できるのだろうか。どう活用すればよいのだろうか。シリーズ「事業経営へのAI活用」ではそれらについて考察していく。ぜひ、これらのレポートについて参照されたい。

### 【シリーズ「事業経営へのAI活用」タイトル一覧】

- ①AIによる退会抑止
- ②AIによる来店客数（売上）予測
- ③AIによる商品購買予測
- ④AIによる商品購買予測とアソシエーション分析 ……<本稿>
- ⑤AIによる新規出店集客数・売上予測

本稿では「AIによる商品購買予測とアソシエーション分析」を取り上げる。顧客への商品・サービスのおすすめが鍵となるビジネスモデルにとって、顧客の購買行動予測は大きな課題の1つである。AIを活用してデータを分析評価し、マス（集団）としてではなく個々の顧客の傾向を把握し、先んじて個別性の高い対策を講じることができれば大きなビジネス効果が見込める。

本稿の後半ではサンプルデータを活用したAIでの分析評価の例を紹介する。どの程度の分析ができるのか、施策への活用が可能なのかについて考察するとともに、ぜひ貴社での活用イメージを高めていただきたい。

## I. AIを活用した商品購買予測とアソシエーション分析

### 1. 要旨

- (1) AIによる商品購買予測では、会員情報や過去の購買履歴を利用して、それぞれの会員が将来に購入する商品を予測する。またアソシエーション分析は、たとえば「商品Aが買われるときに同時に買われやすい商品Bの組み合わせ」を見つけ出す手法である。
- (2) 今回は、機械学習コンペのKaggleのインターネットスーパーのデータセット<sup>1</sup>を使い、AIによる商品購買予測について考察した。データセットの分析後、LightGBMで商品購買予測を行い、並行してアソシエーション分析も実施した。データセットの分析・調整後、AIの作成・評価を行い、各種予測・分析

<sup>1</sup> <https://www.kaggle.com/c/instacart-market-basket-analysis>（2020年9月3日確認。Kaggleを見るためのブラウザについては、Internet Explorerでは閲覧できなかったが、Chromeでは可能だった。以下同じ。）

を行うという流れとなる。

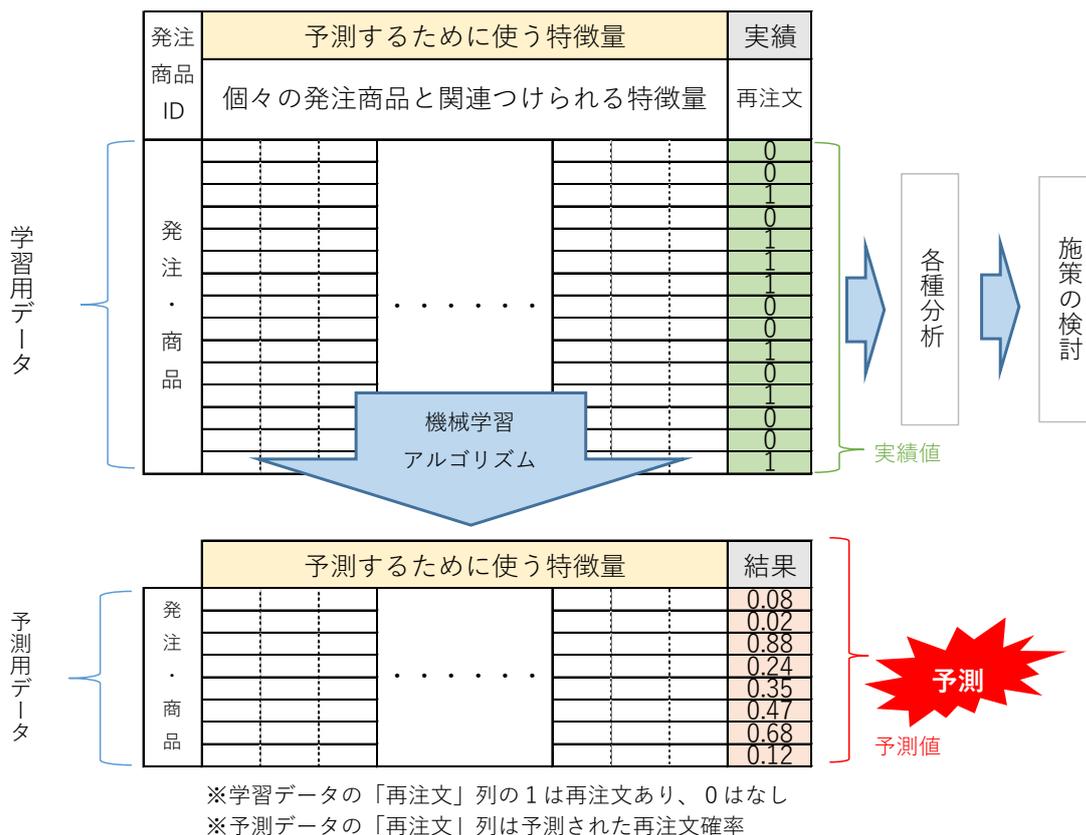
- (3) AI による個々のオーダーごとの商品購買予測確率を利用すると、販売促進策などの施策の費用対効果が高まる。アソシエーション分析を利用して、組み合わせ商品の推薦・近接に設置・セット販売などを行うと、増収につながる見込みがある。

## 2. AI による商品購買予測

### (1) 概要

- ① 会員情報や過去の商品購買履歴を利用して、AI でそれぞれの会員が将来に購入する商品を予測する。
- ② 分析に必要なデータは主に「個々の会員に関連した情報」・「過去の個人別・商品別の購買履歴」である。
- ③ 一般的に予測・分析の精度が高くなるのは、会員数が多い場合、会員に関連した情報・商品購買に影響のある情報が多い場合、データとアルゴリズムの相性が良い場合などである。

【図表 1】 AI による商品購買予測 イメージ



(出所) 当社作成

### (2) 業種応用例

- ① ダイレクトメールでの商品販売の場合、顧客情報がきっちりと紐づけされており、顧客にどの商品をどのタイミングでおすすめするのかの判断に役立つ。
- ② アパレルなどの個別商品の売上予測と在庫抑制が必要なビジネスの場合に、商品の発注・見切り販

売などに活用ができる。

- ③ インターネットでの商品販売などの場合は、ウェブ上の顧客の行動履歴なども特徴量（AI が学習・分析に使うデータの種類）に加えることができる。
- ④ 基本的には BtoC の小売業全体で利用できる可能性がある。

### 3. アソシエーションルール

#### (1) 概要

- ① 商品購買履歴を利用して、「アソシエーションルール」というセット販売の販促可能性の高い商品の組み合わせ「ルール」を見つける。
- ② 分析に必要なデータは主に「トランザクション別にどんな商品が販売されたか」の情報である。
- ③ 支持度（ビジネス上のインパクト）、確信度（組み合わせ商品の関連度）、リフト（組み合わせの目新しさ）などの指標基準の組み合わせによって、大量のデータの中から「ルール」を選別する。

#### (2) 業種応用例

- ① インターネットでの商品販売の商品推薦で利用できる。
- ② スーパーマーケットなどでの商品の棚割りやセット商品の検討に利用できる。
- ③ 基本的には BtoC の小売業全体で利用できる可能性がある。

## II. サンプルデータによる AI 分析評価

### 1. 【サンプルデータによる AI 分析評価】ステップ[1]データセット

#### (1) 内容(特徴量)

- ① 使用するデータセットは、個人向けスーパーマーケットの購買配達代行会社のオーダー情報である。
- ② 情報の中身は大きく「商品」に関する情報と「個別のオーダー」に関する情報に分けられる。

#### [1] 商品に関する情報

- a. データは「商品 ID」「department カテゴリ」「aisle カテゴリ」からなる。
- b. 商品点数は約 5 万種あり、それぞれ「department」（21 カテゴリ）と「aisle」（134 カテゴリ）に分類されている。
- c. たとえば「バナナ」は、「department カテゴリ」の”produce”で、「aisle カテゴリ」の”fresh fruits”となる。

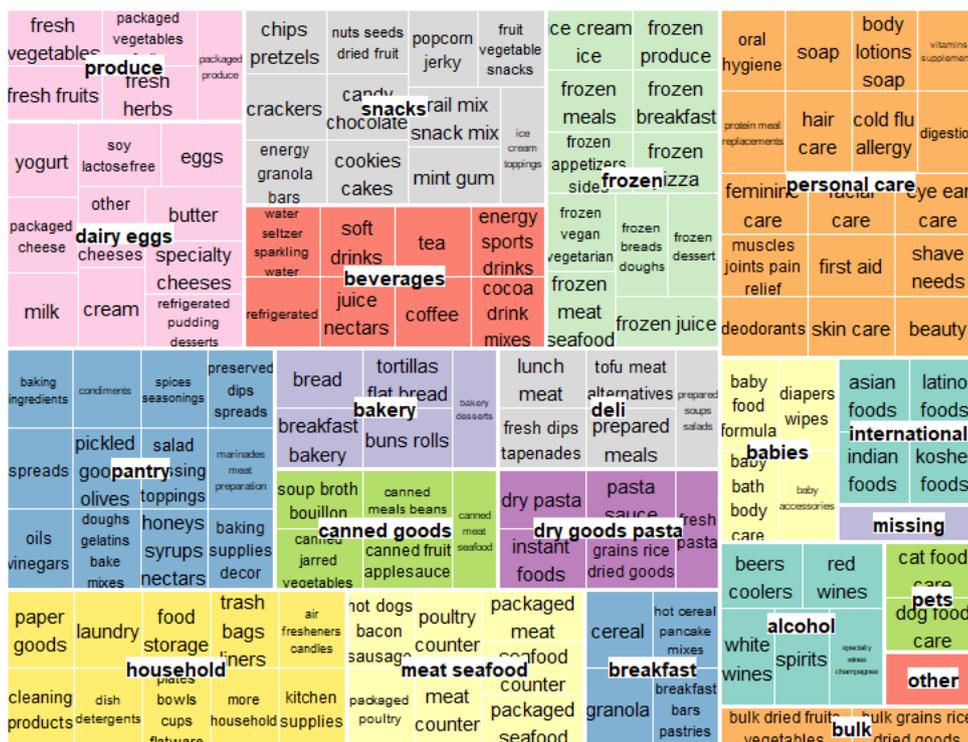
【図表 2】「商品 ID とカテゴリ」に関するデータ

product_id	department_id	aisle_id
1	19	61
2	13	104
3	7	94
4	1	38
5	13	5

(出所)「Instacart Market Basket Analysis」Data より<sup>2</sup>当社作成

<sup>2</sup> <https://www.kaggle.com/c/instacart-market-basket-analysis/data> (2020 年 9 月 3 日確認)

【図表 3】「商品カテゴリ」の全体像



(出所)「Exploratory Analysis – Instacart<sup>3</sup>」転載

## [2] 注文した「個々の商品」に関する情報

- 学習用データ数は約 3,500 万列ある。
- データは「注文 ID」「商品 ID」「商品を何番目にカートに入れたか」「次に同じ商品の注文をしたか否か」からなる。

【図表 4】「個々の注文商品」に関するデータ

No	order_id	product_id	add_to_cart_order	reordered
0	2	33120		1
1	2	28985		2
2	2	9327		3
3	2	45918		4
4	2	30035		5

(出所)「Santander Product Recommendation」Data<sup>4</sup>より当社作成

## [3] 個々の注文(トランザクション)に関する情報

- AI 学習用データは約 342 万列ある。
- データは「注文 ID」「顧客 ID」「同一顧客の注文回数」「注文した曜日(“order\_dow”列で 0~6 の数字で曜日を表現 (たとえば土曜日は 0))」「注文した時間 (24 時間の中)」「前回の発注から経過した日数」からなる。

<sup>3</sup> <https://www.kaggle.com/philippesp/exploratory-analysis-instacart> (2020 年 9 月 3 日確認)を転載

<sup>4</sup> <https://www.kaggle.com/c/santander-product-recommendation/data> (2020 年 9 月 3 日確認)

【図表 5】「注文」に関するデータ

order_id	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order
2539329	1	prior	1	2	8	NA
2398795	1	prior	2	3	7	15
473747	1	prior	3	3	12	21
2254736	1	prior	4	4	7	29
431534	1	prior	5	4	15	28

 (出所)「Santander Product Recommendation」Data<sup>5</sup>より当社作成

 (2) EDA<sup>6</sup>(探索的データ解析)

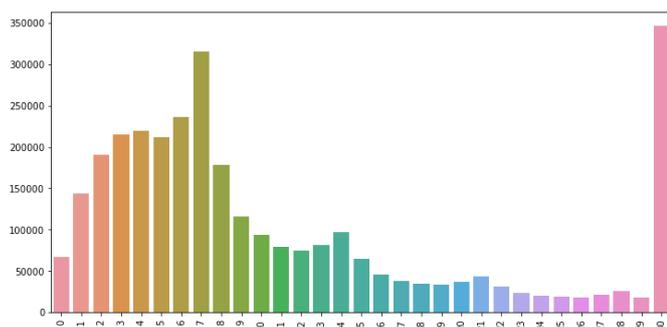
## ① 再購入までの日数

7日ごとにピークがあるが、日数経過ごとに徐々に少なくなる。

## ② 「商品の購入点数」と「再注文率」

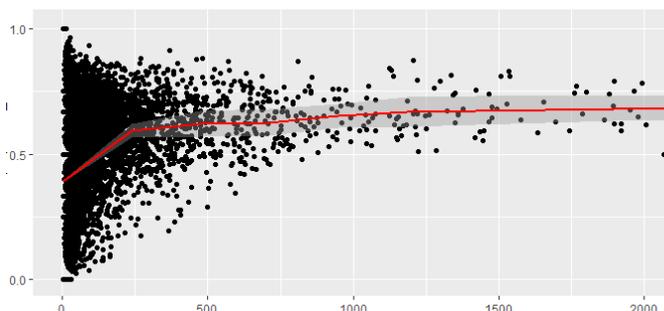
商品の購入点数が 200 を超えると、再注文率は約 0.6 (60%) となる。

【図表 6】 (左図)再購入までの日数



(縦軸:購入点数 横軸:再購入までの日数)

(右図)「商品の購入点数」と「再注文率」



(縦軸:再注文率 横軸:注文点数)

 (左図出所)「Instacart Simple Data Exploration<sup>7</sup>」転載

 (右図出所)「Exploratory Analysis – Instacart<sup>8</sup>」転載

## 2. 【サンプルデータによる AI 分析評価】ステップ[2]AI の作成と評価と分析

 (1) AI の作成<sup>9</sup>と評価・分析

- ① 今回使用するアルゴリズムは LightGBM である。これは他のアルゴリズムと比べて性能がよい場合が多く、データの前処理の手間も少ない為を選択した。
- ② 顧客 75,000 人の注文・商品ごとに、再注文の有無を予測した。
- ③ テストデータの正解(再発注の有無)はわからないので、学習用データの一部をテストデータと見立てて、予測を行い、実績と比較した。

## (2) 個別の顧客が新たに購入するサービスの予測確率

<sup>5</sup> <https://www.kaggle.com/c/santander-product-recommendation/data> (2020 年 9 月 3 日確認)

<sup>6</sup> Exploratory Data Analysis

<sup>7</sup> <https://www.kaggle.com/serigne/instacart-simple-data-exploration> (2020 年 9 月 3 日確認)を転載

<sup>8</sup> <https://www.kaggle.com/philippsp/exploratory-analysis-instacart> (2020 年 9 月 3 日確認)を転載

<sup>9</sup> <https://www.kaggle.com/paulantoine/light-gbm-benchmark-0-3692> (2020 年 9 月 3 日確認)を参考にして分析・加工

- ① 前述のデータセットから特徴量を作り出し、顧客別・商品別の再注文の予測確率を算出し、実績と比べた。
- ② 顧客に販売促進する場面では、顧客が商品を購入しようとする場面で、過去の商品注文の中で再注文の可能性が高い商品をリコメンドとして提示することが考えられる。

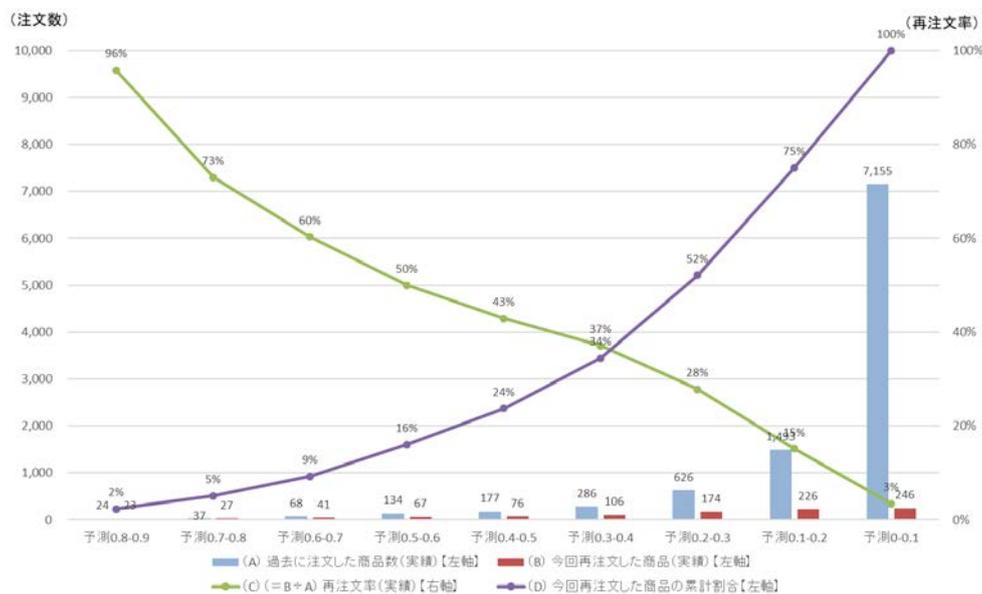
**【図表 7】 AIによる顧客別 サービス購入予測確率**

特徴量														再注文	
注文ID	商品ID	顧客の全注文数	顧客の全注文商品数	顧客注文商品の種類	再注文までの日数	顧客の注文あたりの商品点数	注文時間	前回注文からの日数	商品カテゴリ(aisle)	商品カテゴリ(department)	全商品の注文数	全商品の再注文数	全商品の再注文率	予測確率	実績
663263	25	37	309	109	10.1	8.4	16	9	3	19	2,166	1,391	0.64	0.01	0
3232985	25	5	54	30	21.3	10.8	11	20	3	19	2,166	1,391	0.64	0.37	1
1436428	26	86	1315	396	4.3	15.3	16	4	41	8	334	247	0.74	0.01	0
461898	34	7	90	63	13.8	12.9	11	11	121	14	6,536	3,928	0.60	0.06	0
1318886	34	9	84	59	21.1	9.3	13	30	121	14	6,536	3,928	0.60	0.07	0

(出所) 当社作成

**(3) AIの個別商品ごとの再注文率予測**

- ① 全データの中から1万サンプルを抽出し、予測確率と実績を比較してみた。(下記図表)
- ② 予測確率が0.8~0.9の商品の場合、24件の商品が再注文されるとの予測に対して、23件の予測が的中(0.96の再注文率)した。
- ③ 再注文される商品の52%へ提示するには、予測確率0.2~0.3までの商品に対して実施すればよい。

**【図表 8】 再注文の予測と実績**


(出所) 当社作成

#### (4) アソシエーション分析の実施<sup>10</sup>

##### ① アソシエーションルールの絞り込み

- [1] アソシエーションルールの組み合わせは大量にあるので、「支持度 (support)」「確信度 (confidence)」「リフト (lift)」での指標で一定の基準を設けて、分析対象を限定する。
- [2] 「支持度」は、全注文の中でこのルールが含まれる注文のある割合である。「支持度」が高いと収益インパクトが高くなる。「確信度」は、A または B を含むデータの中でこの「ルール」がある割合である。確信度が高いと、A と B との関連性は高いといえる。「リフト」は、「確信度 ÷ 全データの中でこの商品を含む割合」である。リフトが高ければ、この「ルール」は魅力的と言える。

##### ② やや新鮮味に欠けるが販売量の多い商品の「組み合わせ」

###### [1] 分析対象の限定

- a. ある程度売れている商品ペアを選ぶために、商品 A と B の組み合わせの支持度を 0.1 以上として、収益インパクトを高めた。
- b. 顧客への推薦効果を高めるために、既に同じコーナーにあると思われる同一カテゴリ (“aisle”, “department”) 商品を対象から除外した。
- c. 「リフト値」上位 5 位を抽出した。

###### [2] 商品の組み合わせ

- a. No1 と No2 はどちらもフルーツの組み合わせだが、「ラズベリー」「いちご」は生鮮食品で、「ブルーベリー」は冷凍食品である。「ベリー系のケーキ」や「フルーツ盛り合わせ」などのレシピで使われるだろう。No3 の「アボカド」と「塩分なしの黒豆缶詰」の組み合わせは、「黒豆、コーン、アボカド、プチトマトをオリーブオイルであえたサラダボウル」のようなレシピが想像される。No4/No5 の「フムス<sup>11</sup> (中東の料理でひよこ豆とゴマをペースト状にしたもの)」と「プチ人参」の組み合わせなどが出てきた。
- b. リフト値は高くないのでありきたりな組み合わせだが、レシートで購入品目から顧客の家庭のレシピが想像でき、購買意欲をそそるレシピ提案ができると面白いかもしれない。

<sup>10</sup> <https://www.kaggle.com/datatheque/association-rules-mining-market-basket-analysis> (2020 年 9 月 3 日確認)を参考に分析・加工

<sup>11</sup> <https://ja.wikipedia.org/wiki/%E3%83%95%E3%83%A0%E3%82%B9>

【図表 9】「違う売り場での組み合わせ(リフト(lift)値上位 5 位)」と「家庭で作られるレシピイメージ写真」

No	A			B		
	A_product_name	A_aisle	A_department	B_product_name	B_aisle	B_department
1	Raspberries	packaged produce	produce	Blueberries	frozen produce	frozen
2	Strawberries	fresh fruits	produce	Blueberries	frozen produce	frozen
3	Organic Avocado	fresh fruits	produce	No Salt Added Black Beans	canned meals beans	canned goods
4	Original Hummus	fresh dips tapenades	deli	Organic Peeled Whole Baby Carrots	packaged vegetables fruits	produce
5	Original Hummus	fresh dips tapenades	deli	Organic Baby Carrots	packaged vegetables fruits	produce

(出所) 当社作成


 (出所) 「iStock<sup>12</sup>」より当社が上記図表に合う写真を選択

### ③ 販売量は少ないが新鮮味のある「組み合わせ」

#### [1] 分析対象の限定

- 商品 A と B の組み合わせの支持度が 0.001 以上で収益インパクトの低いものも含めた。
- 顧客への推薦効果を高めるために、既に同じコーナーにあると思われる同一カテゴリ (“aisle”, “department”) 商品を対象から除外した。
- アソシエーションルールを適用し、「リフト値」上位 5 位を抽出した。

#### [2] 商品の組み合わせ

- No1 はクランベリーアップルジュースとクイックオーツ、No2 がダブルミントガムとダイエットライムソーダ、No3 は無漂白中力粉とオリーブオイルスプレー、No4 はツナとシーザーサラダ、No5 はチョコレートバーとライトエナジードリンクである。
- リフト値は高いので、組み合わせの訴求力は高いと言える。違う売り場に置いてあるが、セット等にしておすすめできるようにしたい。

【図表 10】 違う売り場でそれほど売れない商品の組み合わせ(リフト(lift)値上位 5 位)

No	A			B		
	A_product_name	A_aisle	A_department	B_product_name	B_aisle	B_department
1	Cranberry Apple Juice Cocktail	juice nectars	beverages	Quick Oats	hot cereal pancake mixes	breakfast
2	Slim Pack Doublemint Gum	mint gum	snacks	Diet Tangerine Lime Premium Soda, Zero Calories, Sugar Free, No Caffeine	soft drinks	beverages
3	All-Purpose Unbleached Flour	baking ingredients	pantry	Extra Virgin Olive Oil Spray	oils vinegars	pantry
4	Hickory Smoked Tuna Creations	canned meat seafood	canned goods	Salad Kit, Light Caesar	packaged vegetables fruits	produce
5	Chocolate Love Bar	candy chocolate	snacks	Lite Energy Drink	energy sports drinks	beverages

(出所) 当社作成

<sup>12</sup> <https://www.istockphoto.com/jp> より写真を購入

### 3. 【サンプルデータによる AI 分析評価】ステップ[3]施策への展開

#### (1) 商品推薦による売上の増加

- ① 顧客 ID が特定できる場合、購入予測に基づき個々の顧客に対して商品推薦を行うと、顧客の商品サービスの購入率を高め、売上・利益を増加させる可能性がある。
- ② 顧客 ID が特定できない場合、アソシエーション分析で抽出した組み合わせ商品を選択するとよい。

#### (2) 商品推薦の効果の検証

- ① AI を導入することによるビジネス上の評価は、商品推薦による売上・利益の増加と、AI 導入に伴い発生する商品推薦のコストの増額を比較すればよい。たとえばダイレクトメールや、対面販売や電話による販売促進を行う場合はコストがかかる。ウェブ上であっても、コストは無料というわけではない。
- ② 費用対効果の高め方としては、予測の確率が高い場合だけコストをかけた販促を行うのもよい。また、利益額の高い商品に限って、コストをかけた推薦を行う方法もある。

#### (3) データセット特徴量の増強

- ① 今回のケースでは過去の購買実績だけを使って、商品の購入可能性を予測している。同じデータセットを使っている限り、AI 精度の向上には限りがある。
- ② 分析顧客数を増やす方法もあるが、顧客に関する情報などをデータセットに統合し、特徴量の種類を増やすことは有効であると思われる。

#### (4) 商品のおすすめ方法の工夫

「アソシエーションルールを使った」の結果の使い方は重要だ。商品のおすすめ方法は、店頭の「POP」、「チラシ」、ウェブでの「画像・映像」、商品の「セット化」、「割引」などだが、おすすめ方法によっても、売上が大きく変化する。A/B テストのように複数の方法を試して、売上の高さを比べるとよいだろう。

#### (5) 新商品は早期に分析

新商品と相性のよい組み合わせ商品は、「アソシエーションルール」を活用して、商品投入段階で別途素早く見つけたい。新商品のあるレシート（注文）を分析することで、よい「組み合わせ」ルールが見つかる場合がある。

#### — ご利用に際して —

- 本資料は、信頼できると思われる各種データに基づいて作成されていますが、当社はその正確性、完全性を保証するものではありません。
- また、本資料は、執筆者の見解に基づき作成されたものであり、当社の統一的な見解を示すものではありません。
- 本資料に基づくお客様の決定、行為、及びその結果について、当社は一切の責任を負いません。ご利用にあたっては、お客様ご自身でご判断くださいますようお願い申し上げます。
- 本資料は、著作物であり、著作権法に基づき保護されています。著作権法の定めに従い、引用する際は、必ず出所：三菱UFJリサーチ&コンサルティングと明記してください。
- 本資料の全文または一部を転載・複製する際は著作権者の許諾が必要ですので、当社までご連絡ください。