

課題およびデータ説明

(アブストラクト)

2023年7月25日

課題

個人購買データ (SCI) からの

2015年3月に

サンプル 「メリット」 の

購入有無を予測して下さい

※花王様の商品パッケージより引用

課題

「モデル作成用データ」と「モデル予測用データ」を提供。
モデル作成とモデル検証でデータを使い分けて下さい。

モデル作成

「**モデル作成用データ**」を用いて、**3月メ
リット購入者の予測モデル**を作成

モデル検証

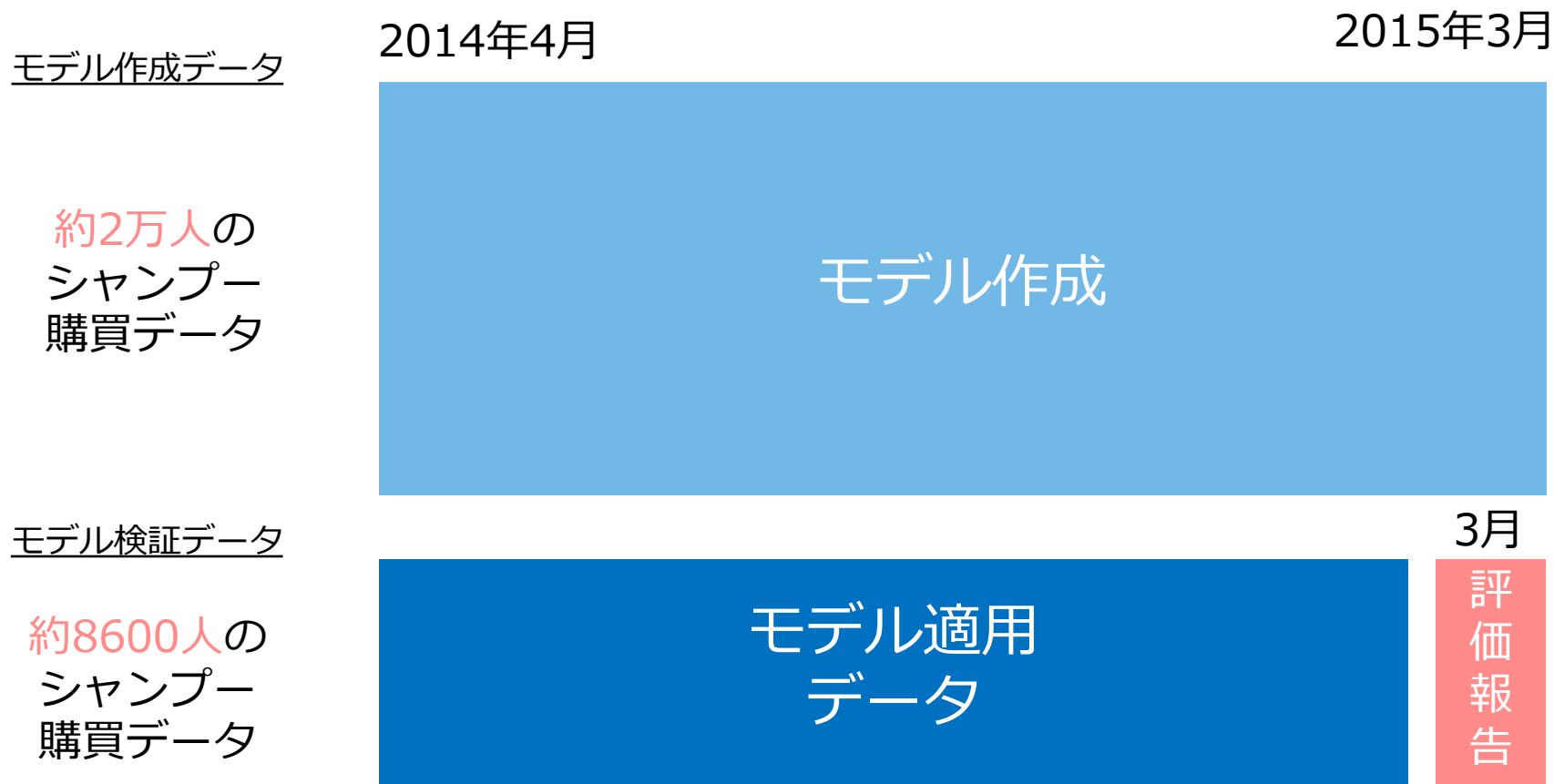
「**モデル検証用データ**」に対して、作成
した**予測モデル**で**3月購入者を予測**

評価

精度評価は**F-score**でお願いします。
精度だけでなく**モデルの解釈・説明**も
お願いします。

データ

「モデル作成用データ」と「モデル予測用データ」を提供。
モデル作成とモデル検証でデータを使い分けて下さい。



データ

提供データ（モデル作成用データ）

データ：SCI

期間：2014年4月～2015年3月

地域：全国

対象商品群：シャンプー

文字コード：UTF-8

レコード数：94,840レコード

人数：20,073人

提供データ（モデル検証用データ）

データ：SCI

期間：2014年4月～2015年3月

地域：全国

対象商品群：シャンプー

文字コード：UTF-8

レコード数：40,862レコード

人数：8,603人

データテーブル

変数名	データ型	データ内容	データ例
receipt_num	Int	レシートのユニーク番号	1,2,3
receipt_buydate	Date	購入日	2014-08-21
receipt_buytime	Time(HMS)	購入時間	14:55:00
G_gyoutai_name	Character	業態名称	スーパー、薬局・ドラッグストア、…
receipt_value	Int	レシート合計金額	1,234
cat1	Character	商品分類1	日用品雑貨
cat2	Character	商品分類2	パーソナルケア
cat3	Character	商品分類3	ヘアケア用品
cat4	Character	商品分類4	シャンプー
product_name	Character	商品名	メリットシャンプー - 200ML
merit_brand ※	Character	メリット判定	メリット、メリット以外
unit	Int	購入個数	1
price	Int	購入金額	380
volume	Int	購入量	200,000
tanka	Int	単価	331
id	Character	購入者ユニークID	31121
gender	Character	性別	男性、女性
age	Character	年代 (10才刻み)	10代、20代、…、60代
married	Character	未既婚	既婚、未婚
job	Character	職業	正社員・公務員、パート・アルバイト、…
merit_flag	Int	メリット商品:1、それ以外:0	0,1

※メリットが商品名に含まれ商品

ご参加お待ちしております



Create Consumer-centric Values

お客様企業のマーケティングに寄り添い、共に生活者の幸せを実現する

株式会社インテージ