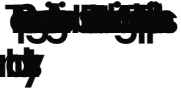




BOOST YOUR CAREER WITH BIG DATA ANALYTICS

LEARNING CORNER

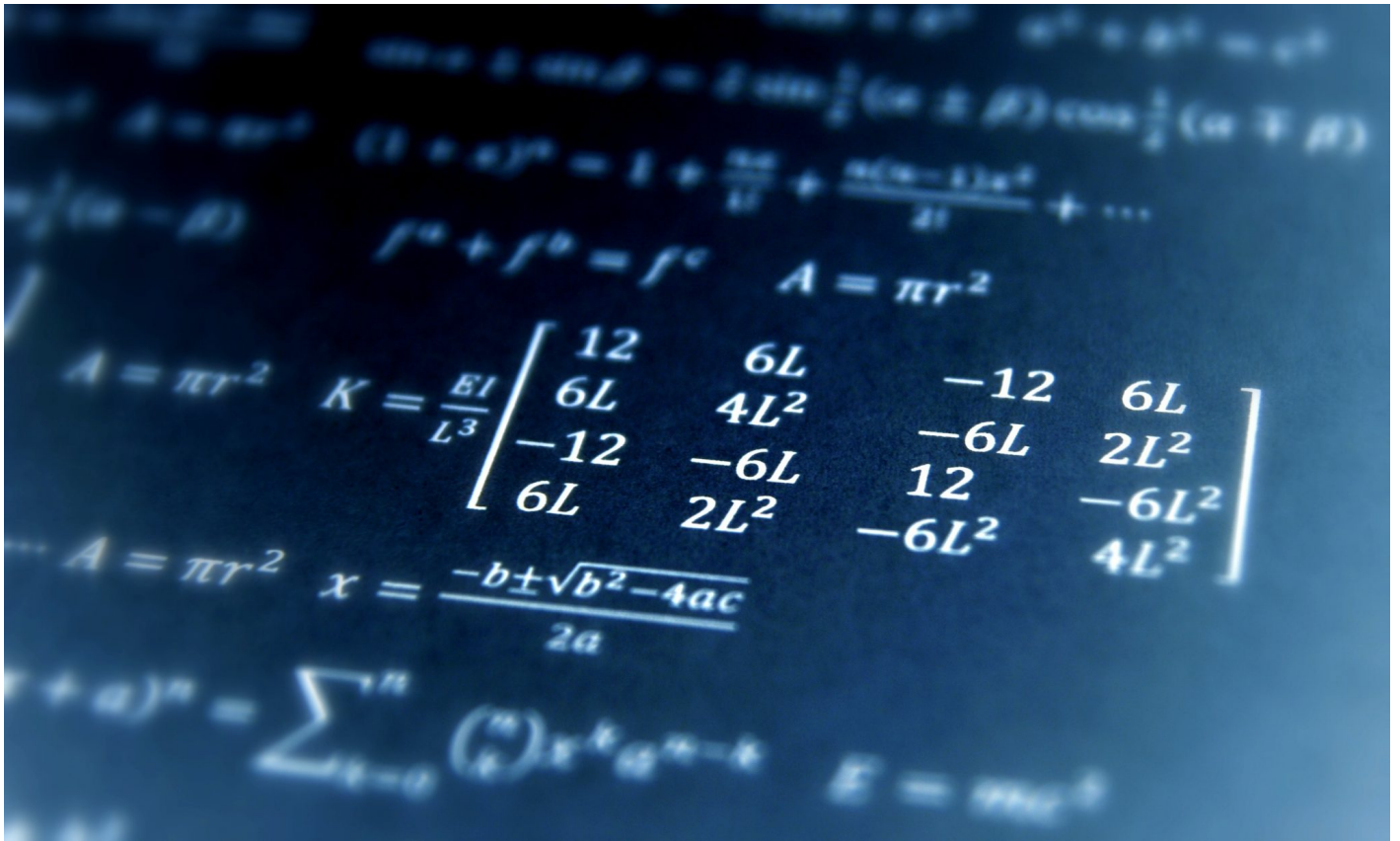


RICHA BHATIA - SEP 19, 2017



**PROVEN RECORD OF
CAMPUS PLACEMENT**

Advertisement

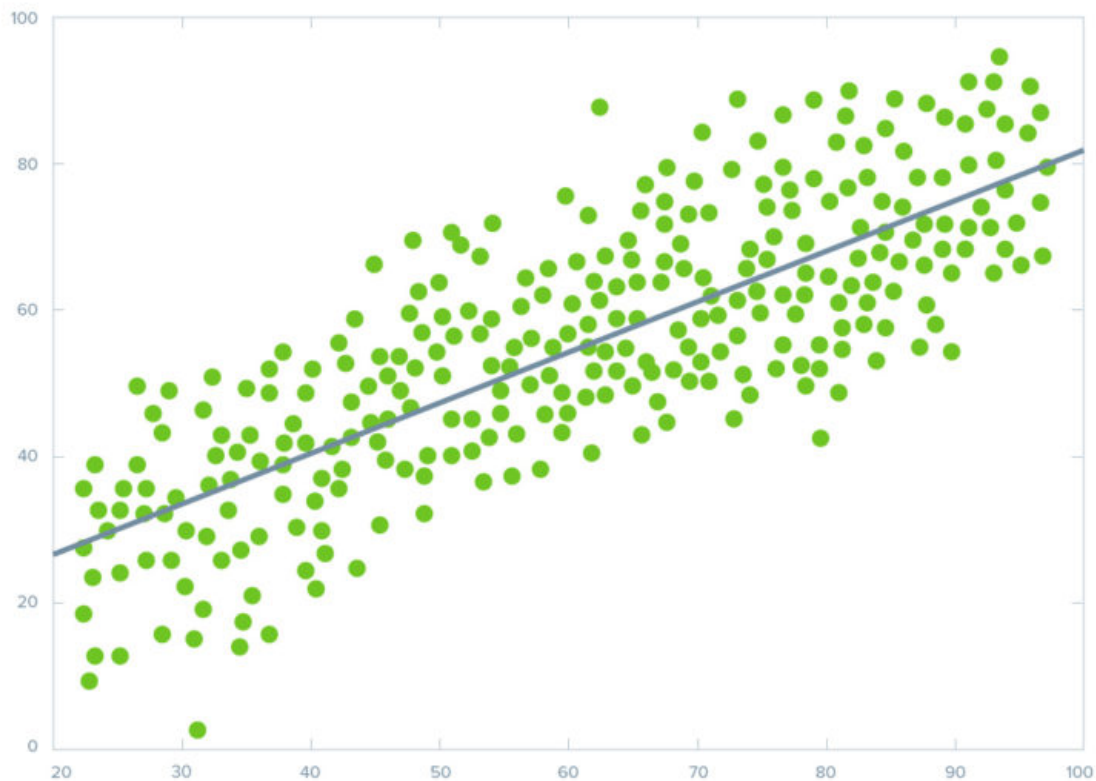


Regression algorithms fall under the family of Supervised Machine Learning algorithms which is a subset of machine learning algorithms. One of the main features of supervised learning algorithms is that they model dependencies and relationships between the target output and input features to predict the value for new data. Regression algorithms predict the output values based on input features from the data fed in the system. The go-to methodology is the algorithm builds a model on the features of training data and using the model to predict value for new data.

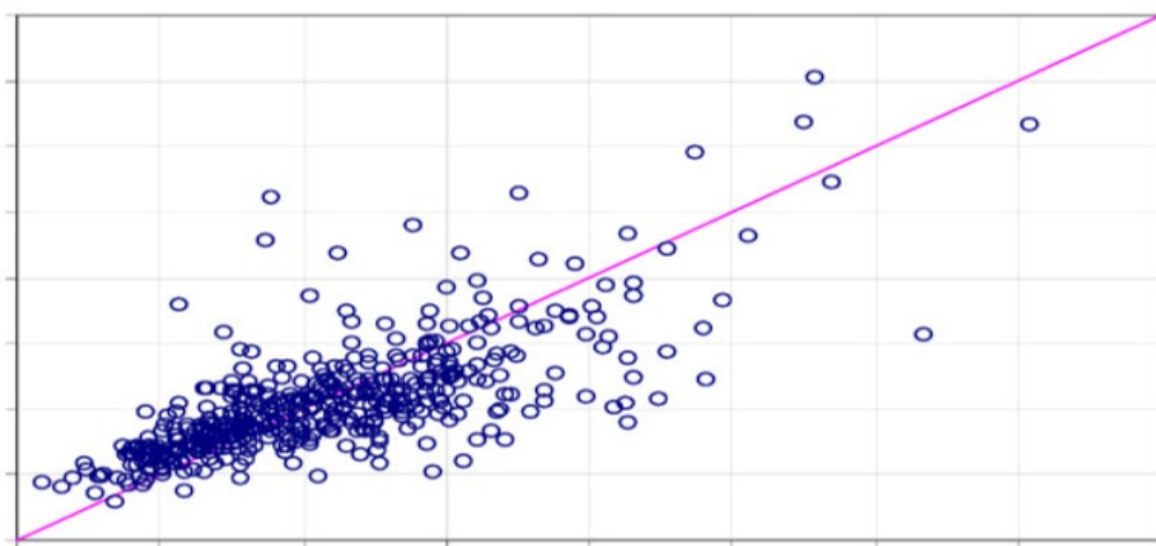
According to Oracle, here's a great definition of Regression – a data mining function to predict a number. Case in point, how regression models are leveraged to predict real estate value based on location, size and other factors. Today, regression models have many applications, particularly in financial forecasting, trend analysis, marketing, time series prediction and even drug response modeling. Some of the popular types of regression algorithms are linear regression, regression trees, lasso regression and multivariate regression.

Over 100,000 people subscribe to our newsletter.

See stories of Analytics and AI in your inbox.

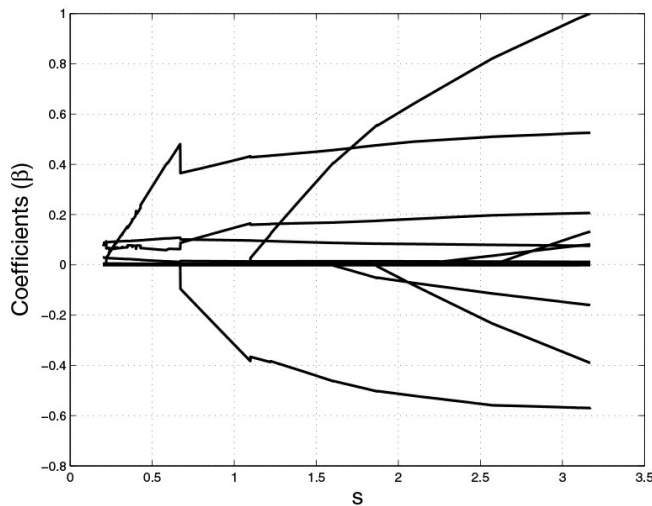


Analytics India Magazine lists down the most popular regression algorithms

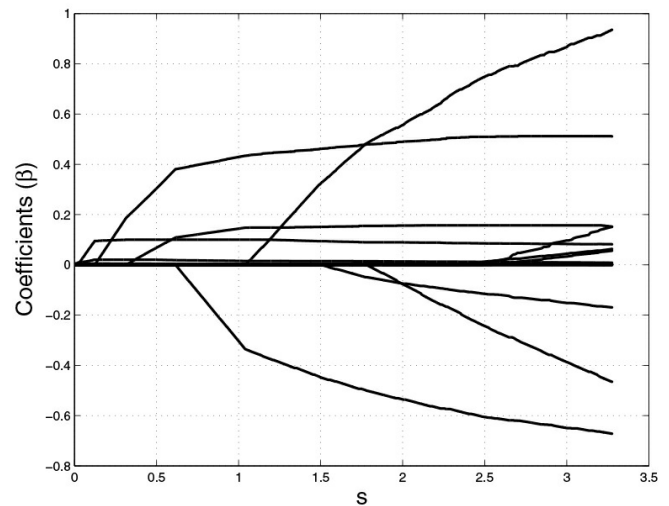


1. Simple Linear Regression model: Simple linear regression is a statistical method that enables users to summarise and study relationships between two continuous (quantitative) variables. Linear regression is a linear model wherein a model that assumes a linear relationship between the input variables (x) and the single output variable (y). Here the y can be calculated from a linear combination of the input variables (x). When there is a single input variable (x), the method is called a simple linear regression. When there are multiple input variables, the procedure is referred as multiple linear regression.

Application: some of the most popular applications of Linear regression algorithm are in financial portfolio prediction, salary forecasting, real estate predictions and in traffic in arriving at ETAs.



(a) original Lasso



(b) Forward-Lasso

2. Lasso Regression: LASSO stands for Least Absolute Selection Shrinkage Operator wherein shrinkage is defined as a constraint on parameters. The goal of lasso regression is to obtain the subset of predictors that minimize prediction error for a quantitative response variable. The algorithm operates by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward a zero.

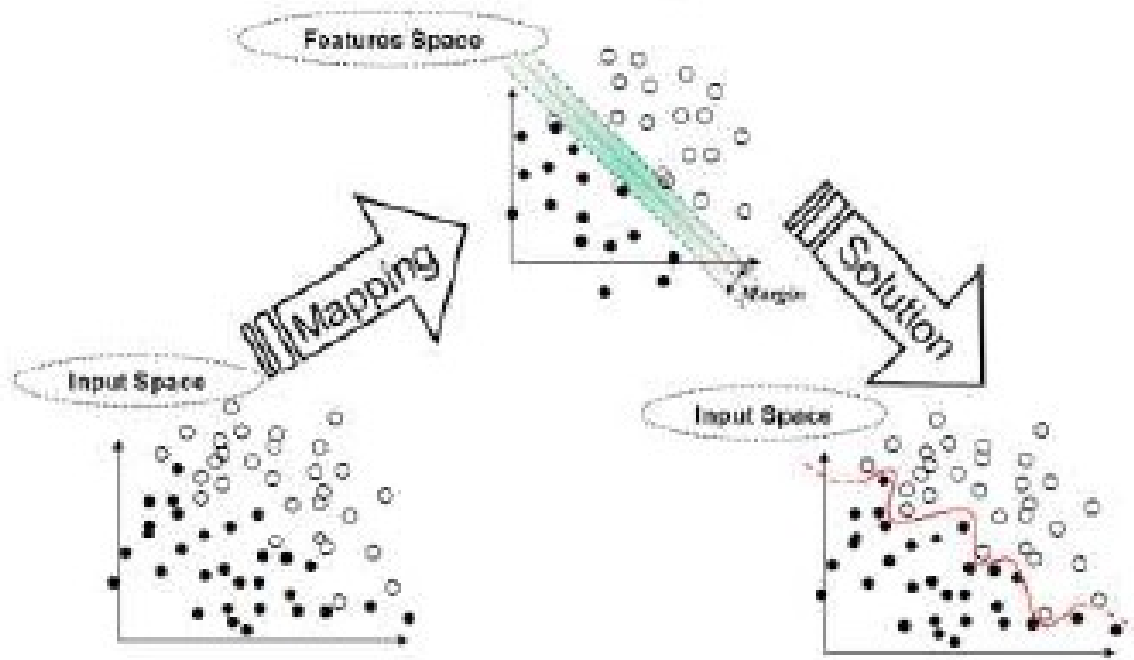
Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficients variables are most strongly associated with the response variable. Explanatory variables can be either quantitative, categorical or both. This lasso regression analysis is basically a shrinkage and variable selection method and it helps analysts to determine which of the predictors are most important.

Application: Lasso regression algorithms have been widely used in financial networks and economics. In finance, its application is seen in forecasting probabilities of default and Lasso-based forecasting models are used in assessing enterprise wide risk framework. Lasso-type regressions are also used to perform stress test platforms to analyze multiple stress scenarios.

3. Logistic regression: One of the most commonly used regression techniques in the industry which is extensively applied across fraud detection, credit card scoring and clinical trials, wherever the response is binary has a major advantage. One of the major upsides is of this popular algorithm is that one can include more than one dependent variable which can be continuous or dichotomous. The other major advantage of this supervised machine learning algorithm is that it provides a quantified value to measure the strength of association according to the rest of variables. Despite its popularity, researchers have drawn out its limitations, citing a lack of robust technique and also a great model dependency.

Application: Today enterprises deploy Logistic Regression to predict house values in real estate business, customer lifetime value in the insurance sector and are leveraged to produce a continuous outcome such as whether a customer can buy/will buy scenario.

The SVM algorithm

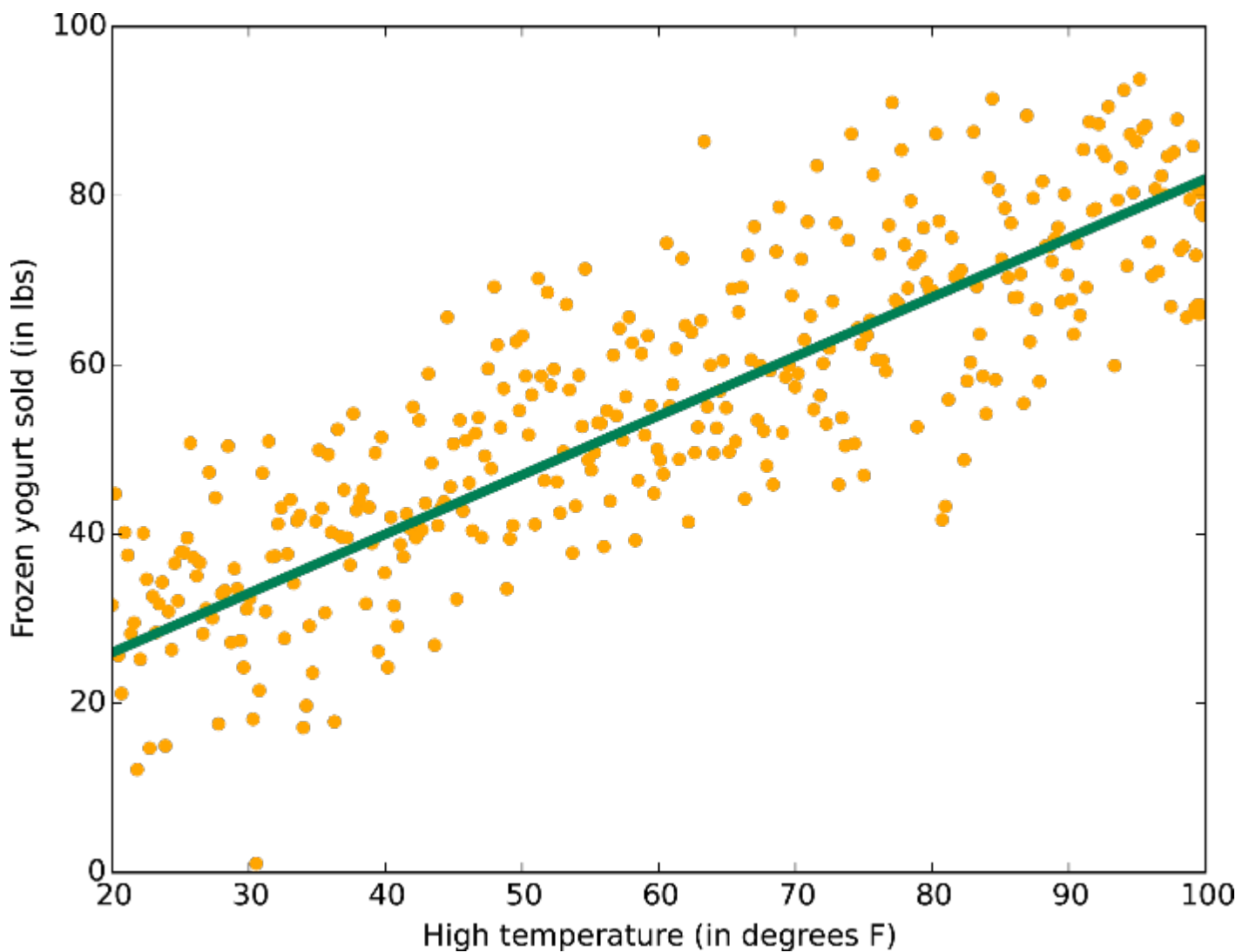


4. Support Vector Machines: Support Vector Machine (SVM) is another most powerful algorithm with strong theoretical foundations based on the Vapnik-Chervonenkis theory, as defined by Oracle docs. This supervised machine learning algorithm has strong regularization and can be leveraged both for classification or regression challenges. They are characterized by usage of kernels, the sparseness of the solution and the capacity control gained by acting on the margin, or on number of support vectors, etc. The capacity of the system is controlled by parameters that do not depend on the dimensionality of feature space. Since the SVM algorithm operates natively on numeric attributes, it uses a z-score normalization on numeric attributes. In regression, Support Vector Machines algorithms use epsilon-insensitivity (margin of tolerance) loss function to solve regression problems.

Application: support vector machines regression algorithms has found several applications in the oil and gas industry, classification of images and text and hypertext categorization. In the oilfields, it is specifically leveraged for exploration to understand the position of layers of rocks and create 2D and 3D models as a representation of the subsoil.

5. Multivariate Regression algorithm: This technique is used when there is more than one predictor variable in a multivariate regression model and the model is called a multivariate multiple regression. Termed as one of the simplest supervised machine learning algorithms by researchers, this regression algorithm is used to predict the response variable for a set of explanatory variables. This regression technique can be implemented efficiently with the help of matrix operations and in Python, it can be implemented via the "numpy" library which contains definitions and operations for matrix object.

Application: Industry application of Multivariate Regression algorithm is seen heavily in the retail sector where customers make a choice on a number of variables such as brand, price and product. The multivariate analysis helps decision makers to find the best combination of factors to increase footfalls in the store.



6. Multiple Regression Algorithm: This regression algorithm has several applications across the industry for product pricing, real estate pricing, marketing departments to find out the impact of campaigns. Unlike linear regression technique, multiple regression, is a broader class of regressions that encompasses linear and nonlinear regressions with multiple explanatory variables.

Application: Some of the business applications of multiple regression algorithm in the industry are in social science research, behavioural analysis and even in the insurance industry to determine claim worthiness.

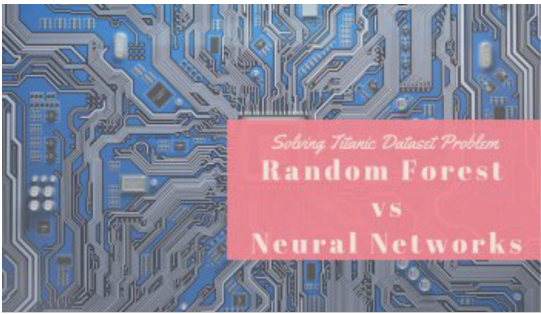
Related



10 Machine Learning Algorithms every Data Scientist should know

Jan 20, 2017

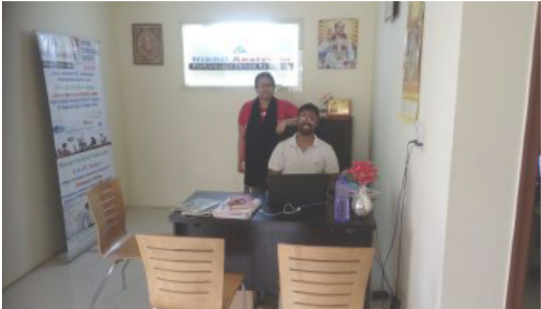
In "Learning Corner"



Solving The Titanic ML Survival Problem Using Random Forest vs Neural Networks on Tensorflow. Which One is Better?

Jan 8, 2018

In "Learning Corner"



In Review: Nikhil Analytics Certificate Program in Analytics Data Science & Machine Learning

Mar 14, 2017

In "Education"

Provide your comments below

0 comments

0 Comments

Sort by Newest

Add a comment...

Facebook Comments Plugin

Also Read [When Will Car Insurance in India Become Analytics Driven?](#)

Richa Bhatia

Richa Bhatia is a seasoned journalist with six-years experience in reportage and news coverage and has had stints at Times of India and The Indian Express. She is an avid reader, mum to a feisty two-year-old and loves writing about the next-gen technology that is shaping our world.

SHARE THIS

f

0

t

1

G+

0

in



PREVIOUS ARTICLE

Can Micro-segmentation Save CPG Companies Find Niche Customers In The Age Of Personalization

NEXT ARTICLE

In conversation with Shankar Viswanathan, Partner At ZS Associates

SEARCH



**BITS Pilani**
Pilani | Dubai | Goa | Hyderabad

**PG Certification in
Big Data
Engineering**
200+ hrs of learning

 **UpGrad**



APPLY NOW



Become a Data Scientist
with India's Best

POST GRADUATE
DIPLOMA IN DATA SCIENCE

[APPLY NOW](#)

 **MANIPAL**
ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the DGC Act, 1956)

ANALYTIX LABS

Your Gateway to great
career in Analytics!

[REGISTER](#)

Sign up for a free live online class

 **greatlearning**

LEADERS IN ANALYTICS



LATEST VIDEO

Behind The Scenes: Data Visualisation In Business Intelligence ft BDB



THE MACHINE CONFERENCE 2018

11th MAY 2018
HOTEL HYATT REGENCY, MUMBAI
WWW.THEMACHINECON.COM

Over 100,000 people subscribe to our newsletter.

See stories of Analytics and AI in your inbox.



CAREER WITH

[ABOUT US](#)

[US](#)

[ADVERTISE](#)

[COPYRIGHT](#)

[PRIVACY](#)

[TERMS OF USE](#)

[CONTACT US](#)

Copyright 2018 Analytics India Magazine Pvt Ltd. All RIGHTS RESERVED.

