

[f](https://www.facebook.com/AnalyticsVidhya) (<https://www.facebook.com/AnalyticsVidhya>) | [Twitter](https://twitter.com/analyticsvidhya) (<https://twitter.com/analyticsvidhya>)

[g+](https://plus.google.com/+Analyticsvidhya/posts) (<https://plus.google.com/+Analyticsvidhya/posts>)

[in](https://www.linkedin.com/groups/Analytics-Vidhya-Learn-everything-about-5057165) (<https://www.linkedin.com/groups/Analytics-Vidhya-Learn-everything-about-5057165>)

Home (<https://www.analyticsvidhya.com/>) | Blog (<https://www.analyticsvidhya.com/blog/>)

Jobs (<https://www.analyticsvidhya.com/jobs/>) | Trainings (<https://trainings.analyticsvidhya.com>)

Learning Paths (<https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/>)

Discuss (<https://discuss.analyticsvidhya.com>) | Corporate (<https://www.analyticsvidhya.com/corporate/>)



(<https://www.analyticsvidhya.com>)



(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVhome_top)

Home (<https://www.analyticsvidhya.com/>) > Machine Learning (<https://www.analyticsvidhya.com/blog/category/machine-learning/>) > How to build Ensemble Models in machine learning? (with code in R) (<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/>)

How to build Ensemble Models in machine learning? (with code in R)

MACHINE LEARNING ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/MACHINE-LEARNING/](https://www.analyticsvidhya.com/blog/category/machine-learning/)) R ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/R/](https://www.analyticsvidhya.com/blog/category/r/))

?com/sharer.php?u=<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/>

le%20Models%20in%20machine%20learning%20(with%20code%20in%20R)) [Twitter](https://twitter.com/analyticsvidhya)

TOP ANALYTICS VIDHYA USERS

Rank	Name
1	vopani (https://datahack.analyticsvidhya.com/user/profile/Rohit)
2	SRK (https://datahack.analyticsvidhya.com/user/profile/SHRIKANT)
3	binga (https://datahack.analyticsvidhya.com/user/profile/binga)

1%20Ensemble%20Models%20in%20machine%20learning?%20(with%20code%20in%20R)+https:

017/02/introduction-to-ensembling-along-with-implementation-in-r/)

Analytics Vidhya HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/)

idhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/)

button?utm_source=AVBlog_sidebottom (https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-

https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/02/15040351



Over the last 12 months I have been participating in a number of machine learning hackathons on Analytics Vidhya and /data-engineering-talent-hunt-hackathon



(https://datahack.analyticsvidhya.com/contest

/the-young-optimization-crackerjack-contest/?utm_source=AVBlog_sidebottom)

Most of the time I was able to crack the feature engineering part but probably didn't use the ensemble of multiple models. If you are a beginner, it's even better to get familiar with ensembling as early as possible. Chances are that you are already applying it without knowing!

In this article, I'll take you through the basics of ensemble modeling. Then I will walk you through the advantages of ensembling. Also, to provide you hands-on experience on ensemble modeling, we will use ensembling on a hackathon problem using R.

Aayushmnit LEARN ENGAGE COMPETE

(https://datahack.com/user/profile/aayushmnit)

Mark Landry (https://datahack.com/user/profile/marklandry)

5 (https://datahack.com/user/profile/marklandry)

More Rankings (http://datahack.analyticsvidhya.com

management.corporatetrainingmate

/users)

Table of Content

Analytics Vidhya
Learn everything about analytics

(<https://www.analyticsvidhya.com>)
2. Types of ensembling [GET HIRED](#)

HOME ([HTTPS://WWW.ANALYTICSVIDHYA.COM/](https://WWW.ANALYTICSVIDHYA.COM/)) LEARN ENGAGE COMPETE



ensembles
ensembling in R

t you can build individual
tart your journey with our
[dhya.com/learning-
st-business-intelligence](https://www.analyticsvidhya.com/learning-st-business-intelligence)

(<https://www.analyticsvidhya.com/campus-ambassador-program>)

/data-engineering-talent-hunt-hackathon



the process of combining two or more different types called base learners. A system which incorporates multiple learners. It can be understood as when multiple traders to make a prediction if the stock will go up or not.

(<https://www.analyticsvidhya.com/tutorial/ensemble-learning-machine-learning>) Since all of them have a different understanding of the stock market and utilize a different mapping function from the problem statement to the outcome. Therefore, they are supposed to make varied predictions on the stock price based on their own understandings of the market.

Now we can take all of these predictions into account while making the final decision. This will make our final decision more robust, accurate and less likely to be biased. The final decision would have been opposite if one of these traders would have made this decision alone.

You can consider another example of a candidate going through multiple rounds of job interviews. The final decision of candidate's ability is generally taken based on the feedback of

Hawking's Bold Prediction

Stephen Hawking Predicts "Big Bang" in the History of Civilization
The Motley Fool

POPULAR POSTS

- Essentials of Machine Learning Algorithms (with Python and R Codes)
(<https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>)
- A Complete Tutorial to Learn Data Science with Python from Scratch
(<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>)
- 7 Types of Regression Techniques you should know!
(<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)
- 11 most read Deep Learning Articles from

all the interviewers. Although a single interviewer might not be able to test the candidate for each required skill and trait. But the combined feedback of multiple interviewers usually helps in better assessment of the candidate.

Analytics Vidhya in 2017
 LEARN ✓ ENGAGE ✓ COMPETE
<https://www.analyticsvidhya.com/blog/2017/12/11-deep-learning-analytics-vidhya-2017/>

Understanding Support Vector Machine algorithm from examples (along with code)

<https://www.analyticsvidhya.com/blog/2017/09/understaing-campush-ambassador-machine-example-code/>

10 Data Science, Machine Learning and AI Podcasts You Must Listen To
<https://www.analyticsvidhya.com/blog/2018/01/10-data-science-machine-learning-ai-podcasts-must-listen/>

A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python)

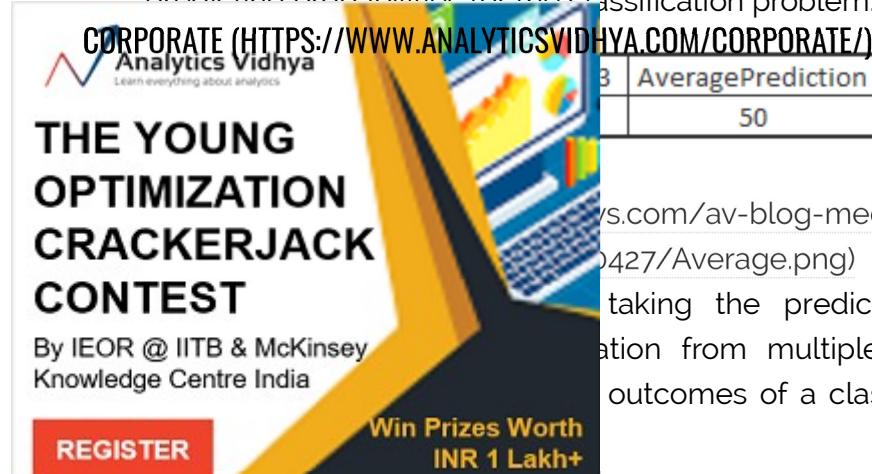
<https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/>

6 Easy Steps to Learn Naive Bayes Algorithm (with codes in Python and R)

<https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>



Averaging: It's defined as taking the average of predictions from models in case of regression problem or while predicting probabilities for the classification problem.



https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_sidebottom

[\(https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/02/14160531/voting.png\)](https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/02/14160531/voting.png)

- **Weighted average:** In this, different weights are applied to predictions from multiple models then taking the average which means giving high or low importance to specific model output.

	Model1	Model2	Model3	VotingPrediction
Weight	0.4	0.3	0.3	
Prediction	45	40	60	48

[\(https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/02/14161019/Wtaverage1.png\)](https://s3-ap-south-1.amazonaws.com/av-blog-media/wp-content/uploads/2017/02/14161019/Wtaverage1.png)

Practically speaking, there can be a countless number of ways in which analytics videos resemble different models. But these are some techniques that are mostly used:

LEARN ✕ ENGAGE ✕ COMPETE ✕



ferred to as bootstrap
gging we first need to
OM/BLOG/CATEGORY/AVBYTES/)
strapping is a sampling
observations or rows out of
VALCOM/CONTACT) is that each
from the original dataset so
e selected in each iteration
DHYA.COM/ABOUT-ME/CAREER-A
01 2 and 3

(https://www.analyticsvidhya.com/campus-ambassador-program (https://www.analyticsvidhya.com/campus-ambassador/)) /data-engineering-talent-hunt-hackathon RECENT POSTS

Hawking's Bold P

Stephen Hawking Predicts "Big VIBHVAJ) the History of Civilization The Motley Fool

RECENT POSTS

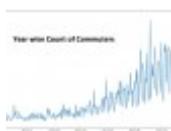


Choose one out of these
two?

(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVhbsidebottom)

You see that even though Row 2 is chosen from the data to the bootstrap sample, it's still present in the data. Now, each of the three:

Data	Bootstraped Sample
Row 1	Row 2
Row 2	Row 1
Row 3	



Natural Language Processing for Beginners: Using TextBlob
[\(https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/\)](https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/)

Rows have the same probability of being selected again.
 Let's say we choose Row 1 this time.

(<https://www.analyticsvidhya.com>)

Again, each row in the **DATA** has the same probability to be



selected. Let's say we randomly
 Row 2
 Row 1
 Row 1

(<https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/>)

7 methods to perform Time Series forecasting (with Python codes)
[\(https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/\)](https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/)

GURCHETAN SINGH ,...

(<https://www.analyticsvidhya.com/campus-ambassador-program>)

/data-engineering-talent-hunt-hackathon



strapped samples from the population. These multiple bootstrapped samples are used to fit trees for each of these samples. Finally, we use the majority vote or average of all the trees to make the final prediction. This is how

Random Forest works. The idea is that it's done mainly to reduce variance. A Random Forest actually uses this technique to further reduce the variance.

(<https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest>)

?utm_source=AVblog_sidebottom)



(<https://www.analyticsvidhya.com/blog/2018/02/introductory-guide-regularized-greedy-forests-rgf-python/>)

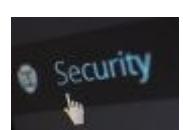
An Introductory Guide to Regularized Greedy Forests (RGF) with a case study in Python
[\(https://www.analyticsvidhya.com/blog/2018/02/introductory-guide-regularized-greedy-forests-rgf-python/\)](https://www.analyticsvidhya.com/blog/2018/02/introductory-guide-regularized-greedy-forests-rgf-python/)

ANKIT CHAUDHARY ...

2. Boosting: Boosting is a sequential technique in which, the first algorithm is trained on the entire dataset and the subsequent algorithms are built by fitting the residuals of the first algorithm, thus giving higher weight to those observations that were poorly predicted by the previous model.

It relies on creating a series of weak learners each of which might not be good for the entire dataset but is good for some part of the dataset. Thus, each model actually boosts the performance of the ensemble.

It's really important to note that boosting is focused on



reducing the bias. This makes the boosting algorithms prone to overfitting. Thus, parameter tuning becomes a crucial part of boosting algorithms to make them avoid overfitting.

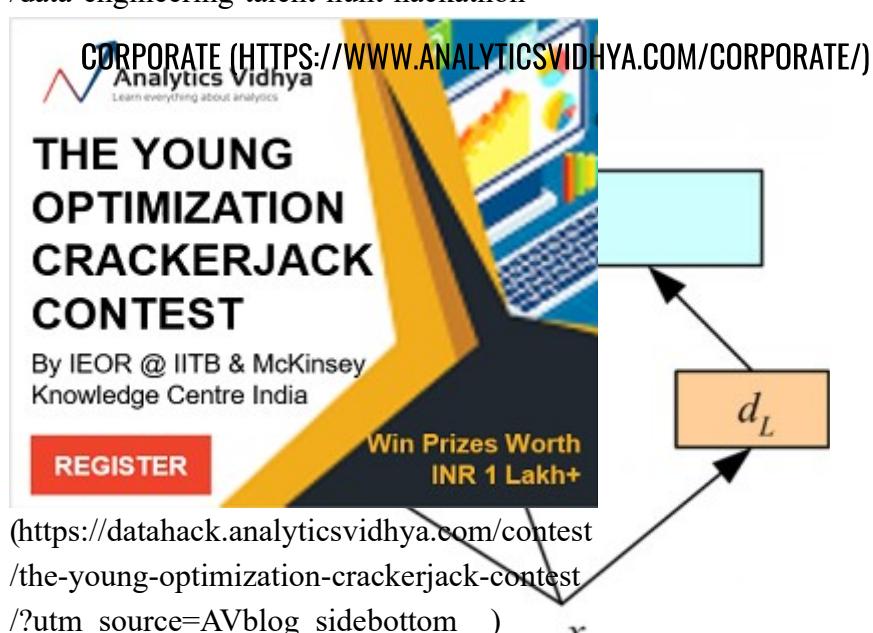
[\(https://www.analyticsvidhya.com/learn/101/engaging-with-data-science/\)](https://www.analyticsvidhya.com/learn/101/engaging-with-data-science/)

Demystifying Information Security Using Data Science (<https://www.analyticsvidhya.com/blog/2018/02/demystifying-security-data-science/>)

GUEST BLOG FEBR



(<https://www.analyticsvidhya.com/campus-ambassador-program>) Let's understand it with an example:



Here, we have two layers of machine learning models:

- Bottom layer models (d_1, d_2, d_3) which receive the original input features(x) from the dataset.
- Top layer model, $f()$ which takes the output of the bottom layer models (d_1, d_2, d_3) as its input and predicts the final output.
- One key thing to note here is that out of fold predictions are used while predicting for the training data.

Here, we have used only two layers but it can be any number of layers and any number of models in each layer.



Ama
S



http://www.edvancer.in/certified-data-scientist-with-python-course?utm_source=AV&utm_medium=AVads&

Two of the key principles for selecting the models:
Analytics Vidhya [HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com/)



- Ensembling makes the model more robust and stable thus ensuring decent performance on the test cases in most scenarios

- You can use ensembling to capture linear and simple as well non-linear complex relationships in the data. This can be done by using two different models and forming an ensemble of two.

3.2 Disadvantages

- Ensembling reduces the model interpretability and makes it very difficult to draw any crucial business insights at the end.
 - It is time-consuming and thus might not be the best idea for real-time applications.
 - The selection of models for creating an ensemble is an art

utm_campaign=AVadsnonfc&
LEARN ENGAGE COMPETE 

GET CONNECTED

 14,228 FOLLOWERS

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

42,813
Followers

(<http://www.facebook.com/YAGIUNIVERSITYAMBASSADOR/>)
2,547
FOLLOWERS

(<https://plus.google.com>
/+AnalyticsEmailhya)
 [SUBSCRIBE](#)

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

which is really hard to master.



[HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com/) LEARN ENGAGE COMPETE

(<https://www.analyticsvidhya.com>)

A Practical guide to Implementing



[AVBYTES \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/\)](https://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)

[Data Engineering](#)

[CONTACT US \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT-US/\)](https://WWW.ANALYTICSVIDHYA.COM/CONTACT-US/)

[Hackathon](#)

[WE ARE HIRING! \(HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/\)](https://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

the basics of ensembling, then

([YOU CAN CALL THEM TO GET DOWN TO THE GUTS OF THIS RESOURCE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR-PROGRAM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/\)\)](https://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR-PROGRAM))

(<https://www.analyticsvidhya.com/blog/2015/08/introduction-data-engineering-talent-hunt-hackathon>)

[CORPORATE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/\)](https://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

[REGISTER](#)

[Win Prizes Worth
INR 1 Lakh+](#)

grasp on ensembling

get down to

whether it can help us

in building the model

the basics of ensembling, then

you can read more about the

AnalyticsVidhya.com/contest

(ii/).

training various individual

models, it's the goto package for modeling in R. Don't worry if

you are not familiar with the caret package, you can get

through this article (<https://www.analyticsvidhya.com>

?utm_source=AVblog_sidebottom

) (<https://www.analyticsvidhya.com>

/blog/2016/12/practical-guide-to-implement-machine-

learning-with-caret-package-in-r-with-practice-problem/) to

get the comprehensive knowledge of caret package. Let's get

done with getting the data and data cleaning part.

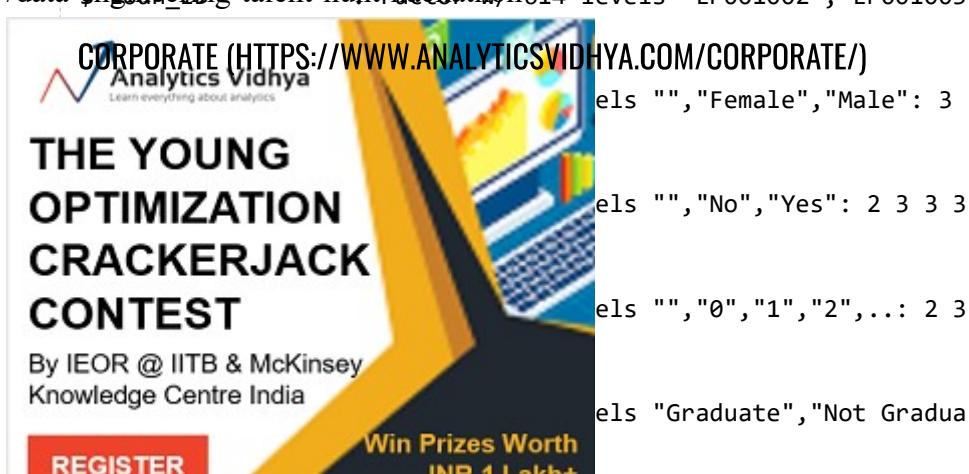
```
#Loading the required libraries
library('caret')
library('dplyr')
```

```
(https://www.analyticsvidhya.com )
#Setting the random seed
GET HIRED
```



LEARN ▾ ENGAGE ▾ COMPETE ▾

[ANALYTICS VIDHYA CAMPUS AMBASSADOR PROGRAM](https://www.analyticsvidhya.com/campus-ambassador-program) (<https://www.analyticsvidhya.com/campus-ambassador/>)
[DATA ENGINEERING TALENT HUNT HACKATHON](https://www.analyticsvidhya.com/data-engineering-talent-hunt-hackathon) (<https://www.analyticsvidhya.com/data-engineering-talent-hunt-hackathon>) 614 levels "LP001002", "LP001003



<https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest>

?utm_source=AVblog_mainbottom5849 4583 3000 2583 6000 5417 2333 3
 036 4006 12841 ...

```
$ Self_Employed: Factor w/ 3 levels "", "No", "Yes": 2 2 3 2
($ CoapplicantIncome: num 0 1508 0 2358 0 ...
$ LoanAmount : int NA 128 66 120 141 267 95 158 168 349
...
$ Loan_Amount_Term : int 360 360 360 360 360 360 360 360 360
360 ...
$ Credit_History : int 1 1 1 1 1 1 0 1 1 ...
$ Property_Area : Factor w/ 3 levels "Rural", "Semiurban", ..
: 3 1 3 3 3 3 2 3 2 ...
$ Loan_Status : Factor w/ 2 levels "N", "Y": 2 1 2 2 2 2 2
1 2 1 ...
```



HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/)

LEARN ▾ ENGAGE ▾ COMPETE ▾



#Does Analytics Vidhya missing values

Learn everything about analytics

(https://www.analyticsvidhya.com/)

GET HIRED ▾

[1] 86

ABInBev

AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)

Data Engineering

CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Hackathon

WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

(https://www.analyticsvidhya.com/campus-ambassador-program (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/))
 /data-engineering-talent-hunt-hackathon

CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)

Analytics Vidhya
Learn everything about analytics

**THE YOUNG
OPTIMIZATION
CRACKERJACK
CONTEST**

By IEOR @ IITB & McKinsey
Knowledge Centre India

REGISTER

Win Prizes Worth
INR 1 Lakh+

parts which I'll be using to

(https://datacrackerjack.analyticsvidhya.com/test/cooperations. We now define the
 /the training operation and other jacks to our and outcome variables:
 /?utm_source=AVblog_sidebottom)

#Defining the training controls for multiple models
fitControl <- trainControl(
(https://www.analyticsvidhya.com)
method = "cv",
GET HIRED ▾
number of folds = 10
ABInBev
AVBYTES ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/](https://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/))
Data Engineering
CONTACT US ([HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/](https://WWW.ANALYTICSVIDHYA.COM/CONTACT/))
Hackathon
WE ARE HIRING! ([HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/](https://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/))

([HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/](https://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/))
/data-engineering-talent-hunt-hackathon

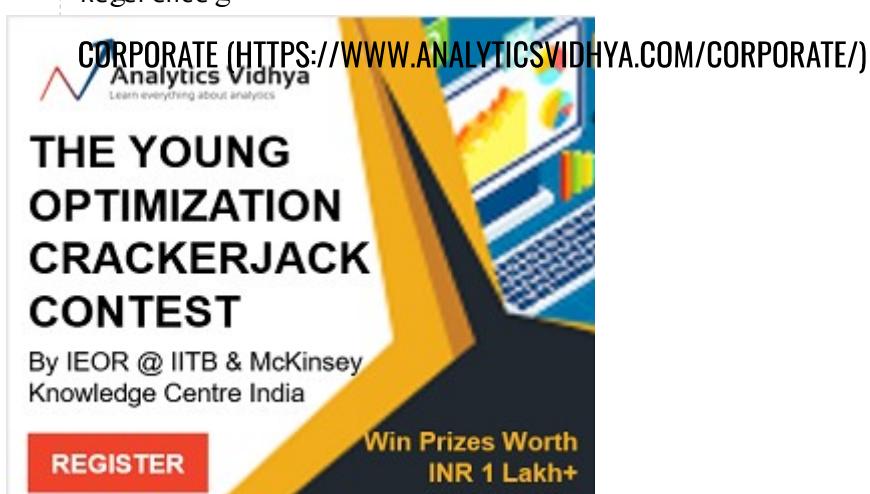
CORPORATE ([HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE](https://WWW.ANALYTICSVIDHYA.COM/CORPORATE)) test its
Analytics Vidhya Learn everything about analytics
e created:
THE YOUNG
OPTIMIZATION
CRACKERJACK
CONTEST
By IEOR @ IITB & McKinsey
Knowledge Centre India
REGISTER Win Prizes Worth
INR 1 Lakh+
Win Prizes Worth
INR 1 Lakh+

([HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/CONTEST/THE-YOUNG-OPTIMIZATION-CRACKERJACK-CONTEST/?UTM_SOURCE=AVBLOG_SIDEBOTTOM](https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_sidebottom))

```
#Training the random forest model
Analytics Vidhya HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM)
model_rf<-train(trainSet[,predictors],trainSet[,outcomeName],m
(https://www.analyticsvidhya.com)
ethod= rf, trControl=fitControl,tuneLength=3)
GET HIRED
```

LEARN ▾ ENGAGE ▾ COMPETE ▾ 

(https://www.analyticsvidhya.com/campus-ambassador/ /data-engineering-talent-hunt-hackathon)



(https://datahack.analyticsvidhya.com/contest /the-young-optimization-crackerjack-contest /?utm_source=AnalyticsVidhya_blog_sdebotom)

Neg Pred Value : 0.9143

Prevalence : 0.2418

Detection Rate : 0.1830

Detection Prevalence : 0.3137

Balanced Accuracy : 0.7922

'Positive' Class : N

Well, as you can see, we got 0.81 accuracy with the individual random forest model. Let's see the performance of KNN:

#Training the knn model

```
Analytics Vidhya HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/)  
model_knn<-train(trainSet[,predictors],trainSet[,outcomeName],  
(https://www.analyticsvidhya.com/),  
method=knn,trControl=fitControl,tuneLength=3)  
GET HIRED
```

LEARN ✓ ENGAGE ✓ COMPETE ✓ 



AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)
tel_knn,testSet[,predicto

CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)
Data Engineering
Hackathon

WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

(https://www.analyticsvidhya.com/campus-ambassador/
/data-engineering-talent-hunt-hackathon

CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

REGISTER

Win Prizes Worth
INR 1 Lakh+

(https://datahack.analyticsvidhya.com/contest
/the-young-optimization-crackerjack-contest
/?utm_source=AnalyticsVidhya_blog_sde6042018)

Neg Pred Value : 0.9810

Prevalence : 0.2026

Detection Rate : 0.1895

Detection Prevalence : 0.3137

Balanced Accuracy : 0.8899

'Positive' Class : N

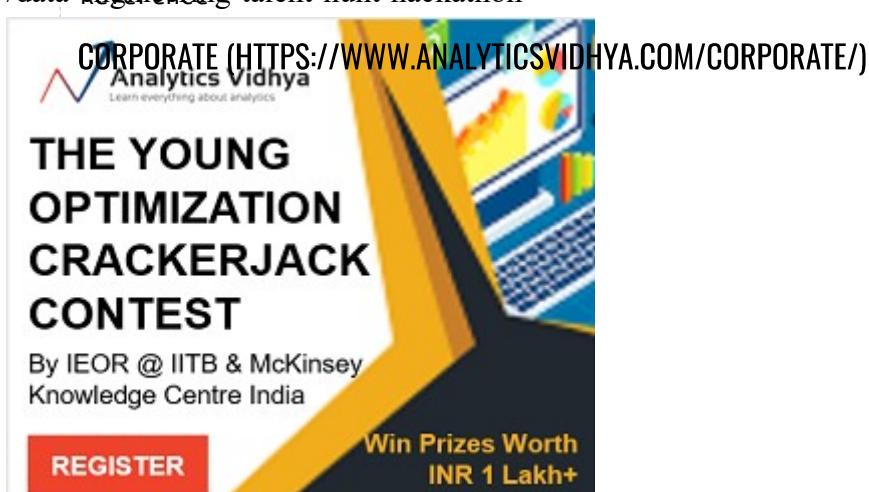
It's great since we are able to get 0.86 accuracy with the individual KNN model. Let's see the performance of Logistic regression as well before we go on to create ensemble of these three.

```
#Training the Logistic regression model
Analytics Vidhya HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/)

model_lr<-train(trainSet[,predictors],trainSet[,outcomeName],m
(https://www.analyticsvidhya.com/...ethod=glm,trControl=fitControl,tuneLength=3)
GET HIRED
```

LEARN ▾ ENGAGE ▾ COMPETE ▾ 

(https://www.analyticsvidhya.com/campus-ambassador/.../data-engineering-talent-hunt-hackathon)



(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AnalyticsVidhya_blog_sdebotom)

Neg Pred Value : 0.9810

Prevalence : 0.2026

Detection Rate : 0.1895

Detection Prevalence : 0.3137

Balanced Accuracy : 0.8899

'Positive' Class : N

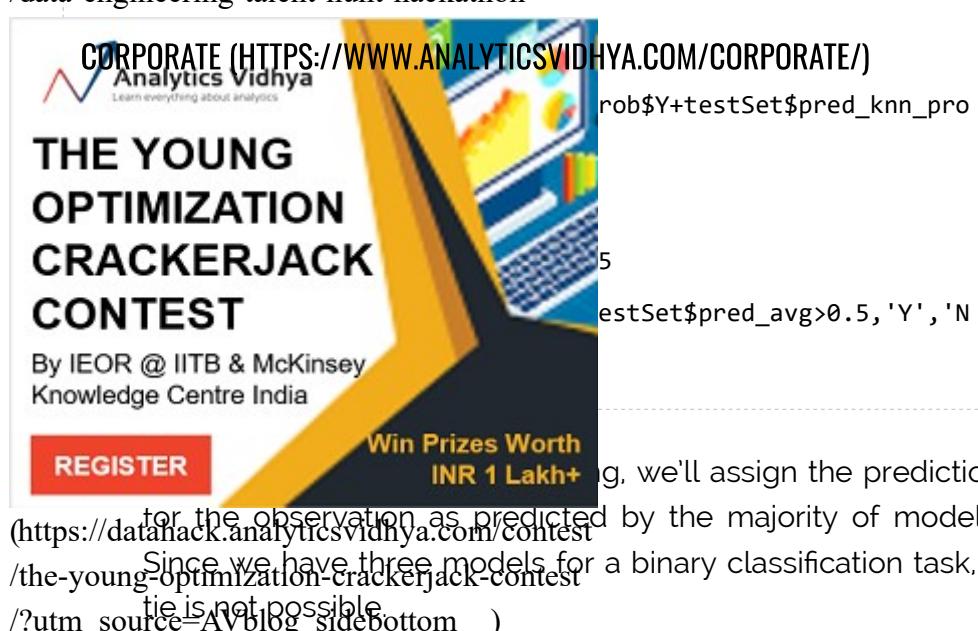
And the logistic regression also gives us the accuracy of 0.86.

Now, let's try out different ways of forming an ensemble with these models as we have discussed:

- **Averaging:** In this, we'll average the predictions from the three models. Since the predictions are either 0 or 1, averaging doesn't make much sense for this binary classification.
(<https://www.analyticsvidhya.com>) However, we can do averaging on the probabilities of observations to be in either of those binary classes.



(<https://www.analyticsvidhya.com/campus-ambassador/>)
[/data-engineering-talent-hunt-hackathon](https://www.analyticsvidhya.com/campus-ambassador/)



```
#The majority vote

testSet$pred_majority<-as.factor(ifelse(testSet$pred_rf=='Y' &
testSet$pred_knn=='Y','Y',ifelse(testSet$pred_rf=='Y' & testSe
t$pred_lr=='Y','Y',ifelse(testSet$pred_knn=='Y' & testSet$pred
_lr=='Y','Y','N'))))
```

- **Weighted Average:** Instead of taking simple average, we can take weighted average. Generally, the weights of predictions are high for more accurate models. Let's assign 0.5 to logistic regression and 0.25 to KNN and random forest each.

#Taking weighted average of predictions
Analytics Vidhya [HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com/)
`testSet$pred_weighted_avg<- (testSet$pred_rf_prob$Y*0.25)+(testSet$pred_knn_prob$Y*0.25)+(testSet$pred_lr_prob$Y*0.5)`
[GET HIRED](#)

LEARN ▾ ENGAGE ▾ COMPETE ▾ 



We are you to recall about two discussed on individual prediction correlation which

[WE ARE HIRING! \(HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/\)](https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/)

[\(HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/\)](https://www.analyticsvidhya.com/campus-ambassador-program)

/data-checking-for-the-correlation-between-the-predictions-of-the

[CORPORATE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/\)](https://www.analyticsvidhya.com/corporate)
 Learn everything about analytics

on these three models for a the predictions are highly might not give better results the point. Right?

as at the top layer. Instead, learning model which is n use linear regression for making the predictions in g bottom layer model

[predictions to the outcome or logistic regression similarly in /the-young-optimization-crackerjack-contest](https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest)
 case of classification problem.
 /?utm_source=AVblog_sidebottom)

Moreover, we don't need to restrict ourselves here, we can also use more complex models like GBM, neural nets to develop a non-linear mapping from the predictions of bottom layer models to the outcome.

On the same example let's try applying logistic regression and GBM as top layer models. Remember, the following steps that we'll take:

1. Train the individual base layer models on training data.
2. Predict using each base layer model for training data and test data.

3. Now train the top layer model again on the predictions of the bottom layer models that has been made on the training data.
- HOME (<https://www.analyticsvidhya.com/>) LEARN ◁ ENGAGE ◁ COMPETE ◁ 
4. Finally, predict using the top layer model with the predictions (<https://www.analyticsvidhya.com/>) of bottom layer models that has been made for testing data.
- GET HIRED**



(<https://www.analyticsvidhya.com/campus-ambassador/>)
(<https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/>)
DATA SCIENCE BLOGS (<https://www.analyticsvidhya.com/blogs/>)
BLOG CATEGORIES (<https://www.analyticsvidhya.com/categories/>)
CONTACT US (<https://www.analyticsvidhya.com/contact/>)
WE ARE HIRING! (<https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/>)

Note in step 2 is that you

will be able to predict only the

layer models will only be

able to recall the

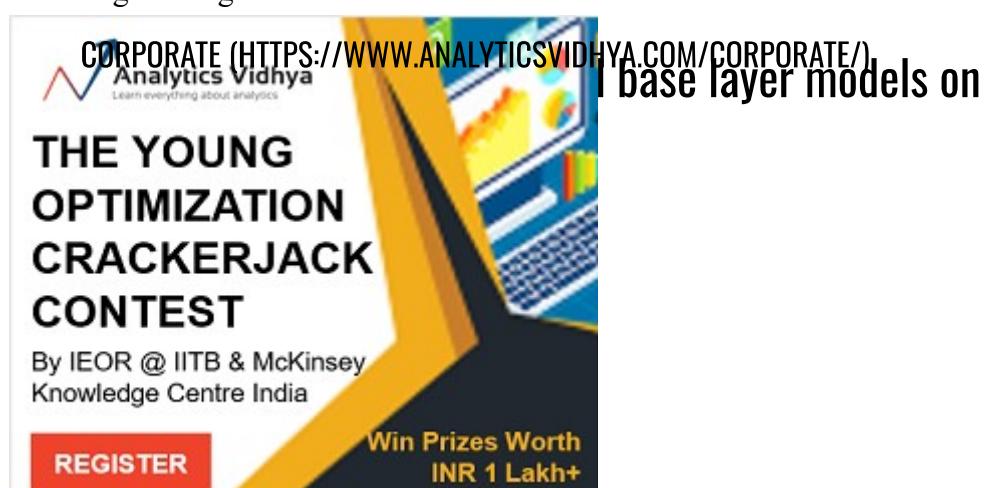
earlier model can recall the

one again.

already done previously, but

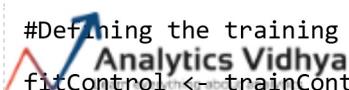
one again.

(<https://www.analyticsvidhya.com/campus-ambassador/>)
(<https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/>)
/data-engineering-talent-hunt-hackathon



(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_sidebottom)

#Defining the training control



HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/) LEARN ENGAGE COMPETE SEARCH

```
fitControl<-trainControl(
  method = "cv",
  number = 10)
```

(https://www.analyticsvidhya.com) **GET HIRED** ▾

number = 10

ABInBev **AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)**

Data Engineering **CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)**

Hackathon

WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

(https://www.analyticsvidhya.com/campus-ambassador-program,(https://www.analyticsvidhya.com/campus-ambassador/)/data-engineering-talent-hunt,hackathon

CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)



Learn everything about analytics

**THE YOUNG
OPTIMIZATION
CRACKERJACK
CONTEST**

By IEOR @ IITB & McKinsey Knowledge Centre India

REGISTER **Win Prizes Worth INR 1 Lakh+**

(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AnalyticsVidhya&utm_medium=banner&utm_campaign=CrackerjackContest) **model_lr<-train(trainSet[,predictors],trainSet[,outcomeName],method='glm',trControl=fitControl,tuneLength=3)**

[,trainSet[,outcomeName],m
Length=3

[,trainSet[,outcomeName],
s],trainSet[,outcomeName],

- Step 2: Predict using each base layer model for training data and test data



#Predicting the out of fold prediction probabilities for train

HOME (<https://www.analyticsvidhya.com/>) LEARN ▾ ENGAGE ▾ COMPETE ▾

([https://www.analyticsvidhya.com/trainSet\\$0OF_pred_rf\\$<-model_rf\\$pred\\$Y\[order\(model_rf\\$pred\\$rowI...](https://www.analyticsvidhya.com/trainSet$0OF_pred_rf$<-model_rf$pred$Y[order(model_rf$pred$rowI...))

GET HIRED ▾

AVBYTES (<https://www.analyticsvidhya.com/blog/category/abytes/>)CONTACT US (<https://www.analyticsvidhya.com/contact/>)

Hackathon

WE ARE HIRING! (<https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/>)

testSet[predictors],type

(<https://www.analyticsvidhya.com/campus-ambassador-program/>)

(<https://www.analyticsvidhya.com/data-engineering-talent-hunt-hackathon/>)



Learn everything about analytics

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

REGISTER

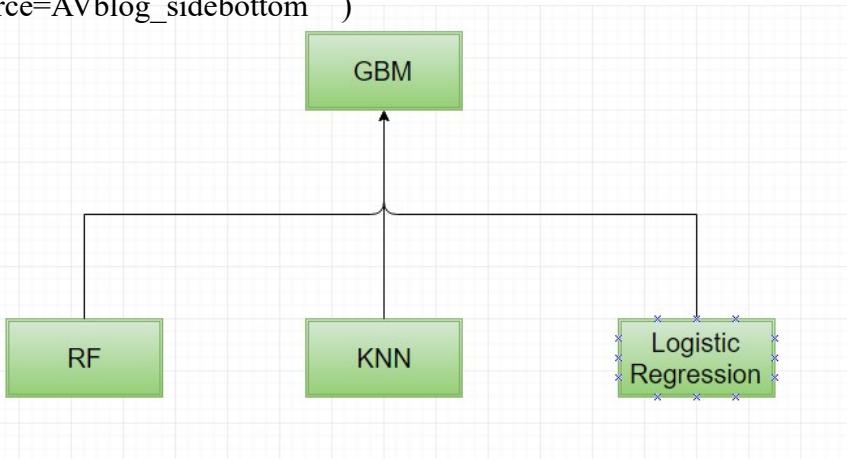
Win Prizes Worth
INR 1 Lakh+

ayer model again on
om layer models that
ning data

(<https://datahack.analyticsvidhya.com/contest/>)

First, let's start with the GBM model as the top layer model.

[/the-young-optimization-crackerjack-contest
/?utm_source=AVblog_sidebottom](https://www.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_sidebottom))



#Predictors for top layer models
Analytics Vidhya HOME (<https://www.analyticsvidhya.com/>)
 predictors_top<-c('OOF_pred_rf','OOF_pred_knn','OOF_pred_lr')
 (https://www.analyticsvidhya.com)
 GET HIRED ▾
 #CRM as top layer model

LEARN ▾ ENGAGE ▾ COMPETE ▾ 

(<https://www.analyticsvidhya.com/campus-ambassador-program>)
 /data-engineering-talent-hunt-hackathon

CORPORATE (<https://www.analyticsvidhya.com/corporate/>)
 Analytics Vidhya
 Learn everything about analytics

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

REGISTERWin Prizes Worth
INR 1 Lakh+

Logistic
Regression

#Logistic regression as top layer model
 (https://datahack.analyticsvidhya.com/contest
 /the-young-optimization-crackerjack-contest
 /?utm_source=AVBlog&utm_medium=

```
model.glm<-  
train(trainSet[,predictors_top],trainSet[,outcomeName],method=  
'glm',trControl=fitControl,tuneLength=3)
```

- Step 4: Finally, predict using the top layer model with the predictions of bottom layer models that has been made for testing data

```
#predict using GBM top layer model
Analytics Vidhya HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM)
testSet$gbm_stacked<-predict(model_gbm,testSet[,predictors_top
(https://www.analyticsvidhya.com )
]} GET HIRED ▾
```

LEARN ▾ ENGAGE ▾ COMPETE ▾ 

use the models for the ensemble will do get the best out of the ensemble. The two thumb rules that we discussed will greatly help you in that.



an in-depth conceptual as
ensembling. I would like to
machine learning hackathons
you can find here
[m/contest/all/](https://datahack.analyticsvidhya.com/contest/m/contest/all/).

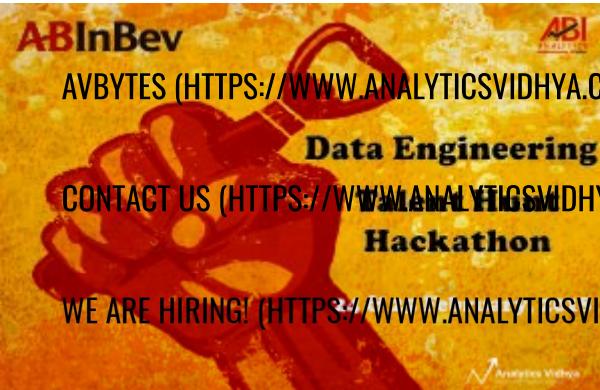
(<https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest>) You'll probably find this article (<https://www.analyticsvidhya.com/blog/2015/09/questions-ensemble-modeling/>) on top five questions related to ensembling helpful.

Also, if you missed out on the skilltest on ensembling, you can check your understanding of ensembling concepts here (<https://www.analyticsvidhya.com/blog/2017/02/40-questions-to-ask-a-data-scientist-on-ensemble-modeling-techniques-skilltest-solution/>).

End Notes



Ensembling is a very popular and effective technique that is very frequently used by data scientists for beating the accuracy benchmark of even the best of individual algorithms. More often than not it's the winning recipe in hackathons. The more



admire its beauty.

- AVBYTES ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/](https://www.analyticsvidhya.com/blog/category/abytes/))
 - o share your views in the comments below / any doubts / questions
- CONTACT US ([HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/](https://www.analyticsvidhya.com/contact/))

[AnalyticsVidhya.com](https://www.analyticsvidhya.com)

([HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR-PROGRAM/](https://www.analyticsvidhya.com/campus-ambassador-program/))
([HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/](https://datahack.analyticsvidhya.com/))

[AnalyticsVidhya.com/jobs/#](https://datahack.analyticsvidhya.com/jobs/#)

7/02/introduction-to-ensembling-along-with-implementation-in-r?share=facebook&nb=1) 605

7/02/introduction-to-ensembling-along-with-implementation-in-r?share=google-plus-1&nb=1) 332

([HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/CONTEST-GPLUS/](https://datahack.analyticsvidhya.com/contest-gplus/))
G+ ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/?SHARE=GOOGLE-PLUS-1&NB=1](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/?share=google-plus-1&nb=1))

?utm_source=AVblog_sidebottom)

Twitter ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/?SHARE=TWITTER&NB=1](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/?share=twitter&nb=1))

Pocket ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/?SHARE=POCKET&NB=1](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/?share=pocket&nb=1))

Reddit ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/?SHARE=REDDIT&NB=1](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/?share=reddit&nb=1))

COULD A ROBOT BECOME
YOUR BFF?





LEARN ▴ ENGAGE ▴ COMPETE ▴

[\(https://www.analyticsvidhya.com/blog/category/avbytes/\)](https://www.analyticsvidhya.com/blog/category/avbytes/)on-
/12/improve-machine-
learning-results/)CONTACT US (<https://www.analyticsvidhya.com/contact/>)WE ARE HIRING! (<https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/>)from (<https://www.analyticsvidhya.com/campus-ambassador-program/>)

/dataengineering/2016-hunt-the-ultimate-student

[\(https://www.analyticsvidhya.com/campus-ambassador/\)](https://www.analyticsvidhya.com/campus-ambassador/)

dhya.com/blog/2015

/12/improve-machine-
learning-results/)

December 29, 2015

In "Machine Learning"

TAGS:

RELATED

ADVANTAGES OF ENSEMBLE

By IEOR @ IITB & McKinsey
Knowledge Centre IndiaWin Prizes Worth
INR 1 Lakh+

(<https://www.analyticsvidhya.com/blog/tag/advantages-of-ensemble/>), BAGGING
(<https://www.analyticsvidhya.com/blog/tag/bagging/>), BOOSTING
(<https://www.analyticsvidhya.com/blog/tag/boosting/>), ENSEMBLE MODELING
(<https://www.analyticsvidhya.com/blog/tag/ensemble-modeling/>), ENSEMBLE
(<https://www.analyticsvidhya.com/blog/tag/ensemble/>), ENSEMBLE MODELING IN R
(<https://www.analyticsvidhya.com/blog/tag/ensemble-modeling-in-r/>)



HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/) LEARN ENGAGE COMPETE

[40 Questions to ask a Data Scientist on GET HIRED](https://www.analyticsvidhya.com/40-questions-to-ask-a-data-scientist-on-get-hired/)

Top 28 Cheat Sheets for

Machine Learning, Data

Science, Probability, SQL

& Big Data

AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)

[/blog/2017/02/top-28-cheat-sheets-for-machine-learning-data-science-probability-sql-big-data/](https://www.analyticsvidhya.com/blog/2017/02/top-28-cheat-sheets-for-machine-learning-data-science-probability-sql-big-data/)

CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Hackathon

WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

[/sql-big-data/](https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/sql-big-data/)

(https://ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/PROGRAM (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/)/data-engineering-talent-hunt-hackathon



CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

[REGISTER](#)

Win Prizes Worth
INR 1 Lakh+

(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_SauravKaushik)

Saurav Kaushik is a Data Science enthusiast, currently in the final year of his graduation at MAIT, New Delhi. He loves to use machine learning and analytics to solve complex data problems.

✉ (mailto:sauravkaushik8@gmail.com)

🌐 (<http://seasoninsights.blogspot.in/>)

ƒ (<https://www.facebook.com/sauravkaushik8>)

🐦 (<https://twitter.com/sauravkaushik8>)

📷 (<https://www.instagram.com/sauravkaushik8/>)



(https://www.analyticsvidhya.com)
GET HIRED

LEARN ▾ ENGAGE ▾ COMPETE ▾ SEARCH



Data Engineering
Hackathon

[CONTACT US](https://www.analyticsvidhya.com/contact/) ([HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/](https://www.analyticsvidhya.com/contact/))

might not get a prompt
 request you to post this
<https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/>
 to get your queries

(https://www.analyticsvidhya.com/campus-ambassador-program) ([HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/](https://www.analyticsvidhya.com/campus-ambassador/))
 /data-engineering-talent-hunt-hackathon

78 TITMFFNTS



INTRODUCTION-TO-ENSEMBLING-ALONG-
[/WWW.ANALYTICSVIDHYA.COM
 EMBLING-ALONG-WITH-IMPLEMENTATION-](https://www.analyticsvidhya.com/introduction-to-ensembling-along-with-implementation-in-r/#comment-122702)

explained!

REGISTER

Win Prizes Worth
 INR 1 Lakh+

(https://datahack.analyticsvidhya.com/replies/1674599#comment-122702)
 REPLY ([HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/REPLIES/1674599#COMMENT-122702#REPLY](https://datahack.analyticsvidhya.com/replies/1674599#comment-122702))
 /the-young-optimization-crackerjack-contest/
 ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122702](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122702))
 /?utm_source=AVblog_sidebottom

Thanks!



Ankit Gupta (https://www.facebook.com/app_scoped_user_id/1107394225967405/) says:

FEBRUARY 15, 2017 AT 6:11 AM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122698](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122698))

Nicely written!



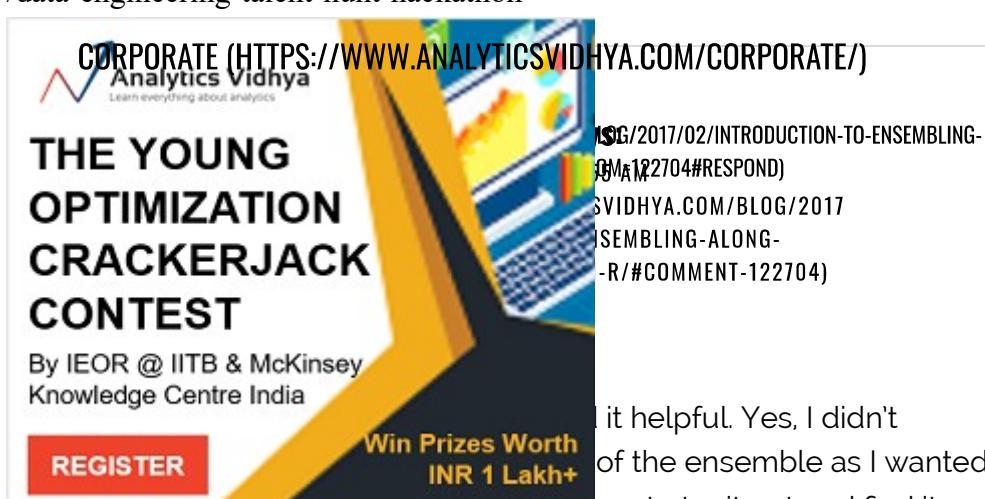
Saurav Kaushik says:
 REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-
 ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122703)
 HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017
 (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-
 WITH-IMPLEMENTATION-IN-R/#COMMENT-122703)
GET ENRAGED

LEARN ▾ ENGAGE ▾ COMPETE ▾



ed it!
[AVBYTES \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/\)](https://www.analyticsvidhya.com/blog/category/abytes/)
Data Engineering
[CONTACT US \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/\)](https://www.analyticsvidhya.com/contact/)
Hackathon
[WE ARE HIRING! \(HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/\)](https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/)

(https://www.analyticsvidhya.com/campus-ambassador-program (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/)
 /data-engineering-talent-hunt-hackathon



[CORPORATE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/\)](https://www.analyticsvidhya.com/corporate/)
**THE YOUNG
OPTIMIZATION
CRACKERJACK
CONTEST**
 By IEOR @ IITB & McKinsey
 Knowledge Centre India

REGISTER

**Win Prizes Worth
INR 1 Lakh+**

it helpful. Yes, I didn't
 of the ensemble as I wanted
 to encourage readers to try it out and find it
 themselves. Whether it gave better
 performance than any individual model. If it did,
 great. But if it didn't, then you'll need to think
 about why it didn't and what could be done to
 overcome it by thinking on the lines of the
 important criterias for ensembling that I have
 mentioned.

I think this curiosity will make you try it as well!

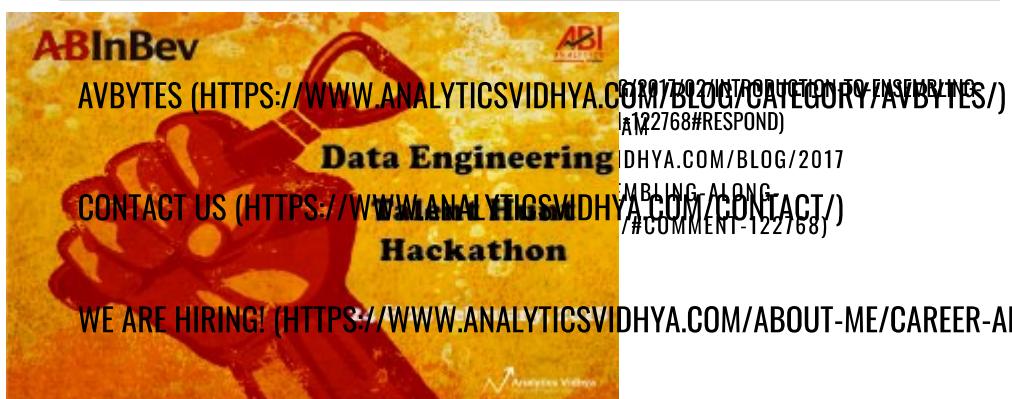
Best,
 Saurav.

Kumar (http://hadooptraininginhyderabad.com/) says:
 REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-
 ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122715)
 HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017
 (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-
 WITH-IMPLEMENTATION-IN-R/#COMMENT-122715)
GET ENRAGED

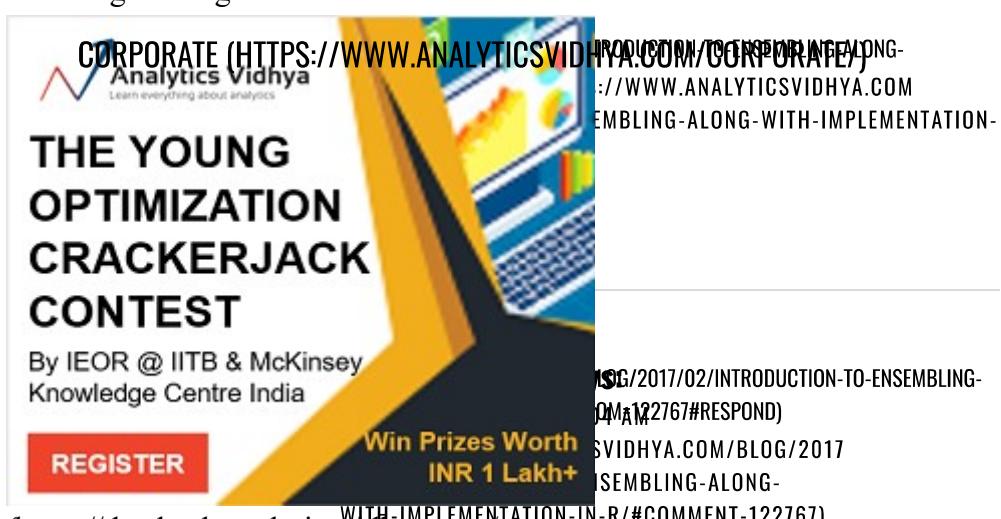
[/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122715](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122715)) [HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com) LEARN ▾ ENGAGE ▾ COMPETE ▾ [SEARCH](#)

Learn everything about analytics
Simply written and easy to understand.
(<https://www.analyticsvidhya.com>)

[GET HIRED](#) ▾



(<https://www.analyticsvidhya.com/campus-ambassador-program>) (<https://www.analyticsvidhya.com/campus-ambassador/>) /data-engineering-talent-hunt-hackathon



Thanks. I'm glad you found this helpful.

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#REPLY?REPLTOPCOM=122768#RESPOND) (<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122768>)

Gonzalo Moreno says: thank you. Very good. However I would like to check the accuracy of the ensembled model and you did not show it.



(https://www.analyticsvidhya.com/campus-ambassador-program (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/)/data-engineering-talent-hunt/hackathons) important criterias for ensembling that I have



(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_sidebottom) I really wanted to compliment in being so transparent in sharing this knowledge and learning platform. You are strengthening the often overlooked notion when in reality sharing knowledge makes us better.

I am an avid reader of your posts as it helps me understand the changing world of advanced analytics and have informed discussions with data scientists.

Thanks again
Yogi





(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122766)

(https://www.analyticsvidhya.com/)
GET HIRED ▾



WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

(https://www.analyticsvidhya.com/campus-ambassador-program (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/))
(BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122774)



By IEOR @ IITB & McKinsey
Knowledge Centre India

REGISTER

Win Prizes Worth
INR 1 Lakh+

(https://datahack.analyticsvidhya.com/contest
/the-young-optimization-crackerjack-contest

?utm_source=AVblog_sidebottom)
REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-
WITH-IMPLEMENTATION-IN-R/?REPLYTOCOM=122807#REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-
IN-R/#COMMENT-122807)

Thanks Saurav. This article came at the right time. I was looking for stacking and blending example and read many article on the subject but didn't find complete hands on example on R. So struggling thru but this article initialized me with the basic. That's why I love this site!!!!.

REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-
WITH-IMPLEMENTATION-IN-R/?REPLYTOCOM=122818#REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM
/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-

IN-R/#COMMENT-122813)



HOME (<https://www.analyticsvidhya.com/>) LEARN ▾ ENGAGE ▾ COMPETE ▾ [SEARCH](#)
 Analytics Vidhya, I have been searching for something like this to learn ensemble modeling and glad i found it.
[\(https://www.analyticsvidhya.com/\)](https://www.analyticsvidhya.com/) Kudos to your work! ▾ [GET HIRED](#) ▾



(<https://www.analyticsvidhya.com/campus-ambassador-program>)

/REPLY (<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122813>)

/REPLY (<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122813>)

CORPORATE (<https://www.analyticsvidhya.com/corporate/>)

Analytics Vidhya Learn everything about analytics

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

[REGISTER](#)

Win Prizes Worth
INR 1 Lakh+

about ensemble modelling. I

look at the disadvantages of
ensemble modelling in the sense like no use of
real time application an other

(<https://datahack.analyticsvidhya.com/contest>)

/the-young-optimization-crackerjack-contest

Correct me if I am wrong.

?utm_source=AVblog_sidebottom)

Regards Arun

Saurav Kaushik
 REPLY (<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122987>)

(<https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122987>)

Hi Arun,

See these are the practical challenges that
you're bound to face when you use
ensembling. It is definitely more time



consuming than using a single model.

[HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com/) LEARN ▾ ENGAGE ▾ COMPETE ▾

They way I see it is as a trade-off between
[\(https://www.analyticsvidhya.com/\)](https://www.analyticsvidhya.com/) accuracy and the training time. You can

GET HIRED

optimise the parameter search, use higher

etc to reduce the time if

[AVBYTES \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/\)](https://www.analyticsvidhya.com/blog/category/avbytes/)

you only good results

[CONTACT US \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/\)](https://www.analyticsvidhya.com/contact/)

el. This always boils down

old time you want to

use within for your real time

[WE ARE HIRING! \(HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/\)](https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/)

helpful.

(<https://www.analyticsvidhya.com/campus-ambassador/>)
[CAMPUS AMBASSADOR PROGRAM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/\)](https://www.analyticsvidhya.com/campus-ambassador/)

Best,
Saurav

[CORPORATE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/\)](https://www.analyticsvidhya.com/corporate/)

INTRODUCTION-TO-ENSEMBLING-ALONG-
[://WWW.ANALYTICSVIDHYA.COM
EMBLING-ALONG-WITH-IMPLEMENTATION-](https://www.analyticsvidhya.com/introduction-to-ensembling-along-with-implementation/)

you!!

could easily understand..

By IEOR @ IITB & McKinsey

Knowledge Centre India

[REGISTER](#)

Win Prizes Worth
INR 1 Lakh+

(<https://datahack.analyticsvidhya.com/contest>

/the-young-optimization-crackerjack-contest

/?utm_source=Avblog_sidebottom_)
[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-
ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-122990#RESPOND\)](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122990#respond)

([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017
/02/INTRODUCTION-TO-ENSEMBLING-ALONG-
WITH-IMPLEMENTATION-IN-R/#COMMENT-122990](https://www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-122990))

Hi Saurav,

I completely agree with you. There are more possibility to improve the accuracy using Ensemble modelling.

We are building model only to draw crucial business insights.

But as you mentioned in the disadvantages of



ensemble modelling "It very difficult to draw

[HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com/) LEARN ENGAGE COMPETE



any crucial business insights at the vidhya.com

(<https://www.analyticsvidhya.com/>) how ensemble modelling will benefits us,even have higher accuracy.

[GET HIRED](#)



[AVBYTES \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/\)](AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/))

Data Engineering

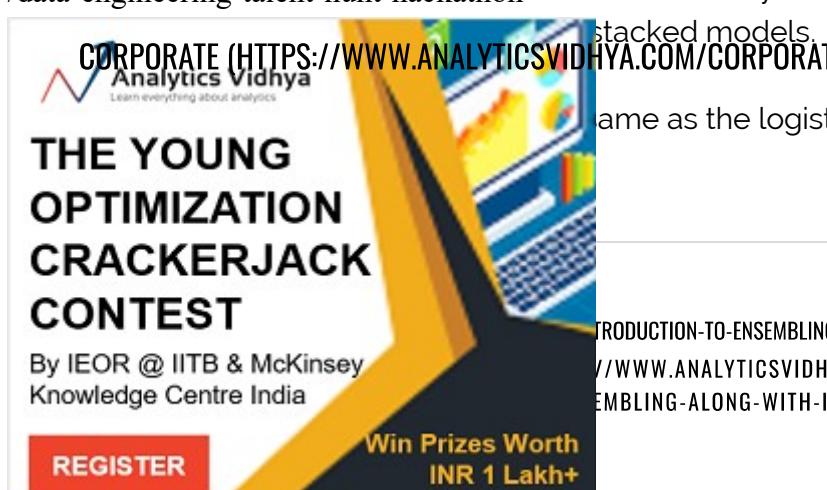
[CONTACT US \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/\)](CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/))

Hackathon

[WE ARE HIRING! \(HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/\)](WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/))

sample.

(<https://ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR>) [PROGRAM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/\)](https://ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/PROGRAM)
[/data-engineering-talent-hunt-hackathon](https://ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/) Well I find that KNN, in terms of accuracy, is the best



[CORPORATE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/\)](CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/))

stacked models,

same as the logistic

INTRODUCTION-TO-ENSEMBLING-ALONG-
[//WWW.ANALYTICSVIDHYA.COM](https://WWW.ANALYTICSVIDHYA.COM)
 EMBLING-ALONG-WITH-IMPLEMENTATION-

(<https://datahack.analyticsvidhya.com/contest>

/the-young-optimization-crackerjack-contest

Thank you for the article

?utm_source=AnalyticsVidhya

?utm_medium=bottom

use :

```
testSet$pred_rf_prob<-predict(object =
model_rf,testSet[,predictors],type='prob')
```

I am getting zero probabilities. Why it is giving zero probabilities and what it actually means?

**REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-
 WITH-IMPLEMENTATION-IN-R/2017/02/28/REPLY-123621#REPLY)** [www.analyticsvidhya.com/blog/2017/02/introduction-to-ensembling-along-with-implementation-in-r/#comment-123621](https://WWW.ANALYTICSVIDHYA.COM/BLOG/2017/02/INTRODUCTION-TO-ENSEMBLING-ALONG-WITH-IMPLEMENTATION-IN-R/#COMMENT-123621)

For the part : model_rf<-train(trainset[,
predictors],trainset[,outcomeName],method='rf',trControl=fi



I am getting error msg as :

Error in trainset[, predictors]: incorrect number of

(https://www.analyticsvidhya.com)

GET HIRED ▾

View 2



AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)

Data Engineering

CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Hackathon

WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

Learning overdose here..!

initially thought that a

(https://ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR-PROGRAM (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/)/data-engineering would do the job. And I just found out the reason that I



CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)

positions in the leaderboard by

to study some more I

very much for your wonderful

contribution of the topic, you have

THE YOUNG
OPTIMIZATION
CRACKERJACK
CONTEST

By IEOR @ IITB & McKinsey

Knowledge Centre India

REGISTER

Win Prizes Worth
INR 1 Lakh+

INTRODUCTION-TO-ENSEMBLING-ALONG-
/WW.ANALYTICSVIDHYA.COM/BLOG/2017
ONG-WITH-IMPLEMENTATION-

(https://datahack.analyticsvidhya.com/contest
Very Nicely explained
/the-young-optimization-crackerjack-contest
/?utm_source=AVblog_sidebottom)

LEAVE A REPLY

Your email address will not be published.

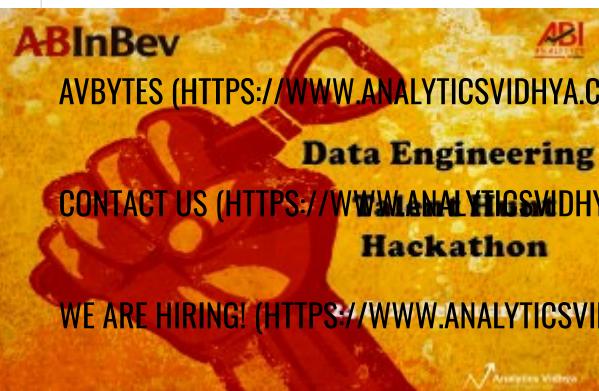


[HOME \(HTTPS://WWW.ANALYTICSVIDHYA.COM/\)](https://www.analyticsvidhya.com)

LEARN ▾ ENGAGE ▾ COMPETE ▾



(<https://www.analyticsvidhya.com>)
[GET HIRED](#) ▾



[WE ARE HIRING! \(HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/\)](https://www.analyticsvidhya.com/about-me/career-analytics-vidhya/)

(<https://www.analyticsvidhya.com/campus-ambassador>)
[/data-engineering-talent-hunt-hackathon](https://www.analyticsvidhya.com/campus-ambassador/)

[CORPORATE \(HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/\)](https://www.analyticsvidhya.com/corporate)

SUBMIT COM

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey
Knowledge Centre India

[REGISTER](#)

Win Prizes Worth
INR 1 Lakh+

(https://datahack.analyticsvidhya.com/contest/the-young-optimization-crackerjack-contest/?utm_source=AVblog_sidebottom)



Analytics Vidhya
Learn everything about analytics
(https://www.analyticsvidhya.com/)

GET HIRED
(http://www.analyticsvidhya.com/get-hired-us/)

ABInBev

AVBYTES (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/AVBYTES/)

CONTACT US (HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/)

Data Engineering Hackathon

WE ARE HIRING! (HTTPS://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/CAREER-ANALYTICS-VIDHYA/)

ANALYTICS VIDHYA CAMPUS AMBASSADOR PROGRAM (HTTPS://WWW.ANALYTICSVIDHYA.COM/CAMPUS-AMBASSADOR/)

WAITS FOR US

CORPORATE (HTTPS://WWW.ANALYTICSVIDHYA.COM/CORPORATE/)

THE YOUNG OPTIMIZATION CRACKERJACK CONTEST

By IEOR @ IITB & McKinsey Knowledge Centre India

REGISTER

Win Prizes Worth INR 1 Lakh+

HOME (HTTPS://WWW.ANALYTICSVIDHYA.COM/) **LEARN** **ENGAGE** **COMPETE** **JOIN OUR COMMUNITY :**

ANALYTICS VIDHYA

DATA SCIENTISTS COMPANIES

Blog [Post Jobs](https://www.analyticsvidhya.com/blog/)
(https://www.analyticsvidhya.com/blog/)

Hackathon [Trainings](https://www.analyticsvidhya.com/hackathon/)
(https://datahack.analyticsvidhya.com/hackathon/)

Discussions [Discussions](https://www.analyticsvidhya.com/discussions/)
(https://discuss.analyticsvidhya.com/)

Analytics [Hiring Hackathons](https://www.analyticsvidhya.com/analytics-jobs/)
(https://discuss.analyticsvidhya.com/hiring-hackathons/)

Impact [Advertising](https://www.analyticsvidhya.com/analytics-jobs/)
(https://www.analyticsvidhya.com/analytics-jobs/)

Jobs [Leaderboard](https://www.analyticsvidhya.com/jobs/)
(https://www.analyticsvidhya.com/jobs/)

Followers [Reach Us](https://www.analyticsvidhya.com/followers/)
(https://datahack.analyticsvidhya.com/contact/)

FOLLOWERS [Followers](https://www.analyticsvidhya.com/followers/)
(https://www.analyticsvidhya.com/followers/)

Leaderboard [Contact](https://www.analyticsvidhya.com/contact/)
(https://www.analyticsvidhya.com/contact/)

Reach Us [Contact](https://www.analyticsvidhya.com/contact/)
(https://www.analyticsvidhya.com/contact/)

Followers [Followers](https://www.analyticsvidhya.com/contact/)
(https://www.analyticsvidhya.com/contact/)

Followers [Followers](https://www.analyticsvidhya.com/contact/)
(https://www.analyticsvidhya.com/contact/)

Subscribe to emailer >

Policy (https://www.analyticsvidhya.com/privacy-policy/)
of Use (https://www.analyticsvidhya.com/terms/)
Refund Policy (https://www.analyticsvidhya.com/refund-policy/)

Don't have an account? Sign up (https://www.analyticsvidhya.com/sign-up/)