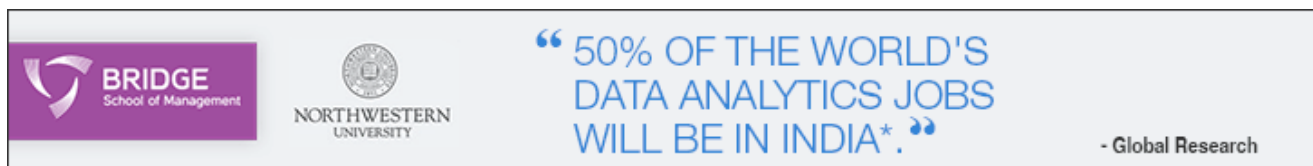


f (<https://www.facebook.com/AnalyticsVidhya>)t (<https://twitter.com/analyticsvidhya>)g+ (<https://plus.google.com/+Analyticsvidhya/posts>)in (<https://www.linkedin.com/groups/Analytics-Vidhya-Learn-everything-about-5057165>)(<http://www.analyticsvidhya.com>)(<http://datahack.analyticsvidhya.com/contest/date-your-data>)Home (<http://www.analyticsvidhya.com/>) Business Analytics (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/>) [REPLY \(HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#RESPOND\)](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/#respond)

# Trick to enhance power of Regression model

BUSINESS ANALYTICS ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/BUSINESS-ANALYTICS/](http://www.analyticsvidhya.com/blog/category/business-analytics/)) SAS([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/SAS/](http://www.analyticsvidhya.com/blog/category/sas/))

SHARE f (<http://www.facebook.com/sharer.php?u=http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/&t=Trick%20to%20enhance%20power%20of%20Regression%20model>) t (<https://twitter.com/home?status=Trick%20to%20enhance%20power%20of%20Regression%20model-http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/>) g+ (<https://plus.google.com/share?url=http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/>) p (<http://pinterest.com/pin/create/button?url=http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/&media=http://www.analyticsvidhya.com/wp-content/uploads/2013/10/decision-tree1.png&description=Trick%20to%20enhance%20power%20of%20Regression%20model>)



([http://admissions.bridgesom.com/pba-new/?utm\\_source=AV&utm\\_medium=Banner&utm\\_campaign=AVBanner](http://admissions.bridgesom.com/pba-new/?utm_source=AV&utm_medium=Banner&utm_campaign=AVBanner))

**We, as analysts, specialize in optimization of already optimized processes.** As the optimization gets finer, opportunity to make the process better gets thinner. One of the predictive modeling technique used frequently use is regression (Linear or Logistic). Another equally competing technique (typically considered as a challenger) is Decision tree.

**What if we could combine the benefits of both the techniques to create powerful predictive models?**

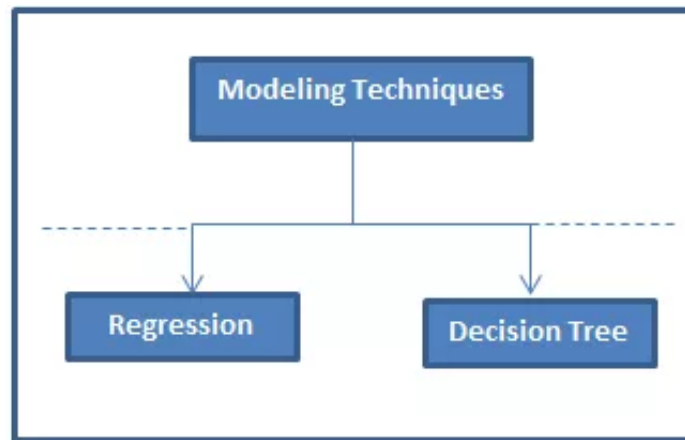


(<http://i2.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/leap.jpg>)

The trick mentioned in this article does exactly that and can help you improve model lifts by as high as 120%.

### **Overview of both the modeling techniques**

Before getting into details of this trick, let me touch up briefly on pros and cons of the two mentioned techniques. If you want to read basics of predictive modeling, click here (<http://www.analyticsvidhya.com/blog/2013/04/predictive-modeling-what-case-study-part-1/>).



(<http://i1.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/modelling-org.png>)

- **Regression** assumes continuous variable as is and generates a prediction through fitting curves for each combination input variables.
- **Decision tree** (CART/CHAID), on the other hand, converts these continuous variables into buckets and thereby segments the overall population.

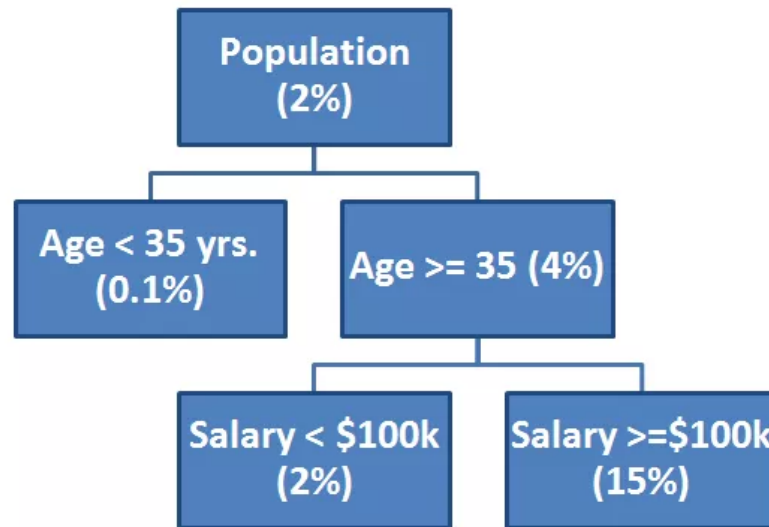
Converting variables into buckets might make the analysis simpler, but it makes the model lose some predictive power because of its indifference for data points lying in the same bucket.

### A simple case study to understand pros and cons of the two techniques

If decision tree loses such an important trait, how come it has a predictive power similar to that of a regression model? It is because it captures the covariance term effectively, which makes a decision tree stronger. Say, we want to find probability of a person to buy a BMW.

#### Decision tree:

A decision tree simply segments the population in as discrete buckets as possible. This is how a typical decision tree would look like:



**\* Numbers in bracket are the response rate of that segment**

(<http://i1.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/decision-tree1.png>)

Even though the tree does not distinguish between Age 37 yrs. and 90 yrs. or salary \$150k and \$1M, the covariance between Age and Salary is making the decision tree's prediction powerful.

### Regression model:

On the other hand, logistic regression makes use of Logit function (shape below) to create prediction. A typical equation would look like this:

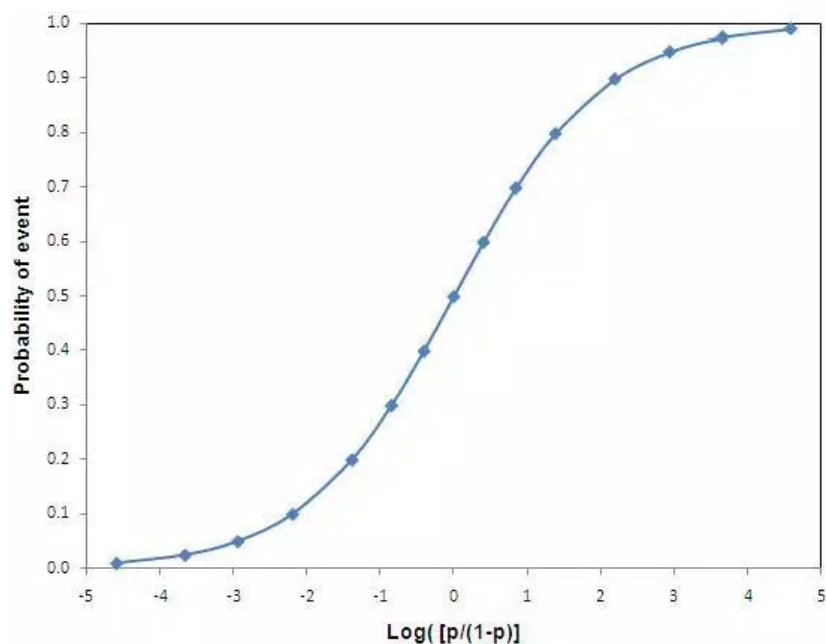
$$\text{Logit } f = a * f_{\text{Salary}} + b * g_{\text{Age}} + c$$

(<http://i2.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/Equation.png>)

where a, b and c are constants

### Shape of Logit function:

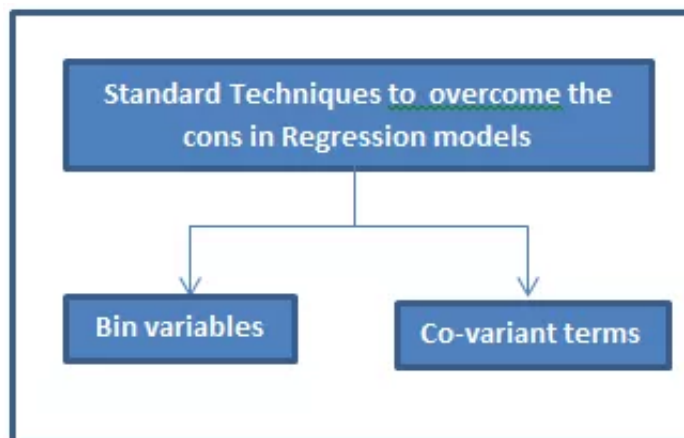
(<http://www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/Formula-1.png>)



([http://i1.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/the\\_logistic\\_function.jpg](http://i1.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/the_logistic_function.jpg))

### Industry standard techniques to address shortcomings of regression modeling:

There are two basic techniques to capture covariance and discontinuity of the target variable:

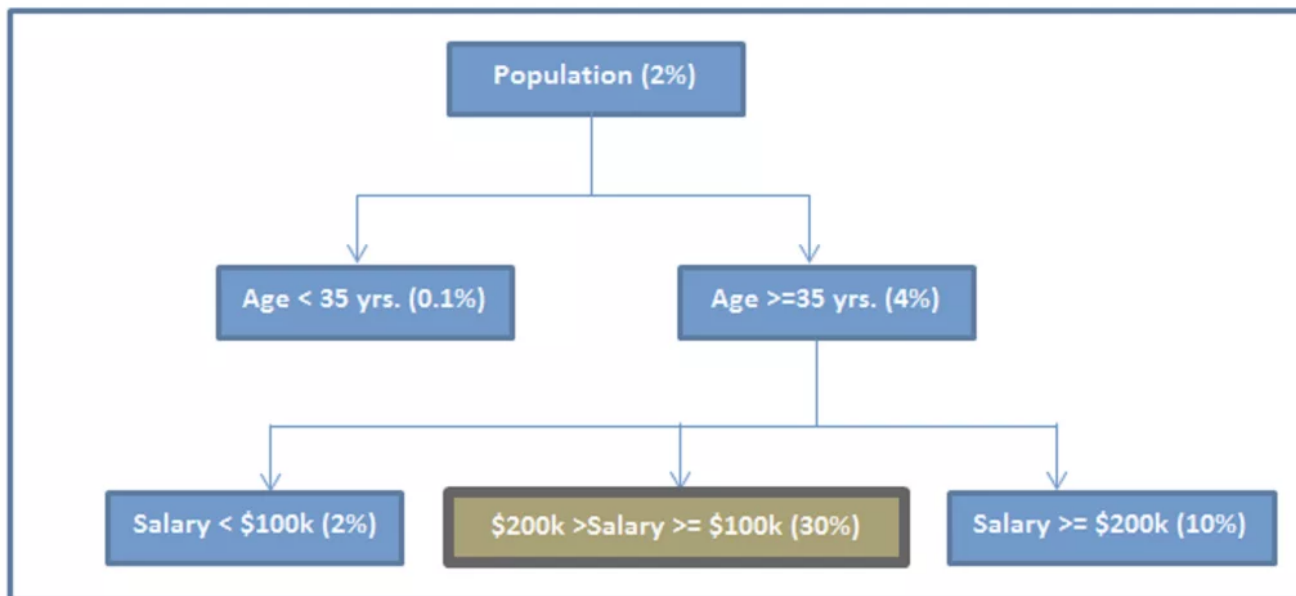


(<http://i2.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/Std-tech-org.png>)

1. **Bin variable** with discontinuous relation: This is a technique used in almost all the models. If you are not familiar with this technique, it is nothing but flagging variables in the interval where a strong input variable shows a discontinuous relationship with the output variable. For example, 10% people break at least one traffic signal in US everyday. Only 3% of households with salary between \$70k and \$100k break traffic signal. Whereas, 11% of the rest of the population break traffic signal everyday, and this is almost uniformly distributed. In this case, to predict the propensity to break the signal, a good variable can be the salary bin \$70k to \$100k.
2. Introduce **covariance variables**: This is a technique used rarely. The reason being such variables are very difficult to comprehend and difficult to explain business.

Each of these techniques capture co-variance and discontinuous variable well.

However, consider the following scenario where these approach have a high propensity to fail:



(<http://i2.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/DT-22.png>)

Say in the population discussed in last section, people with salary between \$100 K and \$200 K and age above 35 years form a segment with exceptionally high BMW take up rate (30%). If we use the two discussed techniques, will the model capture this exceptionally high take-up segment? Will regression still be a better model compared to decision tree?



(<http://io.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/graph1.png>)

The answer is **NO** and **NO**, the regression will not be able to effectively capture this segment. Why does bin technique not capture this? The reason is that binning is done on a one-dimensional variable and in overall population salary bracket \$100k to \$200k might not even be different from rest. As you can see in the figure above, the response rate of the income bucket \$100k to \$200k is not differentiated when analyzed on overall population. But this bucket becomes very different when an initial cut of age  $\geq 35$  yrs is added.

Why does covariance term fail as well? The overall covariance between age and salary might not be significant for the overall population. Hence to have higher predictive power, the model needs an input that the trends of a particular segment are significantly different from rest of the population. Therefore, when the two problems i.e. discontinuity and covariance, exist simultaneously, regression model fails to capture the hidden segment. Decision tree, on the other hand works very well in these scenarios.

**Have you guessed the trick?**

Having worked on many of such problems, I find the following solution very handy. Both regression and decision tree have pros. ***Why not combine the pros of the two methods?*** I have used this technique in a number of models. And I was pleasantly surprised by the additional predictive power I got every time. There are two ways to combine the two methods:

1. **Introduce a new Covariant Variable:** A faster and an effective way to use the tool. Simply add this as one of the input variable for the logistic model. This Covariant term (a bi-variant bin) can be defined as:

$$Z = \begin{cases} 1 & \text{if } 200k > \text{Salary} > 100k \text{ \& Age } > 35 \\ 0 & \text{otherwise} \end{cases}$$

(<http://i1.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/eq3.png>)

2. **Make two alternative models:** A time taking but more effective method in case the exceptional bin has reasonable size. In this method we build two regression models separately for the identified bin (Age > 35yrs. and \$200k > Salary > \$100k ) and the rest of population. And add the two function by following logic.

$$H(x) = (1 - z) * f_x + z * g_x$$

(<http://i2.wp.com/www.analyticsvidhya.com/blog/wp-content/uploads/2013/10/eq2.png>)

Here,  $g(x)$  is the equation for the identified bin and  $f(x)$  is the equation for rest of the population.  $Z$  is same as defined in the last block.

I am sure you are wondering "What makes the lift go even higher than the lifts you saw in the two models?" If you are not I will really like to know your opinions on the reason.

**How does it work? How do this trick create such impactful results?**



Now let's try to think about what did we just do? Decision tree was coming out to be a better model because of a hidden pocket, which was two-dimensional bin. Even with the limitation of not using the continuous behavior of interval variable decision tree became very efficient to reduce false positive in a particular segment. By introducing the flag of this segment in logistic regression we have given the regression the additional dimension decision tree was able to capture. Hence by additionally using the continuous behavior of interval variables such as age, salary the new logistic regression becomes stronger than the decision tree.

### **Constraints of this trick:**

There are two major constraints of using this technique,

1. Multi-collinearity: For models, where the VIF factor becomes unacceptable, the number of variables used to create the new input function should be reduced.
2. High Covariance: When overall covariance between two terms is high, this technique simply fails. This is because we will have to create too many buckets and , therefore, too many variables to be introduced in the regression model. This will introduce very high collinearity to the regression.

In general, I follow a thumb rule of not making more than 6 leaves in the parent tree. This, first of all, captures the most important co-variant buckets and does not introduce the two mentioned problems. Also, make sure the final bucket makes sense with business and is not merely a noise.

### **Final notes:**

This trick helped me make my lifts rise by as high as 120% of the original lift. The best part of this trick is that it gives you a good starting point for your regression where you start with a couple of already proved significant variables.

What do think of this technique? Do you think this provides solution to any problem you face? Are there any other techniques you use to improve performance of your models (prediction or stability)? Do let us know your thoughts in comments below.

**If you like what you just read & want to continue your analytics learning, subscribe to our emails (<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>) or like our facebook page (<http://facebook.com/analyticsvidhya>).**

Share this:


 (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?share=linkedin&nb=1>) 107

 (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?share=facebook&nb=1>) 39

 (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?share=google-plus-1&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?share=twitter&nb=1>)

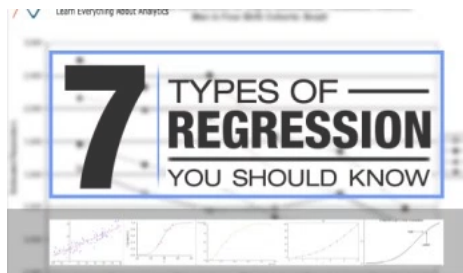
 (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?share=pocket&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?share=reddit&nb=1>)

## RELATED



(<http://www.analyticsvidhya.com/blog/2015/06/hackathon-practice-guide-analytics-vidhya/>)  
**The Hackathon Practice Guide by Analytics Vidhya**  
 (<http://www.analyticsvidhya.com/blog/2015/06/hackathon-practice-guide-analytics-vidhya/>)



(<http://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)  
**7 Types of Regression Techniques you should know!**  
 (<http://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)



(<http://www.analyticsvidhya.com/blog/2014/12/caret-package-stop-solution-building-predictive-models/>)  
**The "caret" Package - One stop solution for building predictive models in R**

**practice-guide-analytics-vidhya/)**  
In "Analytics Vidhya"

**guide-regression/)**  
In "Business Analytics"

**([http://www.analyticsvidhya.com  
/blog/2014/12/caret-package-  
stop-solution-building-predictive-  
models/](http://www.analyticsvidhya.com/blog/2014/12/caret-package-stop-solution-building-predictive-models/))**  
In "Business Analytics"

TAGS: BUSINESS ANALYTICS ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BUSINESS-ANALYTICS/](http://www.analyticsvidhya.com/blog/tag/business-analytics/)), DECISION TREE  
([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DECISION-TREE/](http://www.analyticsvidhya.com/blog/tag/decision-tree/)), PREDICTIVE MODELING  
([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/PREDICTIVE-MODELING/](http://www.analyticsvidhya.com/blog/tag/predictive-modeling/)), REGRESSION  
([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/REGRESSION/](http://www.analyticsvidhya.com/blog/tag/regression/))

#### Previous Article

**Must read books (and blogs) on Web  
Analytics**

([http://www.analyticsvidhya.com/blog/2013/10/  
read-books-web-analytics/](http://www.analyticsvidhya.com/blog/2013/10/read-books-web-analytics/))

#### Next Article

**Analytics training recommendations  
from last 2 months**

([http://www.analyticsvidhya.com/blog/2013/10/  
analytics-training-recommendations/](http://www.analyticsvidhya.com/blog/2013/10/analytics-training-recommendations/))



(<http://www.analyticsvidhya.com/blog/author/tavish1/>)

Author

**Tavish Srivastava**

**(<http://www.analyticsvidhya.com/blog/author/tavish1/>)**

I am Tavish Srivastava, a post graduate from IIT Madras in Mechanical Engineering. I have more than two years of work experience in Analytics. My experience ranges from hands on analytics in a developing country like India to convince banking partners with analytical solution in matured market like US. For last two and a half years I have contributed to various sales strategies, marketing strategies and Recruitment strategies in both Insurance and Banking industry.

## 14 COMMENTS

---

**Samarendra Behera says:**

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=534#RESPOND)  
OCTOBER 23, 2013 AT 4:28 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-534)

Hi Kunal,

If I follow the equation given in step 2 to compute  $H(x)$  then is it not the same as combining  $f(x)$  and  $g(x)$ ? Because,  $z$  value for  $g(x)$  will always be 1 and same for  $f(x)$  will always be 0 meaning  $1-z$  is again always 1. Is my understanding correct?

Best Regards,  
Samarendra

**Kunal Jain (<http://analyticsvidhya.com>) says:**

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=543#RESPOND)  
OCTOBER 23, 2013 AT 5:57 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-543)

Samarendra,

Let me try and explain. When  $\text{age} > 35$  for specific income,  $z = 1$ . In this case  $H(x) = g(x)$

When becomes 0,  $H(x) = f(x)$ .

Hence,  $H(x)$  ends up becoming a combination of 2 models.

Hope this becomes clear.

Kunal

**Samarendra Behera says:**

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=551#RESPOND)  
OCTOBER 24, 2013 AT 12:30 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-551)

Kunal,

Thanks for clarifying. Then if my understanding is correct, it will not be a combination of both the model which is fine considering that a case will exclusively fall in any of the two group. Sounds interesting to me. I will try to implement it in some real-life data. Thanks again for sharing this.

Samarendra



**Thomas Speidel says:**

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=561#RESPOND)  
OCTOBER 25, 2013 AT 12:11 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-561)

Tavish, you are assuming discontinuity, but in reading the article it seems to me discontinuity was artificially imposed by the analyst. CART (and its ilk) is inefficient not just because of issues you mentioned, but also because it hardly provides stable results. There's absolutely no reason to bin continuous variables which will lead to oversimplifications, besides implicitly assuming that the misclassification cost is the same. I suspect that might be one of the reasons why you are getting similar prediction errors when using logistic vs. CART. A sensible logistic model with no linearity assumptions and appropriate penalization and shrinkage will almost always produce superior results to CART and the likes.



**Tavish says:**

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=580#RESPOND)  
OCTOBER 26, 2013 AT 12:30 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-580)

Thomas, I completely agree that regression with non-linearity assumption will give superior result because for obvious reason, Linear regression is a subset of non-linear regression. But the motive of the article was that even non linear regression can be enhanced by introducing multi-variant bin variable. This multi-variant bin variables can be generated by plotting a simple decision tree.



**Hongbing Zhang says:**

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=569#RESPOND)  
OCTOBER 25, 2013 AT 4:15 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-569)

I think you can achieve this in SAS EM in similar way. First you fit a decision tree and then output the leaf node number (nominal) as one input variables to the sub-sequent regression model. So you model would like  $Y = X \cdot \text{Beta} + x_l \cdot \text{Lead\_number}$ . My worry is same as yours, original X matrix is highly correlated with the  $x_l$ . Multicollinearity would become a problem.



**Tavish Srivastava says:**

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=582#RESPOND](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?replytocom=582#RESPOND)  
OCTOBER 26, 2013 AT 12:43 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-582](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/#comment-582))

Hongbing, You are right. SAS EM does have that functionality. I use to do the same but in general it leads to multi-collinearity because the tree is too exhaustive. To combat this problem you can make an interactive tree only to 4-6 leafs and then add the regression node. Nonetheless, you are correct, the same can be achieved directly by E-Miner.



**Barbara Naydenk says:**

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=676#RESPOND](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?replytocom=676#RESPOND)  
OCTOBER 30, 2013 AT 6:56 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-676](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/#comment-676))

Thank you for your very cogent explanation here. Some may know this phenomenon as "testing for interaction terms", very commonly done in epidemiology and public health statistics (use your example, instead of "buying a BMW, think "having a good health outcome", based on an interaction between gender and age, or age and a chronic disease. When the interaction is significant, final models can be created within each age-sex group, or stratified outcomes can be presented to the audience to make it apparent what is going on. Interaction terms can easily be specified in any SAS model, too. I liked seeing this comparison to the differences between using regression and decision trees. Maybe this is helpful, maybe not, depending on training and subject matter.

## Diagnosing residual plots in linear regression model

(<http://www.analyticsvidhya.com/blog/2013/12/residual-plots-regression-model/>) says:

DECEMBER 1, 2013 AT 7:17 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-3357](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/#comment-3357))

[...] Every linear regression model should be validated on all the residual plots . Such regression plots directionally guides us to the right form of equations to start with. You might also be interested in the previous article on regression ( <http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/> ) [...]

**We just turned 1! - Analytics Vidhya Anniversary (<http://www.analyticsvidhya.com/blog/2014/04/turned-1/>) says:**

APRIL 20, 2014 AT 12:34 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-5758](http://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-5758))

[...] others in this list includes – Regression trick, Qlikview set analysis, tips to solve guess estimate. Do go through these articles if you have [...]



**Anup says:**

REPLY ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=10252#RESPOND](http://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=10252#RESPOND))  
JUNE 10, 2014 AT 7:41 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-10252](http://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-10252))

Interesting post Tavish! But a point to note is that binning continuous variables is not the most efficient solution and I don't think it will provide superior information. From a purely statistical perspective you are coarsening the information and this invariably should not be helpful. If you suspect the behavior of the variable is kinky then using a spline regression may be a good solution. Alternatively if you suspect wider non linearity then rather than use logit you could think of alternative approaches like Generalized Additive Models (GAM's).My two cents!!



**neetu saraf says:**

REPLY ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=78961#RESPOND](http://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=78961#RESPOND))  
FEBRUARY 25, 2015 AT 2:16 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-78961](http://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-78961))

Hi – I have build a linear regression as well as a logistic regression model using the same dataset. In both cases, i have changed the definition of the target. As we are still not sure how we would be implementing the final model. Now the results from both models are very close. Is there a way wherein i can combine the 2 models in order to increase the predictive power of the resultant model? thanks in advance for the reply.



**Amzee says:**

REPLY ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=82113#RESPOND](http://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=82113#RESPOND))



MARCH 22, 2015 AT 5:28 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-82113](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/#comment-82113))

hi, do you have any experience with support vector machines (AI)?? I need some help,thanks



**Saurabh says**  
REPLY ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/?REPLYTOCOM=91936#RESPOND](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/?replytocom=91936#respond))  
AUGUST 5, 2015 AT 2:25 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2013/10/TRICK-ENHANCE-POWER-REGRESSION-MODEL-2/#COMMENT-91936](http://www.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/#comment-91936))

Great article, Tavish !!

However, I have a couple of questions here.


How is the income bracket (  $> 200\text{ k}$  &  $\leq 100\text{ k}$  ) deduced, considering the fact that this income bracket did not turn up while creating a decision tree.

Also, how would the results be interpreted after executing logistic regression with an interaction variable ( Z, here).

How do we explain output of the model in business terms with respect to a unit change in interaction variable.

## LEAVE A REPLY

Connect with:

 ([http://www.analyticsvidhya.com/wp-login.php?action=wordpress\\_social\\_authenticate&mode=login&provider=Facebook&redirect\\_to=http%3A%2F%2Fwww.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/](http://www.analyticsvidhya.com/wp-login.php?action=wordpress_social_authenticate&mode=login&provider=Facebook&redirect_to=http%3A%2F%2Fwww.analyticsvidhya.com/blog/2013/10/trick-enhance-power-regression-model-2/))

Your email address will not be published.

Comment

Name (required)

Email (required)



Website

☐ Notify me of follow-up comments by email.

☐ Notify me of new posts by email.

**SUBMIT COMMENT**

## TOP USERS

Rank	Name	Points
1	 Nalin Pasricha ( <a href="http://datahack.analyticsvidhya.com/user/profile/Nalin">http://datahack.analyticsvidhya.com/user/profile/Nalin</a> )	3478
2	 SRK ( <a href="http://datahack.analyticsvidhya.com/user/profile/SRK">http://datahack.analyticsvidhya.com/user/profile/SRK</a> )	3364
3	Aayushmnit ( <a href="http://datahack.analyticsvidhya.com/user/profile/aayushmnit">http://datahack.analyticsvidhya.com/user/profile/aayushmnit</a> )	3075



4

binga (<http://datahack.analyticsvidhya.com/user/profile/binga>)

2623

5

vikash (<http://datahack.analyticsvidhya.com/user/profile/vikash>)

2190

[More Rankings \(http://datahack.analyticsvidhya.com/users\)](http://datahack.analyticsvidhya.com/users)

<http://pgpba.greatlakes.edu.in/?>
[utm\\_source=AVM&utm\\_medium=Banner&utm\\_campaign=Pgpba\\_decjan](#)

## POPULAR POSTS

- Free Must Read Books on Statistics & Mathematics for Data Science  
(<http://www.analyticsvidhya.com/blog/2016/02/free-read-books-statistics-mathematics-data-science/>)
- A Complete Tutorial to Learn Data Science with Python from Scratch  
(<http://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>)
- Essentials of Machine Learning Algorithms (with Python and R Codes)  
(<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>)

- A Complete Tutorial on Time Series Modeling in R  
(<http://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>)
- Complete guide to create a Time Series Forecast (with Codes in Python)  
(<http://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>)
- 4 tricky SAS questions commonly asked in interview  
(<http://www.analyticsvidhya.com/blog/2013/11/4-sas-tricky-analytics-interview/>)
- SAS vs. R (vs. Python) – which tool should I learn?  
(<http://www.analyticsvidhya.com/blog/2014/03/sas-vs-vs-python-tool-learn/>)
- 7 Important Model Evaluation Error Metrics Everyone should know  
(<http://www.analyticsvidhya.com/blog/2016/02/7-important-model-evaluation-error-metrics/>)



([http://imarticus.org/programs/business-analytics-](http://imarticus.org/programs/business-analytics-professional/)

[professional/](http://imarticus.org/programs/business-analytics-professional/))

## RECENT POSTS

---



(<http://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/>)

### **A Complete Tutorial to learn Data Science in R from Scratch**

(<http://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/>)

MANISH SARASWAT , FEBRUARY 28, 2016



(<http://www.analyticsvidhya.com/blog/2016/02/guide-build-predictive-models-segmentation/>)

### **Guide to Build Better Predictive Models using Segmentation**

(<http://www.analyticsvidhya.com/blog/2016/02/guide-build-predictive-models-segmentation/>)

GUEST BLOG , FEBRUARY 26, 2016



(<http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/>)

### **Quick Insights: India Analytics and Big Data Salary Report 2016**

(<http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/>)

KUNAL JAIN , FEBRUARY 24, 2016



(<http://www.analyticsvidhya.com/blog/2016/02/analytics-big-data-salary-report-2016/>)

### **India Exclusive: Analytics and Big Data Salary Report 2016**

(<http://www.analyticsvidhya.com/blog/2016/02/analytics-big-data-salary-report-2016/>)

KUNAL JAIN , FEBRUARY 22, 2016



([http://www.edvancer.in/certified-business-analytics?](http://www.edvancer.in/certified-business-analytics?utm_source=AV&utm_medium=AVads&utm_campaign=AVads1&utm_content=cbapavad)

[utm\\_source=AV&utm\\_medium=AVads&utm\\_campaign=AVads1&utm\\_content=cbapavad](http://www.edvancer.in/certified-business-analytics?utm_source=AV&utm_medium=AVads&utm_campaign=AVads1&utm_content=cbapavad))

## GET CONNECTED



**4,159**

FOLLOWERS

(<http://www.twitter.com/analyticsvidhya>)



**915**

FOLLOWERS

(<https://plus.google.com/+Analyticsvidhya>)



**11,950**

FOLLOWERS

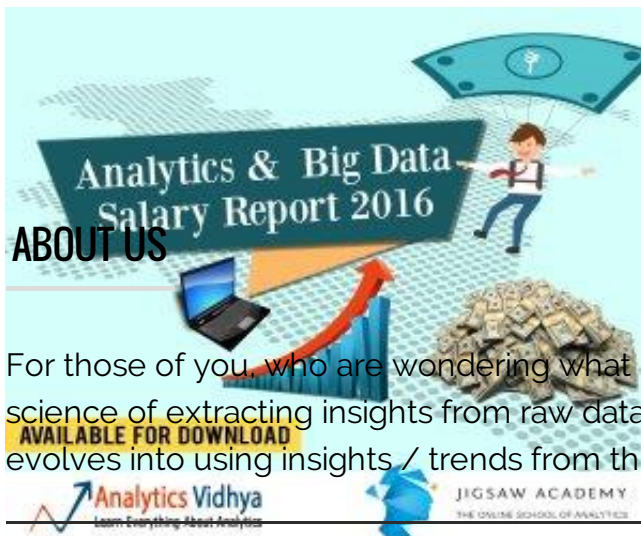
(<http://www.facebook.com/Analyticsvidhya>)



**Email**

SUBSCRIBE

(<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>)



([http://www.analyticsvidhya.com/blog/2016/02/quick-](http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/)

For those of you, who are wondering what is "Analytics Vidhya", "Analytics" can be defined as the science of extracting insights from raw data. The spectrum of analytics starts from capturing data and evolves into using insights / trends from this data to make informed decisions.

[insights-analytics-big-data-salary-report-2016/](http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/))

## STAY CONNECTED

---



**4,159**

FOLLOWERS

(<http://www.twitter.com/analyticsvidhya>)



**915**

FOLLOWERS

(<https://plus.google.com/+Analyticsvidhya>)



**11,950**

FOLLOWERS

(<http://www.facebook.com/Analyticsvidhya>)



**Email**

SUBSCRIBE

(<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>)

## LATEST POSTS

---



([http://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-](http://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/)

[science-scratch/](http://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/))

### **A Complete Tutorial to learn Data Science in R from Scratch**

(<http://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/>)

MANISH SARASWAT , FEBRUARY 28, 2016



([http://www.analyticsvidhya.com/blog/2016/02/guide-build-predictive-models-](http://www.analyticsvidhya.com/blog/2016/02/guide-build-predictive-models-segmentation/)

[segmentation/](http://www.analyticsvidhya.com/blog/2016/02/guide-build-predictive-models-segmentation/))

### **Guide to Build Better Predictive Models using Segmentation**

(<http://www.analyticsvidhya.com/blog/2016/02/guide-build-predictive-models-segmentation/>)

GUEST BLOG , FEBRUARY 26, 2016



([http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-](http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/)

[salary-report-2016/](http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/))

### **Quick Insights: India Analytics and Big Data Salary Report 2016**

(<http://www.analyticsvidhya.com/blog/2016/02/quick-insights-analytics-big-data-salary-report-2016/>)

KUNAL JAIN , FEBRUARY 24, 2016

### **India Exclusive: Analytics and Big Data Salary Report 2016**



2016/)

([http://www.analyticsvidhya.com/blog/2016/02/analytics-big-data-salary-report-](http://www.analyticsvidhya.com/blog/2016/02/analytics-big-data-salary-report-2016/)

**(<http://www.analyticsvidhya.com/blog/2016/02/analytics-big-data-salary-report-2016/>)**

KUNAL JAIN , FEBRUARY 22, 2016

## QUICK LINKS

---

Home (<http://www.analyticsvidhya.com/>)

About Us (<http://www.analyticsvidhya.com/about-me/>)

Our team (<http://www.analyticsvidhya.com/about-me/team/>)

Privacy Policy  
(<http://www.analyticsvidhya.com/privacy-policy/>)

Refund Policy  
(<http://www.analyticsvidhya.com/refund-policy/>)

Terms of Use  
(<http://www.analyticsvidhya.com/terms/>)

---

## TOP REVIEWS

---

---

© Copyright 2015 Analytics Vidhya