

2020年11月30日

コンサルティングレポート

AIによる商品購買予測

シリーズ「事業経営へのAI活用」③

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 ……<本稿>
- ④AIによる商品購買予測とアソシエーション分析
- 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」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. 商品購買予測におけるAI活用

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将来に購買する商品（または利用するサービス）を予測する。
- (2) 今回は、機械学習コンペで有名なKaggleの「銀行のサービス利用」のデータセット¹を使い、AIによる将来に購買する商品（今回は利用するサービス（以下同じ））について考察する。入手したデータセットの分析後、XGBoostで会員別の商品購買予測を行った。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・分析を行うという流れとなる。

¹ <https://www.kaggle.com/c/santander-product-recommendation/overview>（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では可能であった。以下同じ。）

(3) プロジェクト成功のポイント

① 商品購入率を高める「施策」

- [1] 利用予測確率ごとに投入できる「施策」を用意する。予測確率の違い等によって投入すべき「施策」は異なるからだ。
- [2] 「施策」結果の検証を行い、「施策」がどれだけ顧客のサービス利用率を高めたのか、そしてその費用対効果を検証する。
- [3] 検証を踏まえて、「施策」の更新をしつつ、新たな「特徴量」を追加し、AIの予測精度を上げるサイクルは必要である。

② プロジェクト体制

- [1] プロジェクトのメンバーには、情報系だけではなく、営業系・企画系など、幅広い人材が必要になる。
- [2] 情報担当者以外の社内のキーマンが、取り組みを深く理解し、必要なフォローができるとよい。

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、2015年1月～2016年5月までの、スペインの銀行の顧客約96万人からなる約1,364万行分のサービス利用情報である。
- ② 顧客情報は、「会員ID、性別、就労状況、居住地域、世帯所得、顧客セグメント」など24項目からなる。
- ③ 利用サービスは、「普通預金、長期預金、クレジットカード、年金」など24種類からなる。

【図表2】(左図)顧客の特徴量一覧

No	特徴量名 (スペイン語)	特徴量の説明
1	fecha dato	The table is partitioned for this column
2	ncodpers	Customer code
3	ind_empleado	Employee index: A active, B ex employed, F filial, N not employee, P pasive
4	pais_residencia	Customer's Country residence
5	sexo	Customer's sex
6	age	Age
7	fecha_alta	The date in which the customer became as the first holder of a contract in the bank
8	ind_nuevo	New customer Index. 1 if the customer registered in the last 6 months.
9	antiguedad	Customer seniority (in months)
10	indrel	1 (First/Primary), 99 (Primary customer during the month but not at the end of the month)
11	ult_fec_cli_1t	Last date as primary customer (if he isn't at the end of the month)
12	indrel_1mes	Customer type at the beginning of the month .1 (First/Primary customer), 2 (co-owner), P (Potential), 3 (former primary), 4 (former co-owner)
13	tiprel_1mes	Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer), R (Potential)
14	indresi	Residence index (S (Yes) or N (No) if the residence country is the same than the bank country)
15	indext	Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country)
16	conyuemp	Spouse index. 1 if the customer is spouse of an employee
17	canal_entrada	channel used by the customer to join
18	indfall	Deceased index. N/S
19	tipodom	Address type. 1, primary address
20	cod_prov	Province code (customer's address)
21	nomprov	Province name
22	ind_actividad_cliente	Activity index (1, active customer; 0, inactive customer)
23	renta	Gross income of the household
24	segmento	segmentation: 01 - VIP, 02 - Individuals 03 - college graduated

(右図) 利用サービス名

No	サービス名 (スペイン語)	サービスの説明
1	ind_ahor_fin_ult1	Saving Account
2	ind_aval_fin_ult1	Guarantees
3	ind_cco_fin_ult1	Current Accounts
4	ind_cder_fin_ult1	Derivada Account
5	ind_cno_fin_ult1	Payroll Account
6	ind_ctju_fin_ult1	Junior Account
7	ind_ctma_fin_ult1	Más particular Account
8	ind_ctop_fin_ult1	particular Account
9	ind_ctpp_fin_ult1	particular Plus Account
10	ind_deco_fin_ult1	Short-term deposits
11	ind_deme_fin_ult1	Medium-term deposits
12	ind_dela_fin_ult1	Long-term deposits
13	ind_ecue_fin_ult1	e-account
14	ind_fond_fin_ult1	Funds
15	ind_hip_fin_ult1	Mortgage
16	ind_plan_fin_ult1	Pensions
17	ind_pres_fin_ult1	Loans
18	ind_reca_fin_ult1	Taxes
19	ind_tjcr_fin_ult1	Credit Card
20	ind_valo_fin_ult1	Securities
21	ind_viv_fin_ult1	Home Account
22	ind_nomina_ult1	Payroll
23	ind_nom_pens_ult1	Pensions
24	ind_recibo_ult1	Direct Debit

(出所)「Santander Product Recommendation」Data²より当社作成

(2) EDA³(探索的データ解析)⁴⁵

① 年齢の分布

25歳前後に大きな山があり、40歳前後に次の山があり、100歳に向けてすそ野が広がる。

② 所得の分布

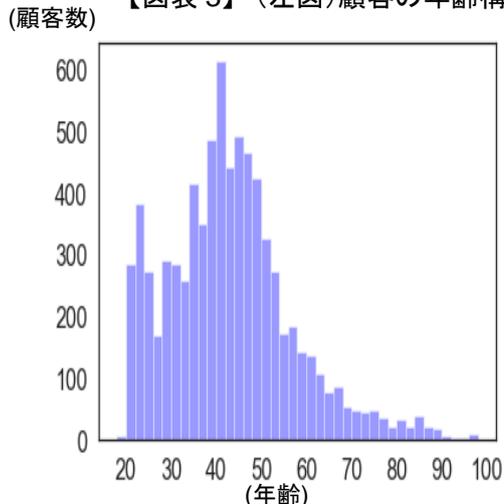
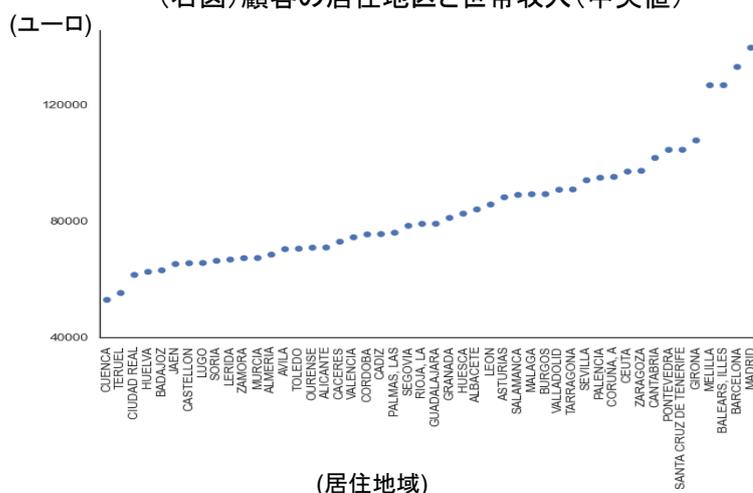
居住地区別の中央値を見ると低いところで約5~6万ユーロ、高いところで約13~14万ユーロに分布している。

² <https://www.kaggle.com/c/santander-product-recommendation/data> (2020年9月3日確認)

³ Exploratory Data Analysis

⁴ <https://www.kaggle.com/apryor6/detailed-cleaning-visualization-python> (2020年9月3日確認)を参考に分析・加工

⁵ <https://www.kaggle.com/apryor6/detailed-cleaning-visualization> (2020年9月3日確認)を参考に分析・加工

【図表 3】 (左図)顧客の年齢構成

(右図)顧客の居住地区と世帯収入(中央値)

 (出所)「Detailed Cleaning/Visualization (Python)⁶」を参考に当社作成

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⁷⁸と評価と分析
(1) 使用する AI の絞り込み

- ① データセットとアルゴリズムの相性の良さを示す指標を「MAP@7 (顧客別に利用可能性が高いサービスを予測確率の高いものから順に 7 位まで計算)」とした。
- ② 今回使用するアルゴリズムは XGBoost とした。

(2) 個別の顧客が新たに利用するサービスの予測確率

- ① 「MAP@7」では 7 位まで予測するが、本レポートではわかりやすさの観点から 3 位までの予測で分析している。
- ② 下記のように顧客が予測対象月にどのサービスを利用する確率が高いかが算出される。

【図表 4】 AI による顧客別 サービス利用予測確率(一部表示)

顧客ID	Current Accounts	Direct Debit	Pensions	Payroll	Credit Card	Taxes	Payroll Account	e-account	Long-term deposits	Short-term deposits	Más particular Account	Funds
211060	0.62	0.06	0.02	0.02	0.07	0.02	0.02	0.02	0.09	0.00	0.00	0.00
1330337	0.00	0.51	0.19	0.22	0.01	0.10	0.27	0.01	0.00	0.00	0.00	0.00
909730	0.00	0.46	0.03	0.02	0.12	0.08	0.01	0.02	0.01	0.00	0.00	0.01
1039871	0.00	0.53	0.15	0.16	0.03	0.14	0.14	0.01	0.00	0.00	0.00	0.00

 購入サービス予測順位 ⇒

1位	2位	3位
----	----	----

(出所) 当社作成

⁶ <https://www.kaggle.com/apryor6/detailed-cleaning-visualization-python> (2020 年 9 月 3 日確認)を参考に分析・加工

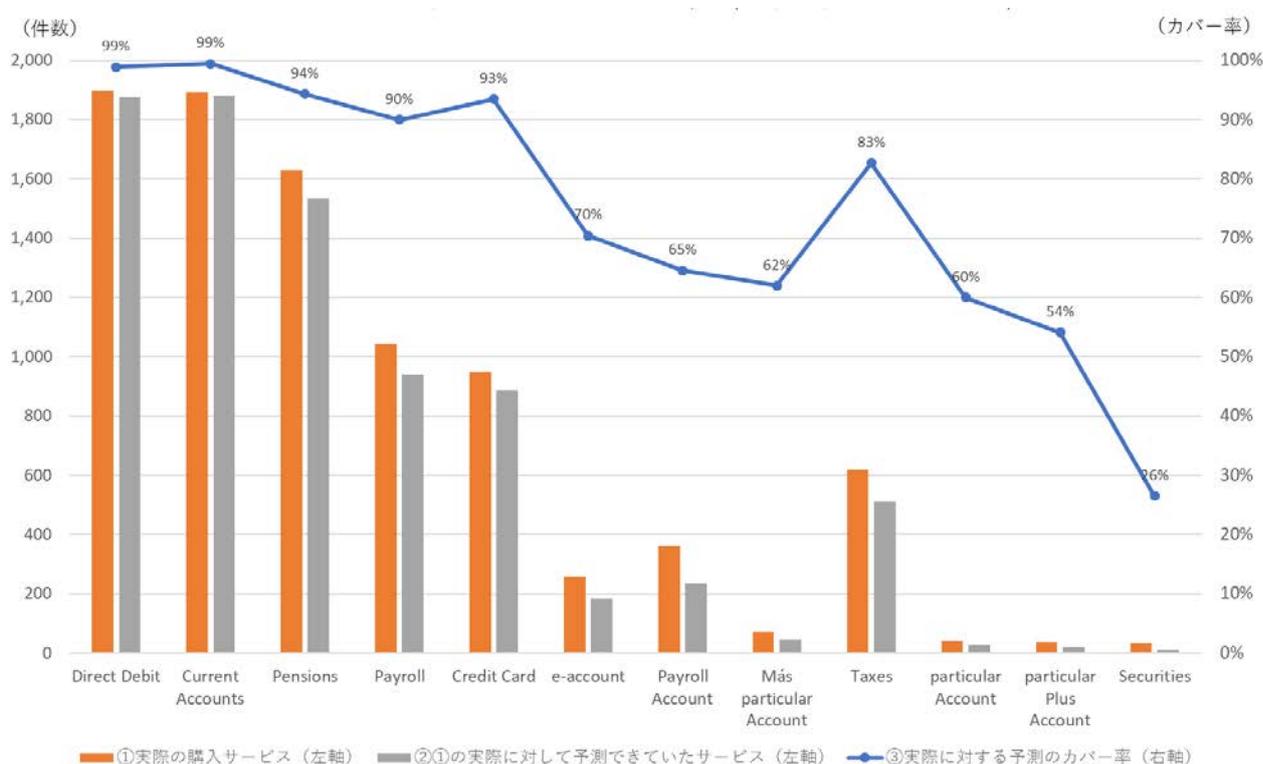
⁷ <https://www.kaggle.com/sudalairajkumar/when-less-is-more> (2020 年 9 月 3 日確認)を参考に分析・加工

⁸ <https://www.kaggle.com/angelmm/santander-product> (2020 年 9 月 3 日確認)を参考に分析・加工

(3) AIが予測する顧客別購買確率の利用

- ① 顧客にサービスを販売促進する場合を想定してみる。個々の顧客の中でAIの予測確率が高い上位3位までのサービスを、ウェブサイトなどですべての顧客に対してお勧め・お知らせするものとする。
- ② 実際の利用サービス数 8,827 個に対して、予測が実際の利用サービスの中に含まれていたのは 8,139 個で、カバー率は 92.2%であった。サービス別には下記の図表のようになる。
- ③ 予測対象者数 126,422 人に対し、実利用者数は 7,130 人(8,827 個の中には複数利用した人がいる)でありサービスの利用率は 5.6%である。大半の顧客が新たにサービスを利用をしていない。しかし、この 5.6%の少数の人々に対して、彼らが利用したサービスが 92.2%の割合でお勧め・お知らせの対象に含まれていることになり、漏れが少なく推薦・お知らせができていているといえる。

【図表 5】 実際の利用に対する予測のカバー率(上位 3 位までの予測)



(出所) 当社作成

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

たとえば、アマゾンでは推薦によって顧客の購買意欲を高め、大きく売上を伸ばしてきた。利用予測に基づき商品推薦を行うと、顧客の商品サービスの利用率を高め、売上・利益を増加できる可能性がある。

(2) 商品推薦の効果の検証

- ① AIのビジネス上の評価は、商品推薦による売上・利益の増加と、商品推薦のコストを比較すればよい。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) セグメント分析

- ① 顧客情報から顧客セグメントを作り出し、顧客セグメントにあった施策ができる。
- ② AI でセグメントを新たに作れば新しい示唆を得られるかもしれないが、その場合前述の EDA（探索的データ解析）の結果を組み合わせると有用であろう。

(4) データセットの特徴量の増強

- ① 今回のケースでは与えられた顧客の 24 の特徴量と過去の購買実績を使って、商品の利用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 の予測精度の一層の向上には限界がある。
- ② 分析顧客数を増やす方法もあるが、たとえばサービス利用の回数や金額や預金残高など特徴量の種類の増加を考えることは有効であろう。

－ ご利用に際して －

- 本資料は、信頼できるとさ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、当社の統一した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。