Targeted offers for marketing

Analysis of different optimization algorithms

Team

Rajesh Sathya Kumar (rsathyak@asu.edu)
Chandrakant Vankayalapati (cvankaya@asu.edu)
Pramukh Belam (pbelam@asu.edu)

CONTENTS

Motivating Question...

- In the dynamic modern world, companies across all domains are adapting to serve and reach out to their customers in a much better way every single day. To carry out the activity, marketing campaigns are of paramount importance and millions, if not billions, of dollars are invested to effectuate a lucrative marketing campaign.

- Considering a country like USA, which has a population of c.330 million, sending marketing emails/texts/pamphlet/etc. to everyone or a randomly chosen set of people would be a waste of time and money.

- In this research piece we try to answer the question : **"Which products should be targeted to which customers to maximize profits, under the constraints that only a limited number can be targeted, and each product has a minimum sales target?"**

- We first formulate a naive objective function followed by set covering formulation. We then perform extensive computational experiments to choose the optimal model using integer programming.

Methodologies

1. Basic Formulation

2. Set-Covering Formulation:
   a) Tactical Model Formulation
   b) Operational Model Formulation

Basic Formulation:

Sets and Indices

$i \in I$: Customer Index and set of customers.

$j \in J$: Index of products and set of products.

Decision Variables

$y_{i,j} \geq 0$: Customer $i \in I$ is offered product $j \in J$.

Parameters

$r_{i,j}$: Expected profit given that customer $i \in I$ is offered product $j \in J$.

$v_{i,j}$: Average variable cost associated with the offer of product $j \in J$ to customer $i \in I$.

Q_j: Minimum number of offers of product $j \in J$.

R: Corporate hurdle rate. Hurdle rate is defined as the minimum return on investment for the marketing campaign to be vaiable.

B: Total marketing campaign budget.

Basic Formulation: Constraints

- Budget. The total marketing budget is limited by a fixed value B

$$\sum_{i \in I} \sum_{j \in J} v_{i,j} \cdot y_{i,j} \leq B$$

- Number of offers. Maximum number of products for each customer should be less than or equal to one (only one product can be offered to customer).

$$\sum_{j \in J} y_{i,j} \leq 1 \quad \forall i \in i$$

- Offers limit. Minimum number of offers of each product.

$$\sum_{i \in I} y_{i,j} \geq Q_j \quad \forall j \in J$$

- ROI. The minimum ROI constraint ensures that the ratio of total profits over cost is at least one plus the corporate hurdle rate.

$$\sum_{i \in I} \sum_{j \in J} r_{i,j} \cdot y_{i,j} \geq (1 + R) \cdot \sum_{i \in I} \sum_{j \in J} v_{i,j} \cdot y_{i,j}$$

where, xi,j∈{0,1} This variable is equal to 1, if product j∈J is offered to customer i∈I, and 0 otherwise.

Basic/Naive Formulation: Objective Function

Total profit.

Maximize total expected profit from marketing campaign limited by the budget.

$$\textbf{Max} \quad Z = \sum_{i \in I} \sum_{j \in J} r_{i,j} \cdot y_{i,j}$$

Why do we need a better method than the Naïve Model?

- The ideal formulation is difficult to solve because of its scale. For 1 million customers and 10 products there are 10-million integer variables y_{ij}, this yields $2^{10,000,000}$ possible customer-offer combinations.
- Using standard optimization methods, a problem of this size can, in principle, result in a branch and cut tree of as many nodes.
- Thus, problems of this size are extremely difficult to solve, so we need a better solution than the naïve method. So, we use set-covering formulation method to approach this problem in a more efficient manner.

Set-Covering Formulation Approach

- We divide the problem into two individual problems.
  a. Tactical Model Formulation
  b. Operational Model Formulation

Tactical Formulation

- We create clusters of customers based on individual expected profit parameters, which can be computed using predictive models. We use K-Means for clustering the customers.
- By clustering, we can formulate the problem as a Linear Programming Problem by identifying the proportion within each cluster for each product that will maximize the return considering the constraints, instead of assigning offers to individual customers

Operational Model Formulation

- Here, we determine the product offers for individual customers
- We use estimated expected profits and output of tactical formulation as the inputs to operational formulation. We use this information to assign the products to individual customers within each cluster.

Tactical Formulation

Sets and Indices

$k \in K$: Index and set of clusters.

$j \in J$: Index and set of products.

Decision Variables

$y_{k,j} \geq 0$: Number of customers in cluster $k \in K$ that are offered product $j \in J$.

Parameters

$r_{k,j}$: Expected profit to the bank from the offer of product $j \in J$ to an average customer of cluster $k \in K$.

$v_{k,j}$: Average variable cost associated with the offer of product $j \in J$ to an average customer of cluster $k \in K$.

N_k: Number of customers in cluster $k \in K$.

Q_j: Minimum number of offers of product $j \in J$.

R: Corporate hurdle rate. This hurdle rate is used for the ROI calculation of the marketing campaign.

B: Marketing campaign budget.

Tactical Formulation: Constraints

- **Number of offers.** Maximum number of offers of products for each cluster is limited by the number of customers in the cluster

$$\sum_{j \in J} y_{k,j} \leq N_k \quad \forall k \in K$$

- **Offers limit.** Minimum number of offers of each product

$$\sum_{k \in K} y_{k,j} \geq Q_j \quad \forall j \in J$$

- **Budget.** The marketing campaign budget constraint enforces that the total cost of the campaign should be less than the budget campaign.

$$\sum_{k \in K} \sum_{j \in J} v_{k,j} \cdot y_{k,j} \leq B$$

- **ROI.** The minimum ROI constraint ensures that the ratio of total profits over cost is at least one plus the corporate hurdle rate.

$$\sum_{k \in K} \sum_{j \in J} r_{k,j} \cdot y_{k,j} \geq (1 + R) \cdot \sum_{k \in K} \sum_{j \in J} v_{k,j} \cdot y_{k,j}$$

Tactical Formulation: Objective Function

Total profit.

Maximize total expected profit from marketing campaign which is constrained by the total budget.

$$\text{Max} \quad Z = \sum_{k \in K} \sum_{j \in J} r_{k,j} \cdot y_{k,j}$$

Where

$y_{k,j} \geq 0$: Number of customers in cluster $k \in K$ that are offered product $j \in J$.

$r_{k,j}$: Expected profit from product $j \in J$ to customer of cluster $k \in K$.

Operational Model Formulation

Sets and Indices

$i \in I^k$: Index and set of customers in cluster $k \in K$.

$j \in J^k$: Index and subset of products offered to customers in cluster $k \in K$, where $J^k = \{j \in J : y_{k,j} > 0\}$.

Decision Variables

$x_{k,i,j} \in \{0, 1\}$: This variable is equal to 1, if product $j \in J^k$ is offered to customer $i \in I^k$, and 0 otherwise.

Parameters

$r_{k,i,j}$: Expected individual profit of customer $i \in I^k$ from offer of product $j \in J^k$.

$Y_{k,j} = \lfloor y_{k,j} \rfloor$: Number of customers in cluster k that will get an offer of product $j \in J^k$.

Operational Model Formulation: Constraints

- **Product offers.** Allocate offers of a product to customers of each cluster.

$$\sum_{i \in I^k} x_{k,i,j} = Y_{k,j} \quad \forall j \in J^k, k \in K$$

- **Binary constraints.** Either a product offer is given to a customer of cluster k or not.

$$x_{k,i,j} \in \{0, 1\} \quad \forall i \in I^k, j \in J^k, k \in K$$

- **Offers limit.** At most one product may be offered to a customer of a cluster.

$$\sum_{j \in J^k} x_{k,i,j} \leq 1 \quad \forall i \in I^k, k \in K$$

Where

$x_{k,i,j} \in \{0, 1\}$: This variable is equal to 1, if product $j \in J^k$ is offered to customer $i \in I^k$, and 0 otherwise.

$Y_{k,j} = \lfloor y_{k,j} \rfloor$: Number of customers in cluster k that will get an offer of product $j \in J^k$.

Operational Model Formulation: Objective Function

Total profit.

Objective Function

- **Total profit**. Maximize total individual expected profit.

$$\text{Max} \quad Z = \sum_{k \in K} \sum_{i \in I^k} \sum_{j \in J^k} r_{k,i,j} \cdot x_{k,i,j}$$

Where

$x_{k,i,j} \in \{0, 1\}$: This variable is equal to 1, if product $j \in J^k$ is offered to customer $i \in I^k$, and 0 otherwise.

$r_{k,i,j}$: Expected individual profit of customer $i \in I^k$ from offer of product $j \in J^k$.

Experimental Parameter Generation

- Budget.

  We set budgets as the average cost times the number of customers

- Product Minimum Quantity. We generated this number as a random number such that the sum of all the products is at least 75% of the customers count

- Cost per Product per Customer. Generated a random number from 80 to 450 with 40 spacing

- Expected Profit per Product per Customer. Generated a random number from 1000 to 4000 with 200 spacing

Note: The cost and expected profit are not as per the experiments provided in the paper [1] but we made sure that the solution is feasible

Business Experiments: Description

- An experiment was conducted to analyze the impact of budget on Return on Investment, Expected profits and Expected costs on the methodologies – Naïve Method and Set-Formulation Method.
- The experiment was run while relaxing the "Return on Investment" constraint for the two methods. The experiment was conducted using the following set of parameters.

Parameters:

Number of customers: 10,000
Number of products offered: 15
Number of clusters: 10

The budget was varied between $80,000 and $2,941,698.

Business Experiments: Results

The results from the experiments are shown as below:

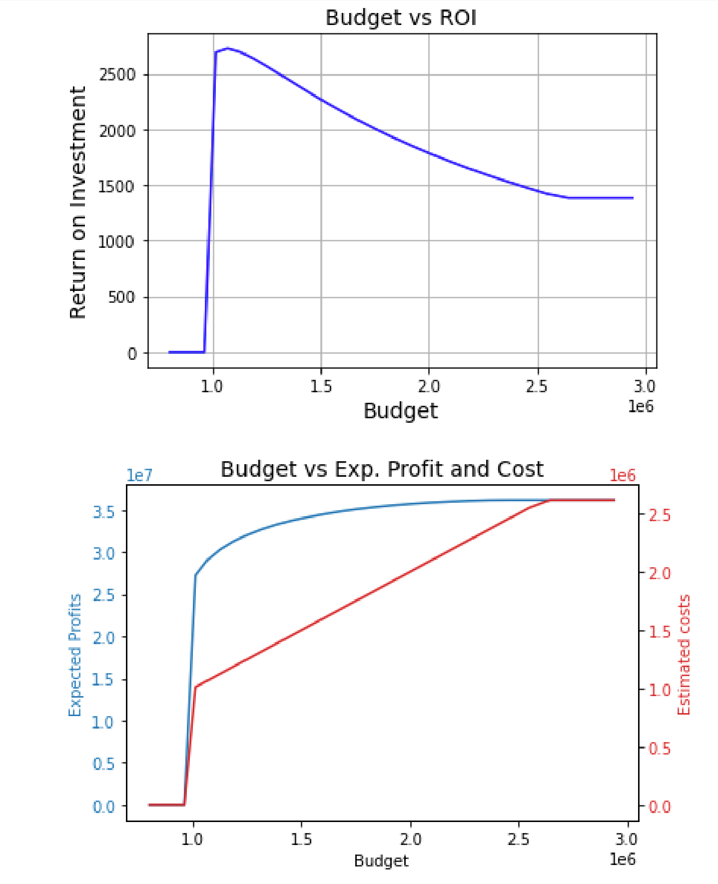

Fig 1: Campaign budget vs business metrics for the Naïve Model

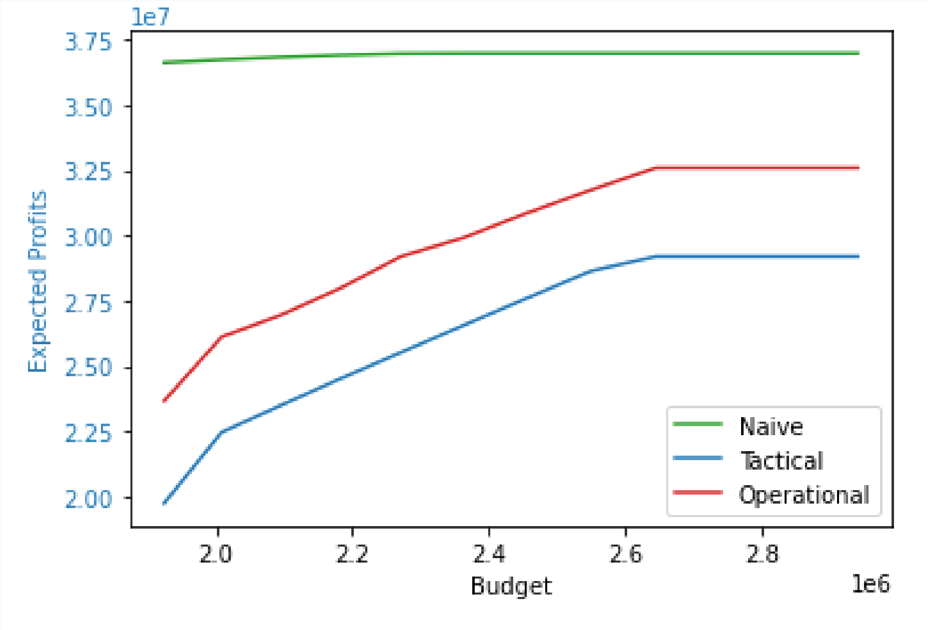

Fig 2: Expected profits estimated by the Naïve and Set-Formulation Models.

Business Experiments: Observations and Conclusion

- From Fig. 1, the expected ROI (100 * Total Expected Profit / Total Expected Loss) seems to increase steeply as we increase our campaign budget and then, decreases gradually. The initial ROI of 0 is when the the model is infeasible, and the allotted campaign budget is insufficient for the campaign.

- The above observation shows us that after a particular budget, increase in the expected cost of campaign does not give a proportionate increase in profits.

- The ROI flattens after a particular budget since only one product can be offered to each customer and once the product with most expected profit is offered given the constraints, we can no longer increase the profits. This is also evident from the expected profits and costs plot in Fig. 1.

- Fig 2. shows us that the Naïve model gives us the best estimate of expected profits. This is expected since, we relax the binary constraint Y_{ij} in the set-covering formulation method. The Operational model gives us a better estimate since we are introducing the binary constraints again and the provided solution is closer to the optimal objective value given by the Naïve Model.

Computational Experiments: Description

- We analyzed the algorithmic run times of the Basic Formulation and the Set-Covering Formulation with a given budget and varying the other parameters like Number of Customers, Number of Products and Corporate Rate

- We also determine the Optimality Gap of the Basic formulation and the Set-Covering Formulation (with K-Means and Random Clustering)

- The Experiments are run by varying with 3 product counts, 6 customer counts and 3 corporate rate. In total, we ran 3 * 6 * 3 = 54 computational experiments for basic formulation and varying only product and customer with 18 experiments with the set-covering formulation.

- All other fixed constraints including budget are computed as per the paper [1] in such a way that the results are feasible and consistent with the real-time behavior.

Parameters

i. Number of Products = [5, 10, 15]
ii. Number of Customers = [100, 200, 300, 1000, 2000, 10000]
iii. Corporate rate = [0.05, 0.10, 0.15]

Computational Experiments: Results

Basic Formulation: Corporate Rate = 0.05

Number of Products	Number of customers	Corporate Rate	Algorithm Running time	Budgets	Best Objective
5	100	0.05	0.02	26616.00	336200.00
5	200	0.05	0.04	51576.00	676600.00
5	300	0.05	0.05	78856.00	1025000.00
5	1000	0.05	0.18	258112.00	3399400.00
5	2000	0.05	0.28	518688.00	6795200.00
5	10000	0.05	1.63	2605928.00	33970600.00
10	100	0.05	0.64	26216.00	360600.00
10	200	0.05	0.76	51752.00	723200.00
10	300	0.05	0.67	77344.00	1084200.00
10	1000	0.05	0.93	260052.00	3610600.00
10	2000	0.05	1.39	519988.00	7234400.00
10	10000	0.05	4.45	2606792.00	36162600.00
15	100	0.05	2.18	26186.67	369400.00
15	200	0.05	2.08	52538.67	740600.00
15	300	0.05	2.22	77440.00	1107400.00
15	1000	0.05	2.41	259072.00	3692200.00
15	2000	0.05	3.24	518936.00	7411000.00
15	10000	0.05	8.56	2599925.33	36943200.00

Computational Experiments: Results

Basic Formulation: Corporate Rate = 0.1

Number of Products	Number of customers	Corporate Rate	Algorithm Running time	Budgets	Best Objective
5	100	0.1	4.30	25448.00	335000.00
5	200	0.1	4.36	52744.00	676800.00
5	300	0.1	4.35	78296.00	1030200.00
5	1000	0.1	4.43	260680.00	3405400.00
5	2000	0.1	4.47	522816.00	6793800.00
5	10000	0.1	6.70	2600112.00	33836200.00
10	100	0.1	5.06	26016.00	362000.00
10	200	0.1	5.23	51996.00	725200.00
10	300	0.1	5.14	77524.00	1084200.00
10	1000	0.1	5.54	259572.00	3614800.00
10	2000	0.1	6.04	519912.00	7225800.00
10	10000	0.1	9.48	2599504.00	36159000.00
15	100	0.1	6.94	25789.33	369000.00
15	200	0.1	6.73	51738.67	738200.00
15	300	0.1	6.90	77893.33	1109800.00
15	1000	0.1	8.48	258357.33	3700600.00
15	2000	0.1	7.68	518133.33	7388600.00
15	10000	0.1	17.18	2603560.00	36932200.00

Computational Experiments: Results

Basic Formulation: Corporate Rate = 0.15

Number of Products	Number of customers	Corporate Rate	Algorithm Running time	Budgets	Best Objective
5	100	0.15	9.18	26400.00	348000.00
5	200	0.15	9.02	51704.00	662000.00
5	300	0.15	9.06	77000.00	1017400.00
5	1000	0.15	9.16	260552.00	3421000.00
5	2000	0.15	10.25	518424.00	6779800.00
5	10000	0.15	11.78	2614504.00	33902800.00
10	100	0.15	8.14	26228.00	361200.00
10	200	0.15	8.85	51576.00	725000.00
10	300	0.15	9.03	78528.00	1078600.00
10	1000	0.15	8.65	260388.00	3611000.00
10	2000	0.15	11.24	521008.00	7219400.00
10	10000	0.15	12.09	2603260.00	36159800.00
15	100	0.15	11.25	25858.67	367400.00
15	200	0.15	11.87	52192.00	738000.00
15	300	0.15	11.93	77973.33	1107000.00
15	1000	0.15	12.55	258466.67	3701200.00
15	2000	0.15	12.58	523194.67	7393600.00
15	10000	0.15	19.31	2603194.67	36975400.00

Computational Experiments: Results

Set-Covering Formulation with random clustering: Keeping Corporate Rate constant at 0.15

Number of Products	Number of customers	corporate rate	Algorithm Running time	Budget	Objective Value	Gap %
5	100	0.15	8.18	26400.00	342340	1.63
5	200	0.15	8.40	51704.00	599192	9.49
5	300	0.15	8.90	77000.00	1011096	0.62
5	1000	0.15	9.10	260552.00	3523348	2.99
5	2000	0.15	9.40	518424.00	7356400	8.50
5	10000	0.15	10.10	2614504.00	36344030	7.20
10	100	0.15	8.20	26228.00	326202	9.69
10	200	0.15	8.40	51576.00	706229	2.59
10	300	0.15	8.40	78528.00	1078194	0.04
10	1000	0.15	8.70	260388.00	3933182	8.92
10	2000	0.15	10.20	521008.00	7311942	1.28
10	10000	0.15	12.20	2603260.00	34728581	3.96
15	100	0.15	10.21	25858.67	357522	2.69
15	200	0.15	10.40	52192.00	695928	5.70
15	300	0.15	10.20	77973.33	1058769	4.36
15	1000	0.15	12.99	258466.67	3736773	0.96
15	2000	0.15	14.70	523194.67	7188666	2.77
15	10000	0.15	16.40	2603194.67	40568455	9.72

Computational Experiments: Results

Set-Covering Formulation with K-means clustering: Keeping Corporate Rate constant at 0.15

Number of Products	Number of customers	Corporate Rate	Algorithm Running time	Budget	Objective Value	Gap %
5	100	0.15	10.18	26400.00	343489	1.3
5	200	0.15	10.20	51704.00	726560	9.8
5	300	0.15	10.60	77000.00	903053	11
5	1000	0.15	10.80	260552.00	3511567	2.6
5	2000	0.15	10.90	518424.00	6815645	0.5
5	10000	0.15	11.20	2614504.00	33237144	2
10	100	0.15	9.60	26228.00	333231	7.7
10	200	0.15	9.90	51576.00	784510	8.2
10	300	0.15	9.03	78528.00	1128544	4.6
10	1000	0.15	10.40	260388.00	3282964	9.1
10	2000	0.15	12.50	521008.00	6722006	6.9
10	10000	0.15	14.50	2603260.00	35704886	1.3
15	100	0.15	12.21	25858.67	314564	14
15	200	0.15	12.60	52192.00	788081	6.8
15	300	0.15	12.80	77973.33	1072872	3.1
15	1000	0.15	14.99	258466.67	3497399	5.5
15	2000	0.15	16.70	523194.67	8143881	10
15	10000	0.15	20.40	2603194.67	37531844	1.5

Computational Experiments: Observation and Conclusion

- Based on the Basic Formulation Results, we can conclude that the running time increases as we increase any of the parameters such as Number of products, Number of Customers and Corporate Rate.

- There is also a proof in the paper[1] that shows that the Basic Formulation is an NP-Hard Problem

- The Basic formulation gave us exact solutions and hence, there was no optimality gap

- We observed that in the set-covering formulations, there was some optimality gap, which increases as we increased each of the parameters

- We also observe that the Set-covering algorithm with K-Means takes longer time than the one with random clustering

Problem Extension Experiment: Budget Correction

- The marketing campaign problem that we formulated is a generic version and companies can add extensions, change the constraints or change objective functions tailored to their business requirements.

- We explored one such extension to experiment with the optimization model. We used a hard constraint to limit the budget for the campaign and this gave us infeasible solution for low budgets.

- In order to mitigate this problem, we introduced a new variable "z" which acts as a budget correction variable. However, we penalize the budget correction heavily with a Scalar "M".

Budget Correction: Formulation Changes

The changes in the problem formulation for the naïve model is shown below.

Constraints

- **Budget**. The marketing campaign budget constraint enforces that the total cost of the campaign should be less than the budget campaign. There is the possibility of increasing the budget to ensure the feasibility of the model, the minimum number of offers for all the product may require this increase in the budget.

$$\sum_{k \in K} \sum_{j \in J} v_{k,j} \cdot y_{k,j} \leq B + z$$

Where

$y_{i,j} \geq 0$: Customer $i \in I$ that is offered product $j \in J$.

$z \geq 0$: Increase in budget in order to have a feasible campaign.

$v_{k,j}$: Average variable cost associated with the offer of product $j \in J$ to customer $i \in I$.

B: Marketing campaign budget.

Objective Function

- **Total profit**. Maximize total expected profit from marketing campaign and heavily penalize any correction to the budget.

$$\text{Max} \quad Z = \sum_{k \in K} \sum_{j \in J} r_{k,j} \cdot y_{k,j} - M \cdot z$$

Where

$y_{i,j} \geq 0$: Customer in cluster $i \in I$ that is offered product $j \in J$.

$z \geq 0$: Increase in budget in order to have a feasible campaign.

$r_{i,j}$: Expected profit to the bank from the offer of product $j \in J$ to an average customer of cluster $i \in I$.

Budget Correction: Results

- We run an experiment with 100 customers, 5 products and an insufficient budget of $8000 on the naïve model with budget correction.
- We provide a heavy penalty of 10000 to the budget correction factor M.
- The expected profit after the correction is $252,400.00.
- Optimal total expected cost is $12,800.00 with a budget of $8,000.00 and an extra amount of $4,800.00.
- Optimal ROI is 1971.88% with a minimum ROI of 110.0%.
- Thus, this extension can be used when the budget is insufficient.

Conclusion

- In this project, we explored the various optimization problem formulations for providing targeted offers in marketing
- We experimented with a Basic Formulation and a Set Covering Formulation
- From the business experiments, we can conclude that the business stake holders should investigate the insights from the models and take a decision from the data. For example, we can see that ROI decreases while Profits increase as the campaign budget increases. The business should decide if they need to prioritize profits or the ROI.
- We concluded that the Basic formulation from our experiments but provided exact solutions for the LP problem
- Then, We solved with a Set-Covering Formulation by random clustering and K-Means clustering
- We saw that random clustering performed better in terms of running time that with K-means Clustering
- Both the methods had some optimality gap, that increased as we increased the data

References

1. Optimization models for targeted offers in direct marketing: exact and heuristic algorithms.
   https://feb.kuleuven.be/public/u0004371/published%20papers/promotion.pdf

2. Exploiting response models: optimizing cross-sell and up-sell opportunities in banking.
   https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.12.7598&rep=rep1&type=pdf

Thank You