Uplift Modeling: Predictive Analytics Can’t Optimize Marketing Decisions Without It

To drive business decisions for maximal impact, analytical models must predict the *marketing influence* of each decision on customer buying behavior.

Eric Siegel, Ph.D.
Prediction Impact, Inc.

This Prediction Impact white paper is sponsored by:

Pitney Bowes Software
To drive business decisions for maximal impact, analytical models must predict the marketing influence of each decision on customer buying behavior. Uplift modeling provides the means to do this, improving upon conventional response and churn models that introduce significant risk by optimizing for the wrong thing. This shift is fundamental to empirically driven decision making. This convention-altering white paper reveals the why and how, and delivers case study results that multiply the ROI of predictive analytics by factors up to 11.

**Standard Predictive Analytics Optimizes the Wrong Thing**

Predicting customer behavior is not enough, and conventional predictive analytics suffers because of this. It predictively scores each customer for expected buying, defection or other behavior, but these scores cannot drive marketing decisions optimally because they do not predict the *impact* of each decision.

Instead, a business decision is optimized when based on the predicted *marketing influence* it will have on the customer’s future behavior. For many customers, the choice of which offer, creative or price – or even whether to make contact at all – will make the difference between sale and no sale. Enter uplift modeling.

**Uplift modeling:** Analytically modeling to *predict the influence on a customer’s buying behavior* that results from choosing one marketing *treatment* (customer-facing action) over another. The secondary treatment is often passive – make no contact – as evaluated over a control group. The uplift model answers the question, “How much more likely is this treatment to generate the desired outcome than the alternative treatment?” For each customer, the model’s prediction drives the decision of which treatment to apply.

Uplift modeling is also known as *differential response, impact, incremental impact, incremental lift, net lift, net response, true-lift, or true response modeling.*
Conventional Predictive Analytics: The Good and the Bad

Predictive analytics optimizes enterprise decisions by driving them with a predictive model, which, in turn, has been optimized across relevant data.

A standard predictive model scores each customer or other organizational element according to the likelihood of behavior such as response or defection. These predictive scores drive operational decisions, typically on a per-customer basis. In this way, predictive analytics often delivers a hefty bottom line impact.

Conventional analytical models that predict customer behavior – although intrinsically limited – can deliver significant business value. For example, model scores that predict how likely a customer is to buy after receiving direct mail drive the decision of whether to contact, spiking campaign ROI by saving on the cost of contacting those customers unlikely to respond. For a rigorous introduction and an overview of value-driven business applications, see “Seven Reasons You Need Predictive Analytics Today” (http://www.pawcon.com/sevenreasons). For articles, case studies and other resources, see the Predictive Analytics Guide (http://www.pawcon.com/guide).

However, conventional predictive analytics misses a key opportunity and introduces a significant danger: in some cases, precisely the wrong customers are targeted, causing more damage than good. The following two sections uncover how these weaknesses imperil the most pervasive business applications of predictive analytics, i) targeting direct marketing and ii) targeting retention efforts.

i) The weakness of response modeling for direct marketing

“You didn't have to be so nice; I would have liked you anyway.”
– The Lovin' Spoonful (a rock band), 1965

Standard response modeling is designed to maximize the wrong thing: response rate. This is not an appropriate measure of a direct marketing campaign's success since it does not match the objectives of the business. Instead, incremental impact – the additional revenue resulting from the campaign that would not have come without it – is central to evaluating the campaign's true ROI.
Response Modeling

Which **green** respondents would have purchased anyway?

Even if a campaign targeted with response modeling is showing a high rate of response, there remains a fundamental, unanswered question: *What about those respondents who would have made a purchase anyway?* In some cases, up to half of them – or even more – are so prone to purchasing, they will do so via another channel even if not contacted. This is where many experts, including even the most senior analytics practitioners, exclaim, “Doh!”

Response rate (top) does not predict incremental response (bottom). Graph courtesy of Kane et al (2011).1
A higher response rate does not correspond with marketing success, as illustrated in the figure above. Response model scores can place customers into 10 “deciles.” As shown in the top bar graph, the higher deciles (on the left) have a higher response rate than the lower deciles. However, the same is true for those in a control group of customers who were not contacted. As this graph shows, the customers more prone to respond were more prone to buy even if not contacted. The bottom bar graph shows the incremental improvement gained by marketing contact, which is calculated by subtracting the lighter bars from the darker bars in the top bar graph. This difference does show that marketing contact results in more purchases, but the amount of increase gained in no way corresponds with the response model’s ranking. In fact, the deciles ranked lowest (on the right) are where there is the most to be gained from marketing contact.

Response modeling is falsely named. “Response models” don’t model the response to contact; they model the propensity to buy, overall. (Radcliffe & Surry 2011) Customers who buy might be responding to marketing contact, but they might have been prone to buy in any case; there is not necessarily the causal relationship implied by the word “response”.

**Missed opportunity.** Marketing campaigns adjusted to avoid purchasers who will buy “no matter what” stand to gain significantly by averting the waste of contacting these “sure things”.

**Danger.** Moreover, customers predicted likely to respond are, in some conditions, exactly the wrong ones to contact. They should be suppressed from the contact list, since, although they are likely to buy if contacted, that is only the case because they are likely to buy either way. While the campaign’s “response” rate may be high, the positive impact of the campaign is low, and potentially not worth its cost (no ROI achieved). Even worse, some customers already dead-set on buying are more prone to an adverse response – they are bothered or otherwise turned away by the unnecessary contact and, possibly feeling it is “pushy,” decide to not buy. A campaign that pushes these customers away has done more damage than good. Each campaign targeted with a response model faces the clear and present danger that most dollars spent are wasted, as in the US Bank example described by Radcliffe & Surry (2011), or even harmful.

**Solution.** Predict which customers will buy only if contacted.
ii) The weakness of churn modeling for targeted retention

A similar story applies to what's emerged, arguably, as the hottest business application of predictive analytics: Targeting retention activities with a churn model that predicts which customers are most likely to defect (if not contacted). Although targeted to customers most likely to leave, the impact of the retention offer itself is traditionally not modeled and predicted.

“Let sleeping dogs lie.”

Contact that attempts to retain the customer may be ineffective or – much worse – may trigger an adverse response from “sleeping dog” customers who would have stayed if left alone, but now get up and leave.

Many businesses face adverse responses to retention outreach. Cell phone carriers often offer a free phone to consumers when their one- or two-year commitment is up in exchange for signing another contract. For some subscribers, this serves only to remind them they’re now free to defect to a competitor. Similarly, other subscription services such as online dating sites, video rental services and health clubs have certain customers who don’t use the service at all, but are charged periodically despite this. A retention offer may remind them to finally get around to canceling. Any and all types of business may have customers who respond adversely to what they consider to be bothersome or unnecessary contact.

Missed opportunity. Retention campaigns adjusted to avoid “sleeping dog” customers stand to gain significantly by averting the customer attrition that will otherwise result. Campaigns also gain by saving the cost of contacting “lost cause” customers that are destined to leave whether contacted or not.

Danger. Moreover, customers predicted likely to cancel are, in some conditions, exactly those that should be suppressed from a retention campaign’s contact list – they are positioned at the ready, prone to depart, and are therefore the easiest sleeping dogs to “wake up”.

Solution. Predict which customers will be saved only if contacted.
Analytical Decisions: Selecting the Optimal Customer Treatment

“Look, when you face a decision, you don’t just consider the result of one choice – you compare the outcomes of both options at hand.”

– The author, and pretty much everyone, including Plato and Benjamin Franklin

It's an often-overlooked truth. Response and churn modeling only test, model and predict for one of the two treatments (customer-facing actions) between which the enterprise is deciding: To contact or not to contact. In the case of response modeling, the active choice, to contact, has been tested, and the test results are employed as training data from which to generate a model that predicts whether a customer will respond if contacted. In the case of churn modeling, existing historical data tracking customers that have not been contacted with retention outreach is employed to generate a model that predicts whether a customer will leave if not contacted.

But the decision between two treatments is only fully informed by predicting the outcome of them both. So, rather than predicting:

(i) Will the customer buy if contacted?

Instead predict:

(ii) Will the customer buy only if contacted?

This distinction speaks to the very heart of empirically driven decision making. As Radcliffe (2007) points out, adding the single word “only” to the stated prediction goal makes all the difference. The predictive score for a customer is now actionable since it directly informs the question, “Should we contact this customer?”

Predicting marketing influence. Although question (ii) may appear to be simple at first look, it achieves its actionability only by answering the composite of two questions, i.e., “Will the customer buy if contacted and not buy otherwise?” This two-in-one query reveals the difference in outcome that will result from one treatment over another, so it also provides the answer to, “Will contacting the customer influence her to buy?” Building a model to answer such a composite question requires the collection of multiple data sets, one from the testing of each treatment, as discussed below in Section, “Data-Driven Decision-Making Demands Additional Data Sets.”
Response Uplift Modeling

“Uplift modeling empowers your organization to capture more than 100% of responses by contacting less than 100% of the target population.”

Kathleen Kane
Principle Decision Scientist
Fidelity Investments

Uplift modeling predicts the buying behavior of customers treated both actively and passively, as shown in the figure below. This quad first distinguishes from top to bottom which customers will respond to an offer, which is the job of conventional response modeling. But then it further distinguishes along a second dimension: Which customers will make a purchase even if not contacted?

![Conceptual response segments](image)

Conceptual response segments. The lower-right segment is targeted with uplift modeling. Table derived from Vittal (2008) and Radcliffe (2007).

The lower-right quadrant includes those customers to target, those worthy of expending the cost of contact. These are the “persuadables,” who won't buy if not contacted, but will buy if they are contacted. They are the customers for whom an uplift model aims to answer with an affirmative prediction, indicating that contacting is a better treatment than not contacting.

Kim Larsen, VP Analytical Insights at MarketShare, says of the persuadables, “These ‘swing clients’ are akin to the swing states of a presidential election; data miners could learn a lot from presidential campaigns.” As residents of California, neither Larsen nor I saw many TV ads for the 2008 presidential campaigns.

Uplift modeling delivers a clear opportunity to reduce costs beyond traditional response modeling: Suppress from the contact list those customers in the lower-left quadrant, the “sure things,” who do indeed respond to direct marketing contact, but actually would buy anyway (perhaps a bit later, or via a different channel) even if the cost of contacting them is not expended.
CASE STUDY: US Bank

**Business case:** Direct mail for a home-equity line of credit to existing customers

**Approach:** Target campaign with an uplift model

**Resulting improvements over prior conventional analytical approach:**

- Campaign ROI increased over 5 times previous campaigns (75% to 400%)
- Cut campaign costs by 40%
- Increase incremental cross-sell revenue by over 300%
- Decrease mailings to customers who would purchase whether contacted or not, and customers who would purchase only if not contacted.

**Sources:** Radcliffe & Surry (2011), Tsai (2010), Patrick Surry (Pitney Bowes Software), Michael Grundhoefer (US Bank)

CASE STUDY: Leading financial institution

**Business case:** Direct mail campaign for financial product

**Approach:** Target campaign with an uplift model

**Resulting improvements over prior conventional analytical approach:**

- Increased revenue per contact by a factor of 20
- Incremental conversion up 0.02% to 0.43%

**Source:** Kim Larsen, *Uplift Workshop at Predictive Analytics World* (Larsen 2011)

### Churn Uplift Modeling

It turns out that employing uplift modeling so that a secondary “predictive dimension” can be incorporated into churn modeling also delivers tremendous performance improvements. With churn uplift modeling, each customer is scored according to these four conceptual segments:

<table>
<thead>
<tr>
<th>Sleeping Dogs</th>
<th>Lost Causes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sure Things</th>
<th>Persuadables</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Conceptual churn segments. The lower-right segment is targeted with uplift modeling. Table derived from Vittal (2008) and Radcliffe (2007).
This quad first distinguishes from left to right which customers will defect if not contacted with the retention offer, which is the job of conventional churn modeling. But then it further distinguishes along a second dimension: Which customers will leave if they do receive the retention offer?

As was the case for response modeling, the lower-right quadrant includes those customers to target, those worthy – in this case – of expending the cost not only of contact, but, more significantly, worthy of expending the often much greater cost incurred by a retention offer such as a discount. These are the “persuadables” – or “savables” – who will defect if not contacted, but who can be convinced to stay with the retention offer.

We turn to the upper-left quadrant for an opportunity to reduce losses incurred by traditional retention with churn modeling. These “sleeping dogs” will stick around if left alone, but, if we contact them with the retention offer, it will trigger an adverse, reverse effect: the customer will leave. Selecting the passive treatment where contact would “scare away” customers is certainly a great way to improve the retention campaign’s success.

**CASE STUDY: Telenor, the world’s 7th largest mobile operator**

**Business case:** Retention campaign for cell phone subscribers

**Approach:** Target campaign with an uplift model

**Resulting improvements over prior conventional analytical retention models:**
- Campaign ROI increased by a factor of 11
- Reduced churn a further 36%
- Campaign costs reduced by 40%

**Sources:** Vittal (2008), Patrick Surry (Pitney Bowes Software)

Impact delivered by uplift modeling for Telenor.

Figure courtesy of Pitney Bowes Software.
CASE STUDY: Lloyds TSB

Business case: Retention campaign
Approach: Target campaign with an uplift model
Results:
- £8,000,000 incremental annual profit
- Increased retention rate by 4.09% pt
Source: Nicholas Radcliffe

Business Applications of Uplift Modeling

“The optimal decisions delivered by uplift modeling are key to making analytics actionable – an enterprise must decide before it can act.”

– An AVP and Marketing Statistical Analyst at a top North American bank

Beyond the active-versus-passive choice of whether or not to contact a customer with a particular treatment, uplift modeling boosts the effectiveness of many other marketing decisions. This analytical technique applies broadly, driving the selection – for each customer – of content, ad, offer, channel, and price that is most likely to succeed. In this way, the marketing treatment that is most relevant, personalized or appropriate is decided upon for each customer.

Uplift modeling maximizes impact for any treatment decision where the objective is to exert an influence. With this wide applicability comes a growing range of proven value propositions within marketing operations and beyond, such as those in the table below. While most of these applications may drive decisions between more than two treatments, that process is beyond the scope of this white paper, which focuses on selecting between two treatment options only.

Uplift modeling applies broadly, driving the selection of content, ad, offer, channel, and price that is most likely to succeed.
### Uplift application | Decision between treatments | Business objective | Deployments
--- | --- | --- | ---
Response uplift | Should we contact the customer or not (active or passive treatment)? | Positive impact of direct marketing campaigns | US Bank, a top financial firm, another top financial firm
Churn uplift | Should we provide the customer a retention offer or not (active or passive treatment)? | Positive impact of retention campaigns | Telenor, Lloyds TSB
Content targeting | With which ad, creative, copy, or product should we solicit the customer? | Response rate of direct marketing, cross-sell and online & offline ads | Online ads, Email campaign
Channel selection | Through which channel should we contact the customer, e.g., mail, email, or telephone? | Positive impact of direct marketing campaigns | (No known public case study)
Dynamic pricing and discounting | Should we offer the customer a higher price or a lower price? | Revenue of sales | (No known public case study)
Collections | Should we offer the debtor a deeper write-off or not? | Revenue of accounts receivable | (No known public case study)
Credit risk | Should we offer the customer a higher or lower credit limit? A higher or lower APR? | Revenue from interest payments and savings from fewer defaults | (No known public case study)
Electoral politics | Should we contact the constituent/state (is it a “swing voter”/“swing state”)? | Positive votes resulting from political election campaigns | US presidential campaign targeting of “swing states”
Personalized medicine | Which medical treatment should the patient receive? | Patient outcome in clinical healthcare | Treating heart failure with beta-blockers

**Content targeting.** Uplift modeling selects for each customer the ad, offer, content, or product most likely to elicit a response. This is a great improvement over standard AB testing, which uses random trials to determine which marketing content works best overall, across customers. AB testing in effect “throws out” the treatment with lower average performance, and by doing so misses the opportunity to find pockets of customers that are actually more likely to respond to it than the alternative treatment. Per-customer content targeting replaces AB testing with personalized “AB targeting,” earning a higher response rate than either treatment alone.

**CASE STUDY: A popular online education portal**
- **Business case:** Online ad targeting
- **Approach:** Select the best ad for each customer
- **Resulting improvement over prior conventional analytical approach:**
  - 25% increase in response rate
  - Ad revenue rate increased by $1 million every 14 months
- **Source:** Prediction Impact case study (more info: blog interview)
Dynamic pricing and collections. As for any decision, a certain risk is faced for each candidate treatment when pricing: The higher price may turn a customer away, but the lower price (or deeper discount) unnecessarily sacrifices revenue if the customer would have been willing to pay more. There are presently no known public case studies in uplift modeling applied for price selection, although commercial deployment is in the works.

Credit risk. The balance between risk and potential profitability for each debtor is influenced by both the credit limit and APR offered. Raising one or both may result in higher revenue in the form of interest payments, but may also increase the chance of the debtor defaulting on payments and an ensuing write-off. Lowering APR, e.g., a time-limited zero APR offer, can also increase interest-bearing balance, which in turn may also increase risk. Since these factors vary by individual debtor, there is an advantage to targeting offers accordingly. There are presently no known public case studies in uplift modeling applied for credit risk management, although at least one large creditor has put this to action.

Electoral politics. Just as “swing clients” are potentially persuaded by marketing contact, the same benefit it gained where this term originates: political campaigns that target swing voters. The constituents with the most potential to be influenced by campaign contact are worth the cost of contact. Analogously, the swing states most likely to be persuaded as a whole are targeted.

Personalized medicine. Naturally, healthcare is where the term treatment originates. While one medical treatment may deliver better results on average than another, personalized medicine aims to decide which treatment is best suited for each patient, since a treatment that helps one patient could hurt another. For example, to drive beta-blocker treatment decisions for heart failure, researchers “use two independent data sets to construct a systematic, subject-specific treatment selection procedure.” (Claggett et al 2011) Certain HIV treatment is shown more effective for younger children. (McKinney et al 1998) Cancer treatments are targeted by the patient’s genes. (Winslow 2011)

Emerging applications. Uplift’s range of business applications is growing as new types of marketing and business decisions are driven with this technology. For example, it has been demonstrated that following up on certain website clicks with an automatic email message has potential to increase sales but also to “turn off” customers and decrease sales. Uplift modeling discerns for which customers this treatment provides a boost. (Larsen 2011)

Uplift modeling provides a benefit only for business decisions that are intended to exert an influence on a customer. Therefore, it does not apply for application processing and other “screening” decisions that deny requests in order to avert risk. Uplift modeling generally cannot serve to improve the conventional deployment of predictive analytics for credit screening, algorithmic trading, insurance application processing and fraud detection.
Data-Driven Decision-Making Demands Additional Data Sets

Uplift models must be trained across two or more data sets, one for each customer treatment. These data sets track the results of testing each treatment. For example, for response uplift modeling, one data set must be collected for a set of customers who receive the direct marketing contact, and another for those who do not receive it, i.e., the control set. For ad targeting, a data set must be collected for each candidate ad.

These training sets each have the same general form as a training set for conventional predictive modeling. They are “flat” two-dimensional tables with one row per customer or transaction. Each row includes a set of independent variables (predictor variables) available as input to the model, such as customer profile and behavioral values, which the model is provided in order to compute the predictive score it outputs. And each row has a single dependent variable, related to buying behavior or other business outcome, which defines the prediction goal. The model is optimized to predict according to this target variable.

The introduction of a second training set is a “necessary evil” that must be overcome in order to gain the benefits of uplift modeling. Since an uplift model answers the composite question, “Will treatment A lead to a better outcome than treatment B (often the passive action),” it must be built from data with the results of both treatments. And since, in many cases, a customer cannot be treated in both ways – e.g., you cannot both contact and not contact a person – the two data sets must include unique, non-overlapping samples of customers.

For some projects, additional efforts are required in order to collect a second data set. Response uplift modeling requires the introduction of a control set of customers who are suppressed from receiving direct marketing contact. This new data set provides insight as to which type of customer will buy even if not contacted. The control set must be a uniformly-selected, random sample, and is an important thing to put in place anyway, even if there will be no uplift modeling, in order to measure the incremental benefit of the marketing campaign.
Churn uplift modeling requires the introduction of a uniformly-selected, untargeted test set of customers who are contacted with a retention offer. Since this set is not targeted, it includes customers who are not at risk of attrition, so a cost is incurred by contacting these customers, who may consume a retention discount even if they were not going to leave to start with. However, this cost can be mitigated by keeping the set proportionately small and by eliminating those with a very low expected chance of defection (applying uplift modeling only to the rest).

In some cases, content targeting with uplift modeling requires no additional effort to collect data. For example, if an AB test has been deployed, its results can be employed as the two training data sets required for uplift modeling.

Driving Treatment Decisions with Two Response Models

“Weigh your options.”

Specialized analytical methods are required to create uplift models, since conventional predictive modeling methods are designed to operate across a single training data set that consists of customers all treated the same way, rather than two different ways.

An effective but limited approach is to generate two predictive models, one for each treatment, and use the models together to drive each customer decision:

A customer is scored by two predictive models to decide upon which treatment to apply.

As shown, predictive modeling is applied over the data set for treatment A, resulting in a predictive model that can score an individual customer such as Mary as to the likelihood she will respond positively to treatment A. Likewise the treatment B data set is used to create another model that predicts the outcome of treatment B (which is often the passive action, tested on a control set). For a customer such as Mary, we wind up with two predictive scores, the higher of which indicates the treatment more likely to result in a positive outcome.
As described in the case study above, this approach was employed by an education portal for online ad selection, resulting in an ad revenue rate increase of $1 million every 14 months (this project deployed 291 models, one per candidate ad, rather than just 2 models).

Under the Hood: The Analytics behind Uplift Modeling

“Think before you act.”

Despite this success, the “two model” approach described above has a limitation that compromises its robustness and increases the danger of project failure: neither model is optimized for the objective at hand.

Since each model aims to predict only one treatment's outcome, some or most of what each model captures does not inform the choice of treatment. Instead, in many cases, the models will identify customers likely to respond, regardless of treatment. This is common, since oftentimes there are customers “dead set” on buying — or on not buying — regardless of marketing contact. For example, in many projects, customers that have opted in for email solicitations turn out to be “more engaged” and therefore more likely to click or purchase, given either treatment. Capturing this kind of effect is appropriate for a model that is meant to predict one treatment's outcome, but can serve as a “red herring” that compromises success at the analytical task at hand: uplift modeling.

Instead, the pertinent objective is to identify customers for whom the choice of treatment makes the most difference, i.e., for whom the likelihood of a positive outcome increases the most by selecting one treatment over another. The variables and effects that best characterize (model) these customers may be different than those emphasized by two individual response models; the “two model” approach is not designed to analytically hone in directly on what matters the most.

Two training sets are used together to develop an uplift model.
As shown above, best practice analytical methods for uplift modeling operate over both training sets and generate a single uplift model. This model outputs a predictive score for a customer, Mary, which indicates which treatment is better for her specifically. The score answers the question, “Will treatment A result in a better outcome than treatment B for Mary?” This is equivalent to asking, “Would applying treatment A rather than B influence Mary to respond positively?”

Uplift modeling faces an entirely new kind of analytical challenge over standard propensity models generated from a single training data set. Although the two training data sets at hand for treatments A and B are similar in overall form, the dependent variable (prediction target variable) columns of these two tables have distinct meanings: one for positive outcome after treatment A, and the other for treatment B. We have not tested both A and B on any one customer, nor can we observe and record the influence of a treatment on any one customer, since this literally involves the inner workings of the customer’s central nervous system. Therefore there are simply no example customers in the training data for whom influence has been observed.

How Models Identify Which Customers Are Influenceable

If influence cannot be observed for any single customer, how can we possibly analyze it and learn to predict it? The answer is, by identifying a segment of customers—a group of customers that share characteristics, which can be thought of as a “type of customer”—across which the likelihood of being influenced can then be measured.

A segment is defined by one or more customer variables, and this group of customers can then be evaluated for uplift (i.e., the successful influence of marketing contact). For a simple example, consider the segment of customers who have purchased their last car between 4 and 6 years ago as a target for marketing a new car. Lo (2002) conjures as a hypothetical example that customers who have purchased a car recently may not be responsive to marketing outreach since they’re not ready for a new car, and that those who’ve purchased too long ago may be responsive, but only because they had already decided to buy a new car soon—they would have bought one anyway, so contacting them has no incremental effect, no influence.

The likelihood of a segment’s customers being influenced by marketing can be calculated with simple arithmetic. If a higher proportion of customers in the target segment buy a car after marketing contact (counted within the data set for treatment A) than do members of the same segment after no marketing contact (in the data set for treatment B), a positive uplift has been observed, so the value of targeting the segment for contact has been established. The difference between the two buying rates measured on the two data sets is the uplift.

Example Uplift Segments

Hot segment in the mid-range. An “upside-down U” pattern is common for uplift. The figure below depicts a fictional but typical example based on real direct marketing case studies by an
uplift modeling expert practitioner. (Larsen 2011) The horizontal axis is the customer’s number of open revolving accounts, and the vertical axis shows a measure of uplift (resulting from marketing contact), the net weight of evidence, i.e., the evidence of a positive influential effect that led to an incremental gain in comparison with the control set.

![Net weight of evidence (a measure of uplift) varies by a customer's number of open revolving accounts. Graph courtesy of Larsen (2011).](image)

As with the car buying example, there is a mid-range worthy of targeting, which results in an “upside-down U” kind of shape. Customers on the left, with few accounts, are unlikely to be influenced to buy, but those with 3 to 6 accounts are positively influenced to buy from marketing contact. On the other hand, customers on the right, with more accounts, are important to suppress from marketing contact, since it actually decreases the chance of buying as the curve dips down into negative numbers – a veritable downlift. The explanation may be that customers with few accounts are less likely to buy whether contacted or not, yet customers with many accounts are already so engaged they may be more sensitive to, aware of and annoyed by what they consider to be unnecessary marketing contact. An alternative explanation is that customers who have already accumulated so many credit accounts are susceptible to impulse buys, e.g., when they come into a bank branch, but when contacted at home will be prone to respond by considering the decision and possibly researching competing products online. They would have been more likely to make the purchase if left to their own devices.

Customer age. Lo (2002) suggests (by way of a fictional example) that direct mail for certain financial products may have a stronger uplift for customers higher in age. One explanation could be that older clients respond to more traditional marketing contact such as physical mail pieces, or to the act of solicitation in general. In the healthcare industry, certain HIV treatment has been shown more effective for younger children. (McKinney et al 1998)
Sleeping dogs. In targeting retention campaigns, while a customer's decrease in product usage could indicate she is more likely to cancel, if she has had a low usage rate for a while, yet has been paying a non-discounted retail price, she could be a sleeping dog, prone to stay if left alone but adversely responsive if contacted. This includes the health club member who never gets to the gym, and the online subscriber who never logs on to the product.

Uplift Modeling with Segmentation: Uplift Trees

Beyond the simple segments described above, the real win comes when more precise segments that show higher uplift are identified – usually defined with more than just one variable. Uplift trees, a variation of decision trees, do the trick. (Radcliffe & Surry 2011) Decision trees, originally developed for conventional propensity modeling, derive multivariate segments. The analytical objective is to discover customer segments as different from one another as possible with respect to the chance of positive outcome (customer buying behavior).

Uplift trees model for uplift by discovering segments that discern influenceable customers from non-influenceable. The modeling process has succeeded when some are “hot” targetable segments, showing a higher uplift than average, and others are “cold” non-targets, showing a lower uplift or even a downlift (negative effect). This method learns from both training data sets together at once, calculating the uplift of candidate segments with the two data sets, and guiding the discovery of valuable segments accordingly.

Example uplift tree segments. US Bank applied uplift trees to target a direct mail campaign selling a home-equity line of credit to existing customers. (Radcliffe & Surry 2011) (Tsai 2010)

Without a means to select which prospects are worth contacting, this direct mail campaign fails. Although it shows a “response rate” of 1.32%, the purchase rate amongst those not contacted (the control group) came in as equivalent or a margin higher, 1.36%. This indicates a slight downlift.

However, the tree included the following segment with an uplift of 0.76 percentage points. Customers in this segment bought at a rate of 1.83% if contacted, and at 1.07% if not contacted.

CUSTOMER SEGMENT: (Simplified for this illustration)
- Has paid back more than 17.3% of current loan
  AND
- Is using more than 9.0% of revolving credit limit
  AND
- Is designated within a certain set of lifestyle segments

This stands in stark contrast with another single-variable segment from the same uplift tree model with an uplift of -0.97 (i.e., a “do not contact” downlift segment):
CUSTOMER SEGMENT: *(Simplified for this illustration)*  
Has paid back less than 17.3% of current loan

Ironically, while these customers have a much higher buy rate overall, they are exactly the ones *not* to contact, since contacting them influences them to buy *less*. Those customers not contacted buy at a rate of 3.35%, but if contacted this dips to 2.38%.

As with conventional decision trees and predictive modeling in general, the analytical task for which uplift trees are designed is to optimize which variables to use together in what manner, the order in which they are considered and the threshold values applied. This generality enables an assortment of analytical effects to be captured.

Beyond uplift trees, other analytical approaches to uplift modeling include specialized k-nearest neighbor classifiers, a specialized version of Naïve Bayes, and other uplift methods. (Lo 2002) (Larsen 2011)

**A Software Solution for Uplift**

Although uplift modeling has demonstrated proven value for over 10 years, at the time of going to press, only one packaged commercial software solution is available: Pitney Bowes Software’s Portrait Uplift ([http://www.pbinsight.com/products/portrait-customer-interaction-suite/portrait-uplift/](http://www.pbinsight.com/products/portrait-customer-interaction-suite/portrait-uplift/)).

Uplift modeling entails analytical challenges and requirements that go beyond conventional response and churn modeling. This software product specifically addresses these challenges, e.g., by integrating with its uplift capabilities ensemble models and tailored approaches to variable selection. Its analytical approach for uplift trees is described by Radcliffe & Surry (2011).

**Conclusions**

Uplift modeling guides marketing by predicting its *influence* on the customer. It empirically determines the marketing treatment most likely to generate desired buying behavior, driving marketing decisions for maximal impact. This improves over conventional predictive analytics deployment that introduces risk by putting each business decision to action without weighing the alternative. As the case studies show, this technique delivers business value by decreasing wasted marketing costs and by avoiding the adverse response and loss of revenue that can result from customer contact that's not optimally targeted.
Upcoming Conference Sessions and Workshops on Uplift Modeling

- **Persuasion by the Numbers: Optimize Customer Influence by Predicting It**, Keynote presentation at Predictive Analytics World by Eric Siegel, Ph.D., Conference Chair
- **True-Lift Modeling: Mining for the Most Truly Responsive Customers & Prospects**, presentation at Predictive Analytics World by Jane Zheng, Fidelity Investments
- **Uplift Modelling: You Should Not Only Measure But Model Incremental Response**, presentation at Predictive Analytics World by Nicholas Radcliffe
- **Net Lift Models: Optimizing the Impact of Your Marketing**, workshop at Predictive Analytics World instructed by Kim Larsen

Where to Learn More about Predictive Analytics

- **Training seminar**: [Predictive Analytics for Business, Marketing and Web](http://www.businessprediction.com), a two-day intensive seminar brought to you by Prediction Impact, Inc. [www.businessprediction.com](http://www.businessprediction.com)
- **Online e-course**: [Predictive Analytics Applied](http://www.predictionimpact.com/predictive-analytics-online-training.html), immediate access, on demand at any time.
- **Conference**: [Predictive Analytics World](http://www.predictiveanalyticsworld.com), the leading business-focused event for predictive analytics professionals, managers and commercial practitioners. Learn from industry leaders, expert practitioners, case studies and workshops. [www.predictiveanalyticsworld.com](http://www.predictiveanalyticsworld.com)
- **Conference**: [Text Analytics World](http://www.textanalyticsworld.com), the business event covering voice-of-the-customer analytics, decision support, sentiment analysis, and more. [www.textanalyticsworld.com](http://www.textanalyticsworld.com)

About the Author

Eric Siegel, Ph.D. is the president of Prediction Impact, Inc. ([http://www.predictionimpact.com](http://www.predictionimpact.com)), and the conference chair of [Predictive Analytics World](http://www.predictiveanalyticsworld.com) and [Text Analytics World](http://www.textanalyticsworld.com). An expert in predictive analytics and data mining, Dr. Siegel is a former computer science professor at Columbia University, where he won the engineering school award for teaching, including graduate-level courses in machine learning – the academic term for predictive modeling. After Columbia, Dr. Siegel co-founded two software companies for customer profiling and data mining, and then started Prediction Impact in 2003, providing predictive analytics services and training to mid-tier through Fortune 100 companies. Dr. Siegel is the instructor of the acclaimed training program, [Predictive Analytics for Business, Marketing and Web](http://www.businessprediction.com), and its online version, [Predictive Analytics Applied](http://www.businessprediction.com). He has published over 20 papers and articles in applied predictive analytics, data mining research and computer science education.
About Pitney Bowes Software

Pitney Bowes Software is a software & services company that enables organizations to have lifetime relationships with their customers.

Our proven solutions enable organizations to engage with each of their customers as individuals and to connect every customer communication – outbound, inbound, marketing, sales or service – into an ongoing dialogue where customer insight and understanding forms the basis for each and every interaction.

This is achieved through a suite of innovative and compelling capabilities that integrate data management, location intelligence, sophisticated predictive analytics; rules based decision making and cross-channel customer interaction management to deliver enhanced customer profitability and operational efficiency.

For more information
For further information, go to http://www.pbinsight.com.

Acknowledgements

Thanks for much helpful input, discussions, examples and PAW conference presentations go to Patrick Surry, Kim Larsen, Nicholas Radcliffe, and Kathleen Kane. Thanks to Adrien Urbani for helping with the production of this white paper and to my wife Maria de Fatima Callou for patience when I disappeared completely for several days. Thanks to my mom Lisa Schamberg and Elise Johnson for detailed reviews.

References


Graph image reproduced with permission, courtesy of Kane et al (2011), as depicted in their Predictive Analytics World presentation.

Research has shown that being a health club member is not sufficient to achieve a positive outcome - you have to actually go to the gym, buddy. But seriously, some are known to nickname such unintentionally prolonged subscriptions as “health club memberships” even at completely unrelated businesses.

Only informing the expected outcome of one treatment is a form of satisficing.

Fraud aversion could in principle benefit from uplift modeling to decide for which customers to convey a “Do not commit fraud” message, as pointed out by Nicolas Radcliffe.

In some cases the need for non-overlapping training sets of customers may be less absolute. For some applications of uplift modeling, such as content targeting, it is in fact possible to treat one customer in more than one way. However, this could introduce bias to the training data collected, since the outcome for a customer tested with a certain treatment could vary depending on whether an alternative treatment had also already been applied.

Specifically, each model may capture main effects, and emphasize the most important variables, of their data set, and, as a result, not as precisely capture relatively minor effects that could be critical to modeling uplift. This and related factors are explored further by Radcliffe & Surry (2011).

Thanks go to Patrick Surry (Pitney Bowes Software) and Michael Grundhoefer (US Bank) for this detailed example.