

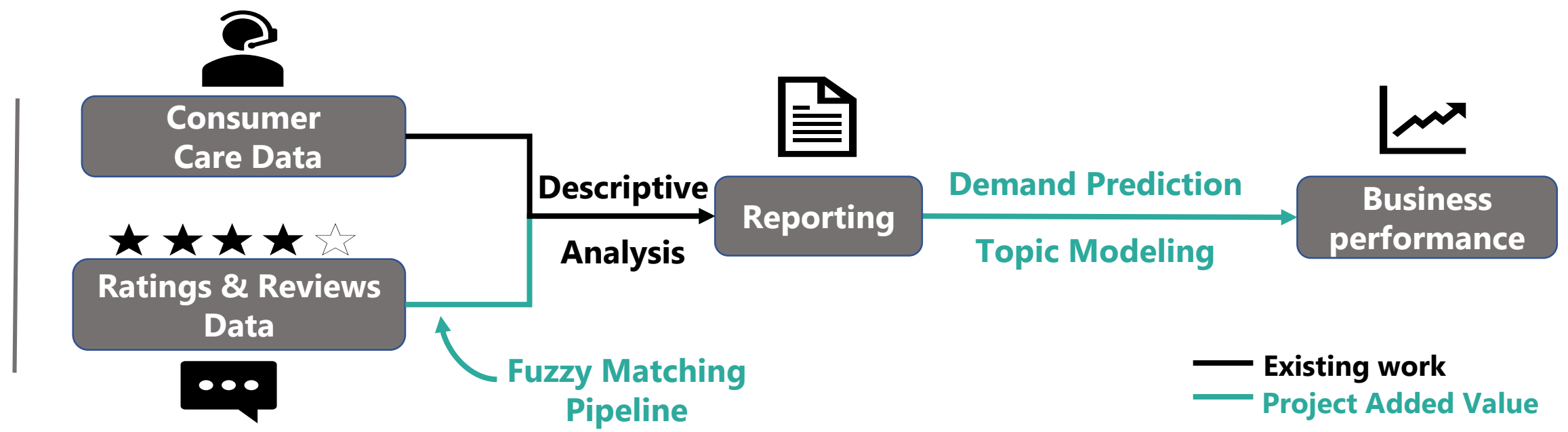


Anis Ben Said & Carrie Fowle

Estee Lauder Manager: Anurag Avlash | MIT Faculty Advisor: Prof. Rama Ramakrishnan

OVERVIEW

Before we joined Estee Lauder Companies (ELC), the Global Consumer Care Analytics Team only used Consumer Care Data and descriptive analytics to generate reports for 21+ Brands and other global teams. The team also acquired a new dataset gathering ratings and reviews from e-market places but couldn't exploit it to its full potential. Our main goal in this project was to create value from these different customer data sources. Our approach was three-fold:



1 Pipeline

Created a pipeline that integrates ratings and reviews to internal system

2 Topics

Identified topics within the reviews

3 Prediction

Of the impact of customer sentiment on demand

DATA



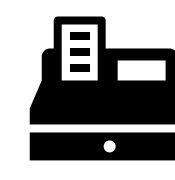
Customer Care

Any complaint made to ELC's care centers via phone, email, or chat is logged in this source. Key variables: product, brand, reason category, reason note.



Ratings & Reviews

This dataset is new to the organization and consists of the 3rd-party scraped data on ecommerce reviews. When we arrived, this dataset was underutilized as it could not be merged with internal sources.



Global Demand

This data records the monthly units sold for each product sold by ELC. For our project, this was our key target variable.

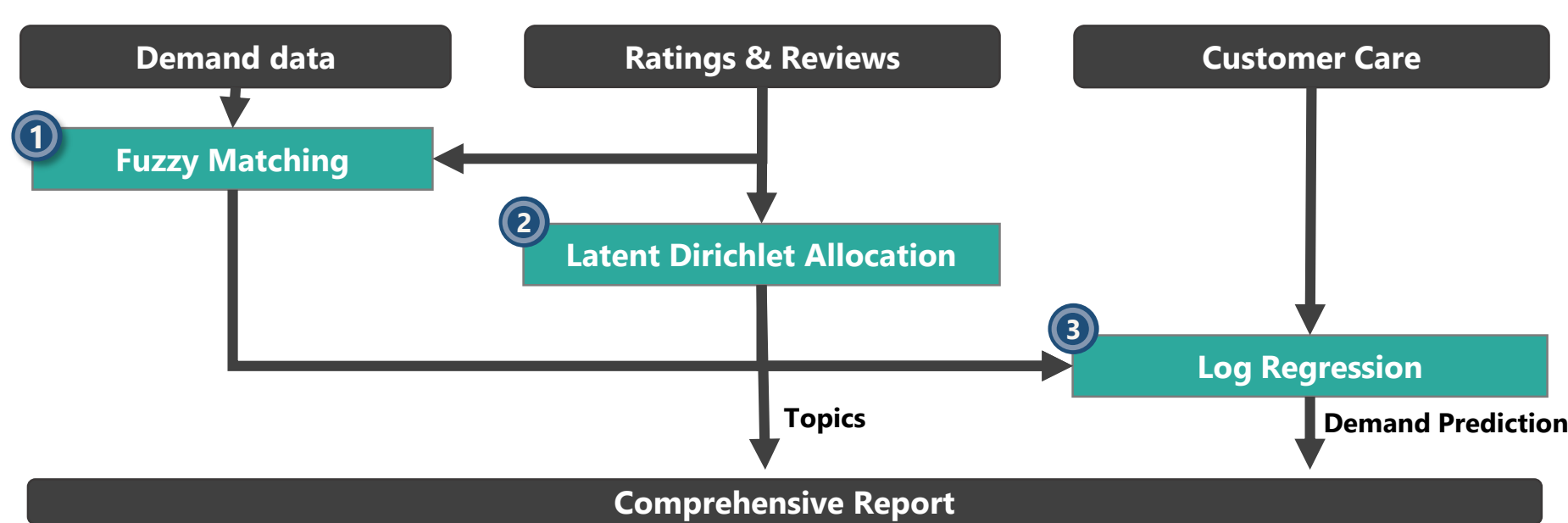


Product Catalog

ELC organizes their products within a five-level product hierarchy. This dataset allowed us to summarize our data at various levels of granularity to be more valuable to our stakeholders.

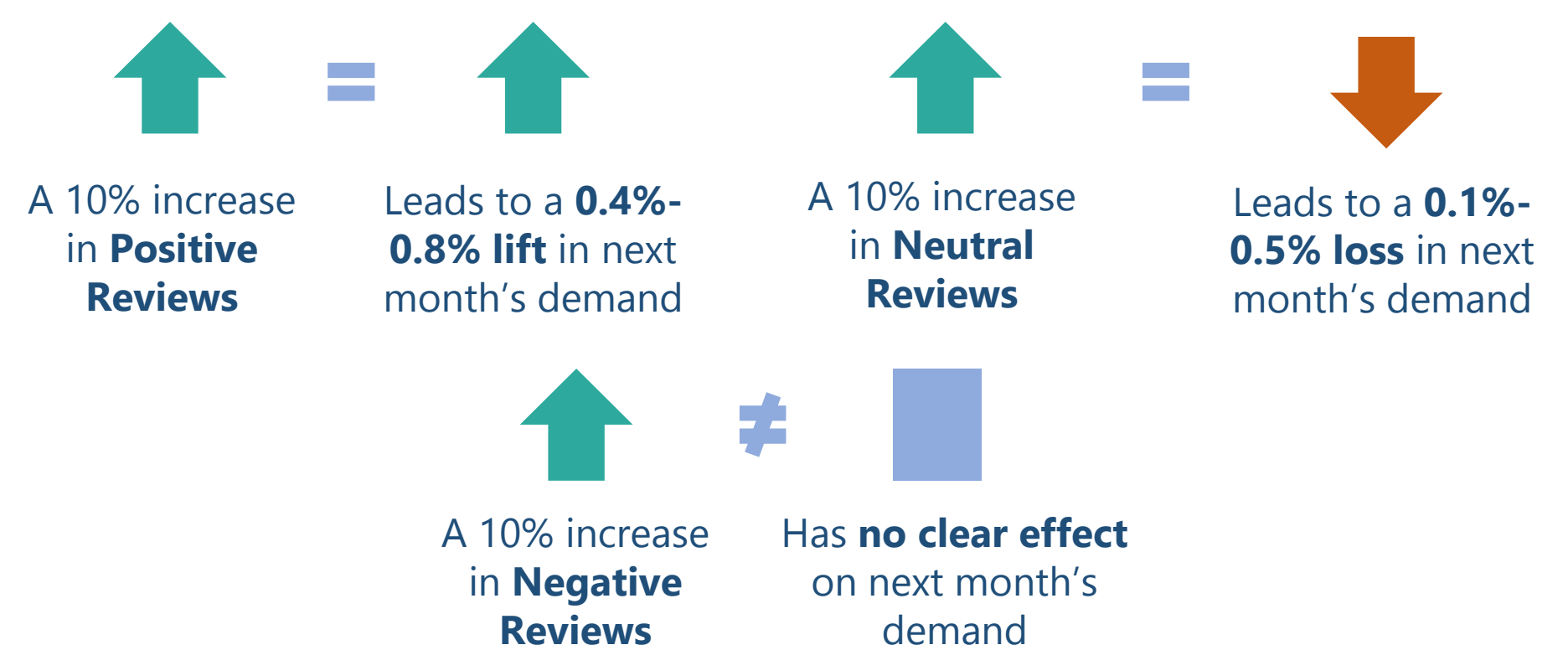
METHODS

- Fuzzy Matching Algorithm:** The key dataset for our project, which contained ratings and reviews for ELC companies from various ecommerce sites, could not be readily joined to the internal datasets since the names of the products were not following the official ELC standard formatting. We used fuzzy matching and the Jaccard distance to create a mapping between Ratings & Reviews product names and the internal product catalogue.
- Topic Modelling:** In meetings with stakeholders, we realized that there was a clear need for a better understanding of what customers were talking about online. We answered that by applying Latent Dirichlet Allocation (LDA), which is a topic modeling technique, to products in 6 different product categories. Training multiple models provided more specificity and resulted in topics that are tailored to each of the product categories.
- Demand Prediction:** The core question underlying the project was, "what feedback matters and what feedback doesn't?" To answer this, we reframed the question as "what feedback is predictive of demand and what feedback isn't?". To ensure that the model was easily interpreted by members of the organization, we chose to focus almost exclusively on Ordinary Least Squares Regression. To tease out signal from an otherwise noisy data set, we tried various transforms of the data, ultimately finding that the log + 1 transform worked best.

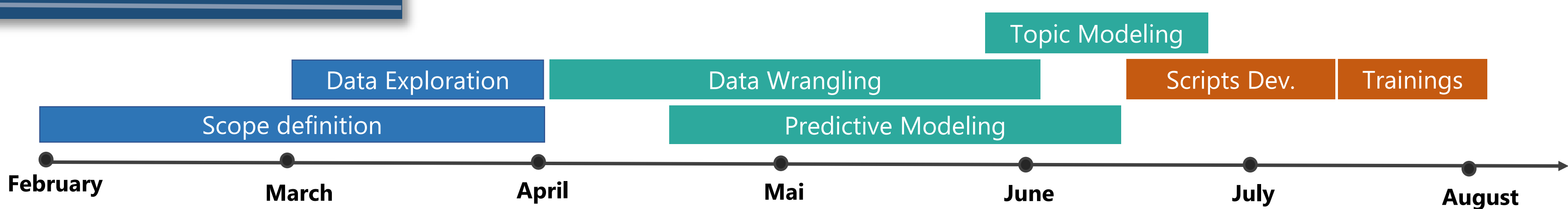


RESULTS

- Mapped the external Ratings and Reviews to the internal ELC official formatting with an **accuracy of 90%** thus allowing to link the ratings and reviews to sales, customer call centers feedback and other internal data.
- Isolated **4 topics** for each product category (face makeup, eye makeup, etc.). Example of topics are **product quality, longevity, color, price**, etc.
- Our customers hold us to a high standard. As a luxury brand, "ok" is not enough to keep our customers coming back



TIMELINE



DELIVERABLES

We provided ELC with **3 Python scripts** with easy-to-use interfaces:

- Reviews Wrangler:** Links Ratings and Reviews to the internal ELC data system
- KPI Builder:** Produces monthly reports featuring the KPI and the topics for each review
- Reports Builder:** A scripts that automates ad-hoc reports production.

NEXT STEPS

Our contribution to the Global Consumer Care Analytics Team was a pioneering work in Analytics. First, we showed in many cases the impact of properly using Machine Learning tools but most importantly, we paved the path for more advanced analytics work to be done in the future.