Cross Channel Attribution and Client Segmentation

Haoqing Mao, Julien Crabié,
Hugo Todeschini, Chethan Balakrishna,
Tandez Sarkaria

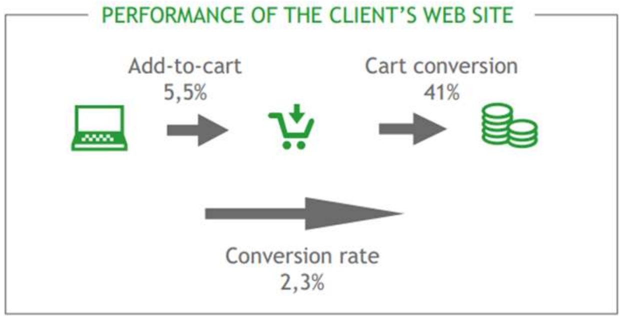
Index

Context and objective

- Cross-channel player in the textile industry
- Originally selling by catalogue and over the phone
- Added an online platform for sales

A website:

- Which doesn't reach the required standards
- But has a very good conversion rate



Context and objective

The goal is to propose explanations for this paradox and to convince the client to invest in their website.

We study only clients who buy on the website, of which there are two types:

- Internet mail order (IMO)
- Pure internet order (PIO)

Context and objective

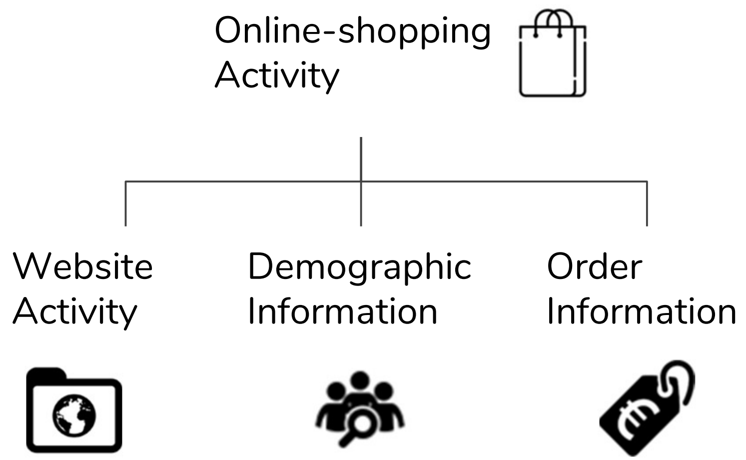
Some customers buy on the website but are motivated by mails to visit the website (IMOs)

Important to identify these customers in order to correctly allocate the marketing budget to different mediums

To identify the IMO from the PIO with a classification or cluster algorithm

Data Cleaning and Merging Tables

Assumption



Steps

Identify web activities that matches purchase

- Merge through
    - Customer Identification
    - Order Date and Web Session Date
- Select payment complete page
- Select unique order number
- Add additional information(Demographic)

Feature Engineering

Seniority - Number of days since a customer became a member of the company

Global Sources - The source from which a customer came to the website such as sponsored links, marketing emails, influential websites etc. We created dummy variables for each type

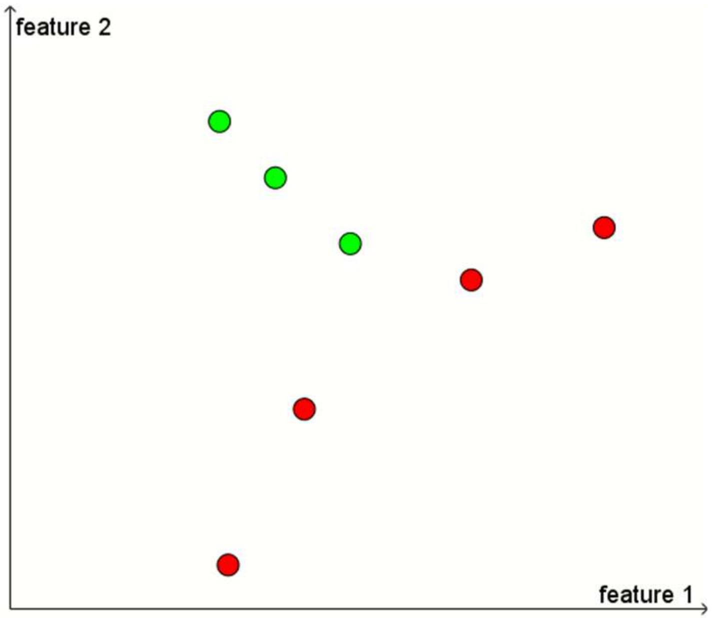
Device Type - The type of device a customer used to access the website such as Laptop or Smartphone. Again we create dummy variables for each type

Title - We created dummy variables for the title of customers

Also, Payment amount, Gender

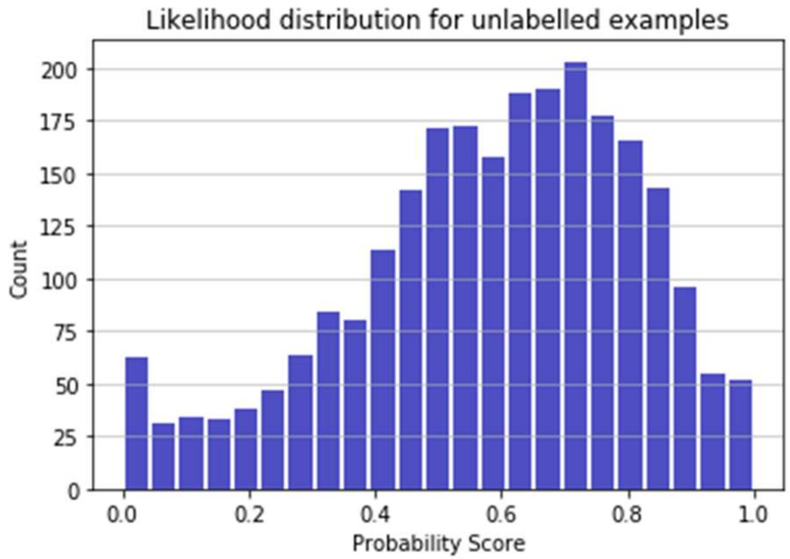
Positive Unlabelled Learning

1) Challenge is to classify wrongly labeled data
2) We used the Method called Positive Unlabeled Bagging
3) The training set is created by combining all the positively labeled data points (IMOs in this case) with a random sample from the unlabeled points
4) The classifier is first trained on this ‘Bootstrap Sample’ treating unlabeled points as negative samples
5) Then we apply the classifier on all the unlabeled data points that were not part of the sample and get a score i.e likelihood of the point being positive
6) Then we repeat the above steps with a different random sample and use an average likelihood score in the end



Testing our Model and Selecting a Threshold

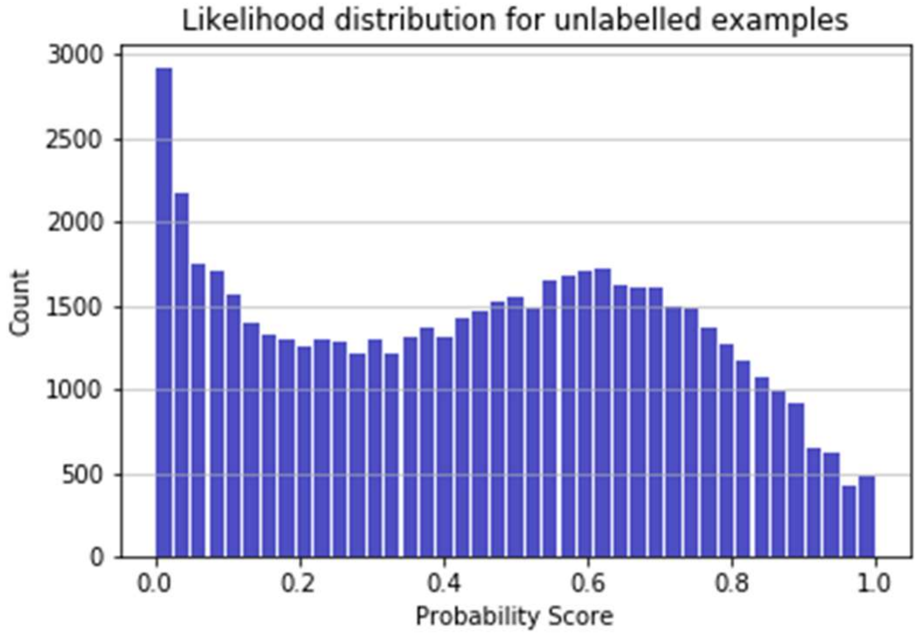
- To test the model we change about 20 % positive labels to negative labels and test the model on these observations
- On average, we got a 58% likelihood for these labels to be positive i.e IMOs
- The distribution being skewed it is preferred to used the median (i.e 61%)



Results

- 18,525 IMO were mislabelled as PIO

- Given that our dataset is composed of about 70,000 clients, we get a new conversion rate of 1,5%.



Conclusion

How to make a better study:

- Enrich the data by asking customers how they learned about the products at the time of purchase on the website
- Create a new marketing segment for these fake PIOs

A weak website:

- Only 68% of customers classified as PIOs are truly PIOs.
- Site performance actually much lower than we thought and lower than the average website in the industry, 1.5% vs 1.8% of conversion rate
- Need to invest on the website to improve performance of the company