

f (<https://www.facebook.com/AnalyticsVidhya>)

🐦 (<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/category/business-analytics/>)

# Tuning the parameters of your Random Forest model

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

f (<http://www.facebook.com/sharer.php?u=http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/>)

🐦 (<https://twitter.com/home?status=Tuning%20the%20parameters%20of%20your%20Random%20Forest%20model>)

(<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/>)

g+ (<https://plus.google.com/share?url=http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/>)

p://pinterest.com/pin/create/button/?url=<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/>

&media=<http://www.analyticsvidhya.com/wp-content/uploads/2015/06/Tuning-the-parameters-of-your-Random-Forest-model.jpg>

&description=Tuning%20the%20parameters%20of%20your%20Random%20Forest%20model)



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

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

# Why to tune Machine Learning Algorithms?

A month back, I participated in a Kaggle competition (<http://www.analyticsvidhya.com/blog/2015/06/start-journey-kaggle/>) called TFI. I started with my first submission at 50th percentile. Having worked relentlessly on feature engineering for more than 2 weeks, I managed to reach 20th percentile. To my surprise, right after tuning the parameters of the machine learning algorithm I was using, I was able to breach top 10th percentile.

This is how important tuning these machine learning algorithms are. Random Forest is one of the easiest machine learning tool used in the industry. In our previous articles, we have introduced you to Random Forest (<http://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/>) and compared it against a CART model (<http://www.analyticsvidhya.com/blog/2014/06/comparing-cart-random-forest-1/>). Machine Learning tools are known for their performance.

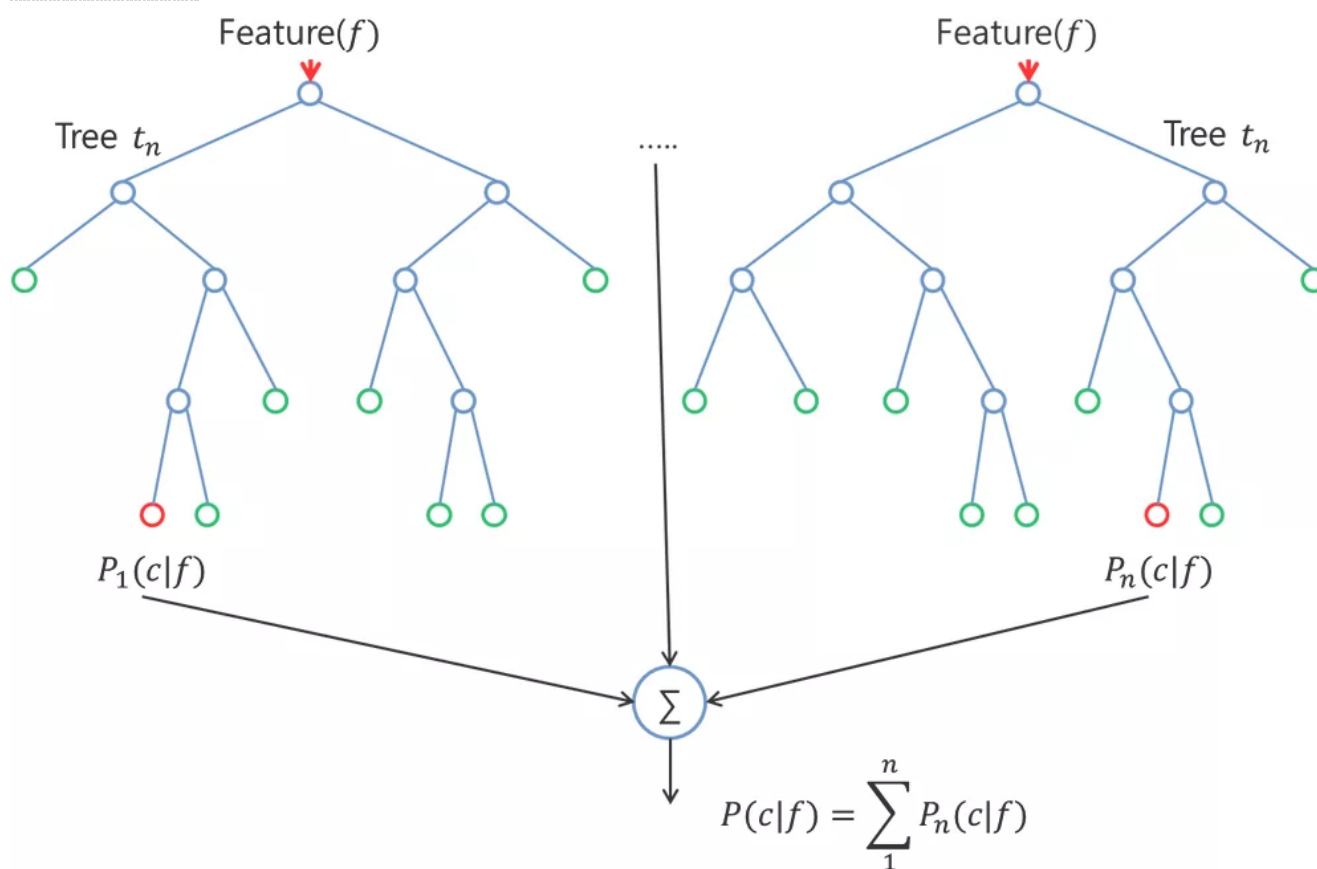


(<http://io.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/06/Tuning-the-parameters-of-your-Random-Forest-model.jpg>)

# What is a Random Forest?

Random forest is an ensemble tool which takes a subset of observations and a subset of variables to build a decision trees. It builds multiple such decision tree and amalgamate them together to get a more accurate and stable prediction. This is direct consequence of the fact that by maximum voting from a panel of independent judges, we get the final prediction better than the best judge.

(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/06/random-forest4.png>)



(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/06/random-forest7.png>)

We generally see a random forest as a black box which takes in input and gives out predictions, without worrying too much about what calculations are going on the back end. This black box itself have a few levers we can play with. Each of these levers have some effect

on either the performance of the model or the resource – time balance. In this article we will talk more about these levers we can tune, while building a random forest model.

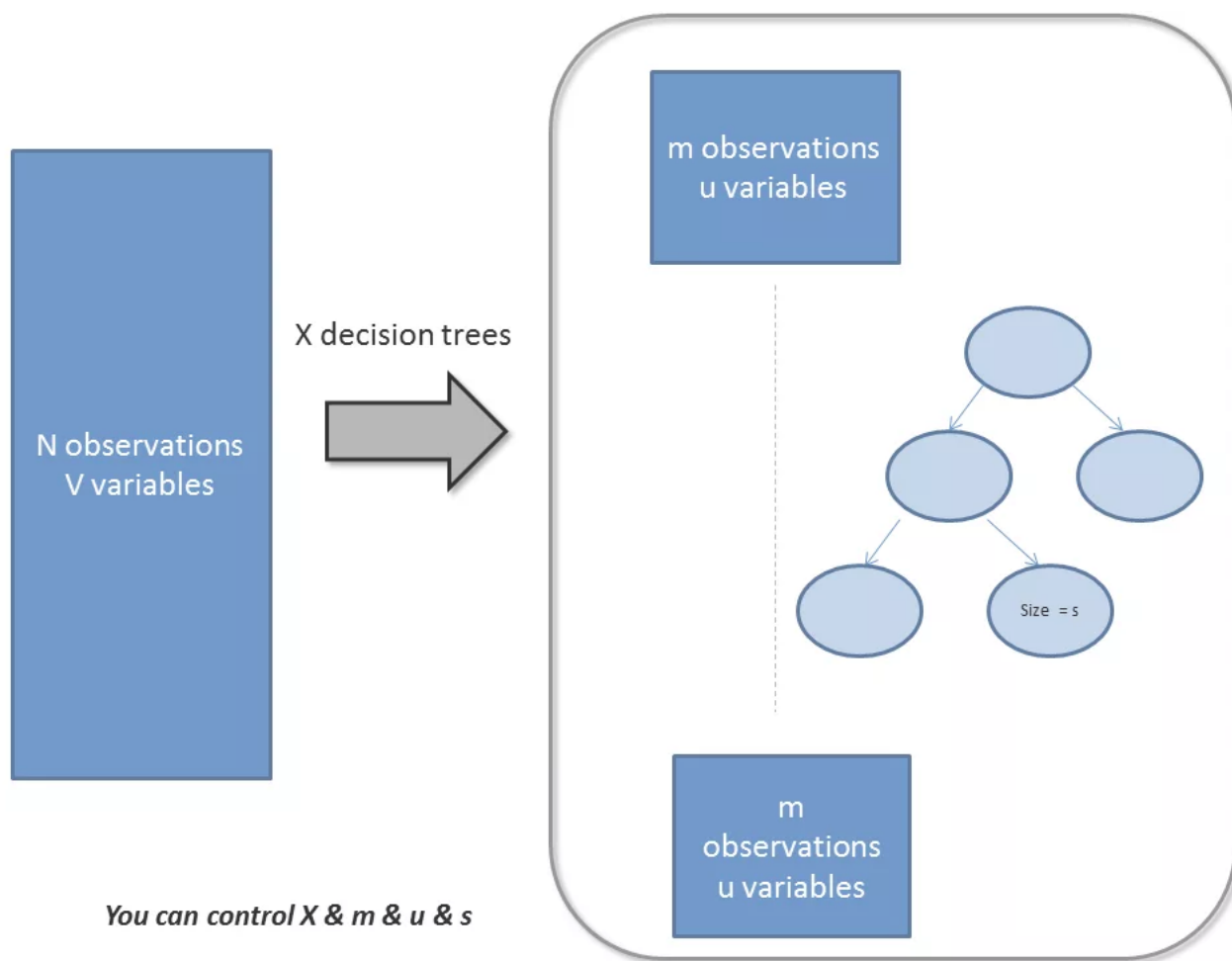
## Parameters / levers to tune Random Forests

Parameters in random forest are either to increase the predictive power of the model or to make it easier to train the model. Following are the parameters we will be talking about in more details (Note that I am using Python conventional nomenclatures for these parameters) :



(<http://io.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/06/knobs.jpg>)

### 1. Features which make predictions of the model better



(<http://i2.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/06/RF.png>)

There are primarily 3 features which can be tuned to improve the predictive power of the model :

### 1.a. max\_features:

These are the maximum number of features Random Forest is allowed to try in individual tree. There are multiple options available in Python to assign maximum features. Here are a few of them :

1. *Auto/None* : This will simply take all the features which make sense in every tree. Here we simply do not put any restrictions on the individual tree.
2. *sqrt* : This option will take square root of the total number of features in individual run. For instance, if the total number of variables are 100, we can only take 10 of them in individual tree. "log2" is another similar type of option for max\_features.

3. *0.2* : This option allows the random forest to take 20% of variables in individual run. We can assign and value in a format "0.x" where we want x% of features to be considered.

### **How does "max\_features" impact performance and speed?**

Increasing max\_features generally improves the performance of the model as at each node now we have a higher number of options to be considered. However, this is not necessarily true as this decreases the diversity of individual tree which is the USP of random forest. But, for sure, you decrease the speed of algorithm by increasing the max\_features. Hence, you need to strike the right balance and choose the optimal max\_features.

#### **1.b. n\_estimators :**

This is the number of trees you want to build before taking the maximum voting or averages of predictions. Higher number of trees give you better performance but makes your code slower. You should choose as high value as your processor can handle because this makes your predictions stronger and more stable.

#### **1.c. min\_sample\_leaf :**

If you have built a decision tree before, you can appreciate the importance of minimum sample leaf size. Leaf is the end node of a decision tree. A smaller leaf makes the model more prone to capturing noise in train data. Generally I prefer a minimum leaf size of more than 50. However, you should try multiple leaf sizes to find the most optimum for your use case.

## **2. Features which will make the model training easier**

There are a few attributes which have a direct impact on model training speed. Following are the key parameters which you can tune for model speed :

#### **2.a. n\_jobs :**

This parameter tells the engine how many processors is it allowed to use. A value of “-1” means there is no restriction whereas a value of “1” means it can only use one processor. Here is a simple experiment you can do with Python to check this metric :

```
%timeit
```

```
model = RandomForestRegressor(n_estimator = 100, oob_score = TRUE,n_jobs = 1,random_state =1)
```

```
model.fit(X,y)
```

Output ——— 1 loop best of 3 : 1.7 sec per loop

```
%timeit
```

```
model = RandomForestRegressor(n_estimator = 100,oob_score = TRUE,n_jobs = -1,random_state =1)
```

```
model.fit(X,y)
```

Output ——— 1 loop best of 3 : 1.1 sec per loop

“%timeit” is an awsum function which runs a function multiple times and gives the fastest loop run time. This comes out very handy while scalling up a particular function from prototype to final dataset.

## 2.b. random\_state :

This parameter makes a solution easy to replicate. A definite value of random\_state will always produce same results if given with same parameters and training data. I have personally found an ensemble with multiple models of different random states and all optimum parameters

sometime performs better than individual random state.

## 2.c. oob\_score :

This is a random forest cross validation method. It is very similar to leave one out validation technique, however, this is so much faster. This method simply tags every observation used in different trees. And then it finds out a maximum vote score for every observation based on only trees which did not use this particular observation to train itself.

Here is a single example of using all these parameters in a single function :

```
model = RandomForestRegressor(n_estimator = 100, oob_score = TRUE, n_jobs = -1, random_state = 50,  
max_features = "auto", min_samples_leaf = 50)
```

```
model.fit(X,y)
```

## Learning through a case study

We have referred to Titanic case study in many of our previous articles. Let's try the same problem again. The objective of this case here will be to get a feel of random forest parameter tuning and not getting the right features. Try following code to build a basic model :

```
from sklearn.ensemble import RandomForestRegressor
```

```
from sklearn.metrics import roc_auc_score
```

```
import pandas as pd
```



```
x = pd.read_csv("train.csv")
```

```
y = x.pop("Survived")
```

```
model = RandomForestRegressor(n_estimator = 100 , oob_score = TRUE, random_state = 42)
```

```
model.fit(x(numeric_variable,y)
```

```
print "AUC - ROC : ", roc_auc_score(y,model.oob_prediction)
```

AUC – ROC : 0.7386

This is a very simplistic model with no parameter tuning. Now let's do some parameter tuning. As we have discussed before, we have 6 key parameters to tune. We have some grid search algorithm built in Python, which can tune all parameters automatically. But here let's get our hands dirty to understand the mechanism better. Following code will help you tune the model for different leaf size.

***Exercise : Try running the following code and find the optimal leaf size in the comment box.***

```
sample_leaf_options = [1,5,10,50,100,200,500]
```

```
for leaf_size in sample_leaf_options :
```

```
    model = RandomForestRegressor(n_estimator = 200, oob_score = TRUE, n_jobs = -1, random_state  
    =50,                                max_features = "auto", min_samples_leaf = leaf_size)
```

```
model.fit(x(numeric_variable,y)
```

```
print "AUC - ROC : ", roc_auc_score(y,model.oob_prediction)
```

## End Notes

Machine learning tools like random forest, SVM, neural networks etc. are all used for high performance. They do give high performance, but users generally don't understand how they actually work. Not knowing the statistical details of the model is not a concern however not knowing how the model can be tuned well to clone the training data restricts the user to use the algorithm to its full potential. In some of the future articles we will take up tuning of other machine learning algorithm like SVM , GBM and neaural networks.

Have you used random forest before? What parameters did you tune? How did tuning the algorithm impact the performance of the model? Did you see any significant benefits by doing the same? 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/2015/06/tuning-random-forest-model/?share=linkedin&nb=1>)

91

 (<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?share=facebook&nb=1>)

194

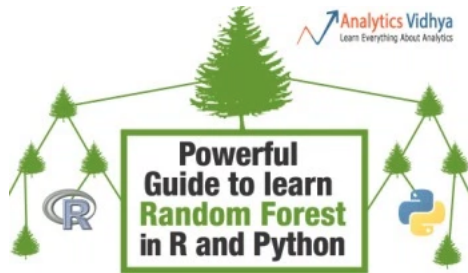
 (<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?share=google-plus-1&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?share=twitter&nb=1>)

✓ (<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?share=pocket&nb=1>)

😊 (<http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?share=reddit&nb=1>)

## RELATED



(<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>)  
Essentials of Machine Learning Algorithms (with Python and R Codes)

(<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>)  
In "Business Analytics"

(<http://www.analyticsvidhya.com/blog/2015/09/random-forest-algorithm-multiple-challenges/>)  
Powerful Guide to learn Random Forest (with codes in R & Python)  
(<http://www.analyticsvidhya.com/blog/2015/09/random-forest-algorithm-multiple-challenges/>)  
In "Business Analytics"

(<http://www.analyticsvidhya.com/blog/2015/01/introduction-online-machine-learning-simplified-2/>)  
Introduction to Online Machine Learning : Simplified  
(<http://www.analyticsvidhya.com/blog/2015/01/introduction-online-machine-learning-simplified-2/>)  
In "Big data"

TAGS: MAX\_FEATURES ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MAX\\_FEATURES/](http://www.analyticsvidhya.com/blog/tag/max-features/)), MIN\_SAMPLE\_LEAF ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MIN\\_SAMPLE\\_LEAF/](http://www.analyticsvidhya.com/blog/tag/min-sample-leaf/)), N\_ESTIMATOR ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/N\\_ESTIMATOR/](http://www.analyticsvidhya.com/blog/tag/n-estimator/)), N\_JOBS ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/N\\_JOBS/](http://www.analyticsvidhya.com/blog/tag/n-jobs/)), PARAMETER\_TUNING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/PARAMETER\\_TUNING/](http://www.analyticsvidhya.com/blog/tag/parameter-tuning/)), PYTHON ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/PYTHON/](http://www.analyticsvidhya.com/blog/tag/python/)), RANDOM FOREST ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/RANDOM\\_FOREST/](http://www.analyticsvidhya.com/blog/tag/random-forest/))

**Previous Article****Cheat Sheet for Exploratory Data Analysis in Python**

(<http://www.analyticsvidhya.com/blog/2015/06/infogtipic/cheat-sheet-data-exploration-python/>)

**Next Article****In Conversation with Mr. Stefan Groschupf, Founder and CEO, Datameer**

(<http://www.analyticsvidhya.com/blog/2015/06/conversa-mr-stefan-groschupf-ceo-datameer/>)



(<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.

## 4 COMMENTS



**Aayush Agrawal says:** [REPLY \(HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/?REPLYTOCOM=88206#RESPOND\)](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?replytocom=88206#respond)  
JUNE 10, 2015 AT 6:50 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/#COMMENT-88206](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/#comment-88206))

Brilliantly written article. Currently I have used all of these techniques in a Data science problem I was working on and it definitely helps in improving model performance and accuracy. Recently, I came across something else also when I was reading some articles on Random Forest, i.e a Regularization of Random Forest. The theme was to only split data with some variables if the splitting is significant enough using Statistical validation, now this is something which can help in taking Random Forest to next level, as It can help in reducing over-fitting. I tried to use it using R caret package but I think this technique is computationally expensive so couldn't run it over my system. I would love to see an article on it to understand it's working and how its performance can be improved.



**KARTHI V says:** [REPLY \(HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/?REPLYTOCOM=88254#RESPOND\)](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?replytocom=88254#respond)  
JUNE 10, 2015 AT 4:00 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/#COMMENT-88254](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/#comment-88254))

Hi Tavish,

Very useful article.



**Ravi says:** [REPLY \(HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/?REPLYTOCOM=94729#RESPOND\)](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?replytocom=94729#respond)  
SEPTEMBER 11, 2015 AT 6:43 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/#COMMENT-94729](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/#comment-94729))

I love AV and am a fan of your articles. I have heard something like Conditional Inference Trees which are similar to Random Forests. Can you share your thoughts on Conditional Inference Trees also? How does it work & its tuning parameters, when does it outcast Random Forests?




**Josh says:** REPLY ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/?REPLYTOCOM=95226#RESPOND](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/?replytocom=95226#respond))  
SEPTEMBER 17, 2015 AT 2:39 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/06/TUNING-RANDOM-FOREST-MODEL/#COMMENT-95226](http://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/#comment-95226))

Great article! I would love to see something similar regarding parameter tuning for the XGBoost package.

## 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/2015/06/tuning-random-forest-model%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/2015/06/tuning-random-forest-model%2F))






Your email address will not be published.

Comment

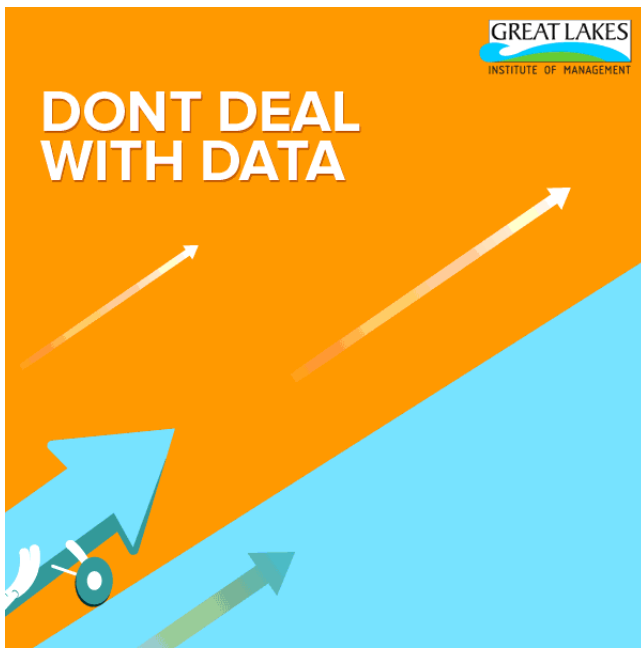
Name (required)

☐ 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 ( <a href="http://datahack.analyticsvidhya.com/user/profile/binga">http://datahack.analyticsvidhya.com/user/profile/binga</a> )	2623
5		vikash ( <a href="http://datahack.analyticsvidhya.com/user/profile/vikash">http://datahack.analyticsvidhya.com/user/profile/vikash</a> )	2190

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



([http://pgpba.greatlakes.edu.in/?](http://pgpba.greatlakes.edu.in/?utm_source=AVM&utm_medium=Banner&utm_campaign=Pgpba_decjan)

[utm\\_source=AVM&utm\\_medium=Banner&utm\\_campaign=Pgpba\\_decjan](http://pgpba.greatlakes.edu.in/?utm_source=AVM&utm_medium=Banner&utm_campaign=Pgpba_decjan))

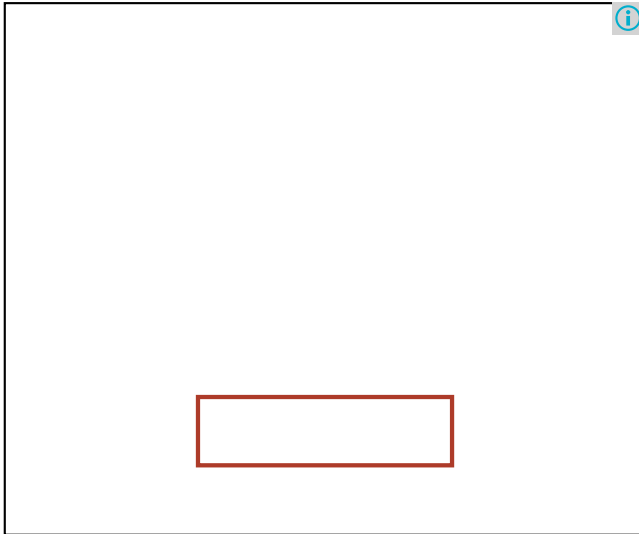
## POPULAR POSTS

---

- 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



**4,162**

FOLLOWERS

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



**915**

FOLLOWERS

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



**11,954**

FOLLOWERS

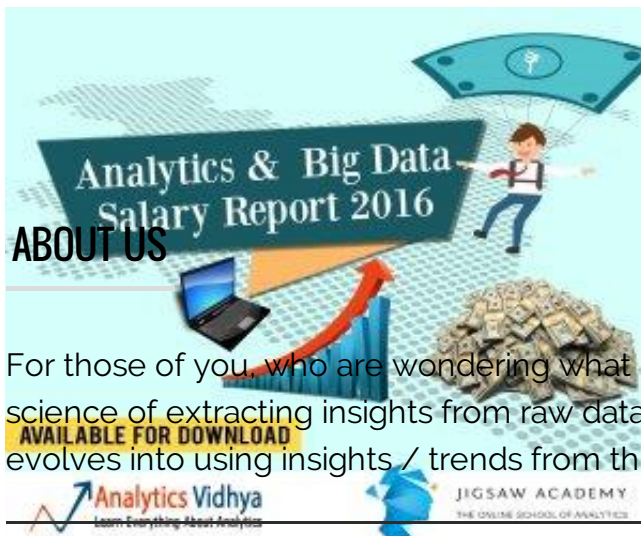
(<http://www.facebook.com/Analyticsvidhya>)



**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