

---

# FUJI DIGITAL CAMERAS

## UNDERSTANDING SENTIMENTS IN EPINIONS

**February 10, 2017**

**Contributors:** Sania Irshad, Daniel Li, Alejandra Salazar Romo, Nikhil Agarwal

**Clients:** Benedikt Boecking, Disha Gupta

### PROJECT BACKGROUND AND DESCRIPTION

Large amounts of customer reviews are available on Epinions.com and Fuji Digital Cameras is interested in determining the potential value of such data in supporting marketing decisions.

The purpose of this project is to determine how we can excavate useful information from such data and understand the sentiments of customers. Using data mining concepts and technology, this project aims to analyze customer reviews and ratings about Fuji Cameras and possibly identify any trends in customer's behavior that can aid data-driven marketing decisions.

### GOAL

The goal of this project is to characterize positive and negative online reviews and determine its potential business adding value to support data-driven marketing decisions for Fuji digital cameras. We intend to build an analytics system that identifies meaningful words as features (like 'great', 'nicely', 'comfortable', 'excellent', 'bad') from customers' comments and monitor trends in consumers' sentiments.

### CRITERIA FOR SUCCESS

- Identify meaningful words in the reviews that create a discriminative value in predicting the consumers' recommendations.
- To build a model that can be effectively used to:
  - Predict sentiments from textual opinions
  - Monitor trends in consumers' sentiments

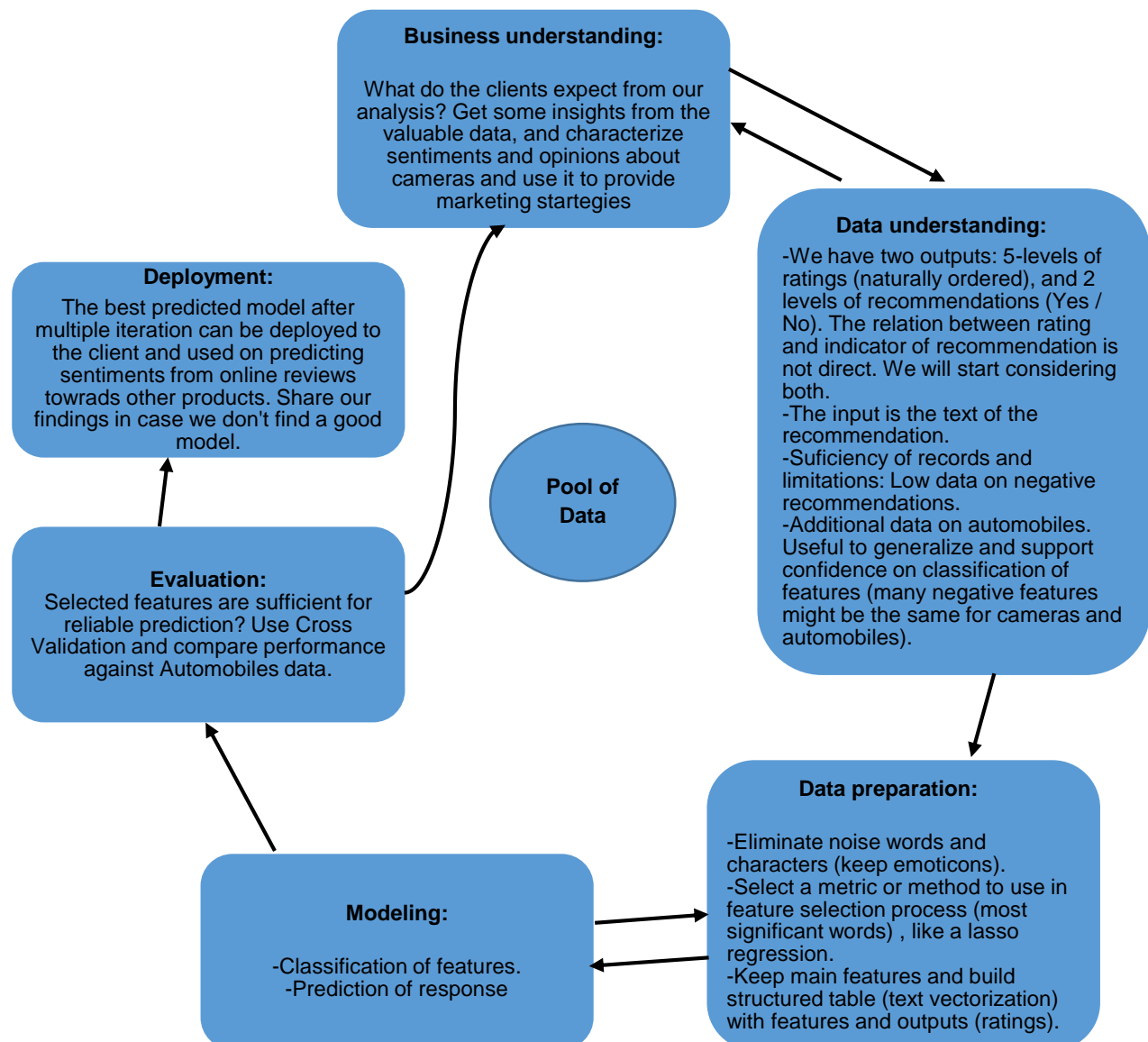
### POTENTIAL CHALLENGES / RISKS

- Collected data from epinions.com has more comments with positive recommendation. Such uneven distribution which could affect the sentiment analysis (Annex 1).
- Dealing with the neutral opinions (Eg. Rating = 3) with either "Yes" or "No" label could be potentially a wrong classification (Annex 2).

## CONTINGENCIES INTENDED TO MITIGATE RISKS

- Calibrate the risk to identify false positives and false negatives.
- If data is not enough, we could use ratings levels to create a direct response between “Yes” and “No”, although we could be losing some information about the sentiment of the client.
- Running models with different specifications and parameters to identify the best possible way to deal with ambiguous data.

## EXECUTION PLAN



## ANNEX: PRELIMINARY ANALYSIS OF DATA

1. Distribution of Customer data for Product “CLASS” against “LABEL” describing if a customer will recommend the product:

	Yes	No
<b>Camera</b>	5274	726
<b>Auto</b>	5105	895

2. Detailed analysis of Customer data for “RATING” about a product “CLASS” against “LABEL” describing if a customer will recommend the product:

CLASS	RATING	LABEL COUNT		Total
		Yes	No	
<b>Auto</b>	1	6	288	294
	2	24	298	322
	<b>3</b>	<b>284</b>	<b>247</b>	<b>531</b>
	4	1804	38	1842
	5	2987	24	3011
<b>Camera</b>	1	6	181	187
	2	13	268	281
	<b>3</b>	<b>345</b>	<b>234</b>	<b>579</b>
	4	2251	32	2283
	5	2659	11	2670