Uplift Modeling in Retail Marketing using SAS

Aashish Arjun Bugalia, Oklahoma State University, Stillwater, OK

Agastya Komarraju, Oklahoma State University, Stillwater, OK

Abstract

Direct marketing campaigns are often targeted to randomly selected customers which results in huge costs for the company. Also, such campaigns result in customer frustration and makes them less likely to react to further communications in future. Companies involved in direct marketing campaigns often use a random response model to target customers for the campaigns. An alternative approach can be through uplift modeling - precise targeting of the beneficial customers resulting in greater return on invested money and resources by the company. Based on the marketing and finance literature, this article looks at high prediction accuracy for the probability of purchase based on a sample of customers, to whom a pilot campaign has been sent. Uplift modeling analyzes the causal effect of an action such as a marketing campaign on a given individual by considering difference in response rate between a treated group and a randomized control group. The resulting model can then be used to select individuals for whom the action will be most profitable. This article aims at predicting beneficial customers to an online retailer with the implementation of several statistical, machine learning and deep learning methods in SAS. Through above-mentioned methods this paper will also help to know effectiveness of the campaign by determining incremental gains, thus resulting in greater return on invested money and resources by the company.

Keywords: Direct marketing, Uplift Modeling, Decision Tree, Logistic Regression, Random Forest, Neural network, Decision tree, Ensemble methods

1 Introduction

A marketing campaign comprises of a time bound offer sent to a specific set of people to observe a specific behavior and has a specific reward attached to it. When a company sends marketing campaigns directly by email or telephone such campaigns are called direct marketing campaigns and their performance is assessed by their effectiveness. However, many factors can affect a marketing campaign, for example seasonality. To avoid these affecting factors, marketers follow a strategy in which the whole population is divided into treatment group and control group. Control group consists of people who have are not given the offers or are not sent the campaign. While on the other hand, the treatment group consists of people who are sent offers. Direct response marketing looks to tempt imminent customers into making a specific move following receiving or reading an advertisement. However, they lack efficient strategies for targeting beneficial customers and thus result in wasting money and resources. The two common modeling techniques used by companies are the traditional response-based modeling or the uplift modeling. Both techniques result in the response probabilities of buying a product if given an offer for an individual. When response probabilities are known, just those customers who are probably going to react over some threshold can be incorporated into the mailing list, while others with response likelihood beneath the threshold would be dropped. This can be potentially gainful as far as expanding response rates to explicit promoting sales and producing additional income. This likewise diminishes generally promoting expenses since volume of mailing lessens, while response rates go up.

In this paper, we discuss the above mentioned two techniques used by marketers for optimizing the direct marketing campaigns. We also look at the various shortcomings of the traditional model and we experimentally verify these claims on real direct marketing data. The data is publicly available and comes from an online retailer offering women's and men's merchandise. We test response-based models as well as uplift approaches described and compare them with the help of various machine learning techniques such as logistic regression, decision trees and neural networks in SAS enterprise miner. The model and uplift results assure that the uplift approach gives much better marketing outcomes. Uplift modelling has applications in customer relationship management for up-sell, cross-sell and retention modelling. It has also been applied to political election and personalized medicine.

2 Problem Statement

We have taken the use case mentioned in Mine That Data challenge by Kevin Hillstrom[3]. In this challenge, an online retailer ran two campaigns one for men and other for women through various channels. In these campaigns the customers were involved in an e-mail test and,

- 1/3 were randomly chosen to receive an e-mail campaign featuring men's merchandise.
- 1/3 were randomly chosen to receive an e-mail campaign featuring women's merchandise.
- 1/3 were randomly chosen to not receive an e-mail campaign.

During a period of two weeks following the e-mail campaign, results were tracked. Here, we are targeting the problem of predicting the people who visited the site within the two-week period because they received the campaign. This was calculated by analyzing the treatment group(customer who received the campaign) and the control group(customers who didn't receive email). Also, we will be looking into how much of incremental gains can be made with respect to control group marketing.

3 Data Exploration

We have used Hillstrom[3] email dataset for an Internet based retailer for performing analysis of predictive response and uplift. This dataset contains 64,000 customers who last purchased within twelve months. The various attributes in the dataset are listed below as follows, *Data Attributes:*

Historical customer attributes include:

- **Recency**: Months since last purchase.
- **History_Segment**: Categorization of dollars spent in the past year.
- **History**: Actual dollar value spent in the past year.
- **Men's**: 1/0 indicator, 1 = customer purchased Men's merchandise in the past year.
- **Women's**: 1/0 indicator, 1 = customer purchased Women's merchandise in the past year.
- **Zip_Code**: Classifies zip code as Urban, Suburban, or Rural.
- **Newbie**: 1/0 indicator, 1 = New customer in the past twelve months.
- **Channel**: Describes the channels the customer purchased from in the past year.

Another variable describes the e-mail campaign the customer received:

- **Segment**
  - Men's E-Mail
  - Women's E-Mail
  - No E-Mail

The following *figure 1* depicts the number of records included in each campaign.

Finally, we have a series of variables describing activity in the two weeks following delivery of the e-mail campaign:

- **Visit**: 1/0 indicator, 1 = Customer visited website in the following two weeks.
- **Conversion**: 1/0 indicator, 1 = Customer purchased merchandise in the following two weeks.
- **Spend**: Actual dollars spent in the following two weeks.

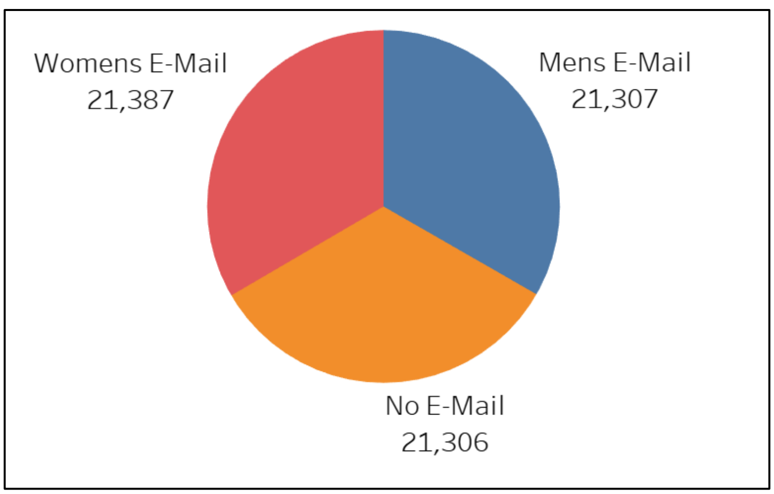


Figure 1: Number of records in each campaign

Since there is a large difference in response between treatment groups who received advertisements for men's and women's merchandise, the two campaign types were analyzed jointly as shown in *figure 2*. In this case, the treatment group consists of all those who received an e-mail and the control group of those who did not.

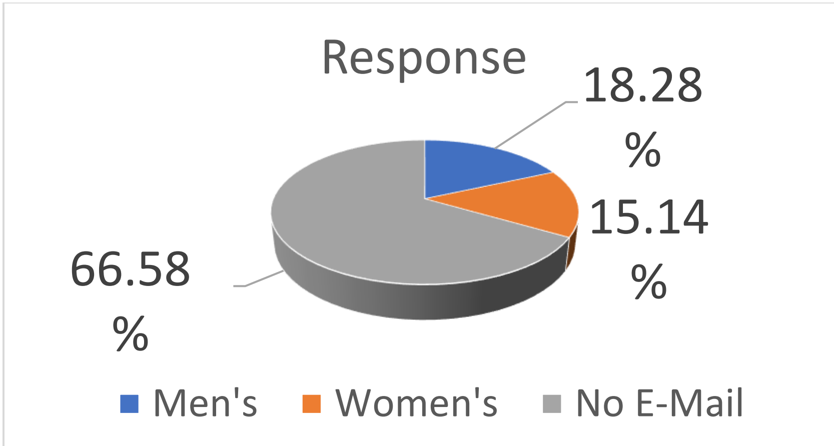


Figure-2: Response from each campaign for variable visits

Because of this some of the variables were rejected, as mentioned below:

- **Men's**
- **Women's**

Also, since we are predicting whether the customer visited the website or not so we are rejecting the variable, **Conversion.**

4 Methodology

4.1 Traditional Response modeling

[1]Traditional response modelling typically takes a group of treated customers and attempts to build a predictive model that separates the likely responders from the non-responders through the use of one of a number of predictive modelling techniques. This model will separate those who are likely to respond from those who are less likely to respond as shown in *figure 3*. This model would only use the treated customers to build the model and hence this approach seems to be eschewing potentially important data. The traditional models predict the conditional class probabilities,

$$P(response \mid treatment).$$

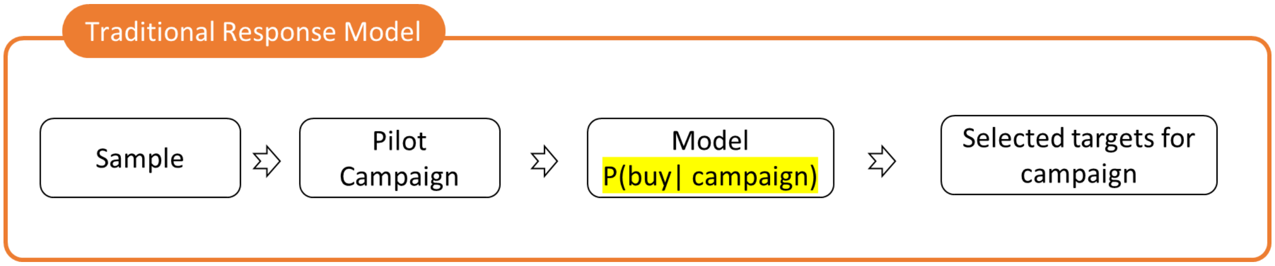


Figure-3: Traditional response modeling process

Since, these models don't take into consideration all the customers, these traditional based modeling were not effective as they didn't tell the effectiveness of the campaign. As an alternative and more efficient approach direct marketing industry evolved the response-based modeling with the introduction of uplift modeling.

4.2 Uplift Modeling

Uplift modelling[1], also known as incremental modelling, true lift modelling, or net modelling is a predictive modelling technique that directly models the incremental impact of a treatment on an individual's behavior. The uplift of a marketing campaign is usually defined as the difference in response rate between a treated group and a randomized control group. Uplift modelling uses a randomized scientific control to not only measure the effectiveness of an action but also to build a predictive model that predicts the incremental response to the action. This allows a marketing team to isolate the effect of a marketing action and measure the effectiveness or otherwise of that individual marketing action.

Uplift modeling process involves dividing the whole population into two groups, treatment and control. Treatment group refers to people who have been sent the offers through a campaign and the control group refers to people who aren't sent any offers and thus their behavior is controlled. The uplift modeling process is defined in *figure 4* and it predicts the change in behavior by calculating the probabilities as follows,

P(response | treatment)–P(response |no treatment).

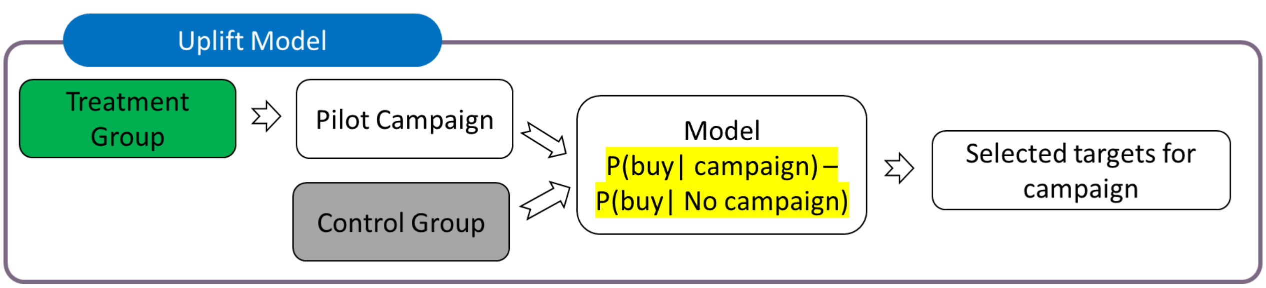


Figure-4: Uplift Modeling process

Due to further developments in uplift modeling, there are three main approaches that the marketers use,

- Two models approach,
- Single model approach and
- Two-stage model approach.

Here, we will be discussing the two model and two model approach in uplift modeling.

4.2.1 Two model approach

[4]The two-model approach is commonly described in uplift modeling literature. It is a simple and intuitive approach. Two separate models are trained: a control model and a treatment model as shown in *figure 5*.

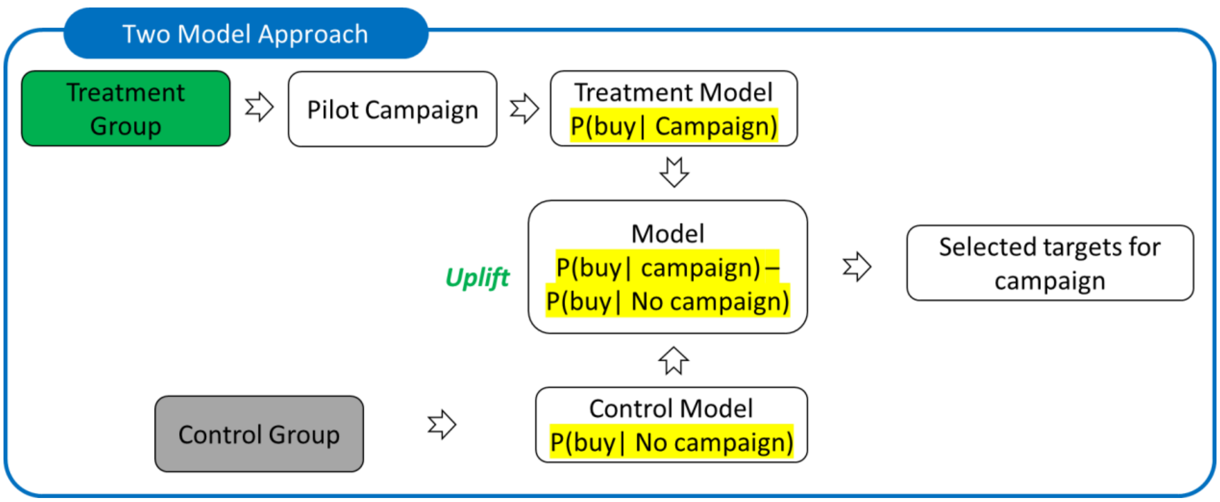


Figure-5: Two model uplift approach

The control model is trained only on the control data which consists of people who did not receive the promotion. It will predict how likely an individual will make a purchase without receiving any offer. The treatment model is trained only on the treatment data which consists of people who did receive the promotion. It will predict how likely an individual will make a purchase when they received offer and gives the uplift by calculating the following probability difference,

Uplift = P(response | treatment)−P(response |control).

Although this model looks good, it has few drawbacks as addressed by Victor Lo[5]. This approach indirectly models uplift i.e., while both models might accurately calculate the probabilities of response in either groups, the difference in the probabilities of the two models may not capture the precise lift. The difference may be only noise involving two times the work and scales may not be comparable. To overcome this issue of predicting lift precisely, we look at the two-stage modeling approach.

4.2.2 Two- stage modeling approach

The two-stage modeling approach can overcome the shortcomings of the two-model approach by modeling over both the datasets. Two different models can be trained on both the treatment group and control group. The creation process for the same is shown in *figure 6* as below.

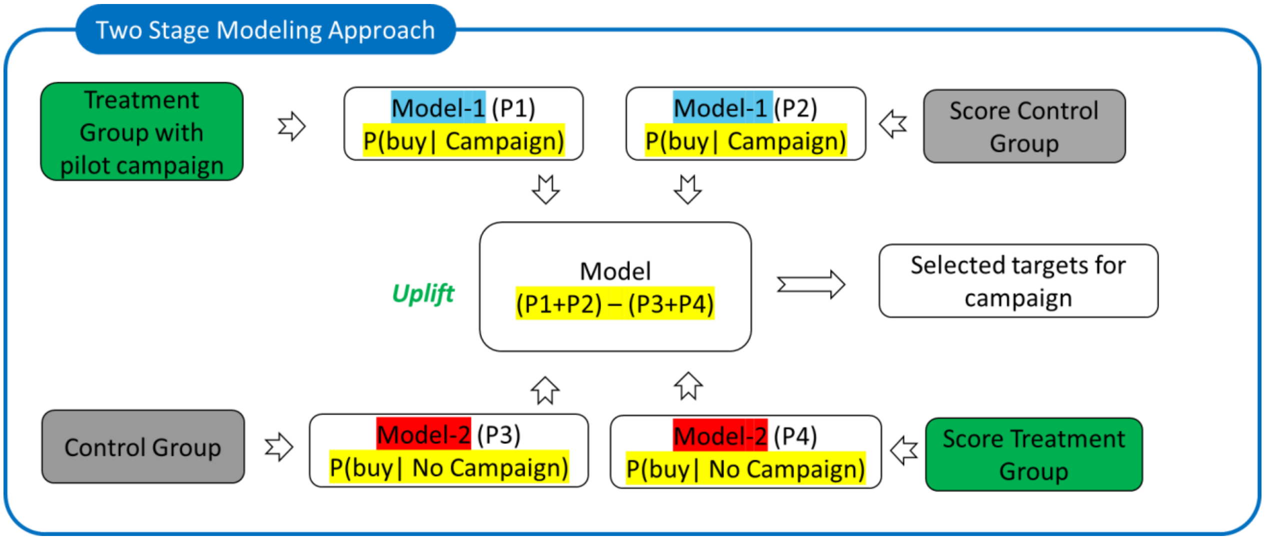


Figure-6: Two-stage uplift modeling approach

In this approach, one classifier is trained on the treatment group and the probability of response is calculated, say P1. This probability identifies how many people from the treatment group if offered an offer through campaign will respond. Then the control group is scored on model-1 and probability of response is calculated, say P2. This way we superimpose the treatment model on the control group, identifying the people who were not sent an offer earlier but if treated they will respond. Thus, identifying the beneficial customers from control group who can be targeted with an offer. Here, P1 and P2 identifies the individuals from the treatment and control group who will respond if targeted.

Now, another classifier is built on the control group identifying individuals who are going to buy from the control group giving us the probability P3. This probability gives us individuals who are going to respond even without an offer. As we did in the model-1, we will score the treatment group on model-2 and calculate the probability of people in the treatment group. This will identify individuals from the treatment group who will respond even without an offer with probability P4. So, basically P3 and P4 will give us the individuals who are surely going to buy the product regardless of the offer.

The difference of these 4 probabilities gives us the uplift modeling of the campaign as shown in the *figure 6*.

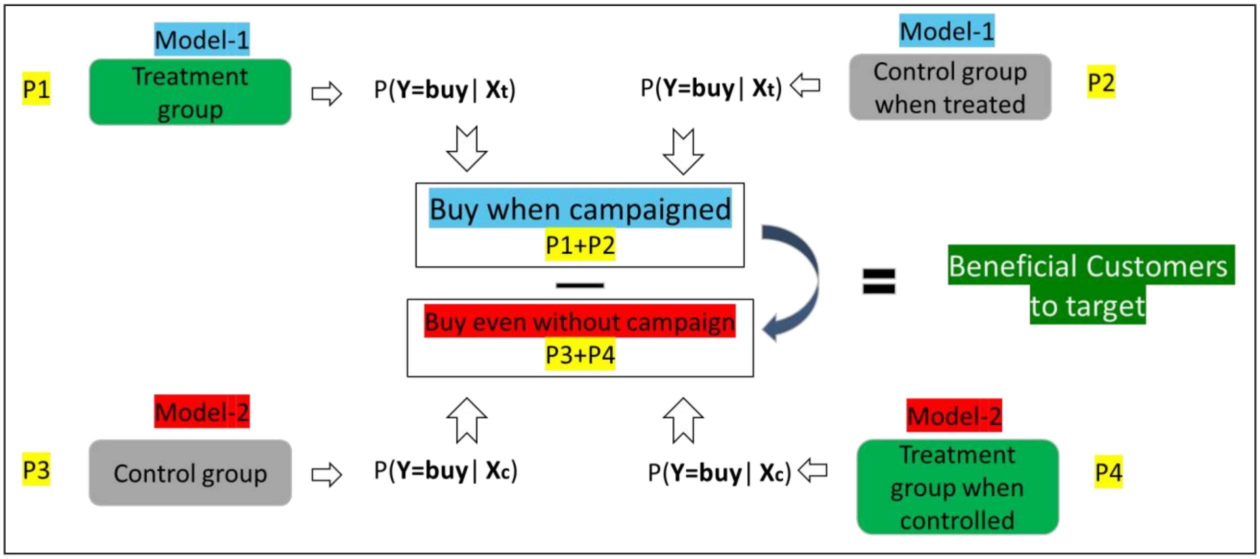


Figure-7: Two stage uplift modeling probability calculation process

*Xc – Controlled group, Xt – Treatment group

*P1, P2 -> P(Y=buy |Xt), probability of respond if treated

* P3, P4 -> P(Y=buy |Xc), probability of respond if controlled

This model separates the people who will buy when they are offered from those who were going to respond even without an offer. The uplift modeling of the campaign can be calculated as follows:

Uplift = [P(response | treatment) + P(response | control group is treated)] –

[P(response |control) + P(response | treatment group is controlled)

OR

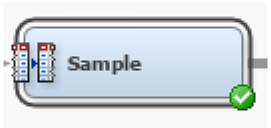
Uplift = [P1 + P2] – [P3 + P4].

5 Two-Stage Uplift Modeling Implementation in SAS® Enterprise Miner™

5.1 Data Preparation

The data had too much imbalance for the target column even after mixing the women's and men's campaign. So, we had to do an under-sampling on the data for both the control data and treatment data. For treatment group, the sample node in the SAS® EMiner was used with the following settings as shown in *figure 8.*

For treatment and control group,



General	
Node ID	Smpl
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Output Type	Data
Sample Method	Stratify
Random Seed	12345
Size	
Type	Computed
Observations	.
Percentage	100.0
Alpha	0.01
PValue	0.05
Cluster Method	Random
Stratified	
Criterion	Proportional
Ignore Small Strata	No
Minimum Strata Size	5
Level Based Options	
Level Selection	Rarest Level
Level Proportion	10.0
Sample Proportion	1.0

Figure-8: Sample node & Settings

.. Property	Value
General	
Node ID	Part2
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Output Type	Data
Partitioning Method	Stratified
Random Seed	12345
Data Set Allocations	
Training	80.0
Validation	20.0
Test	0.0
Report	
Interval Targets	Yes
Class Targets	Yes
Status	
Create Time	10/10/19 7:41 PM
Run ID	a7545157-cd32-4630-9

Figure-9: Data partition node & Settings

After this, we performed data partitioning with a split of 80:20 of the total for training and validation. The outputs are shown in the folloing *figures 10* for control group and *figure 11* for treatment group.

Summary Statistics for Class Targets

Data=DATA

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
visit	0	0	19044	89.3833	visit
visit	1	1	2262	10.6167	visit

Data=TRAIN

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
visit	0	0	15082.85	90.0918	visit
visit	1	1	1658.80	9.9082	visit

Data=VALIDATE

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
visit	0	0	3961.15	86.7845	visit
visit	1	1	603.20	13.2155	visit

Figure-10: Partitioning output control group

Data=DATA

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
visit	0	0	35562	83.2951	visit
visit	1	1	7132	16.7049	visit

Data=TRAIN

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
visit	0	0	24619.85	84.1096	visit
visit	1	1	4651.30	15.8904	visit

Data=VALIDATE

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
visit	0	0	10942.15	81.5189	visit
visit	1	1	2480.70	18.4811	visit

Figure-9: Partitioning output treatment group

5.2 Prediction Models

For modeling purpose, we used 4 different types of classification models for both the groups. The modeling process flow for both the groups has been shown in figure 12(treatment group) and figure 16(treatment group).

<u>Model-1</u>

For performing uplift modeling on the treatment data and scoring on control data following process is used.

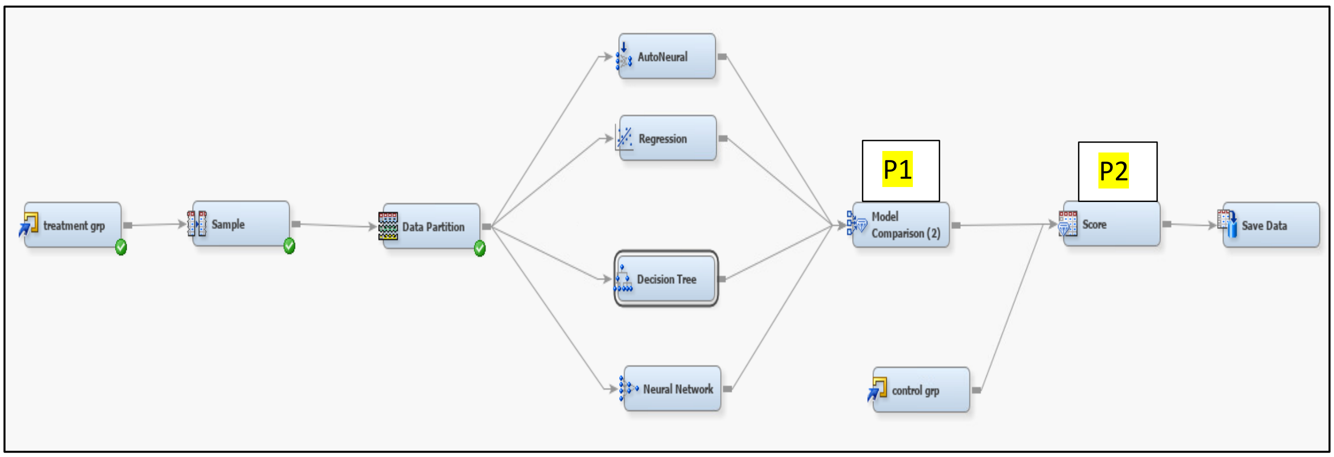


Figure-12: SAS® Enterprise Miner™ process flow diagram for treatment group

From the above process flow, the neural network model was the champion model with a misclassification of 16% and with a lift of 2.96. The misclassification for other models is mentioned in *figure 13* and the lift for all the models is shown in the *figure 14*. Based on these statistics by model comparison node, we choose to go with the neural network model for the treatment group.

Fit Statistics

Selected Model	Predecess or Node	Model Node	Model Descriptio n	Target Variable	Target Label	Selection Criterion: Valid: Misclassifi cation Rate ▲
Y	Neural	Neural	Neural N...	visit	visit	0.160795
	Tree	Tree	Decision ...	visit	visit	0.183897
	AutoNeural	AutoNeural	AutoNeural	visit	visit	0.184811
	Reg	Reg	Regressi...	visit	visit	0.229642

Figure-13: Misclassification statistics

Fit Statistics

Selected Model	Predecess or Node	Model Node	Model Descriptio n	Target Variable	Target Label	Selection Criterion: Valid: Lift
Y	Neural	Neural	Neural N...	visit	visit	2.962003
	Tree	Tree	Decision ...	visit	visit	2.558228
	Reg	Reg	Regressi...	visit	visit	1.249444
	AutoNeural	AutoNeural	AutoNeural	visit	visit	1

Figure-14: Lift statistics

Also, we looked at the ROC curve for all the models and the results have been shown in *figure 15.* In the figure, the neural network model comes out on top and, hence confirming our results.

This neural network is a multilayer perceptron model consisting of 1- layer network with 9 hidden nodes .

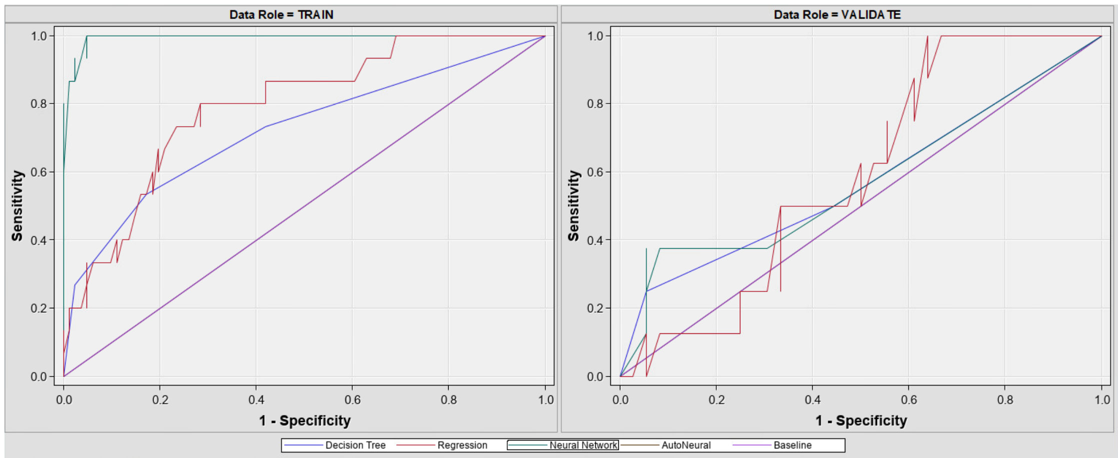


Figure-15: SAS® Enterprise Miner™ ROC curve results

<u>Model-2</u>

For performing uplift modeling on the control data and scoring on treatment data following process is used.

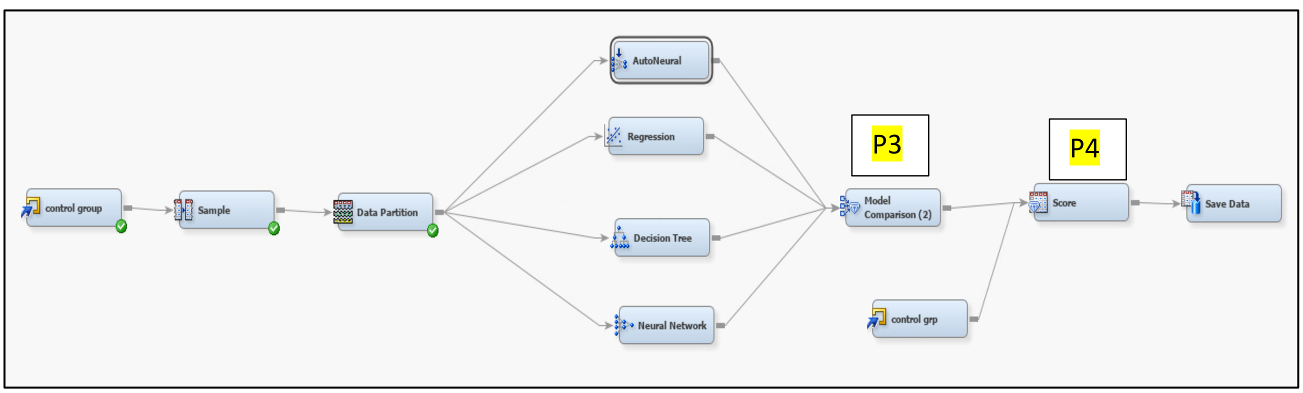
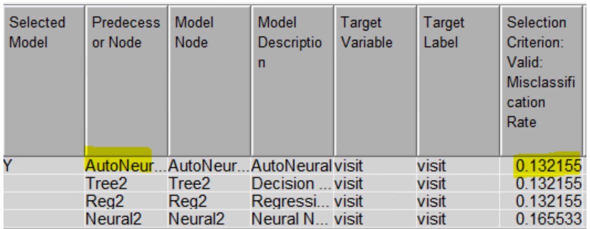


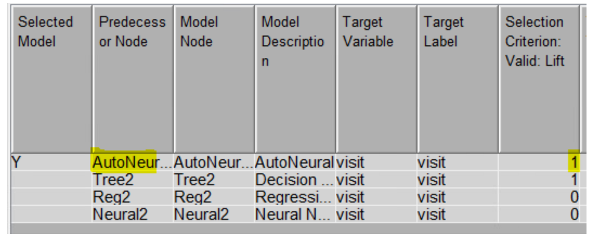
Figure-16: SAS® Enterprise Miner™ process flow diagram for treatment group

From the above process flow, the neural auto neural network model was the champion model with a misclassification of 13.2% and with a lift of 1.1. The misclassification for other models is mentioned in *figure 17* and the lift for all the models is shown in the *figure 18*. Based on these statistics by the model comparison node, we choose to go with the neural network model for the treatment group.



Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Misclassification Rate
Y	AutoNeur...	AutoNeur...	AutoNeural	visit	visit	0.132155
	Tree2	Tree2	Decision ...	visit	visit	0.132155
	Reg2	Reg2	Regressi...	visit	visit	0.132155
	Neural2	Neural2	Neural N...	visit	visit	0.165533

Figure-17: Misclassification statistics



Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Lift
Y	AutoNeur...	AutoNeur...	AutoNeural	visit	visit	1
	Tree2	Tree2	Decision ...	visit	visit	1
	Reg2	Reg2	Regressi...	visit	visit	0
	Neural2	Neural2	Neural N...	visit	visit	0

Figure-18: Lift statistics

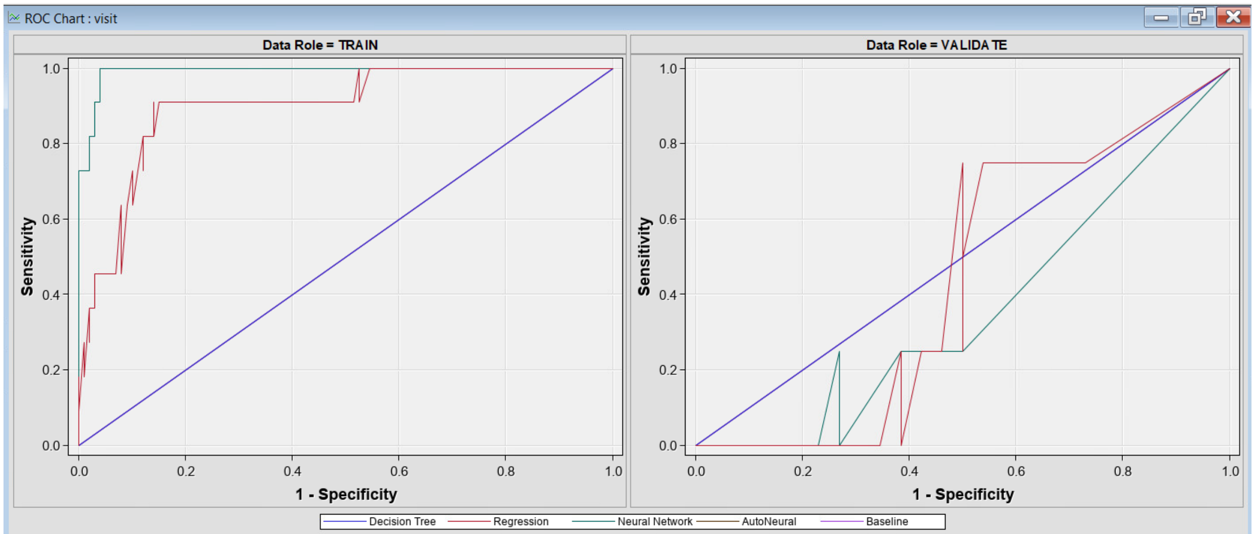


Figure-19: SAS® Enterprise Miner™ ROC curve results

Also, we looked at the ROC curve for all the models and the results have been shown in *figure 19.* In the figure, the auto neural network model comes out on top in the validation group and, hence confirming our results. This auto neural network is a multilayer perceptron model consisting of funnel layers network with 3 hidden nodes and 8 maximum iterations used.

5.3 Uplift Results

The two champion models were used for calculating the uplift probability for the model. After scoring the data, the data was saved as shown in *figure 12 and figure 16* through the save data node in SAS® Enterprise Miner™. The probabilities were calculated through the variable EM_EVENTPROBABILITY for both the input data and scoring data. The results for these probabilities can be seen in the below table 1.

Models	Input data	Scoring data	Input data probability	Scoring data probability
Model-1(Neural)	Treatment	Control	P1 = 13.85	P2 = 15.7
Model-2(Auto neural)	Control	Treatment	P3 = 9.9	P4 = 9.7

Table-1: Uplift modeling results

Based on the above calculated probabilities, we can calculate the uplift modeling probabilities for two-model approach as well as two-stage modeling approach as shown below.

Two-model uplift modeling approach,

Uplift **=** P1-P3 = 13.85 – 9.9 **= 3.95 %**

Two-stage uplift modeling approach**,**

Uplift **= (**P1+P2)-(P3+P4) = (13.85 +15.7)–(9.9+9.7) **= 9.95 %**

As can be seen from the above calculations, our results states that two-stage modeling is way better than two-model approach with a difference of 6%. The individuals are sorted and being placed in deciles starting based on the predicted and observed probabilities as shown in *figure 20.*

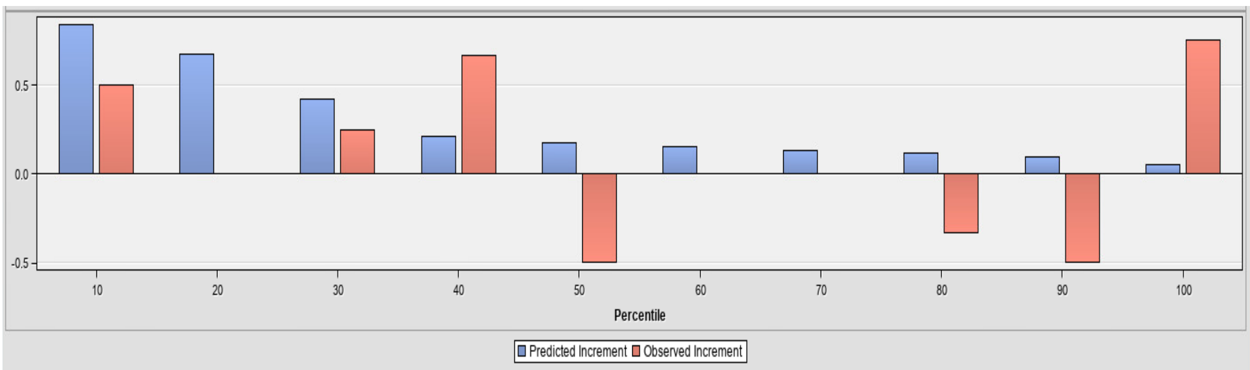


Figure-20: SAS® Enterprise Miner™ response rates

We calculated the average spend by the customers in the respective deciles based on the data mentioned in the spend variable and as shown in *figure 21.* From this figure we can see that our customers to target lie in the first six deciles.

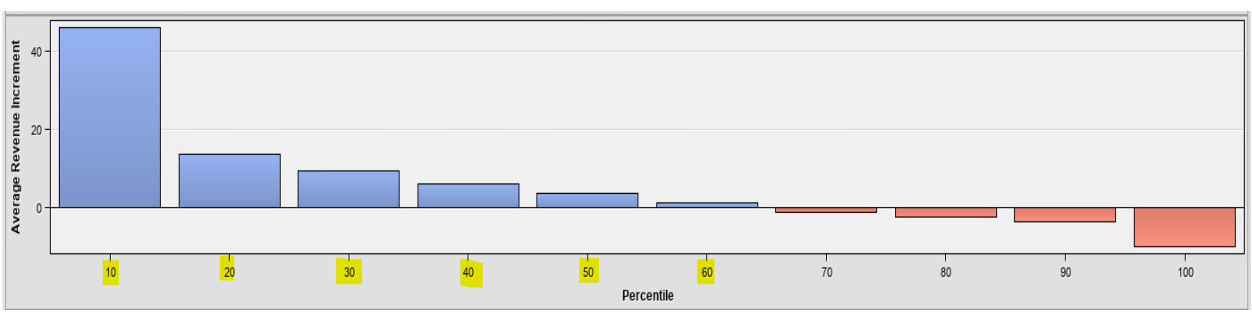


Figure-21: SAS® Enterprise Miner™ average spend results

6 Conclusions

We have seen how two- stage uplift modeling approach can be used to better identify individuals who will respond favorably to your marketing campaigns. We also noticed that neural network models prove to more efficient as compared to other machine learning models. Moreover, [1]uplift modeling separates customers into the following groups:

- The Persuadables: customers who only respond to the marketing action because they were targeted,
- The Sure Things: customers who would have responded whether they were targeted or not,
- The Lost Causes: customers who will not respond irrespective of whether or not they are targeted,
- The Do Not Disturbs or Sleeping Dogs: customers who are less likely to respond because they were targeted.

From the above discussions we can use the decile scoring technique that can separate customers into the groups described above. We know that the only segment that provides true incremental responses is the persuadables and uplift modeling helps identify it. While on the other hand, [1]traditional response modelling often targets the Sure Things being unable to distinguish them from the persuadables.

With uplift modeling, you can reduce the cost of marketing and enhance the value of your marketing campaign. Uplift modelling has applications in customer relationship management for up-sell, cross-sell and retention modelling. It has also been applied to political election and personalized medicine. This is an emerging field in marketing analytics and with the new developments in deep learning it can further be investigated.

7 References

[1] "Uplift modelling." en.wikipedia.org. https://en.wikipedia.org/wiki/Uplift_modelling (accessed March 28, 2019).

[2] Piotr Rzepakowski and Szymon Jaroszewicz, "Uplift Modeling in Direct Marketing", Journal of Telecommunications and Information technology, 02/2012(accessed: March 28, 2019)

[3] K. Hillstrom, "The MineThatData e-mail analytics and data mining challenge", MineThatData blog, 2008 [Online]. Available: http://blog.minethatdata.com/2008/03/minethatdata-e-mailanalytics-and-data.html, accessed: March 20, 2019

[4] Simple Machine Learning Techniques To Improve Your Marketing Strategy: Demystifying Uplift Models. Retrieved from, https://medium.com/datadriveninvestor/simple-machine-learning-techniques-to-improve-your-marketing-strategy-demystifying-uplift-models-dc4fb3f927a2, 2019.

[5] Identifying Individuals Who Are Truly Impacted by Treatment: Introduction to + Recent Advances in Uplift Modeling. Retrieved from, https://www.researchgate.net/publication/270217235_Identifying_Individuals_Who_Are_Truly_Impacted_by_Treatment_Introduction_to_Recent_Advances_in_Uplift_Modeling, 2019.

[6] SAS® Enterprise Miner™ 14.3: Reference Help. Retrieved from, https://documentation.sas.com/?docsetId=emref&docsetTarget=titlepage.htm&docsetVersion=14.3&locale=en, 2019.

8 Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Name: Aashish Arjun Bugalia

Enterprise: M.S. in Business Analytics (2018-2020), Oklahoma State University

Address: Stillwater, OK.

City, State ZIP: Stillwater, OK, 74074.

Work Phone: (+1)-405-780-2125

E-mail: aashish.bugalia@okstate.edu

LinkedIn: www.linkedin.com/in/aashish-arjun-bugalia

Web: https://business.okstate.edu/analytics/msban/index.html