

**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>)



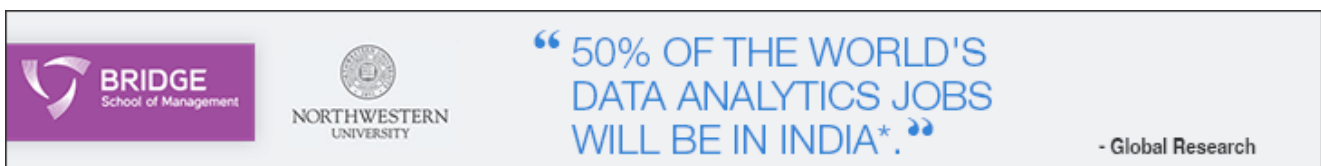
(<http://datahack.analyticsvidhya.com/contest/date-your-data>)

Home (<http://www.analyticsvidhya.com/>) > Business Analytics (<http://www.analyticsvidhya.com/blog/category/business-analytics/>)

# Getting smart with Machine Learning - AdaBoost and Gradient Boost

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

SHARE **f** (<http://www.facebook.com/sharer.php?u=http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/&t=Getting%20smart%20with%20Machine%20Learning%20-%20AdaBoost%20and%20Gradient%20Boost>) **t** (<https://twitter.com/home?status=Getting%20smart%20with%20Machine%20Learning%20-%20AdaBoost%20and%20Gradient%20Boost+http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/>) **g+** (<https://plus.google.com/share?url=http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/>) **p** (<http://pinterest.com/pin/create/button/?url=http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/&media=http://www.analyticsvidhya.com/wp-content/uploads/2015/05/poster-learning-machin.jpg&description=Getting%20smart%20with%20Machine%20Learning%20-%20AdaBoost%20and%20Gradient%20Boost>)



(<http://admissions.bridgesom.com/pba-new/>)

utm\_source=AV&utm\_medium=Banner&utm\_campaign=AVBanner)

## Introduction

Machine Learning algorithms are like solving a Rubik Cube. You grapple at the beginning to figure out the hidden algorithm, but once learnt, some can even solve it in less than 7 seconds (<https://www.youtube.com/watch?v=8xcpaurEZQ>).

Suppose, you are stuck in following situation:

*You are served with legion of data to generate useful insights. To nail this challenge, you got your best team members lined up or you come forward to lead from the front. You begin with building a predictive model. Then, you check the output statistics and gets disheartened. Reason being, the predictive power of the model is very low. Now, you desperately want to figure a way to increase the predictive power. What will you do?*

(Share your answers in the comment section below)

There are some machine learning engines. These engines make use of certain algorithms and help user reach to the output stage. Some of the most popular engines are Decision Tree (<http://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/>) and Regression (<http://www.analyticsvidhya.com/blog/2013/12/extracting-variables-regression-model-2/>).

In this article, we'll introduce you to some of the best practices used to enhance power of these engines to achieve a higher predictability using an additional booster.

## Where are Boosted algorithms required?

Boosted algorithms are used where we have plenty of data to make a prediction. And we seek exceptionally high predictive power. It is used to for reducing bias and variance in supervised learning. It combines multiple weak predictors to a build strong predictor.

If you ever want to participate in Kaggle competitions, I would suggest that **you bookmark this article**. Participants in Kaggle competitions use these boosting algorithms extensively.

The underlying engine used for boosting algorithms can be anything. For instance, **AdaBoost** is a boosting done on Decision stump. There are many other boosting algorithms which use other types of engine such as:

**1. GentleBoost**

**2. Gradient Boosting (Always my first choice for any Kaggle problem)**

**3. LPBoost**

**4. BrownBoost**

Perhaps, I can go on adding more engines to this list. But, I would like to focus on these five boosting techniques which are the most commonly used. **Let's first learn about – AdaBoost.**

## What are Classifier Boosting Algorithms ?

Classification problem is the one where we need to assign every observation to a given set of class. The easiest classification problem is the one with binary class. This problem can be solved using AdaBoost. Let's take a very simple example to understand the underlying concept of AdaBoost. You have two classes : 0's and 1's. Each number is an observation. The only two features available is x-axis and y-axis. For instance (1,1) is a 0 while (4,4) is a 1. Now using these two features you need to classify each observation. Our ultimate objective remains the same as any classifier problem : find the classification boundary. Following are the step we follow to apply an AdaBoost.

**Step 1 : Visualize the data** : Let's first understand the data and find insights on whether we have a linear classifier boundary. As shown below, no such boundary exist which can separate 0's from 1's.

0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0

(<http://i0.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/05/pic1.png>)

**Step 2 : Make the first Decision stump :** You have already read about decision trees in many of our previous articles. Decision stump is a unit depth tree which decides just 1 most significant cut on features. Here it chooses draw the boundary starting from the third row from top. Now the yellow portion is expected to be all 0's and unshaded portion to be all 1's. However, we see high number of false positive post we build this decision stump. We have nine 1's being wrongly qualified as 0's. And similarly eighteen 0's qualified as 1's.

0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0

(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/05/pic2.png>)

**Step 3 : Give additional weight to mis-classified observations:** Once we know the misclassified observations, we give additional weight to these observations. Hence, you see 0's and 1's in bold which were misclassified before. In the next level, we will make sure that these highly weighted observation are classified correct

0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0

(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/05/pic3.png>)

**Step 4 : Repeat the process and combine all stumps to get final classifier:** We repeat the process multiple times and focus more on previously misclassified observations. Finally, we take a weighted mean of all the boundaries discovered which will look something as below.

0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	0	0	0

(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/05/pic4.png>)

### ***More real life examples!***

A classic use case where AdaBoost algorithms is in the problem of **Face Detection**. You can think of this as a more complex boundary detection as we found in last example. Once we have that boundary, we can now create features and classify if the image has a face or not.



(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/05/face-detection-people.jpg>)

However, face recognition is commonly done after a gray scale transformation is done on the RCB image and finally a threshold is assumed to create face boundaries. You can read here (<http://www.analyticsvidhya.com/blog/2015/01/basics-image-processing-feature-extraction-python/>) to make transformation to gray scale or find threshold to create a black and white image. Once we have the transformation, we now analyze each patch of the image.

## Brief introduction to Regression boosters

Similar to classifier boosters, we also have regression boosters. In these problems we have continuous variable to predict. This is commonly done using gradient boosting algorithm. Here is a non-mathematical description of how gradient boost works :

Type of Problem – You have a set of variables vectors  $x_1$  ,  $x_2$  and  $x_3$ . You need to predict  $y$  which is a continuous variable.

Steps of Gradient Boost algorithm

*Step 1* : Assume mean is the prediction of all variables.

*Step 2* : Calculate errors of each observation from the mean (latest prediction).

*Step 3* : Find the variable that can split the errors perfectly and find the value for the split. This is assumed to be the latest prediction.

*Step 4* : Calculate errors of each observation from the mean of both the sides of split (latest prediction).

*Step 5* : Repeat the step 3 and 4 till the objective function maximizes/minimizes.

*Step 6* : Take a weighted mean of all the classifiers to come up with the final model.

We have excluded the mathematical formation of boosting algorithms from this article to keep the article simple.

## End Notes

Boosting is one of the most powerful tool used in machine learning. These models suffer the problem of over-fitting if the data sample is too small. Whenever the training sample is large enough, you should try boosting with many different engines as discussed in this article.

Were you haunted by any questions/doubts while learning this concept? Ask our analytics community (<http://discuss.analyticsvidhya.com>) and never let your learning process stop.

Have you used boosting before for any kind of analysis? Did you see any significant lift compared to the traditional models? Do let us know your thoughts about this guide in the comments section 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), follow us on twitter  
(<http://twitter.com/analyticsvidhya>) or like our facebook page  
(<http://facebook.com/analyticsvidhya>).

#### Share this:

 (<http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?share=linkedin&nb=1>) 67

 (<http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?share=facebook&nb=1>) 67

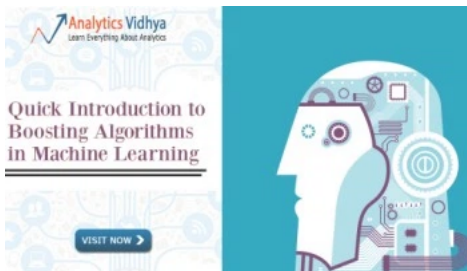
 (<http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?share=google-plus-1&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?share=twitter&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?share=pocket&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?share=reddit&nb=1>)

## RELATED



(<http://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/>)

**Quick Introduction to Boosting Algorithms in Machine Learning**  
(<http://www.analyticsvidhya.com/blog/2015/11/quick-introduction-boosting-algorithms-machine-learning/>)

In "Machine Learning"



(<http://www.analyticsvidhya.com/blog/2015/09/questions-ensemble-modeling/>)

**5 Easy questions on Ensemble Modeling everyone should know**  
(<http://www.analyticsvidhya.com/blog/2015/09/questions-ensemble-modeling/>)  
In "Business Analytics"



(<http://www.analyticsvidhya.com/blog/2016/02/7-important-model-evaluation-error-metrics/>)

**7 Important Model Evaluation Error Metrics Everyone should know**  
(<http://www.analyticsvidhya.com/blog/2016/02/7-important-model-evaluation-error-metrics/>)  
In "Machine Learning"



TAGS: ADABOOST ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ADABOOST/](http://www.analyticsvidhya.com/blog/tag/adaboost/)), BOOSTING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/BOOSTING/](http://www.analyticsvidhya.com/blog/tag/boosting/)), GBM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/GBM/](http://www.analyticsvidhya.com/blog/tag/gbm/)), GRADIENT BOOSTING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/GRADIENT-BOOSTING/](http://www.analyticsvidhya.com/blog/tag/gradient-boosting/)), MACHINE LEARNING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MACHINE-LEARNING/](http://www.analyticsvidhya.com/blog/tag/machine-learning/)), SUPERVISED LEARNING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/SUPERVISED-LEARNING/](http://www.analyticsvidhya.com/blog/tag/supervised-learning/))

#### Previous Article

**Senior Data Analyst - Marketelligent - Bangalore - (2-5 Years of experience)**  
(<http://www.analyticsvidhya.com/blog/2015/05/senior-data-analyst-marketelligent-bangalore-2-5-years-experience/>)

#### Next Article

**In Conversation with Mr. Sanjeev Mishra, CEO, Convergytics**  
(<http://www.analyticsvidhya.com/blog/2015/05/conversation-with-mr-sanjeev-mishra-ceo-convergytics/>)



(<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.

## ONE COMMENT

---




**Sam says:** [REPLY \(HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/05/BOOSTING-ALGORITHMS-SIMPLIFIED/?REPLYTOCOM=98818#RESPOND\)](http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/?replytocom=98818#respond)  
NOVEMBER 2, 2015 AT 2:39 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/05/BOOSTING-ALGORITHMS-SIMPLIFIED/#COMMENT-98818](http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/#comment-98818))

Can you please give the details of how to perform boosting using R?

---

## 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/2015/05/boosting-algorithms-simplified%2F](http://www.analyticsvidhya.com/wp-login.php?action=wordpress_social_authenticate&mode=login&provider=Facebook&redirect_to=http%3A%2F%2Fwww.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified%2F))

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/>)



**Big Data Training Chennai**

Gain in-depth Knowledge on Core ideas of Big Data Hadoop in Chennai

[collaberatact.com/Big-Data-Chennai](http://collaberatact.com/Big-Data-Chennai)



(<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



**4,162**

FOLLOWERS

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



**915**

FOLLOWERS

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



**11,954**

FOLLOWERS

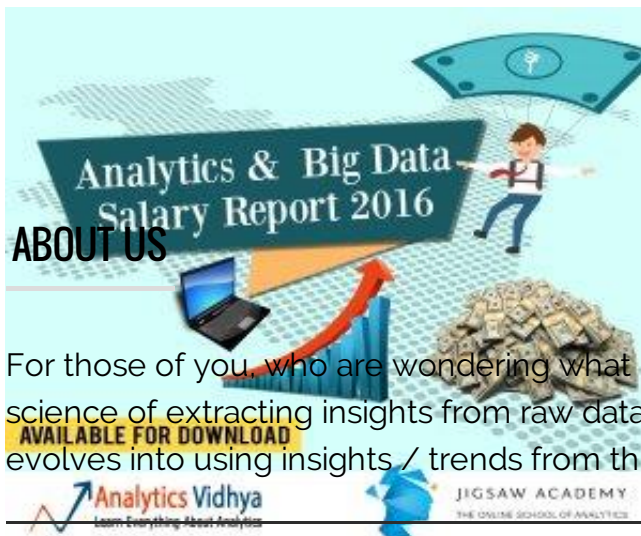
(<http://www.facebook.com/Analyticsvidhya>)



**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

---



**4,162**

FOLLOWERS

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



**915**

FOLLOWERS

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



**11,954**

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