

x

-

(<http://play.google.com/store/apps/details?id=com.analyticsvidhya.android>)



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



([https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?utm\\_source=DS101AVBlogBanner&utm\\_medium=Topbanner&utm\\_campaign=DS101](https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?utm_source=DS101AVBlogBanner&utm_medium=Topbanner&utm_campaign=DS101))

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

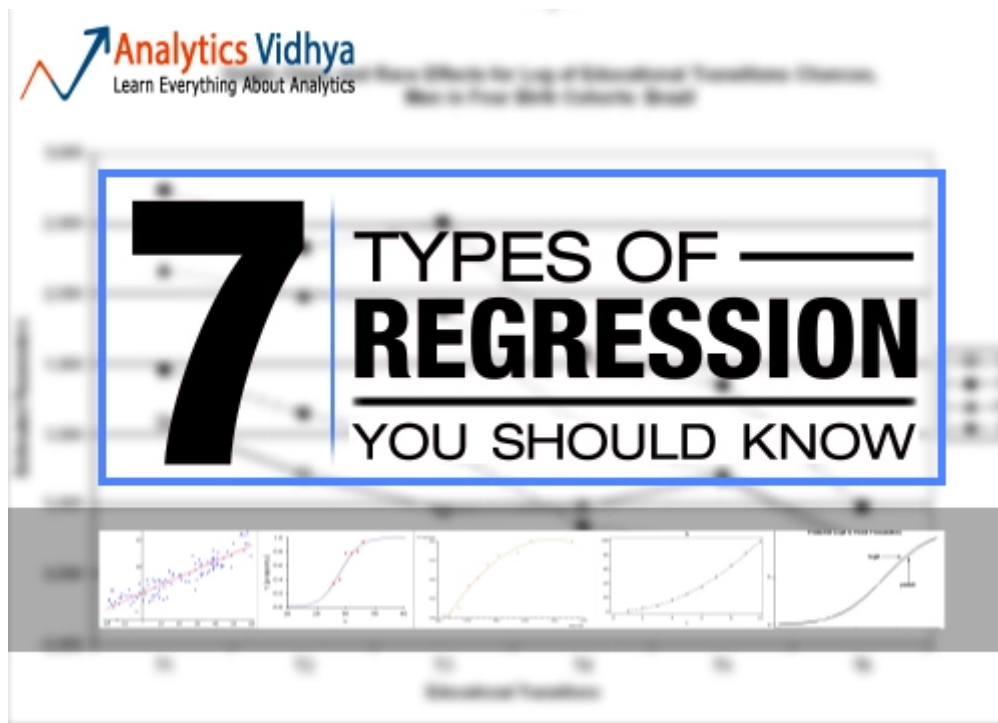
## 7 Types of Regression Techniques you should know!

[SUNIL RAY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/AUTHOR/SUNIL-RAY/\)](https://www.analyticsvidhya.com/blog/author/sunil-ray/), AUGUST 14, 2015

### Introduction

Linear and Logistic regressions are usually the first algorithms people learn in predictive modeling. Due to their popularity, a lot of analysts even end up thinking that they are the only form of regressions. The ones who are slightly more involved think that they are the most important amongst all forms of regression analysis.

The truth is that there are innumerable forms of regressions, which can be performed. Each form has its own importance and a specific condition where they are best suited to apply. In this article, I have explained the most commonly used 7 forms of regressions in a simple manner. Through this article, I also hope that people develop an idea of the breadth of regressions, instead of just applying linear / logistic regression to every problem they come across and hoping that they would just fit!



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/08/regression.jpg>)

## Table of Contents

1. What is Regression Analysis?
2. Why do we use Regression Analysis?
3. What are the types of Regressions?
  - Linear Regression
  - Logistic Regression
  - Polynomial Regression
  - Stepwise Regression
  - Ridge Regression
  - Lasso Regression
  - ElasticNet Regression
4. How to select the right Regression Model?

## What is Regression Analysis?

Regression analysis is a form of predictive modelling technique which investigates the relationship between a **dependent** (target) and **independent variable (s)** (predictor). This technique is used for forecasting, time series modelling and finding the causal effect relationship (<https://www.analyticsvidhya.com/blog/2015/06/establish-causality-events/>) between the variables. For example, relationship between rash driving and number of road accidents by a driver is best studied through regression.

Regression analysis is an important tool for modelling and analyzing data. Here, we fit a curve / line to the data points, in such a manner that the differences between the distances of data points from the curve or line is minimized. I'll explain this in more details in coming sections.



([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Regression\\_Line.png](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Regression_Line.png))

## Why do we use Regression Analysis?

As mentioned above, regression analysis estimates the relationship between two or more variables. Let's understand this with an easy example:

Let's say, you want to estimate growth in sales of a company based on current economic conditions. You have the recent company data which indicates that the growth in sales is around two and a half times the growth in the economy. Using this insight, we can predict future sales of the company based on current & past information.

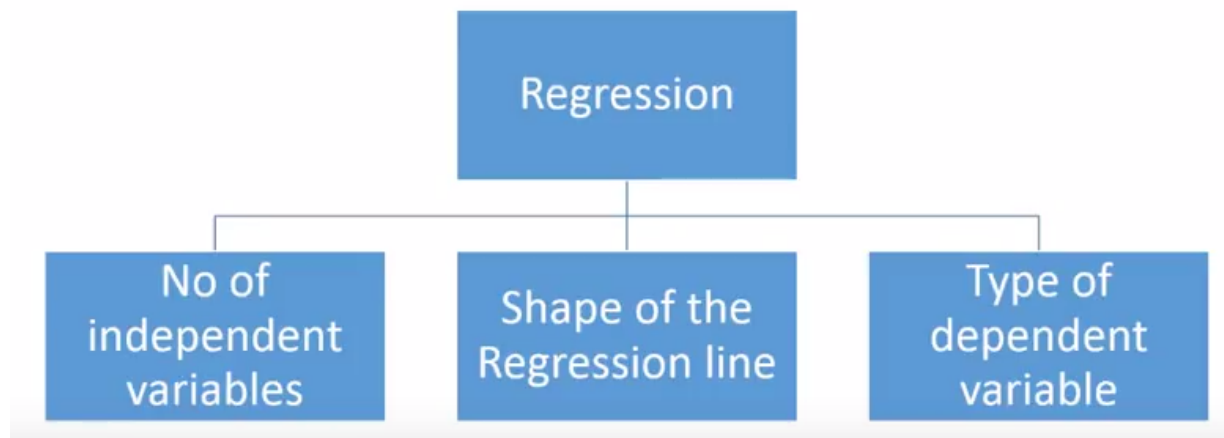
There are multiple benefits of using regression analysis. They are as follows:

1. It indicates the **significant relationships** between dependent variable and independent variable.
2. It indicates the **strength of impact** of multiple independent variables on a dependent variable.

Regression analysis also allows us to compare the effects of variables measured on different scales, such as the effect of price changes and the number of promotional activities. These benefits help market researchers / data analysts / data scientists to eliminate and evaluate the best set of variables to be used for building predictive models.

## How many types of regression techniques do we have?

There are various kinds of regression techniques available to make predictions. These techniques are mostly driven by three metrics (number of independent variables, type of dependent variables and shape of regression line). We'll discuss them in detail in the following sections.



([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Regression\\_Type.png](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Regression_Type.png)).

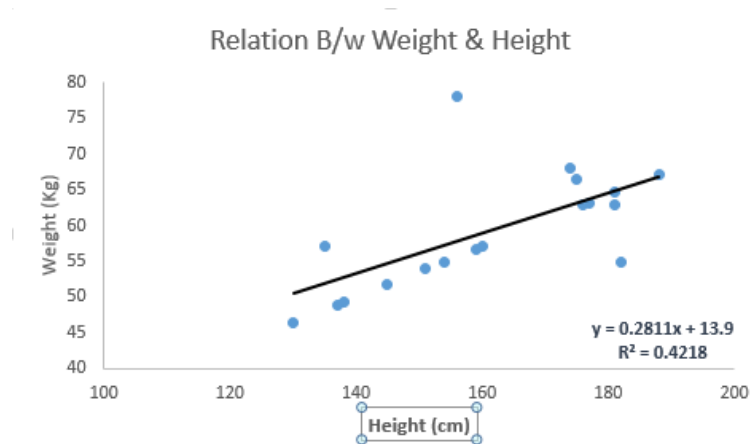
For the creative ones, you can even cook up new regressions, if you feel the need to use a combination of the parameters above, which people haven't used before. But before you start that, let us understand the most commonly used regressions:

## 1. Linear Regression

It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete ([https://en.wikipedia.org/wiki/Continuous\\_and\\_discrete\\_variables](https://en.wikipedia.org/wiki/Continuous_and_discrete_variables)), and nature of regression line is linear.

Linear Regression establishes a relationship between **dependent variable (Y)** and one or more **independent variables (X)** using a **best fit straight line** (also known as regression line).

It is represented by an equation  $Y = a + bX + e$ , where  $a$  is intercept,  $b$  is slope of the line and  $e$  is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).



([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Linear\\_Regression1.png](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Linear_Regression1.png))

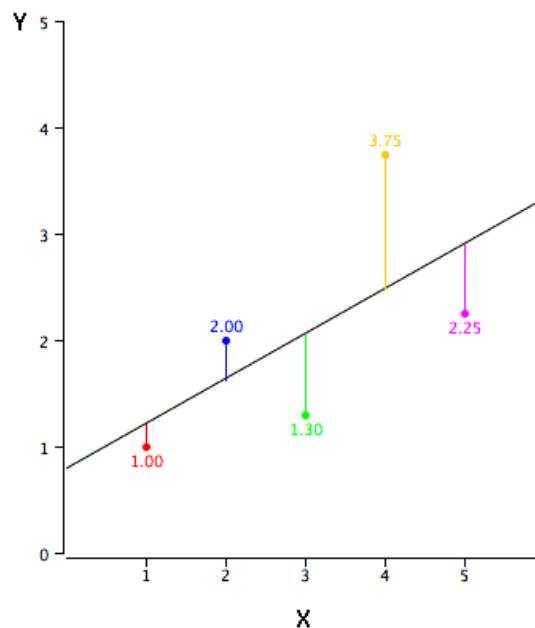
The difference between simple linear regression and multiple linear regression is that, multiple linear regression has ( $>1$ ) independent variables, whereas simple linear regression has only 1 independent variable. Now, the question is “How do we obtain best fit line?”.

### How to obtain best fit line (Value of $a$ and $b$ )?

This task can be easily accomplished by Least Square Method. It is the most common method used for fitting a regression line. It calculates the best-fit line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line. Because the deviations are first squared, when added, there is no cancelling out between positive and negative values.

$$\min_w ||Xw - y||_2^2$$

([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Least\\_Square.png](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Least_Square.png))



([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/reg\\_error.gif](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/reg_error.gif))

We can evaluate the model performance using the metric **R-square**. To know more details about these metrics, you can read: Model Performance metrics Part 1 (<https://www.analyticsvidhya.com/blog/2015/01/model-performance-metrics-classification/>), Part 2 (<https://www.analyticsvidhya.com/blog/2015/01/model-perform-part-2/>).

### Important Points:

- There must be **linear relationship** between independent and dependent variables
- Multiple regression suffers from **multicollinearity, autocorrelation, heteroskedasticity**.
- Linear Regression is very sensitive to **Outliers**. It can terribly affect the regression line and eventually the forecasted values.
- Multicollinearity can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model. The result is that the coefficient estimates are unstable
- In case of multiple independent variables, we can go with **forward selection, backward elimination** and **step wise approach** for selection of most significant independent variables.

## 2. Logistic Regression

Logistic regression is used to find the probability of event=Success and event=Failure. We should use logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature. Here the value of Y ranges from 0 to 1 and it can be represented by following equation.

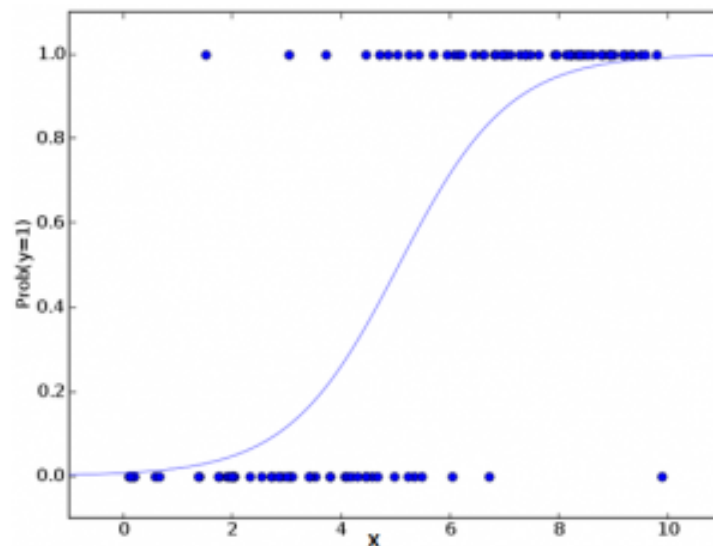
$$\text{odds} = p / (1-p) = \text{probability of event occurrence} / \text{probability of not event occurrence}$$

$$\ln(\text{odds}) = \ln(p/(1-p))$$

$$\text{logit}(p) = \ln(p/(1-p)) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k$$

Above,  $p$  is the probability of presence of the characteristic of interest. A question that you should ask here is “why have we used log in the equation?”.

Since we are working here with a binomial distribution (dependent variable), we need to choose a link function which is best suited for this distribution. And, it is **logit** ([https://en.wikipedia.org/wiki/Logistic\\_function](https://en.wikipedia.org/wiki/Logistic_function)) function. In the equation above, the parameters are chosen to maximize the likelihood of observing the sample values rather than minimizing the sum of squared errors (like in ordinary regression).



([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Logistic\\_Regression.png](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Logistic_Regression.png))

### Important Points:

- It is widely used for **classification problems**
- Logistic regression doesn't require linear relationship between dependent and independent variables. It can handle various types of relationships because it applies a non-linear log transformation to the predicted odds ratio
- To avoid over fitting and under fitting, we should include all significant variables. A good approach to ensure this practice is to use a step wise method to estimate the logistic regression
- It requires **large sample sizes** because maximum likelihood estimates are less powerful at low sample sizes than ordinary least square
- The independent variables should not be correlated with each other i.e. **no multi collinearity**. However, we have the options to include interaction effects of categorical variables in the analysis and in the model.

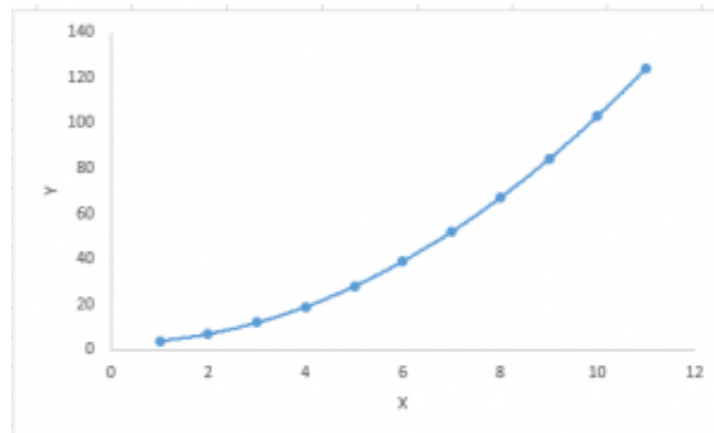
- If the values of dependent variable is ordinal, then it is called as **Ordinal logistic regression**
- If dependent variable is multi class then it is known as **Multinomial Logistic regression**.

### 3. Polynomial Regression

A regression equation is a polynomial regression equation if the power of independent variable is more than 1. The equation below represents a polynomial equation:

$$y=a+b*x^2$$

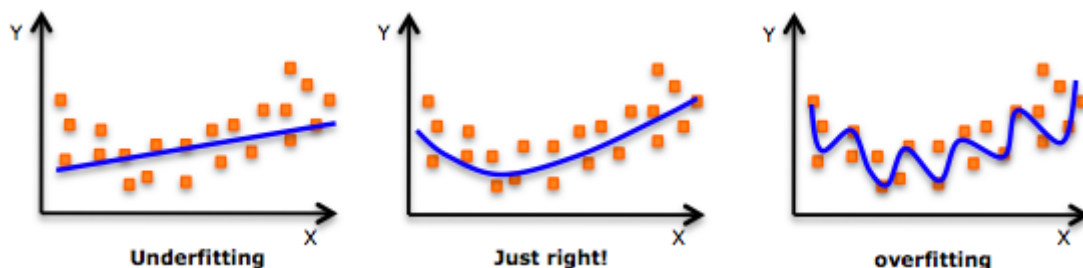
In this regression technique, the best fit line is not a straight line. It is rather a curve that fits into the data points.



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Polynomial.png>)

#### Important Points:

- While there might be a temptation to fit a higher degree polynomial to get lower error, this can result in over-fitting. Always plot the relationships to see the fit and focus on making sure that the curve fits the nature of the problem. Here is an example of how plotting can help:



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/underfitting-overfitting.png>)



- Especially look out for curve towards the ends and see whether those shapes and trends make sense. Higher polynomials can end up producing wierd results on extrapolation.

## 4. Stepwise Regression

This form of regression is used when we deal with multiple independent variables. In this technique, the selection of independent variables is done with the help of an automatic process, which involves *no* human intervention.

This feat is achieved by observing statistical values like R-square, t-stats and AIC metric to discern significant variables. Stepwise regression basically fits the regression model by adding/dropping co-variates one at a time based on a specified criterion. Some of the most commonly used Stepwise regression methods are listed below:

- Standard stepwise regression does two things. It adds and removes predictors as needed for each step.
- Forward selection starts with most significant predictor in the model and adds variable for each step.
- Backward elimination starts with all predictors in the model and removes the least significant variable for each step.

The aim of this modeling technique is to maximize the prediction power with minimum number of predictor variables. It is one of the method to handle higher dimensionality (<https://www.analyticsvidhya.com/blog/2015/07/dimension-reduction-methods/>) of data set.

## 5. Ridge Regression

Ridge Regression is a technique used when the data suffers from multicollinearity ( independent variables are highly correlated). In multicollinearity, even though the least squares estimates (OLS) are unbiased, their variances are large which deviates the observed value far from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors.

Above, we saw the equation for linear regression. Remember? It can be represented as:

$$y = a + b \cdot x$$

This equation also has an error term. The complete equation becomes:

$$y = a + b \cdot x + e \text{ (error term), [error term is the value needed to correct for a prediction error between the observed and predicted value]}$$

=>  $y = a + b_1x_1 + b_2x_2 + \dots + e$ , for multiple independent variables.

In a linear equation, prediction errors can be decomposed into two sub components. First is due to the **biased** and second is due to the **variance**. Prediction error can occur due to any one of these two or both components. Here, we'll discuss about the error caused due to variance.

Ridge regression solves the multicollinearity problem through shrinkage parameter ([https://en.wikipedia.org/wiki/Shrinkage\\_estimator](https://en.wikipedia.org/wiki/Shrinkage_estimator))  $\lambda$  (lambda). Look at the equation below.

$$= \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + \lambda \underbrace{\|\beta\|_2^2}_{\text{Penalty}}$$

(<https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Ridge2.png>)

In this equation, we have two components. First one is least square term and other one is lambda of the summation of  $\beta^2$  (beta- square) where  $\beta$  is the coefficient. This is added to least square term in order to shrink the parameter to have a very low variance.

### Important Points:

- The assumptions of this regression is same as least squared regression except normality is not to be assumed
- It shrinks the value of coefficients but doesn't reaches zero, which suggests no feature selection feature
- This is a regularization method and uses L2 regularization ([https://en.wikipedia.org/wiki/Regularization\\_\(mathematics\)](https://en.wikipedia.org/wiki/Regularization_(mathematics))).

## 6. Lasso Regression

Similar to Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) also penalizes the absolute size of the regression coefficients. In addition, it is capable of reducing the variability and improving the accuracy of linear regression models. Look at the equation below:

$$= \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + \lambda \underbrace{\|\beta\|_1}_{\text{Penalty}}$$

(<https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Lasso.png>) Lasso regression differs from ridge regression in a way that it uses absolute values in the penalty function, instead of squares. This leads to penalizing (or equivalently constraining the sum of the absolute values of the estimates) values which causes some of the parameter estimates to turn out exactly zero. Larger the penalty applied, further the estimates get shrunk towards absolute zero. This results to variable selection out of given n variables.

### Important Points:

- The assumptions of this regression is same as least squared regression except normality is not to be assumed
- It shrinks coefficients to zero (exactly zero), which certainly helps in feature selection
- This is a regularization method and uses L1 regularization  
([https://en.wikipedia.org/wiki/Regularization\\_\(mathematics\)](https://en.wikipedia.org/wiki/Regularization_(mathematics)))
- If group of predictors are highly correlated, lasso picks only one of them and shrinks the others to zero

## 7. ElasticNet Regression

ElasticNet is hybrid of Lasso and Ridge Regression techniques. It is trained with L1 and L2 prior as regularizer. Elastic-net is useful when there are multiple features which are correlated. Lasso is likely to pick one of these at random, while elastic-net is likely to pick both.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} (\|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1).$$

([https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Elastic\\_Net.png](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Elastic_Net.png))

A practical advantage of trading-off between Lasso and Ridge is that, it allows Elastic-Net to inherit some of Ridge's stability under rotation.

### Important Points:

- It encourages group effect in case of highly correlated variables
- There are no limitations on the number of selected variables
- It can suffer with double shrinkage

Beyond these 7 most commonly used regression techniques, you can also look at other models like Bayesian ([https://en.wikipedia.org/wiki/Bayesian\\_linear\\_regression](https://en.wikipedia.org/wiki/Bayesian_linear_regression)), Ecological ([https://en.wikipedia.org/wiki/Ecological\\_regression](https://en.wikipedia.org/wiki/Ecological_regression)) and Robust regression ([https://en.wikipedia.org/wiki/Robust\\_regression](https://en.wikipedia.org/wiki/Robust_regression)).

## How to select the right regression model?

Life is usually simple, when you know only one or two techniques. One of the training institutes I know of tells their students – if the outcome is continuous – apply linear regression. If it is binary – use logistic regression! However, higher the number of options available at our disposal, more difficult it becomes to choose the right one. A similar case happens with regression models.

Within multiple types of regression models, it is important to choose the best suited technique based on type of independent and dependent variables, dimensionality in the data and other essential characteristics of the data. Below are the key factors that you should practice to select the right regression model:

1. Data exploration is an inevitable part of building predictive model. It should be your first step before selecting the right model like identify the relationship and impact of variables
2. To compare the goodness of fit for different models, we can analyse different metrics like statistical significance of parameters, R-square, Adjusted r-square, AIC, BIC and error term. Another one is the Mallow's Cp (<http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/regression-and-correlation/goodness-of-fit-statistics/what-is-mallows-cp/>) criterion. This essentially checks for possible bias in your model, by comparing the model with all possible submodels (or a careful selection of them).
3. Cross-validation is the best way to evaluate models used for prediction. Here you divide your data set into two groups (train and validate). A simple mean squared difference between the observed and predicted values give you a measure for the prediction accuracy.
4. If your data set has multiple confounding variables, you should not choose automatic model selection method because you do not want to put these in a model at the same time.
5. It'll also depend on your objective. It can occur that a less powerful model is easy to implement as compared to a highly statistically significant model.
6. Regression regularization methods (Lasso, Ridge and ElasticNet) works well in case of high dimensionality and multicollinearity among the variables in the data set.

## End Note

By now, I hope you would have got an overview of regression. These regression techniques should be applied considering the conditions of data. One of the best trick to find out which technique to use, is by checking the family of variables i.e. discrete or continuous.

In this article, I discussed about 7 types of regression and some key facts associated with each technique. As somebody who's new in this industry, I'd advise you to learn these techniques and later implement them in your models.

Did you find this article useful ? Share your opinions / views in the comments section below.

**Note – The discussions of this article are going on at AV's Discuss portal. [Join here \(https://discuss.analyticsvidhya.com/t/discussions-for-article-7-types-of-regression-techniques-you-should-know/65230?u=jalfaizy\)](https://discuss.analyticsvidhya.com/t/discussions-for-article-7-types-of-regression-techniques-you-should-know/65230?u=jalfaizy)!**

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

You can also read this article on Analytics Vidhya's Android APP



[http://play.google.com/store/apps/details?id=com.analyticsvidhya.android&utm\\_source=blog\\_article&utm\\_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1](http://play.google.com/store/apps/details?id=com.analyticsvidhya.android&utm_source=blog_article&utm_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1)

**Share this:**

 (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/?share=linkedin&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/?share=facebook&nb=1>)

2K+

 (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/?share=google-plus-1&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/?share=twitter&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/?share=pocket&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/?share=reddit&nb=1>)

---

**Like this:**

Loading...

TAGS : [ELASTICNET REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ELASTICNET-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/elasticnet-regression/), [LASSO REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LASSO-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/lasso-regression/), [LINEAR-REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LINEAR-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/linear-regression/), [LOGISTIC REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/LOGISTIC-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/logistic-regression/), [MULTINOMIAL LOGISTIC REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MULTINOMIAL-LOGISTIC-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/multinomial-logistic-regression/), [MULTIPLE REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/MULTIPLE-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/multiple-regression/), [ORDINAL LOGISTIC REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ORDINAL-LOGISTIC-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/ordinal-logistic-regression/), [POLYNOMIAL REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/POLYNOMIAL-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/polynomial-regression/), [REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/regression/), [REGULARIZATION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/REGULARIZATION/\)](https://www.analyticsvidhya.com/blog/tag/regularization/), [RIDGE REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/RIDGE-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/ridge-regression/), [STEPWISE REGRESSION \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/STEPWISE-REGRESSION/\)](https://www.analyticsvidhya.com/blog/tag/stepwise-regression/).

NEXT ARTICLE

### **11 things you should know as a Data Scientist**

(<https://www.analyticsvidhya.com/blog/2015/08/ready-data-science-resources-common-questions-answered/>)

...

PREVIOUS ARTICLE

### **News – Great Lakes launches Analytics Program in Bangalore, India**

(<https://www.analyticsvidhya.com/blog/2015/08/great-lakes-analytics-program-bangalore/>)



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

## **Sunil Ray (https://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.

## **RELATED ARTICLES**

**KUNAL JAIN (https://www.analyticsvidhya.com/blog/2013/10/rules-creating-insightful-metrics/)**



(<https://www.analyticsvidhya.com/blog/2013/10/rules-creating-insightful-metrics/>)

### **Four rules for creating insightful and actionable reports (and metrics)**

(<https://www.analyticsvidhya.com/blog/2013/10/rules-creating-insightful-metrics/>)



(<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)

**TAVISH SRIVASTAVA (https://www.analyticsvidhya.com/blog/2014/11/applications-sas-iml-analytics/)**



(<https://www.analyticsvidhya.com/blog/2014/11/applications-sas-iml-analytics/>)

### **Applications of SAS Proc IML in Analytics**

(<https://www.analyticsvidhya.com/blog/2014/11/applications-sas-iml-analytics/>)



(<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)

**KUNAL JAIN (https://www.analyticsvidhya.com/blog/2015/12/news-bridge-school-management-launched-predictive-business-analytics-program/)**



(<https://www.analyticsvidhya.com/blog/2015/12/news-bridge-school-management-launched-predictive-business-analytics-program/>)

### **Interview with Dr. Bibek Banerjee – Dean, BRIDGE School of Management**



(<https://www.analyticsvidhya.com/blog/2015/12/news-bridge-school-management-launched-predictive-business-analytics-program/>)

**Kaggle Bike Sharing Demand Prediction – How I got in top 5 percentile of participants?**

**(<https://www.analyticsvidhya.com/blog/2015/06/solution-kaggle-competition-bike-sharing-demand/>)**

**Creating a flawless winning strategy in a Casino (BlackJack) using Data Science?**

**(<https://www.analyticsvidhya.com/blog/2017/04/flawless-winning-strategy-casino-blackjack-data-science/>)**

**Why Business Analytics Degrees Aren't Just a Fad**

**(<https://www.analyticsvidhya.com/blog/2015/05/business-analytics-degrees-arent-fad/>)**

This article is quite old and you might not get a prompt response from the author. We request you to post this comment on Analytics Vidhya's **Discussion portal** (<https://discuss.analyticsvidhya.com/>) to get your queries resolved

## 33 COMMENTS



**PRAKASHPCS**

[Reply](#)

August 14, 2015 at 6:02 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92695>).

Hi Sunil, Really a nice article for understanding the regression models. Especially for novice like me who are stepping into Analytic .



**SUNIL RAY**

[Reply](#)

August 14, 2015 at 7:37 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92707>).

Thanks for the comment ...



**SHIVAM KUSHWAHA**

[Reply](#)

August 14, 2015 at 6:24 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92699>).

Could you please, provide a material ( book/website) where I can understand concept underlying in such regression techniques.



**SUNIL RAY**

[Reply](#)

August 14, 2015 at 7:40 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92708>).

Thanks for the comment ...



You can read book “The Elements of Statistical Learning”, it has detailed explanation of these regression models.

Regards,  
Sunil



**JULIUS MKUMBO**

[Reply](#)

August 14, 2015 at 9:23 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92724>).

I agree with you Sunil, but before reading “The Elements of Statistical Learning”, I would recommend reading An Introduction to Statistical Learning: with application in R, which is more practical because you have to practise with R codes, or you may take Statistical Learning course which is offered by authors of these books, in addition they are inventors of some of these model as well (e.g. Lasso by Tibshirani).

regards,  
Julius



**KUMAR**

[Reply](#)

August 14, 2015 at 8:04 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92712>).

A good refresher on Regression techniques



**TOM**

[Reply](#)

August 14, 2015 at 8:49 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92718>).

Hi Sunil

Thanks for posting this. Very nice summary on a technique used so often but underutilised when looking at the different forms available. You wouldnt be interested in doing something similar for classification techniques..quite a few here as well..

Tom



**SUNIL RAY**

[Reply](#)

August 14, 2015 at 8:57 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92720>).

Thanks Tom...you can refer article on most common machine learning algorithms <http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/> (<http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>). Here I have discussed various types of classification algorithms like decision tree, random forest, KNN, Naive Bayes...

Regards,  
Sunil



**R RAJ KUMAR**

[Reply](#)

August 14, 2015 at 9:14 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92722>).

Dear sir,

Is the regression applied for predicting the results of a college for the coming academic year.



**PRATZJOSHI ([HTTP://TWITTER.COM/PRATZJOSHI](http://twitter.com/pratzjoshi))**

[Reply](#)

August 14, 2015 at 9:34 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92725>).

Hi Sunil,

The difference given between linear regression and multiple regression needs correction. When there is just one independent and one dependent variable, it is called "simple linear regression" not just linear regression.



**SUNIL RAY**

[Reply](#)

August 14, 2015 at 9:43 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92727>).

Hi PratzJoshi,

Thanks for highlighting it...

Regards,  
Sunil



**ROHAN POTA**

[Reply](#)

August 14, 2015 at 11:47 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92735>).

Hey, quite nice article. It did help me broaden my perspective regarding the regression techniques (specially ElasticNet), but still it would be nice to elucidate upon the differences between L1 and L2 regularization techniques. For this, <http://www.quora.com/What-is-the-difference-between-L1-and-L2-regularization> (<http://www.quora.com/What-is-the-difference-between-L1-and-L2-regularization>) will be very helpful. Though it could be incorporated into a new article I think.

**PAUL**[Reply](#)

August 14, 2015 at 1:17 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92743>).

I'm sorry but I am going to complain again.

This is an excellent article. However, when I go to save it or print it, it is a mess!

If I print from IE, the only browser allowed on my network, all the ads and hypertext links cover the article text; you cannot read the article.

I had suggested having a feature where you use a button to convert the article to a PDF, which can then be printed without the ads and hypertext. You did in once, then stopped. Please start again!

**SHASHI**[Reply](#)

August 14, 2015 at 6:27 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92754>).

Good consolidation of concepts... Sunil do you have a comprehensive data set upon which we can apply all/few of the above techniques and see how each regression behaves... thanks again

**SUNIL RAY**[Reply](#)

August 18, 2015 at 9:24 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-93043>).

Shashi,

You can look at scikit-learn example data sets..

Regards,  
Sunil



**LALIT SACHAN**

[Reply](#)

August 17, 2015 at 5:33 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-92929>).

Hi Sunil,

Nice compilation. Suggesting a correction , elastic net penalty has another parameter too and is written as  $\lambda * \text{summation} ([\alpha * L2 \text{ penalty} + (1-\alpha) * L1 \text{ Penalty} ] ) \dots$

Also quoting book by trevor & hastie “The elastic-net selects variables like the lasso, and shrinks together the coefficients of correlated predictors like ridge.”

like your posts, very informative.

-Lalit



**SEEMA**

[Reply](#)

August 18, 2015 at 6:59 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-93036>).

Hi,

Can you please explain this point mentioned under the logistic regression – multi collinearity part “However, we have the options to include interaction effects of categorical variables in the analysis and in the model.”



**SUNIL RAY**

[Reply](#)

August 18, 2015 at 9:22 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-93042>).

Hi Seema,

Read this article to understand the effect of interaction in detail.

<http://www.theanalysisfactor.com/interpreting-interactions-in-regression/>  
(<http://www.theanalysisfactor.com/interpreting-interactions-in-regression/>)

Regards,  
Sunil



**GAURAV** ([HTTP://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/08/COMPREHENSIVE-GUIDE-REGRESSION/](http://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/)) [Reply](#)

September 23, 2015 at 9:34 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-95801>)

Hi sunil,  
The article seems very interesting. Please can you let me know how can we implement Forward stepwise Regression in python as we dont have any inbuilt lib for it.



**SANGAMEAH**

[Reply](#)

October 27, 2015 at 2:02 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-98284>)

good article especially for computer science students. please provide this and further articles in pdf. thank you.



**BA**

[Reply](#)

January 11, 2016 at 7:31 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-103542>)

Thanks for the guide. Just one question : do you really think that “stepwise regression” is a type of regression? In my point of view, it’s just a compilation of methods for selecting relevant variables. And it is performed by making several successive (real) regression techniques (linear, polynomial, ridge or lasso...).



**MUKUND**

[Reply](#)

January 13, 2016 at 9:30 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-103647>)

Very useful article. Are there any specific types of regression techniques which can be used for a time series stationary data?



**BHANUSHREE**

[Reply](#)

February 3, 2016 at 1:41 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-105116>)

Hi,

Very nice article, crisp n neat! Thank you 😊

In 2. Logistic regression

We use log because while calculating MLE ( Maximum Likelihood Estimate) it is easy to differentiate and equate to 0.

Easy to differentiate log (a \* b) rather than just (a \* b)

Where a = P and b = 1-P

Correct me if I m wrong.



**VIMMI**

[Reply](#)

March 12, 2016 at 8:01 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-107158>)

i just love this article 😊 very very useful 😊



**ASESH DATTA (HTTP://WWW.ETECHGUYS.COM)**

[Reply](#)

May 1, 2016 at 1:48 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-110274>)

Sunil, Great feeling to get a modern insight to what I learnt 35 years ago. Excellent. Professional practicing today may have several question to clarify. We had no option to choose Data Analytics as a popular profession in 80's. It was mainly R&D to use regression techniques. Compliment to you for such a vast subject so lucidly worded and explained..

What fascinated me most, is you mention of a tutor teaching students in an institute – if outcome is continuous use linear and if it is binary, use logistics. That's simplistic for a reader to appreciate the importance of Regression. What I want to ask is as under:

In case of multiple independent variables, we can go with forward selection, backward elimination and step wise approach for selection of most significant independent variables.

Please let me know where to get little details on these?

Compliments once again. All is well.

Asesh Datta



**ROHIT**

[Reply](#)

July 3, 2016 at 1:37 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-113012>)

Thanks Sunil ,  
Useful article . Why is poisson regression not mentioned here.

---

**FAZAL AZEEM**[Reply](#)

November 14, 2016 at 12:19 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-118332>)

Hi sunil  
Very nice aticle and really very informative and interesting.

With regard  
Fazal azeem

---

**SMS**[Reply](#)

November 20, 2016 at 4:16 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-118577>)

Amazing article, broadens as once seemingly narrow concept and gives food for thought. Thank you.

---

**SUNIL**[Reply](#)

December 11, 2017 at 10:57 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-147041>)

Thanks SMS

---

**RAWIYAH ALRADDADI**[Reply](#)

January 29, 2018 at 4:00 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-151077>)

Thank you very much!

---

**RYAN ROSARIO**[Reply](#)

March 22, 2018 at 1:49 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-152080>)

This is an awesome article. I just want to point out something important for people that may be new to modeling as many data science students I've mentored get overwhelmed and confused about what the different types of regression truly are, and what the true definition of linear regression is.

These aren't really different types of regression models per se. This is a mix of different techniques with different characteristics, all of which can be used for linear regression, logistic regression or any other kind of generalized linear model.

Linear and logistic are the only two types of base models covered.

Polynomial is just using transformations of the variables, but the model is still linear in the beta parameters. Thus it is still linear regression. Even a weird model like  $y = \exp(a + bx)$  is a generalized linear model if we use the log-link for logistic regression. This yields  $\log y = a + bx$ . This is a concept that bewilders a lot of people.

Stepwise is just a method of building a model by adding and removing variables based on the F statistic.

Ridge, Lasso and ElasticNet aren't really regressions, they're penalties (regularization) on the loss function (OLS), or log-likelihood function (logistic and GLM). Hence, they are useful for other models that are distinct from regression, like SVMs. We usually say "linear regression with a regularization term/penalty."

To be technical, different regression models would be plain linear, logistic, multinomial, poisson, gamma, Cox, etc.

**VIDEO PRODUCTION BANGALORE**[Reply](#)

April 6, 2018 at 3:31 pm (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-152417>)

Hi Sunil,

Wonderfull put down information. Thanks for sharing this in detail.

**JAGADISH**[Reply](#)

May 25, 2018 at 11:29 am (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#comment-153540>)

Hi Sunil ,

Really a good gist on regression techniques. Thanks for sharing the article.

---





(<https://www.analyticsvidhya.com/datahack-summit-2018/?>

utm\_source=AVbannerdhslong)



## Forrester TE Report

Palo Alto Networks

Minimize Risk. Maximize ROI. Uncover Savings With the Security Operations Platform.

DOWNLOAD



## POPULAR POSTS

24 Ultimate Data Science Projects To Boost Your Knowledge and Skills (& can be accessed freely)  
(<https://www.analyticsvidhya.com/blog/2018/05/24-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/>)

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/>)

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

(<https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>)

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

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

A comprehensive beginner's guide to create a Time Series Forecast (with Codes in Python)

(<https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>)

20 Challenging Job Interview Puzzles which every analyst should solve atleast once

(<https://www.analyticsvidhya.com/blog/2016/07/20-challenging-job-interview-puzzles-which-every-analyst-solve-atleast/>)



## Forrester TE Report

Palo Alto Networks

Minimize Risk. Maximize ROI. Unleash Savings With the Security Operations Platform.

DOWNLOAD

## RECENT POSTS

---

**An Intuitive Guide to Interpret a Random Forest Model using fastai library (Machine Learning for Programmers – Part 2)** (<https://www.analyticsvidhya.com/blog/2018/10/interpret-random-forest-model-machine-learning-programmers/>)

OCTOBER 29, 2018

**MADRaS: A Multi-Agent DRiving Simulator for Autonomous Driving Research** (<https://www.analyticsvidhya.com/blog/2018/10/madras-multi-agent-driving-simulator/>)

OCTOBER 29, 2018





Analytics Vidhya

Learn everything about analytics

ANALYTICS VIDHYA

Learn everything about analytics

DATA SCIENCE

Learn everything about analytics

COMPANIES

Learn everything about analytics

JOIN OUR COMMUNITY :

Learn everything about analytics

Blog

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

Hackathon

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

Discussions

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

Apply Jobs

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

Leaderboard

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

Contact Us

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

Write for us

(<https://www.analyticsvidhya.com/about-me/write/>)

Post Jobs

(<https://trainings.analyticsvidhya.com/courses/course->)

Advertising

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

Reach Us

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

Referrals

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

Facebook

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

Twitter

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

LinkedIn

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

Google Plus

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

18460

Followers

3008

Followers

7513

Followers

48336

Followers

Subscribe to emailer

>