Zero-Coding UMAP in Marketing

A Scalable Platform for Profiling and Predicting Customer Behavior by Just Clicking on the Screen

Takuya Kitazawa (Arm Treasure Data)

Customer Data Platform (CDP)

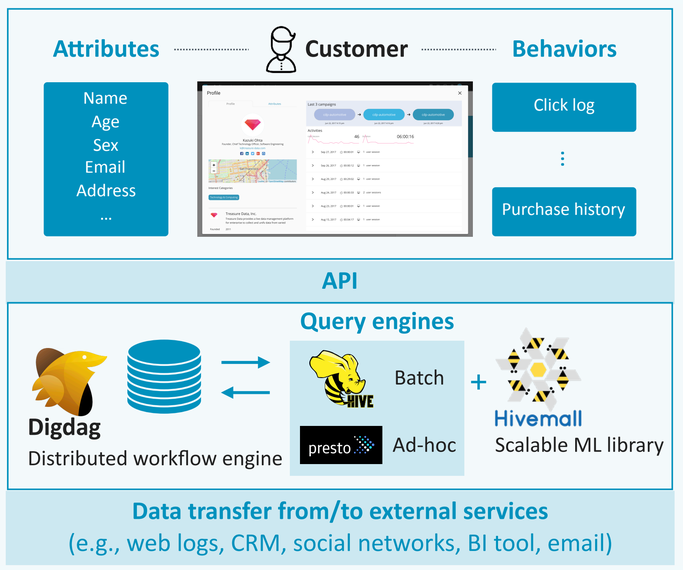
- A rapidly growing need for optimizing day-to-day marketing activities.
- *CDP is a centralized place for:*
  - *creating customer profiles,*
  - *implementing marketing campaigns, and*
  - *predicting customer behavior*

  in connection with a variety of data and signal sources.

  — S. Earley. The Role of a Customer Data Platform. *IT Professional*, 20 (1), pp. 69–76, 2018.
- UMAP techniques play an important role in making deeper insights.

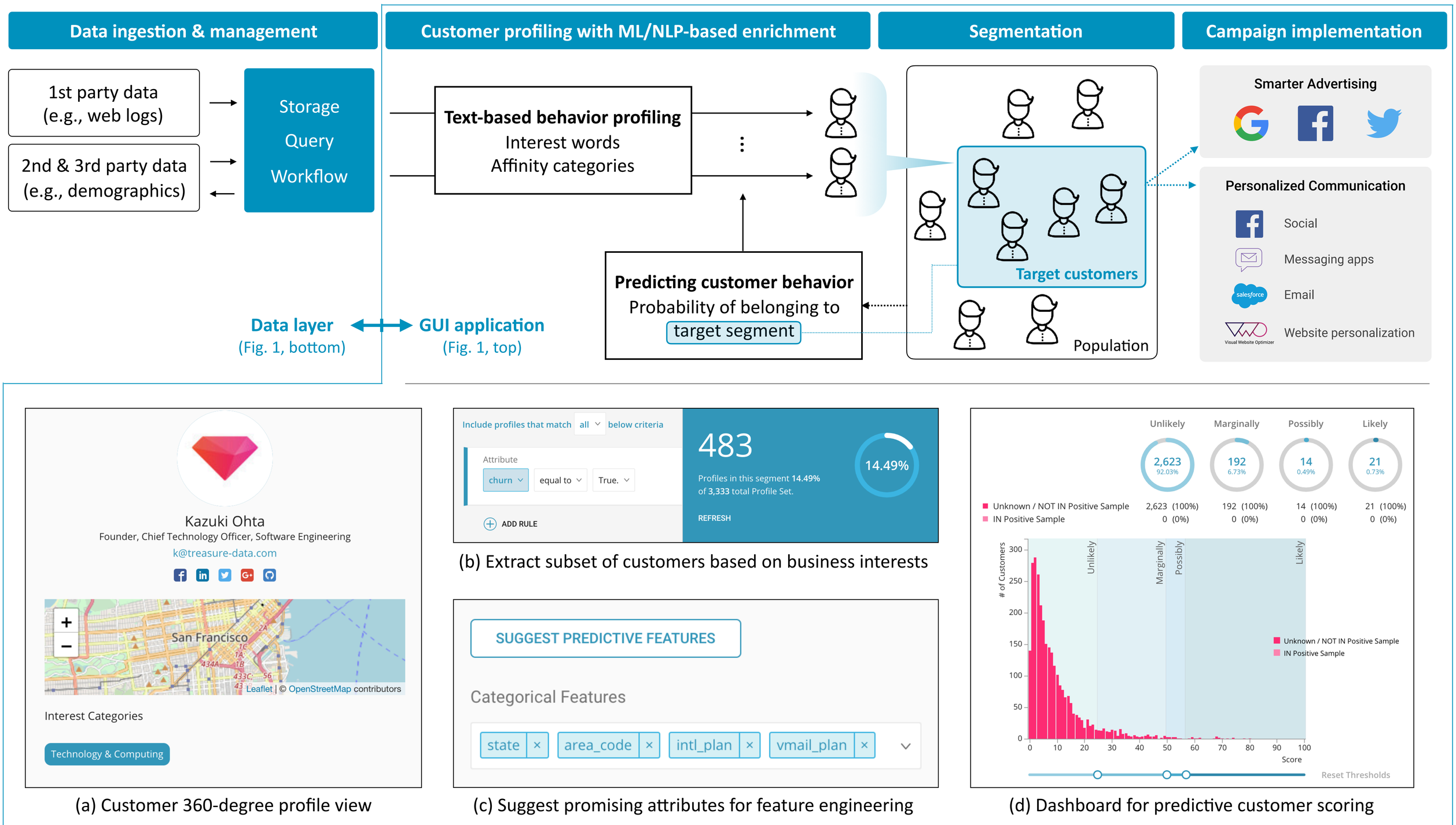
  ↔ Implementation is NOT straightforward due to:
  1. the end user's limited technical expertise, and
  2. complexity of real data.

Demonstrated Solution: Arm Treasure Data Enterprise CDP



CDP: GUI-based UMAP application
- Unified, enriched customer profiles
- Static attributes
- Time-stamped behavioral data

Cloud-based scalable data layer
- ML in query language
- Query execution
- Workflow management



(a) Customer 360-degree profile view

(b) Extract subset of customers based on business interests

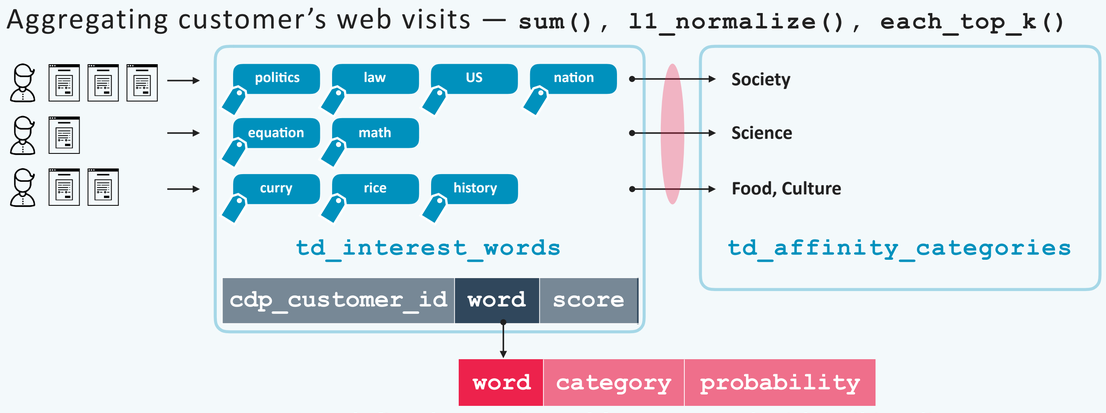
(c) Suggest promising attributes for feature engineering

(d) Dashboard for predictive customer scoring

Zero-Coding UMAP #1: Text-Based Behavior Profiling

TF-IDF-based keyword extraction from web articles (title + description)

Aggregating customer's web visits — `sum()`, `l1_normalize()`, `each_top_k()`



Pre-defined **mapping table** generated with Wikipedia corpus

Zero-Coding UMAP #2: Predicting Customer Behavior

- Implement look-ahead-based marketing campaigns by predicting unseen customer's behavior.
- Solve binary classification problem on customer profiles by just clicking on GUI.

  → Classifier gives the probability of belonging to a specific "target segment."
- Semi-automated ML workflow, including:
  - Logistic regression training and prediction
  - Over- or down-sampling, and probability calibration
  - Min-max normalization
  - 8:2 validation
- Dashboard visualizes probability, evaluation accuracy, and feature importance.

Discussions

- Be as explainable, scalable, and accurate as possible for non-expert users.

  → Limited UMAP techniques like TF-IDF weighting and logistic regression.

↔ More interactive and explainable user modeling e.g., based on topic modeling.

arm TREASURE DATA