भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

IMB 779

IMPROVING LEAD GENERATION AT EUREKA FORBES USING MACHINE LEARNING ALGORITHMS

NANDINI SETH, MANUPRIYA AGRAWAL, MANARANJAN PRADHAN AND U DINESH KUMAR

Nandini Seth, Manupriya Agrawal, Manaranjan Pradhan and U Dinesh Kumar, Professor of Decision Sciences, prepared this case for class discussion. This case is not intended to serve as an endorsement, source of primary data, or to show effective or inefficient handling of decision or business processes.

Shashank Sinha, the Chief Transformation Officer at the Eureka Forbes was keen on identifying an approach to improve their digital efforts and its efficiency of generating sales opportunity. Eureka Forbes offered solutions in water purification, vacuum cleaning, air purification, and security solutions, and continued to be a market leader. Having started in 1982, the company was widely known for its flagship brand, Aquaguard, a water purifier, which was a household name in India. Having worked with the company for over two decades and being well-versed with the functioning of the firm, he was concerned that the company needed to improve marketing efficiency while keeping the costs under control.

Millions of potential customers visited the Eureka Forbes website to understand various products sold by them. Shashank knew that there were a significant number of digital consumers who visited the Eureka Forbes website but were not converted into prospects. With exhaustive details on all the products, post-sales services and a helpful blog, the website was a perfect channel for divulgence of information (Exhibit 1). Shashank was aware that the Digital Marketing team had very rich data on consumers˘ online behavior such as the product pages visited, amount of time spent by them in each page, the various actions taken and so on. While the company was running digital campaigns, they were not leveraging the rich behavioral and visitor acquisition data captured by the firm. Shashank knew that targeting potential customers based on this data would be extremely beneficial for the company. A customer˘s purchase journey broadly goes through four stages: Awareness, Interest, Desire, and Action. The website helped to generate awareness among potential customers about the products and web analytics could be used to gauge interest. Shashank was aware that most customer visits did not end in an action. An action could be defined in two ways here ¯ definitive actions such as making a purchase or sales lead generation actions such as filling interest form or sharing contact details for an in-home demo. It occurred to Shashank that in the present age of Internet, customers conducted a rigorous evaluation of features and functionality along with promotions before making a purchase.

Shashank˘s team had access to all the data collected by Eureka Forbes and he was determined to use analytics for driving sales of Eureka Forbes products.

ABOUT EUREKA FORBES

Founded in 1982, Eureka Forbes was a multi-product organization with a turnover of over INR 30 billion[1] in 2018 and a part of the Shapoorji Pallonji Group. The company sold both household and industrial products, through its multi-channel distribution system. The company started out with a direct sales business model and incorporated retailing in 1986. The company has an impressive direct sales force, franchisee partner network, dealers & distributors, institutional channels and a rural channel across 850+ cities and towns in India. The company˘s mission of `Friend for life_ along with the expansive distribution channel implies that the potential consumers with explicit interest in a product can be individually satisfied to complete the purchase. Eureka Forbes has been one of the world˘s largest direct sales company. In 2018, the company was estimated to have a customer base of 20 million and present across 40+ countries. While Eureka Forbes manufactured and sold products in air purification, vacuum cleaners and ancillary services,

[1] 1 USD was approximately 69 INR in July 2019

it was their water purification brand Aquaguard, which had become a household name in the company˘s domestic market of India. While the company had a near monopoly status in the beginning, it was still a market leader in the water purification business with the highest market share, according to the 2018 company report. Exhibit 2 provides a list of the product offerings by the company. The company has been widely credited for its efforts to understand the changing water conditions and developing suitable purification technologies across India, also helping to create brand awareness for Eureka Forbes. This fusion of business strategy with corporate social responsibility has been a disruptive innovation in the Indian markets. The company was also spending 3-4% of their annual turnover on research and development of new technology.

ABOUT THE INDUSTRY

With water pollution being a serious concern in India, the home water purification industry has been actively growing and is expected to be worth $1.53 billion by the financial year 2019[2] and $4.1 billion in 2024.[3] Water purification industry can cater to three types of segments ¯ households, industrial sector, and commercial sector with the households accounting for more than half of the demand. With a large portion of Indian households still using untreated or traditionally treated (boiled) water, water purification system was a potential but under-penetrated market. The industry received support from policy makers and the government who had launched various policies and campaigns to boost the consumption of clean water. While the penetration was sizable in the urban parts of India, penetration in rural parts was significantly less due to economic factors. The market leaders were Eureka Forbes (Aquaguard), Kent RO Systems, HUL (Pureit), Tata, L.G, A.O Smith and Ion Exchange (Zero B) with Eureka Forbes and Ion Exchange being the first movers in the home water purification market in India. The top five manufacturers accounted for more than 80% of the market share. With very few small players, the Indian market was consolidated by the presence of a few leading players. It was expected that this competition would increase with the arrival of new entrants in the market such as Blue Star and AO Smith.

ABOUT THE PRODUCT

Water purification products based on various technologies are available in the market. Exhibit 3 provides a list of technologies which were used by Eureka Forbes to manufacture water purification systems. While reverse osmosis (RO) purification being a dominant strategy in the market, other products are based on ultraviolet (UV) purification, combination of RO & UV technology, non-electric gravity purifiers, water softeners, and sediment filters [Exhibit 4]. According to Transparency Market Research report[4], RO technology comprised 37% of the total market revenue in 2015. Analysts predict that the gravity water purifiers and water softeners would exhibit steady growth rate in the future. The top 10 players held about 70% market share in the RO technology, with Eureka Forbes, and Kent RO accounting for 50% share in 2013.

[2] According to ValueNotes Research Report titled `Home Water Purifier Industry 2014-2019_
[3] According to Transparency Market Research Report in 2017
[4] Transparency Market Research Report in 2017

ANALYTICS SOLUTION

The more Shashank thought about the problem, the keener he was on discussing it with his team to find an efficient use of the collected data. Shashank convened a meeting which included Vikram Surendran, the President at Eureka Forbes and Kashif Kudalkar, the Head of Digital Marketing & Analytics at the firm. After the customary catch-up with routine work, Shashank came to the point he was waiting to discuss. Shashank said:

> Kashif, we are incurring very high costs by chasing every possible lead physically. We have meaningful customer level data to understand their online behaviour and purchase intent. We can analyse this data to determine which customers are the most likely to make a purchase and target them. We can use predictive analytics models to do it.

Kashif progressively responded:

> I totally agree with you. The digital footprints of potential customers can be used to understand where they are in the purchase funnel and to determine their propensity to make a purchase. We currently don't use behavioral data captured by the website for our remarketing campaigns.
>
> If an individual customer has a relatively higher probability of making a purchase, we can consider him a potential sales lead and send out representative to his door-step to complete the purchase. But in case the customer has a lower probability of making a purchase, we can wait and maybe target him in less expensive ways. This will help save us a lot of money.

Shashank concurred and said:

> This process of short-listing sales leads is valuable in two ways. Along with identifying customers who are likely to make a purchase, we can also make the list longer or shorter depending on our sales budget by using an appropriate probability cut-off.

Vikram added:

> The same probability score can be used to do a one-to-many personalization. This kind of segmentation can give interesting insights which will help us in our journey to build the capability of one-to-one marketing.

Kashif called for another meeting to brainstorm on possible machine learning models. In the next meeting, Kashif met Tatvic Analytics, his business analytics partner to determine the best possible approach to model his problem and to discuss the possible roadblocks they may face.

Kashif: What kind of data do you think we should use for this analysis?

Bismayy, one of the team members said:

> We have streams of online data such as detailed product and event wise data, search engine, marketing channels, etc. We could use all of this data along with any demographics that we might have about the customers to create our model.

Kashif:

> I think the biggest challenge that one faces while dealing with a digital footprint data is the colossal size of the data along with a very high number of variables involved. I think the data might need significant pre-processing before we start the model building process. We would need to determine the importance of variables and prune the data to a manageable quantity. We should be prepared to handle the data challenges posed by the extraordinary quantity of data.
>
> Also along with the large quantity of the data, we are faced with a highly imbalanced data set since the number of consumers finally converting are less than 1%. We would need to be cautious about the models we use.

Kashif looked at Bismayy and asked:

> What type of model do you think will best fit our needs?

Bismayy paused to think and then responded:

> We should run algorithms which can handle highly imbalanced data and at the same time can solve the classification problem we have at hand. I think an ensemble method might be able to give better results than a single model. I believe that we should try out different models and see which one might work the best for our purpose. Let me think more about this.

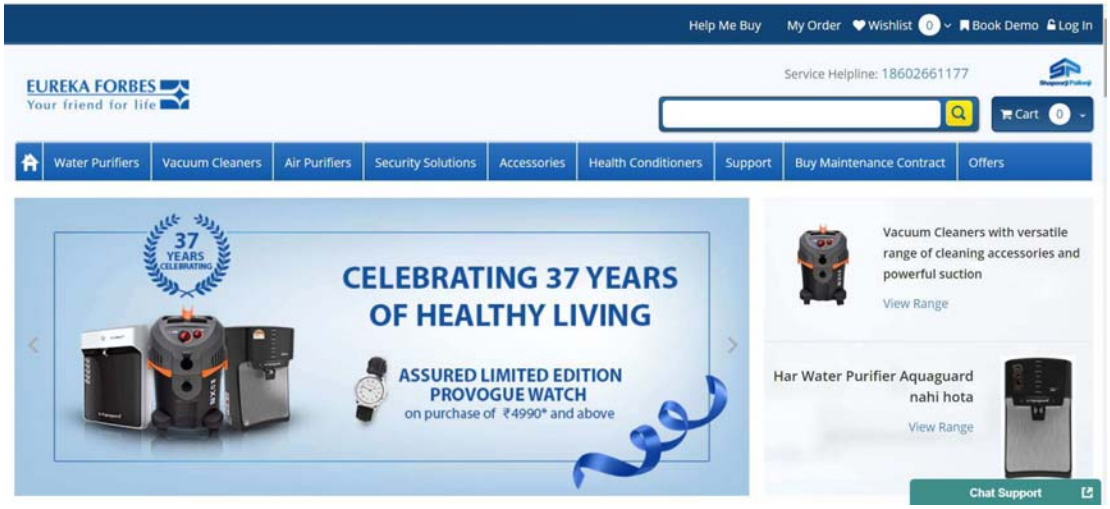
For the next few days, Bismayy started thinking deeply about the various models that he could build. He looked at the extensive list of variables available. He also prepared a list of the most important variables to make the problem tractable [Exhibit 5]. Bismayy also had to consider the imbalance of the data while deciding on an appropriate model keeping in mind the various access routes to the website. [Exhibit 6]. As Bismayy prepared his game-plan, Kashif knew that data-driven decision making was the only way to improve their conversion rate while reducing their costs.

While Eureka Forbes was the current market leader, they were facing stiff competition from new entrants as well as local players. A cost-effective distribution system with the backbone of good data analytics would be their strength in the times to come.

With this thought, Kashif and Bismayy started the work on building the first model.

Exhibit 1

Snapshot from Eureka Forbes Website



Source: Eureka Forbes Website

Exhibit 2

Eureka Forbes Brands (by category)

Water Purifiers	Vacuum Cleaners	Air Purifiers	Security Solutions
Dr. Aquaguard Aquaguard Aquasure	Euroclean Forbes	Dr. Aeroguard Aeroguard	Eurovigil

Source: www.eurekaforbes.com

Exhibit 3

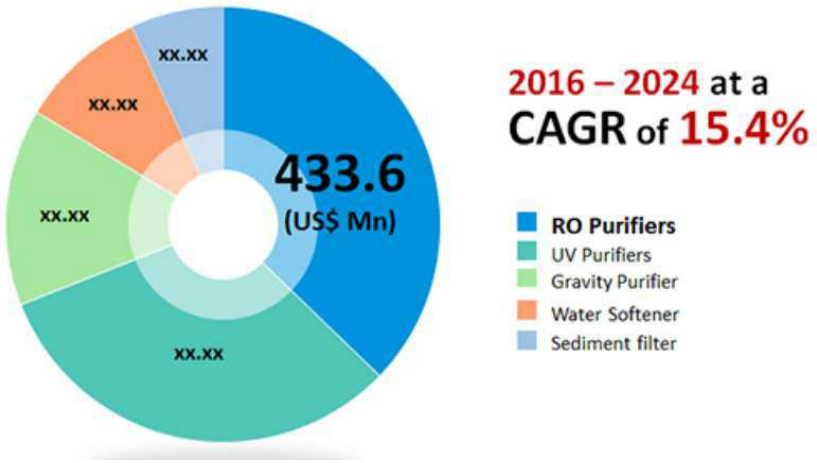
Eureka Forbes Water Purifiers (by technology)

Water Purifiers by Technology
Universal (RO + UV + UF)
Universal (RO + UV)
RO
Green RO
UV
Non-Electric (Gravity Based)
On-the-Go

Source: www.eurekaforbes.com

Exhibit 4

Indian water purification market (by technology)



Source: TMR Analysis, December 2016 (www.transparencymarketresearch.com)

Exhibit 5

Data Description

Feature List	Definition
Pageviews	Number of total pages views per user
Sessions	Number of total sessions per user
air_purifier_page_top	Time spent on air purifier page
Bounces	Total number of single page (or single interaction hit) sessions for the property.
checkout_page_top	Time spent on checkout page
contactus_top	Time spent on contact us page
Country	User's country name
client_id	Unique Identifier of user
customer_service_amc_login_top	Time spent on customer service amc login
customer_service_request_login_top	Time spent on customer service request login
demo_page_top	Time spent on demo page
Device	Users' device category
Dsls	Day since last session
Date	User visit date
fired_DemoReqPg_CallClicks_evt	Event fired for Demo Calls
fired_help_me_buy_evt	Event fired for Help Me Buy CTA
fired_phone_clicks_evt	Event fired for Phone Clicks CTA
goal4Completions	Book Demo Thank You (Goal 4 Completions)
help_me_buy_evt_count	Event count for Help Me Buy CTA
newUser	If userType is New Visitor
offer_page_top	Time spent on offer page
Paid	Users landing from campaign
phone_clicks_evt_count	Event count for Phone Clicks CTA
Region	Users˘ region derived from their IP addresses or Geographical IDs. In U.S., a region is a state, New York, for example.
security_solutions_page_top	Time spend on security solutions page
sessionDuration	Total duration (in seconds) of users' sessions.
sourceMedium	Combined values of ga: source and ga: medium.
storelocator_top	Time spent on storelocator page
successbookdemo_top	Time spent on successbookdemo page
vacuum_cleaner_page_top	Time spent on vacuum cleaner page
visited_air_purifier_page	If user visited Air Purifier Page 1 else 0
visited_checkout_page	If user visited Checkout Page 1 else 0
visited_contactus	If user Visited Contactus Page 1 else 0
visited_customer_service_amc_login	If user Visited Customer Service AMC login page 1 else 0
visited_customer_service_request_login	If user Visited Customer Service Request login 1 else 0
visited_demo_page	If user Visited Demo Page 1 else 0

visited_offer_page	If user Visited Offer Page 1 else 0
visited_security_solutions_page	If user Visited Security Solutions Page 1 else 0
visited_storelocator	If user Visited Store Locator 1 else 0
visited_vacuum_cleaner_page	If user Visited Vacuum Cleaner Page 1 else 0
visited_water_purifier_page	If user Visited Water Purifier page 1 else 0
water_purifier_page_top	Time spent on Water Purifier Page
Dow	Day of week
pageviews_hist	Users 30 days pageviews history
sessions_hist	Users 30 days sessions history
bounces_hist	Users 30 days bounces history
paid_hist	Users 30 days paid history
sessionDuration_hist	Users 30 days sessionDuration history
visited_air_purifier_page_hist	Users 30 days visited Air Purifier Page history
visited_checkout_page_hist	Users 30 days visited Checkout Page history
visited_contactus_hist	Users 30 days Visited Contactus Page history
visited_customer_service_amc_login_hist	Users 30 days Visited Customer Service AMC login history
visited_customer_service_request_login_hist	Users 30 days Visited Customer Service Request login history
visited_demo_page_hist	Users 30 days Visited Demo Page history
visited_offer_page_hist	Users 30 days Visited Offer Page history
visited_security_solutions_page_hist	Users 30 days Visited Security Solutions Page history
visited_storelocator_hist	Users 30 days Visited Store Locator history
visited_vacuum_cleaner_page_hist	Users 30 days Visited Vacuum Cleaner Page history
visited_water_purifier_page_hist	Users 30 days Visited Water Purifier page history
phone_clicks_evt_count_hist	Users 30 days Phone Clicks Event Count history
help_me_buy_evt_count_hist	Users 30 days Help me buy Event Count history
converted_in_7days	Target variable

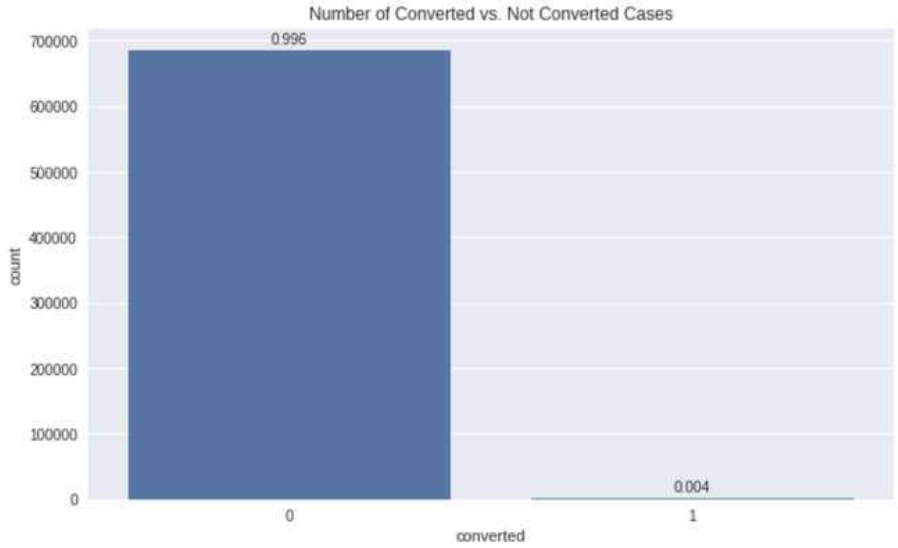
Source: Eureka Forbes
Note: The Dow variable was derived using the date of website access variable

Note: The data can be downloaded from the link: http://hrm.iimb.ernet.in/iimb/Harvard/Eureka/index.htm

Exhibit 6

Discussion on Data

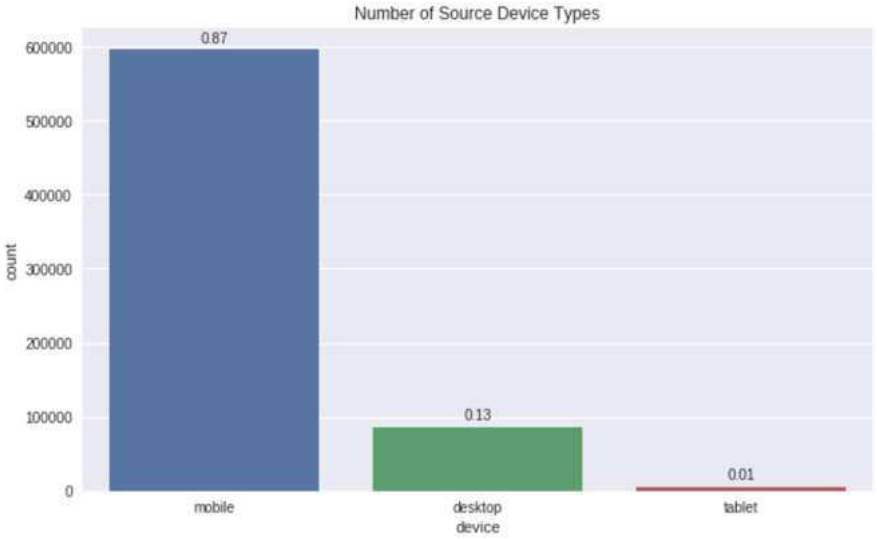
(1) Imbalanced Data



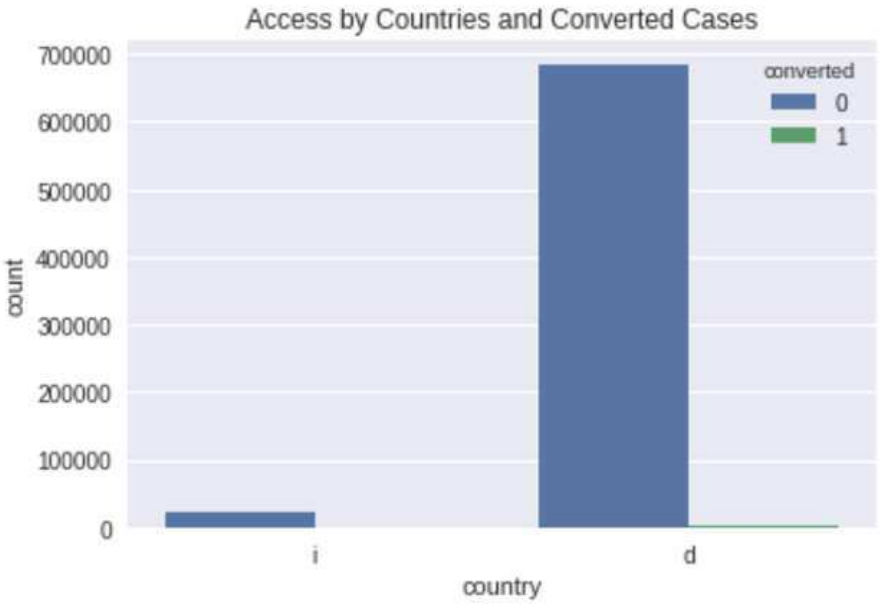
	converted	count
0	0	706497
1	1	2830

0 - Not Converted
1 - Converted

(2) Access (by device type)



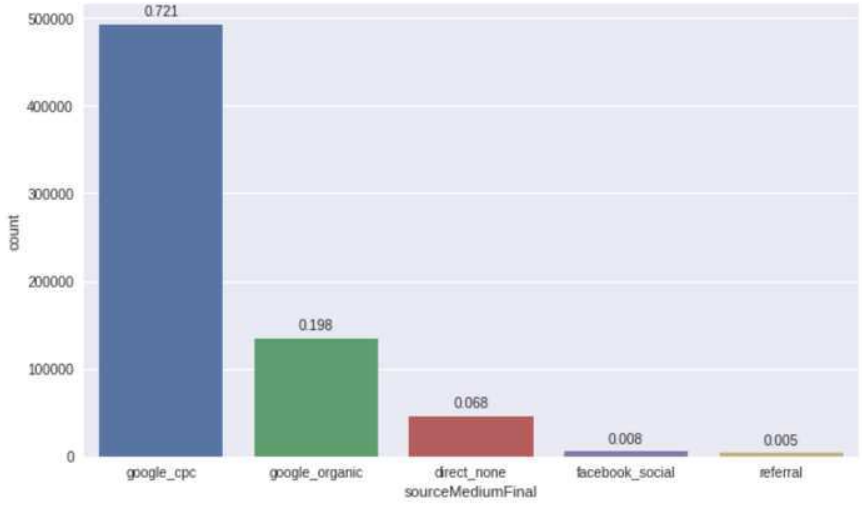
(3) Access by Domestic v/s International users



country	0	1
d	0.996047	0.003953
i	0.994829	0.005171

d - Domestic
i - International

(4) Access by source channel



source	not converted	converted	total
google_cpc	489510	2402	491912
google_organic	134711	248	134959
direct_none	46061	43	46104
facebook_social	5334	5	5339
referral	3706	17	3723

Source: Data provided by Eureka Forbes