

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>)

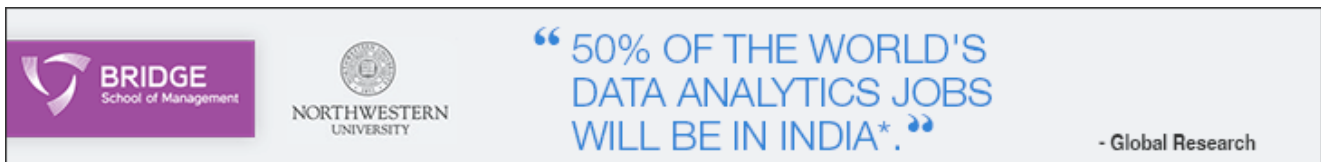
Home (<http://www.analyticsvidhya.com/>) > Business Analytics (<http://www.analyticsvidhya.com/blog/category/business-analytics/>)

Essentials of Machine Learning Algorithms (with Python and R Codes)

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

([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/PYTHON-2/](http://www.analyticsvidhya.com/blog/category/python-2/)) R ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/R/](http://www.analyticsvidhya.com/blog/category/r/))

om/sharer.php?u=<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>
 20Machine%20Learning%20Algorithms%20(with%20Python%20and%20R%20Codes)) **t** (<https://twitter.com/home?query=AnalyticsVidhya>)
 ne%20Learning%20Algorithms%20(with%20Python%20and%20R%20Codes)+<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>
 (<https://plus.google.com/share?url=http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>) **p**
 utton/?url=<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>
 yticsvidhya.com/wp-content/uploads/2015/08/NewI-Machine-Learning-Algorithms-
 tials%20of%20Machine%20Learning%20Algorithms%20(with%20Python%20and%20R%20Codes))



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

utm_source=AV&utm_medium=Banner&utm_campaign=AVBanner)

Introduction

Google's self-driving cars and robots get a lot of press, but the company's real future is in machine learning, the technology that enables computers to get smarter and more personal.

– Eric Schmidt (Google Chairman)

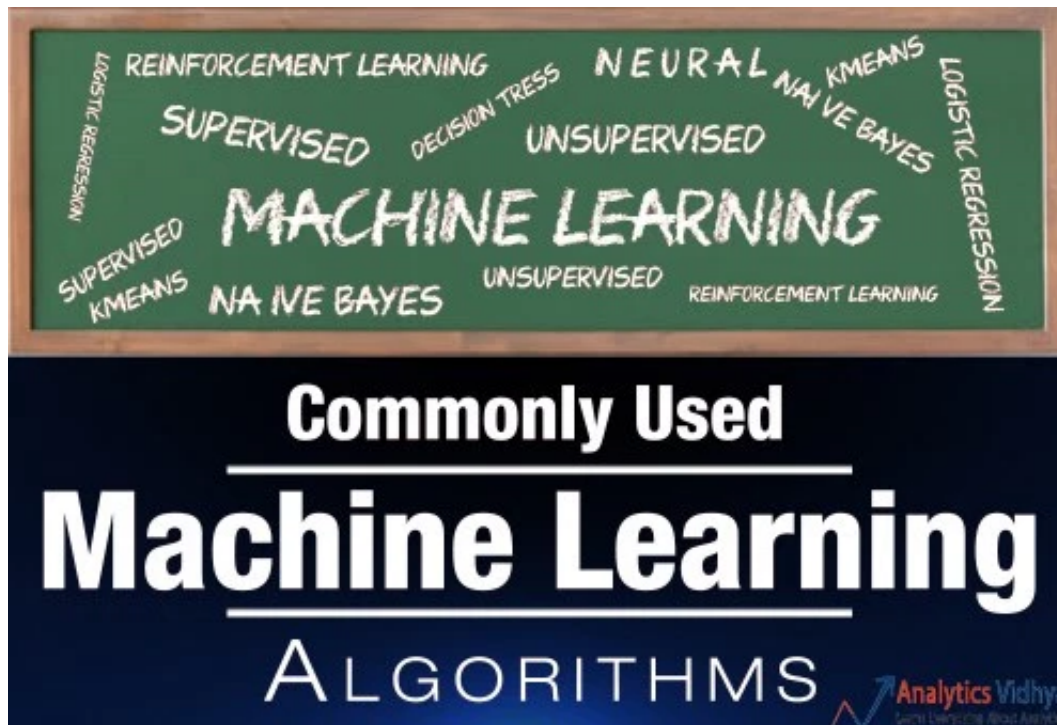
We are probably living in the most defining period of human history. The period when computing moved from large mainframes to PCs to cloud. But what makes it defining is not what has happened, but what is coming our way in years to come.

What makes this period exciting for some one like me is the democratization of the tools and techniques, which followed the boost in computing. Today, as a data scientist, I can build data crunching machines with complex algorithms for a few dollars per hour. But, reaching here wasn't easy! I had my dark days and nights.

Who can benefit the most from this guide?

What I am giving out today is probably the most valuable guide, I have ever created.

The idea behind creating this guide is to simplify the journey of aspiring data scientists and machine learning enthusiasts across the world. Through this guide, I will enable you to work on machine learning problems and gain from experience. **I am providing a high level understanding about various machine learning algorithms along with R & Python codes to run them. These should be sufficient to get your hands dirty.**



(<http://io.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/Newl-Machine-Learning-Algorithms.jpg>)

I have deliberately skipped the statistics behind these techniques, as you don't need to understand them at the start. So, if you are looking for statistical understanding of these algorithms, you should look elsewhere. But, if you are looking to equip yourself to start building machine learning project, you are in for a treat.

Broadly, there are 3 types of Machine Learning Algorithms..

1. Supervised Learning

How it works: This algorithm consist of a target / outcome variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, Decision Tree

(<http://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/>), Random Forest (<http://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/>), KNN, Logistic Regression etc.

2. Unsupervised Learning

How it works: In this algorithm, we do not have any target or outcome variable to predict / estimate. It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

3. Reinforcement Learning:

How it works: Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process

List of Common Machine Learning Algorithms

Here is the list of commonly used machine learning algorithms. These algorithms can be applied to almost any data problem:

1. Linear Regression
2. Logistic Regression
3. Decision Tree
4. SVM
5. Naive Bayes
6. KNN
7. K-Means
8. Random Forest

- 9. Dimensionality Reduction Algorithms
- 10. Gradient Boost & Adaboost

1. Linear Regression

It is used to estimate real values (cost of houses, number of calls, total sales etc.) based on continuous variable(s). Here, we establish relationship between independent and dependent variables by fitting a best line. This best fit line is known as regression line and represented by a linear equation $Y = a * X + b$.

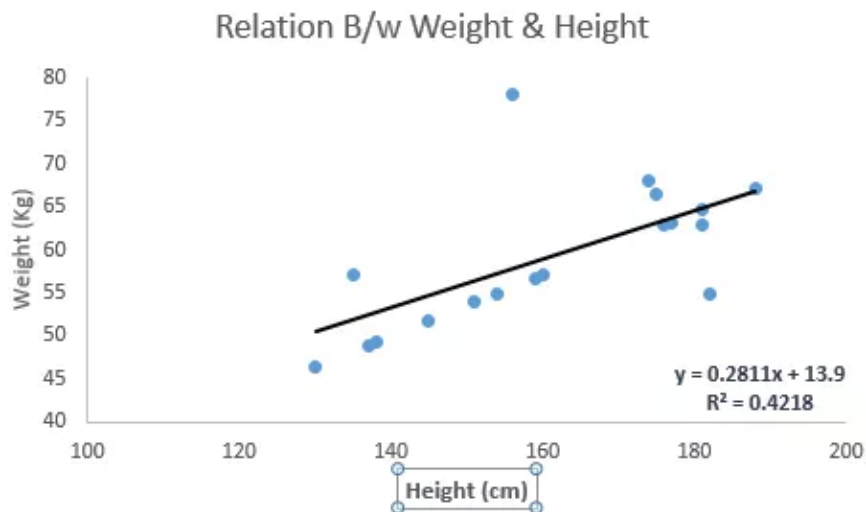
The best way to understand linear regression is to relive this experience of childhood. Let us say, you ask a child in fifth grade to arrange people in his class by increasing order of weight, without asking them their weights! What do you think the child will do? He / she would likely look (visually analyze) at the height and build of people and arrange them using a combination of these visible parameters. This is linear regression in real life! The child has actually figured out that height and build would be correlated to the weight by a relationship, which looks like the equation above.

In this equation:

- Y – Dependent Variable
- a – Slope
- X – Independent variable
- b – Intercept

These coefficients a and b are derived based on minimizing the sum of squared difference of distance between data points and regression line.

Look at the below example. Here we have identified the best fit line having linear equation **$y = 0.2811x + 13.9$** . Now using this equation, we can find the weight, knowing the height of a person.



(http://io.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/Linear_Regression.png)

Linear Regression is of mainly two types: Simple Linear Regression and Multiple Linear Regression. Simple Linear Regression is characterized by one independent variable. And, Multiple Linear Regression(as the name suggests) is characterized by multiple (more than 1) independent variables. While finding best fit line, you can fit a polynomial or curvilinear regression. And these are known as polynomial or curvilinear regression.

Python Code

```
#Import Library
#Import other necessary libraries like pandas, numpy...
from sklearn import linear_model
#Load Train and Test datasets
#Identify feature and response variable(s) and values must be numeric and numpy arrays
x_train=input_variables_values_training_datasets
y_train=target_variables_values_training_datasets
x_test=input_variables_values_test_datasets
# Create linear regression object
linear = linear_model.LinearRegression()
# Train the model using the training sets and check score
linear.fit(x_train, y_train)
linear.score(x_train, y_train)
#Equation coefficient and Intercept
print('Coefficient: \n', linear.coef_)
print('Intercept: \n', linear.intercept_)
#Predict Output
predicted= linear.predict(x_test)
```

R Code

```
#Load Train and Test datasets
#Identify feature and response variable(s) and values must be numeric and numpy arrays
x_train <- input_variables_values_training_datasets
y_train <- target_variables_values_training_datasets
x_test <- input_variables_values_test_datasets
x <- cbind(x_train,y_train)
# Train the model using the training sets and check score
linear <- lm(y_train ~ ., data = x)
summary(linear)
#Predict Output
predicted= predict(linear,x_test)
```

2. Logistic Regression

Don't get confused by its name! It is a classification not a regression algorithm. It is used to estimate discrete values (Binary values like 0/1, yes/no, true/false) based on given set of independent variable(s). In simple words, it predicts the probability of occurrence of an event by fitting data to a logit function (https://en.wikipedia.org/wiki/Logistic_function). Hence, it is also known as **logit regression**. Since, it predicts the probability, its output values lies between 0 and 1 (as expected).

Again, let us try and understand this through a simple example.

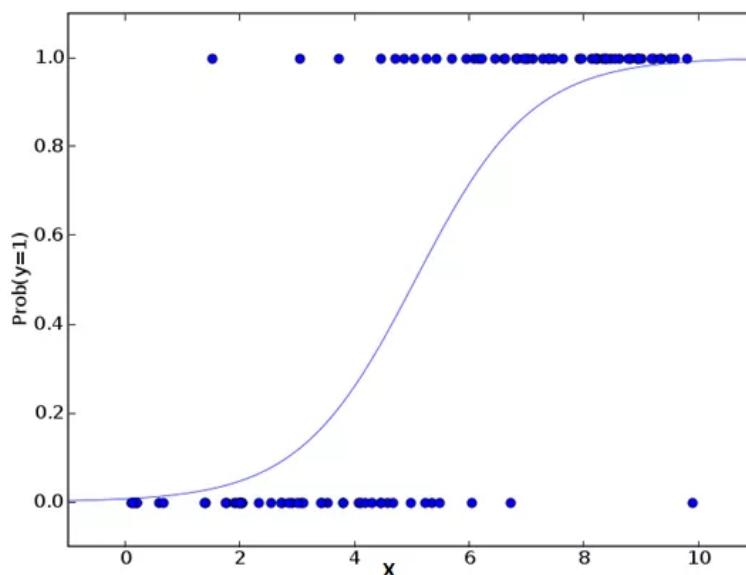
Let's say your friend gives you a puzzle to solve. There are only 2 outcome scenarios – either you solve it or you don't. Now imagine, that you are being given wide range of puzzles / quizzes in an attempt to understand which subjects you are good at. The outcome to this study would be something like this – if you are given a trigonometry based tenth grade problem, you are 70% likely to solve it. On the other hand, if it is grade fifth history question, the probability of getting an answer is only 30%. This is what Logistic Regression provides you.

Coming to the math, the log odds of the outcome is modeled as a linear combination of the predictor variables.

$$\text{odds} = p / (1-p) = \text{probability of event occurrence} / \text{probability of not event occurrence}$$
$$\ln(\text{odds}) = \ln(p/(1-p))$$
$$\text{logit}(p) = \ln(p/(1-p)) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

Above, p is the probability of presence of the characteristic of interest. It chooses parameters that maximize the likelihood of observing the sample values rather than that minimize the sum of squared errors (like in ordinary regression).

Now, you may ask, why take a log? For the sake of simplicity, let's just say that this is one of the best mathematical way to replicate a step function. I can go in more details, but that will beat the purpose of this article.



(http://i2.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/Logistic_Regression.png) **Python Code**

```
#Import Library
from sklearn.linear_model import LogisticRegression

#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset

# Create logistic regression object
model = LogisticRegression()

# Train the model using the training sets and check score
model.fit(X, y)

model.score(X, y)

#Equation coefficient and Intercept
print('Coefficient: \n', model.coef_)
print('Intercept: \n', model.intercept_)

#Predict Output
predicted= model.predict(x_test)
```

R Code

```
x <- cbind(x_train,y_train)

# Train the model using the training sets and check score
logistic <- glm(y_train ~ ., data = x,family='binomial')

summary(logistic)

#Predict Output
predicted= predict(logistic,x_test)
```

Furthermore..

There are many different steps that could be tried in order to improve the model:

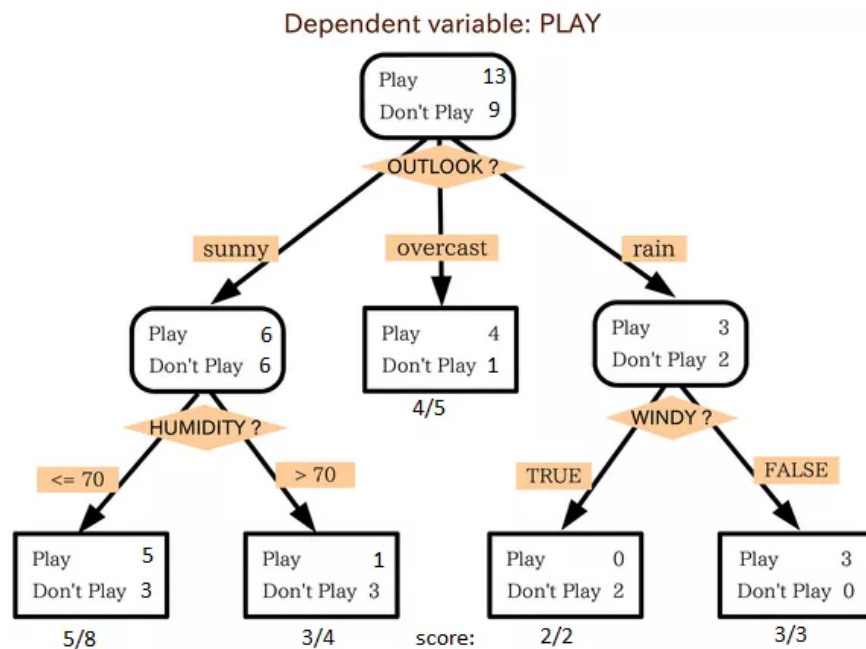
- including interaction terms
- removing features
- regularization techniques (<http://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting->

regularization/)

- using a non-linear model

3. Decision Tree

This is one of my favorite algorithm and I use it quite frequently. It is a type of supervised learning algorithm that is mostly used for classification problems. Surprisingly, it works for both categorical and continuous dependent variables. In this algorithm, we split the population into two or more homogeneous sets. This is done based on most significant attributes/independent variables to make as distinct groups as possible. For more details, you can read: Decision Tree Simplified (<http://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/>).



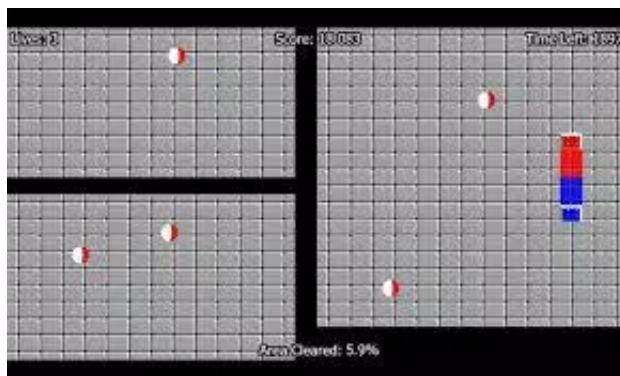
(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/lkBzK.png>)

source: statsexchange (<http://stats.stackexchange.com>)

In the image above, you can see that population is classified into four different groups based on multiple attributes to identify 'if they will play or not'. To split the population into different heterogeneous groups, it uses various techniques like Gini, Information Gain, Chi-square,

entropy.

The best way to understand how decision tree works, is to play Jezzball – a classic game from Microsoft (image below). Essentially, you have a room with moving walls and you need to create walls such that maximum area gets cleared off with out the balls.



(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/download.jpg>)

So, every time you split the room with a wall, you are trying to create 2 different populations with in the same room. Decision trees work in very similar fashion by dividing a population in as different groups as possible.

More: Simplified Version of Decision Tree Algorithms

(<http://www.analyticsvidhya.com/blog/2015/01/decision-tree-simplified/>)

Python Code

```
#Import Library
#Import other necessary libraries like pandas, numpy...
from sklearn import tree
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
# Create tree object
model = tree.DecisionTreeClassifier(criterion='gini') # for classification, here you can change the algorithm as gini or entropy (information gain) by default it is gini
# model = tree.DecisionTreeRegressor() for regression
# Train the model using the training sets and check score
model.fit(X, y)
model.score(X, y)
#Predict Output
predicted= model.predict(x_test)
```

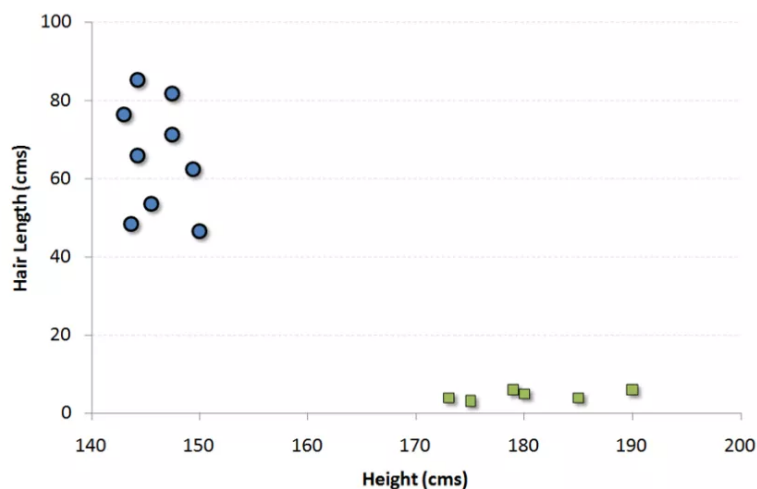
R Code

```
library(rpart)
x <- cbind(x_train,y_train)
# grow tree
fit <- rpart(y_train ~ ., data = x,method="class")
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

4. SVM (Support Vector Machine)

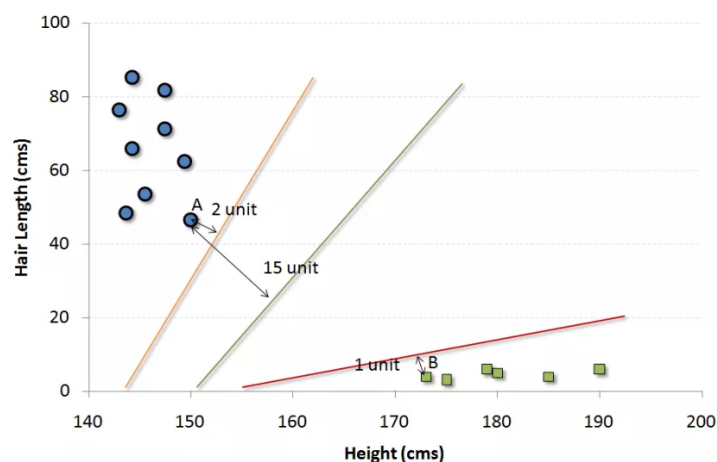
It is a classification method. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate.

For example, if we only had two features like Height and Hair length of an individual, we'd first plot these two variables in two dimensional space where each point has two co-ordinates (these co-ordinates are known as **Support Vectors**)



(<http://io.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/SVM1.png>)

Now, we will find some *line* that splits the data between the two differently classified groups of data. This will be the line such that the distances from the closest point in each of the two groups will be farthest away.



(<http://i2.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/SVM2.png>)

In the example shown above, the line which splits the data into two differently classified groups is the *black* line, since the two closest points are the farthest apart from the line. This line is our classifier. Then, depending on where the testing data lands on either side of the line, that's what class we can classify the new data as.

More: [Simplified Version of Support Vector Machine](http://www.analyticsvidhya.com/blog/2014/10/support-vector-machine-simplified/)
(<http://www.analyticsvidhya.com/blog/2014/10/support-vector-machine-simplified/>)

Think of this algorithm as playing JezzBall in n-dimensional space. The tweaks in the game are:

- You can draw lines / planes at any angles (rather than just horizontal or vertical as in classic game)
- The objective of the game is to segregate balls of different colors in different rooms.
- And the balls are not moving.

Python Code

```
#Import Library
from sklearn import svm
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
# Create SVM classification object
model = svm.svc() # there is various option associated with it, this is simple for classification. You can refer link (http://scikit-learn.org/stable/modules/svm.html), for more detail.
# Train the model using the training sets and check score
model.fit(X, y)
model.score(X, y)
#Predict Output
predicted= model.predict(x_test)
```

R Code

```
library(e1071)
x <- cbind(x_train,y_train)
# Fitting model
fit <-svm(y_train ~ ., data = x)
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

5. Naive Bayes

It is a classification technique based on Bayes' theorem (https://en.wikipedia.org/wiki/Bayes%27_theorem) with an assumption of independence between predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier would consider all of these properties to independently contribute to the probability that this fruit is an apple.

Naive Bayesian model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$. Look at the equation below:

The diagram shows the formula for Bayes' Theorem: $P(c | x) = \frac{P(x | c)P(c)}{P(x)}$. Arrows point from the terms to their labels: $P(c | x)$ is labeled 'Posterior Probability', $P(x | c)$ is labeled 'Likelihood', $P(c)$ is labeled 'Class Prior Probability', and $P(x)$ is labeled 'Predictor Prior Probability'.

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

(http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/Bayes_rule.png)

Here,

- $P(c|x)$ is the posterior probability of *class (target)* given *predictor (attribute)*.
- $P(c)$ is the prior probability of *class*.
- $P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$ is the prior probability of *predictor*.

Example: Let's understand it using an example. Below I have a training data set of weather and corresponding target variable 'Play'. Now, we need to classify whether players will play or not based on weather condition. Let's follow the below steps to perform it.

Step 1: Convert the data set to frequency table

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
All	5	9
	=5/14	=9/14
	0.36	0.64

(http://i2.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/Bayes_41.png)

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Problem: Players will play if weather is sunny, is this statement is correct?

We can solve it using above discussed method, so $P(\text{Yes} | \text{Sunny}) = P(\text{Sunny} | \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$

Here we have $P(\text{Sunny} | \text{Yes}) = 3/9 = 0.33$, $P(\text{Sunny}) = 5/14 = 0.36$, $P(\text{Yes}) = 9/14 = 0.64$

Now, $P(\text{Yes} | \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.

Python Code

```
#Import Library
from sklearn.naive_bayes import GaussianNB
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
# Create SVM classification object model = GaussianNB() # there is other distribution for multinomial classes like Bernoulli Naive Bayes, Refer link
(http://scikit-learn.org/stable/modules/naive\_bayes.html)# Train the model using the training sets and check score
model.fit(X, y)
#Predict Output
predicted= model.predict(x_test)
```

R Code

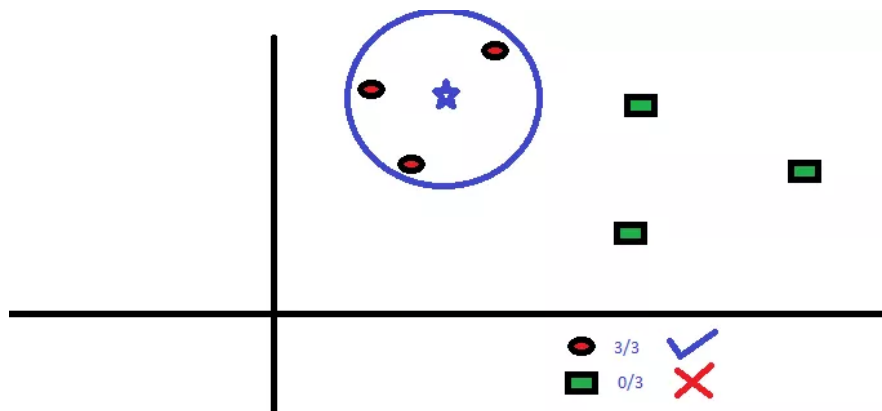
```
library(e1071)
x <- cbind(x_train,y_train)
# Fitting model
fit <-naiveBayes(y_train ~ ., data = x)
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

6. KNN (K- Nearest Neighbors)

It can be used for both classification and regression problems. However, it is more widely used in classification problems in the industry. K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbors. The case being assigned to the class is most common amongst its K nearest neighbors measured by a distance function.

These distance functions can be Euclidean, Manhattan, Minkowski and Hamming distance. First three functions are used for continuous function and fourth one (Hamming) for categorical variables. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor. At times, choosing K turns out to be a challenge while performing KNN modeling.

More: Introduction to k-nearest neighbors : Simplified (<http://Introduction to k-nearest neighbors : Simplified>).



(<http://i0.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/KNN.png>)

KNN can easily be mapped to our real lives. If you want to learn about a person, of whom you have no information, you might like to find out about his close friends and the circles he moves in and gain access to his/her information!

Things to consider before selecting KNN:

- KNN is computationally expensive
- Variables should be normalized else higher range variables can bias it
- Works on pre-processing stage more before going for KNN like outlier, noise removal

Python Code

```
#Import Library
from sklearn.neighbors import KNeighborsClassifier
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of te
st_dataset
# Create KNeighbors classifier object model
KNeighborsClassifier(n_neighbors=6) # default value for n_neighbors is 5
# Train the model using the training sets and check score
model.fit(X, y)
#Predict Output
predicted= model.predict(x_test)
```

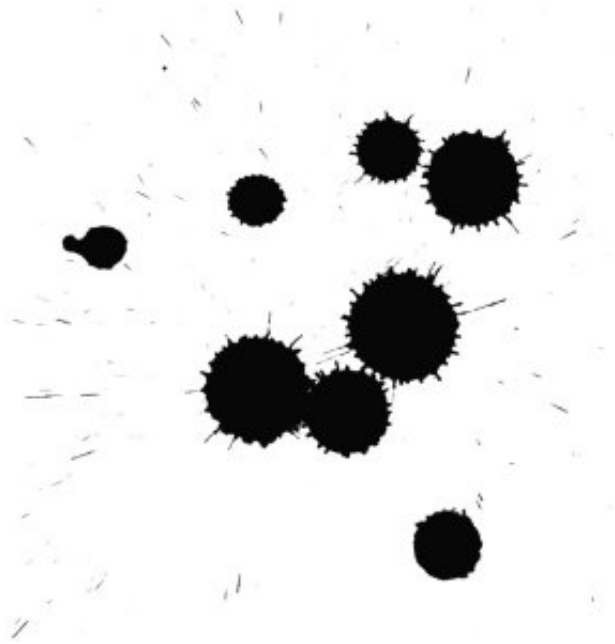
R Code

```
library(knn)
x <- cbind(x_train,y_train)
# Fitting model
fit <-knn(y_train ~ ., data = x,k=5)
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

7. K-Means

It is a type of unsupervised algorithm which solves the clustering problem. Its procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). Data points inside a cluster are homogeneous and heterogeneous to peer groups.

Remember figuring out shapes from ink blots? k means is somewhat similar this activity. You look at the shape and spread to decipher how many different clusters / population are present!



(http://i2.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/splatter_ink_blot_texture_by_maki_tak-d5p6zph.jpg)

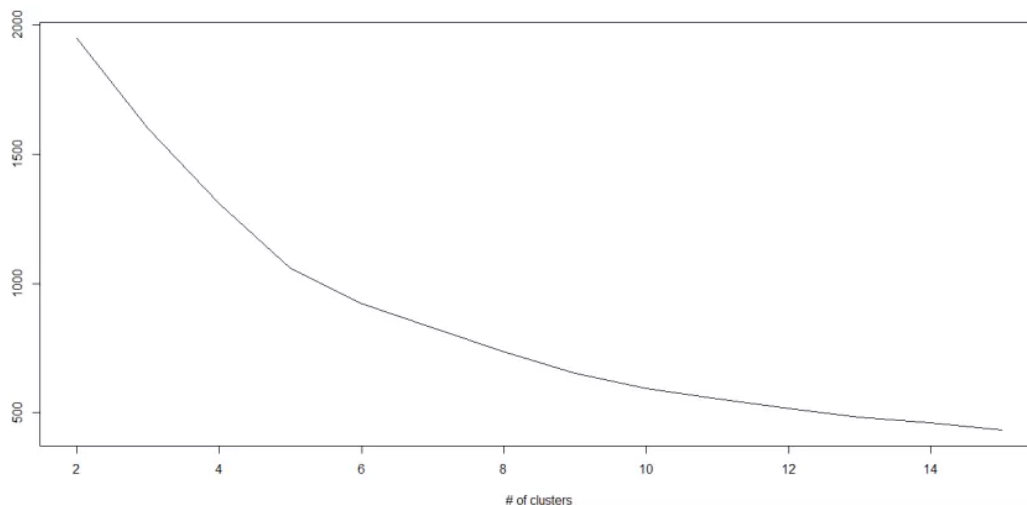
How K-means forms cluster:

1. K-means picks k number of points for each cluster known as centroids.
2. Each data point forms a cluster with the closest centroids i.e. k clusters.
3. Finds the centroid of each cluster based on existing cluster members. Here we have new centroids.
4. As we have new centroids, repeat step 2 and 3. Find the closest distance for each data point from new centroids and get associated with new k-clusters. Repeat this process until convergence occurs i.e. centroids does not change.

How to determine value of K:

In K-means, we have clusters and each cluster has its own centroid. Sum of square of difference between centroid and the data points within a cluster constitutes within sum of square value for that cluster. Also, when the sum of square values for all the clusters are added, it becomes total within sum of square value for the cluster solution.

We know that as the number of cluster increases, this value keeps on decreasing but if you plot the result you may see that the sum of squared distance decreases sharply up to some value of k, and then much more slowly after that. Here, we can find the optimum number of cluster.



(<http://i1.wp.com/www.analyticsvidhya.com/wp-content/uploads/2015/08/Kmenas.png>)

Python Code

```
#Import Library
from sklearn.cluster import KMeans

#Assumed you have, X (attributes) for training data set and x_test(attributes) of test_dataset

# Create KNeighbors classifier object model
k_means = KMeans(n_clusters=3, random_state=0)

# Train the model using the training sets and check score
model.fit(X)

#Predict Output
predicted= model.predict(x_test)
```

R Code

```
library(cluster)

fit <- kmeans(X, 3) # 5 cluster solution
```

8. Random Forest

Random Forest is a trademark term for an ensemble of decision trees. In Random Forest, we've collection of decision trees (so known as "Forest"). To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is planted & grown as follows:

1. If the number of cases in the training set is N , then sample of N cases is taken at random but *with replacement*. This sample will be the training set for growing the tree.
2. If there are M input variables, a number $m \ll M$ is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

For more details on this algorithm, comparing with decision tree and tuning model parameters, I would suggest you to read these articles:

1. Introduction to Random forest – Simplified
(<http://www.analyticsvidhya.com/blog/2014/06/introduction-random-forest-simplified/>)
2. Comparing a CART model to Random Forest (Part 1)
(<http://www.analyticsvidhya.com/blog/2014/06/comparing-cart-random-forest-1/>)
3. Comparing a Random Forest to a CART model (Part 2)
(<http://www.analyticsvidhya.com/blog/2014/06/comparing-random-forest-simple-cart-model/>)

4. Tuning the parameters of your Random Forest model

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

Python

```
#Import Library
from sklearn.ensemble import RandomForestClassifier
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
# Create Random Forest object
model= RandomForestClassifier()
# Train the model using the training sets and check score
model.fit(X, y)
#Predict Output
predicted= model.predict(x_test)
```

R Code

```
library(randomForest)
x <- cbind(x_train,y_train)
# Fitting model
fit <- randomForest(Species ~ ., x,ntree=500)
summary(fit)
#Predict Output
predicted= predict(fit,x_test)
```

9. Dimensionality Reduction Algorithms

In the last 4-5 years, there has been an exponential increase in data capturing at every possible stages. Corporates/ Government Agencies/ Research organisations are not only coming with new sources but also they are capturing data in great detail.

For example: E-commerce companies are capturing more details about customer like their demographics, web crawling history, what they like or dislike, purchase history, feedback and many others to give them personalized attention more than your nearest grocery shopkeeper.

As a data scientist, the data we are offered also consist of many features, this sounds good for building good robust model but there is a challenge. How'd you identify highly significant variable(s) out 1000 or 2000? In such cases, dimensionality reduction algorithm helps us along with various other algorithms like Decision Tree, Random Forest, PCA, Factor Analysis, Identify based on correlation matrix, missing value ratio and others.

To know more about this algorithms, you can read "Beginners Guide To Learn Dimension Reduction Techniques (<http://www.analyticsvidhya.com/blog/2015/07/dimension-reduction-methods/>)".

Python Code

```
#Import Library
from sklearn import decomposition

#Assumed you have training and test data set as train and test

# Create PCA object pca= decomposition.PCA(n_components=k) #default value of k =min(n_sample, n_
features)

# For Factor analysis

#fa= decomposition.FactorAnalysis()

# Reduced the dimension of training dataset using PCA
train_reduced = pca.fit_transform(train)

#Reduced the dimension of test dataset
test_reduced = pca.transform(test)

#For more detail on this, please refer this link (http://scikit-learn.org/stable/modules/decomposition.html#decompositions).
```

R Code

```
library(stats)
pca <- princomp(train, cor = TRUE)
train_reduced <- predict(pca,train)
test_reduced <- predict(pca,test)
```

10. Gradient Boosting & AdaBoost

GBM & AdaBoost are boosting algorithms used when we deal with plenty of data to make a prediction with high prediction power. Boosting is an ensemble learning algorithm which combines the prediction of several base estimators in order to improve robustness over a single estimator. It combines multiple weak or average predictors to a build strong predictor. These boosting algorithms always work well in data science competitions like Kaggle, AV Hackathon, CrowdAnalytix.

More: [Know about Gradient and AdaBoost in detail \(http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/\)](http://www.analyticsvidhya.com/blog/2015/05/boosting-algorithms-simplified/)

Python Code

```
#Import Library
from sklearn.ensemble import GradientBoostingClassifier

#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset

# Create Gradient Boosting Classifier object
model= GradientBoostingClassifier(n_estimators=100, learning_rate=1.0, max_depth=1, random_state=0)

# Train the model using the training sets and check score
model.fit(X, y)

#Predict Output
predicted= model.predict(x_test)
```

R Code

```
library(caret)
x <- cbind(x_train,y_train)
# Fitting model
fitControl <- trainControl( method = "repeatedcv", number = 4, repeats = 4)
fit <- train(y ~ ., data = x, method = "gbm", trControl = fitControl,verbose = FALSE)
predicted= predict(fit,x_test,type= "prob")[,2]
```

GradientBoostingClassifier and Random Forest are two different boosting tree classifier and often people ask about the difference between these two algorithms (<http://discuss.analyticsvidhya.com/t/what-is-the-fundamental-difference-between-randomforest-and-gradient-boosting-algorithms/2341>).

End Notes

By now, I am sure, you would have an idea of commonly used machine learning algorithms. My sole intention behind writing this article and providing the codes in R and Python is to get you started right away. If you are keen to master machine learning, start right away. Take up problems, develop a physical understanding of the process, apply these codes and see the fun!

Did you find this article useful ? Share your views and opinions 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/2015/08/common-machine-learning-algorithms/?share=linkedin&nb=1>) 983

 (<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?share=facebook&nb=1>) 1K+

 (<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?share=google-plus-1&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?share=twitter&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?share=pocket&nb=1>)

 (<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?share=reddit&nb=1>)

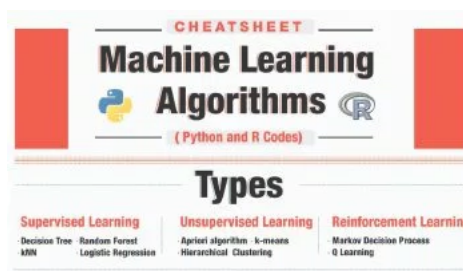
RELATED



(<http://www.analyticsvidhya.com/blog/2015/12/year-review-analytics-vidhya-from-2015/>)

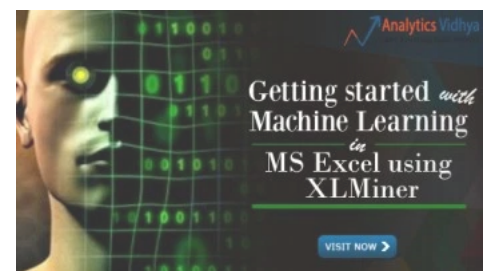
Year in Review: Best of Analytics Vidhya from 2015

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



(<http://www.analyticsvidhya.com/blog/2015/09/full-cheatsheet-machine-learning-algorithms/>)

Cheatsheet - Python & R codes for common Machine Learning Algorithms



(<http://www.analyticsvidhya.com/blog/2015/11/started-machine-learning-ms-excel-xl-miner/>)

Getting started with Machine Learning in MS Excel using XLMiner

[/blog/2015/12/year-review-analytics-vidhya-from-2015/](http://www.analyticsvidhya.com/blog/2015/12/year-review-analytics-vidhya-from-2015/)

In "Business Analytics"

[\(http://www.analyticsvidhya.com/blog/2015/09/full-cheatsheet-machine-learning-algorithms/\)](http://www.analyticsvidhya.com/blog/2015/09/full-cheatsheet-machine-learning-algorithms/)

In "Business Analytics"

[\(http://www.analyticsvidhya.com/blog/2015/11/started-machine-learning-ms-excel-xl-miner/\)](http://www.analyticsvidhya.com/blog/2015/11/started-machine-learning-ms-excel-xl-miner/)

In "Business Analytics"

TAGS: C4.5 ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/C4-5/](http://www.analyticsvidhya.com/blog/tag/c4-5/)), CART ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/CART/](http://www.analyticsvidhya.com/blog/tag/cart/)), DECISION TREE ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DECISION-TREE/](http://www.analyticsvidhya.com/blog/tag/decision-tree/)), GBM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/GBM/](http://www.analyticsvidhya.com/blog/tag/gbm/)), K-MEANS ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/K-MEANS/](http://www.analyticsvidhya.com/blog/tag/k-means/)), KNN ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/KNN/](http://www.analyticsvidhya.com/blog/tag/knn/)), LINEAR-REGRESSION ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LINEAR-REGRESSION/](http://www.analyticsvidhya.com/blog/tag/linear-regression/)), LOGISTIC REGRESSION ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LOGISTIC-REGRESSION/](http://www.analyticsvidhya.com/blog/tag/logistic-regression/)), MACHINE LEARNING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MACHINE-LEARNING/](http://www.analyticsvidhya.com/blog/tag/machine-learning/)), NAIVE BAYES ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/NAIVE-BAYES/](http://www.analyticsvidhya.com/blog/tag/naive-bayes/)), NEURAL NETWORK ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/NEURAL-NETWORK/](http://www.analyticsvidhya.com/blog/tag/neural-network/)), RANDOM FOREST ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/RANDOM-FOREST/](http://www.analyticsvidhya.com/blog/tag/random-forest/)), REINFORCEMENT ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/REINFORCEMENT/](http://www.analyticsvidhya.com/blog/tag/reinforcement/)), SUPERVISED LEARNING ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/SUPERVISED-LEARNING/](http://www.analyticsvidhya.com/blog/tag/supervised-learning/)), UNSUPERVISED ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/UNSUPERVISED/](http://www.analyticsvidhya.com/blog/tag/unsupervised/))

Previous Article

Marketing Analytics: Essentials of Cross-Selling and Upselling (with a case study)

[\(http://www.analyticsvidhya.com/blog/2015/08/learn-guide-learn-content-based-recommender-systems/\)](http://www.analyticsvidhya.com/blog/2015/08/learn-guide-learn-content-based-recommender-systems/)

Next Article

Beginners Guide to learn about Content Based Recommender Engines

[\(http://www.analyticsvidhya.com/blog/2015/08/beginner-guide-learn-content-based-recommender-systems/\)](http://www.analyticsvidhya.com/blog/2015/08/beginner-guide-learn-content-based-recommender-systems/)



(<http://www.analyticsvidhya.com/blog/author/sunil-ray/>)

Author

Sunil Ray (<http://www.analyticsvidhya.com/blog/author/sunil-ray/>)

I am a Business Analytics and Intelligence professional with deep experience in the Indian Insurance industry. I have worked for various multi-national Insurance companies in last 7 years.

26 COMMENTS



Kuber says:

REPLY([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92380#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92380#respond))
AUGUST 10, 2015 AT 11:59 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92380](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92380))

Awesowe compilation!! Thank you.



Karthikeyan says:

REPLY TO <http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92394#respond>
AUGUST 11, 2015 AT 3:13 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92394](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92394))

Thank you very much, A Very useful and excellent compilation. I have already bookmarked this page.



hemanth says:

REPLY TO <http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92400#respond>
AUGUST 11, 2015 AT 4:50 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92400](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92400))

Straight, Informative and effective!!
Thank you



venugopal says:

REPLY TO <http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92402#respond>
AUGUST 11, 2015 AT 6:05 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92402](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92402))

Good Summary airticle



Dr Venugopala Rao says:

REPLY TO <http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92404#respond>
AUGUST 11, 2015 AT 6:27 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92404](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92404))

Super Compilation...



Kishor Basyal says:

REPLY TO <http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92410#respond>
AUGUST 11, 2015 AT 7:30 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92410](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92410))

Wonderful! Really helpful



Brian Thomas says:

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92414#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92414#RESPOND)
AUGUST 11, 2015 AT 9:24 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92414](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92414))

Very nicely done! Thanks for this.



Tesfaye says:

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92416#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92416#RESPOND)
AUGUST 11, 2015 AT 10:30 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92416](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92416))

Thank you! Well presented article.



Tesfaye says:

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92417#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92417#RESPOND)
AUGUST 11, 2015 AT 10:31 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92417](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92417))

Thank you! Well presented.



Huzefa says:

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92436#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92436#RESPOND)
AUGUST 11, 2015 AT 3:53 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92436](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92436))

Hello,

Superb information in just one blog. Can anyone help me to run the codes in R what should be replaced with "~" symbol in codes? Help is appreciated



Huzefa says:

REPLY TO [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92437#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=92437#RESPOND)
AUGUST 11, 2015 AT 3:54 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92437](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-92437))

Hello,

Superb information in just one blog. Can anyone help me to run the codes in R what should be replaced with "~" symbol in codes? Help is appreciated .



Sudipta Basak says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92497#RESPOND)
AUGUST 12, 2015 AT 3:35 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92497)

Enjoyed the simplicity. Thanks for the effort.



Im_utm (http://twitter.com/Im_utm) says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92552#RESPOND)
AUGUST 12, 2015 AT 2:37 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92552)

Great Article... Helps a lot, as naive in Machine Learning.



Sunil Ray says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92706#RESPOND)
AUGUST 14, 2015 AT 7:36 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92706)

Hi All,

Thanks for the comment ...



Dalila says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=92744#RESPOND)
AUGUST 14, 2015 AT 1:35 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-92744)

Very good summary.

Thank!

One simple point. The reason for taking the $\log(p/(1-p))$ in Logistic Regression is to make the equation linear, I.e., easy to solve.



Sunil Ray says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM-93236#RESPOND)
AUGUST 21, 2015 AT 5:21 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-93236)

Thanks Dalila... 😊



Statis says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM-93085#RESPOND)
AUGUST 19, 2015 AT 12:14 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-93085)

Nice summary!

@Huzefa: you shouldn't replace the "~" in the R code, it basically means "as a function of". You can also keep the "." right after, it stands for "all other variables in the dataset provided". If you want to be explicit, you can write $y \sim x_1 + x_2 + \dots$ where $x_1, x_2 \dots$ are the names of the columns of your data.frame or data.table.

Further note on formula specification: by default R adds an intercept, so that $y \sim x$ is equivalent to $y \sim 1 + x$, you can remove it via $y \sim 0 + x$. Interactions are specified with either * (which also adds the two variables) or : (which only adds the interaction term). $y \sim x_1 * x_2$ is equivalent to $y \sim x_1 + x_2 + x_1 : x_2$.

Hope this helps!



Chris says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM-93518#RESPOND)
AUGUST 26, 2015 AT 1:01 AM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-93518)

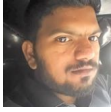
You did a Wonderful job! This is really helpful. Thanks!



Glenn Nelson (<https://www.facebook.com/app-scoped-user-id/10153310687928141/>) says:

REPLY (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM-94650#RESPOND)
SEPTEMBER 10, 2015 AT 7:48 PM (HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-94650)

I took the Stanford-Coursera ML class, but have not used it, and I found this to be an incredibly useful summary. I appreciate the real-world analogues, such as your mention of Jezzball. And showing the brief code snips is terrific.



Shankar Pandala (<https://www.facebook.com/app-scoped-user.id/10153564059874654/>)
 says:

SEPTEMBER 15, 2015 AT 12:09 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-95096](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-95096))

This is very easy and helpful than any other courses I have completed.
 simple. clear. To the point.



markprattley says:

SEPTEMBER 26, 2015 AT 9:29 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-96010](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=96010#respond))

You Sir are a gentleman and a scholar!



whystatistics says:

SEPTEMBER 29, 2015 AT 10:25 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-96181](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=96181#respond))

Hi Sunil,

This is really superb tutorial along with good examples and codes which is surely much helpful. Just, can you add Neural Network here in simple terms with example and code.



Sayan Putatunda says:

NOVEMBER 1, 2015 AT 7:00 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-98659](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=98659#respond))

Errata:- `fit <- kmeans(X, 3) # 5 cluster solution`
 It's a 3 cluster solution.

**Baha says:**

REPLY WITH: [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=100598#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=100598#respond)
NOVEMBER 27, 2015 AT 1:13 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-100598](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-100598))

Well done, Thank you!

**Benjamin says:**

REPLY WITH: [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=101309#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=101309#respond)
DECEMBER 5, 2015 AT 7:00 PM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-101309](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-101309))

This is a great resource overall and surely the product of a lot of work.

Just a note as I go through this, your comment on Logistic Regression not actually being regression is in fact wrong. It maps outputs to a continuous variable bound between 0 and 1 that we regard as probability. It makes classification easy but that is still an extra step that requires the choice of a threshold which is not the main aim of Logistic Regression. As a matter of fact it falls under the umbrella of Generalized Linear Models as the glm R package hints it in your code example.

I thought this was interesting to note so as not to forget that logistic regression output is richer than 0 or 1.

Thanks for the great article overall.


**Bansari Shah says:**

REPLY WITH: [HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/?REPLYTOCOM=103718#RESPOND](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/?replytocom=103718#respond)
JANUARY 14, 2016 AT 6:27 AM ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMMON-MACHINE-LEARNING-ALGORITHMS/#COMMENT-103718](http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/#comment-103718))

Thank you.. reallu helpful article

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/wp-content/uploads/2015/08/common-machine-learning-algorithms%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 (http://datahack.analyticsvidhya.com/user/profile/Nalin)	3478
2		SRK (http://datahack.analyticsvidhya.com/user/profile/SRK)	3364
3		Aayushmnit (http://datahack.analyticsvidhya.com/user/profile/aayushmnit)	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)



**POST GRADUATE PROGRAM
IN BUSINESS ANALYTICS**

Batch starts in **Dec'15/Jan'16**

APPLY NOW

*Admissions open for **Chennai, Gurgaon, Bengaluru & Pune***

(<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/>)

DataFit Curve Fitting

Nonlinear curve fitting,
600+ built in equations,
or create your own





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

professional/)

RECENT 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/)

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/)

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/)

Quick Insights: India Analytics and Big Data Salary Report 2016

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

KUNAL JAIN , FEBRUARY 24, 2016



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

India Exclusive: Analytics and Big Data Salary Report 2016

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

KUNAL JAIN , FEBRUARY 22, 2016



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

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

GET CONNECTED



4,159

FOLLOWERS

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



915

FOLLOWERS

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



11,950

FOLLOWERS

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



Email

SUBSCRIBE

(<http://feedburner.google.com/fb/a/mailverify?>



uri=analyticsvidhya)

(<http://www.analyticsvidhya.com/blog/2016/02/quick->

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

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

STAY CONNECTED



4,159

FOLLOWERS

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



915

FOLLOWERS

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



11,950

FOLLOWERS

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



Email

SUBSCRIBE

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

LATEST POSTS



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

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

A Complete Tutorial to learn Data Science in R from Scratch

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

MANISH SARASWAT , FEBRUARY 28, 2016



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

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

Guide to Build Better Predictive Models using Segmentation

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

GUEST BLOG , FEBRUARY 26, 2016



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

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