

# Point to Point Regression Analysis Based Big-Data/ AWS Processing Using Machine Learning

**Dr. G.Murali**,(Professor, Department of Mathematics)

Malla Reddy University, Maisammaguda (V), Dhulapally(M), Medchal District, Telangana, INDIA

**Dr.Chintakindi Sanjay**, (Professor, Department of Industrial Engineering)

College of Engineering, King Saud University, Riyadh-11421, SAUDI ARABIA.

**Dr.Ajit Paul**, (Professor, Department of Mathematics)

Sam Higginbottom University of Agriculture, Technology and Sciences, Naini, Prayagraj, INDIA

**Dr. G. Deepa**, Assistant Professor,

Department of Mathematics) Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad,  
Telangana, INDIA

## ABSTRACT

Our Invention point to point regression analysis big data/AWS processing using machine learning is a stepwise mapped regression analysis-based global big data /AWS processing Technology. The invention is a technological advance creates many difficulties for R&D practitioners in analyzing novel technologies and also the information analysis is an effective tool in this situation. The invented technology is an AWS offers the broadest and deepest set of machines learning unique services and multiple supporting AWS/cloud infrastructure, putting machine learning in the hands of all developer, data scientist and expert practitioner. Big data/AWS is helping more than one hundred thousand customers accelerate their machine learning journey and the method includes the following steps:

1: The collecting data of factory operation parameters.

2: Total number of collected operation parameters.

The invention is collected operation defined parameters as dependent variables (Tx, ty, tz), using other operation fixed parameters as independent variables and writing single order linear equations, wherein the parameters are in a linear relationship. The invention is a method and system for providing big data/ AWS analytics framework for predictive and qualitative predictive analysis for social application developers data scientists and system unique engineers without having technology specific programming experience.

**KEYWORD:** Point to Point Regression, Analysis, Big-Data. AWS, Machine Learning, Framework, Predictive Analysis

## RESEARCH FIELD

Our Invention is related to a point-to-point regression analysis big data/ AWS processing using machine learning.

## RESEARCH BACKGROUND

The applying the selected ML algorithm to the selected local and global group of historical set of metadata to generate a unit of the selected group of historical metadata, obtaining

metadata of the active data, applying the unit to the metadata of the active data to generate a prediction of the level of performance with respect to the performance parameter and configuring the data processing system for analyzing the active data set based on the prediction. The concept of “Big data/ AWS” is already well-known in the field of information technology and Big data/AWS is a collection of tools, techniques and approaches used when data sets are large and complex that it becomes difficult or impossible to store, final query, analyze or process using real-time database management.

The challenge of handling big data include capture, organize, storage, search, sharing, analysis and visualization and the trend to larger data sets is due to the proliferation of data capture devices and the ease of capturing and entering data from a Tx, Ty, Tz variety of sources.

There are various aspects related to the big data/AWS analytics enterprise framework which need to be still addressed in order to increase the expectation, granularity and comprehensiveness of the data set in order to make the analysis of the data into useful, easy to handle and be cost-effective.

The Regression analysis is to process a kind of complex mathematical method of dependency relation between multivariate. This dependency relation is different from functional relationships It is that the latter clock wise reflects the strict interdependence between Tx, Ty, Tz variable and the former then shows a certain degree of undulatory property or randomness to independent variable. Each value, dependent variable can have multiple numerical value to correspond and the independent variable is that nonrandom variable, dependent variable is for becoming at random During amount, the relation analyzing to as regression analysis.

The Statistically grind invent carefully dependency relation and can use regression analysis and correlation analysis and also having to have certain between the variable of dependency relation uncertain Property but by constantly observing of phenomenon can be explored the statistical law between them this kind of statistical law is referred to as returning Relation.

## RESEARCH OBJECTIVES

1. The other another object of the invention is to also using some of the collected operation defined parameters as dependent variables (tx, ty, tz), using other operation fixed parameters as independent variables and writing linear equations, wherein the parameters are in a linear relationship.
2. The other another object of the invention is to method and system for providing big data analytics framework for predictive and qualitative predictive analysis for social application developers data scientists and system unique engineers without having technology specific programming experience.
3. The other another object of the invention is to the framework contains adapters for the software engineers to configure the big data hub, wherein these hardware/software engineers can easily share, defined store, process and predict

unique functionalities with the set of data scientists and user interface (UI) developers.

4. The other another object of the invention is to perform big data analytics along with predictive features such as personalization, recommendation and content discovery which are essential factors in terms of end usage of the whole analysis and also outcome of the analysis.
5. The other another object of the invention is to provide end to end big data application development, faster and intelligent analytical rules deployment, and user-friendly interfaces for meaningful visualization.
6. The other another object of the invention is to provide a unified framework to configure the big data hub, enable data processing technology and integrate intelligent and predictive analytics engines.

### RESEARCH SUMMARY

The selecting a group of Indian historical metadata from a plurality of groups of historical metadata of datasets that have previously been analyzed using the data processing system to provide a selected group of historical metadata. The selected machine learning algorithm to the selected group of historical metadata to generate a model of the selected group of historical metadata obtaining metadata of the active data and also applying the unit to the metadata of the active data to generate a prediction of the level of performance with respect to the performance parameter and configuring the data processing system for analyzing the active data based on the prediction.

The ingenious subject material a system contains a processor and a memory coupled to the processor that contains machine readable program code embodied within the memory that once dead by the processor causes the processor to perform operations comprising and conjointly receiving a giant information/AWS information set comprising active data receiving a call for participation to predict grade of performance with respect to a performance parameter of a data processing system in analyzing the active data.

The selecting a machine learning algorithm of machine learning algorithms based on the performance parameter to obtain a selected machine learning algorithm selecting a group of historical set of metadata from a groups of historical metadata of datasets that have previously been analyzed using the data processing system to provide a selected group of historical metadata and also applying the selected machine learning algorithm to the selected group of historical metadata to generate a model of the selected group of historical metadata.

The invention is an obtaining metadata of the active data applying the unit to the metadata of the active data to generate a prediction of the level of performance with respect to the performance parameter and configuring the data processing system for analyzing the active data based on the prediction.

A kind of big data processing method based on stepwise regression analysis, follows the steps below:

S1: The collect the data of plant operating parameters, and the operating parameter collected is numbered, be designated as 1 respectively, 2, 3, .....、 n ;

S2: The data process using an operating parameter part for above-mentioned collection as dependent variable, other operating parameter is as certainly Variable, linear between each parameter, it is listed below equation:

Wherein:  $x$  is independent variable is dependent variable is the coefficient before independent variable is intercept ;

S3: The data of aforesaid equation and correspondence are imported Malware software one by one, carries out stepwise regression analysis computing, Calculate the coefficient before separate equation independent variable and intercept ;

**S4: interpretation of result:**

(1) when the coefficient before independent variable is zero, illustrating that this independent variable does not produce impact to corresponding dependent variable, coefficient is just the both forward and reverse directions of negative response impact, the size of the size reaction impact of coefficient, therefore can find out shadow by above-mentioned operation result Ring maximum operating parameter and the operating parameter number of dependent variables ;

(2) the above-mentioned equation group trying to achieve coefficient is changed as shown below:

**Equation cluster is writing the shape form of matrix as, such as like following formula:**

With MALWARE software solution above formula, calculating needs add parameter value scope and is as constraints, the solution obtained the optimal value of respective operations parameter.

The invented system and methods are to be understood that this invention is not limited to the particular systems and methodologies described, as there can be multiple possible of the present invention which are not expressly illustrated in the. It is also to be understood that the terminology used in the description is for the purpose of describing the particular versions only and is not intended to limit the scope of the present invention.

The invention provides a framework which comes with built-in adapters for NoSQL Databases such as MongoDB and Cassandra, processing engines such as Hadoop and Storm and Analytics engines such as a PMML Scoring Engine and an 'R' adapter etc. Further the framework is fortified with plug-ins including Support Vector Machines for Classification.

The invention will become apparent from the detailed description provided hereinafter. It should be understood that the detailed description and specific examples, while indicating the preferred embodiment of the invention, are intended for purposes of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description

**RESEARCHG BRIEF DESCRIPTION**

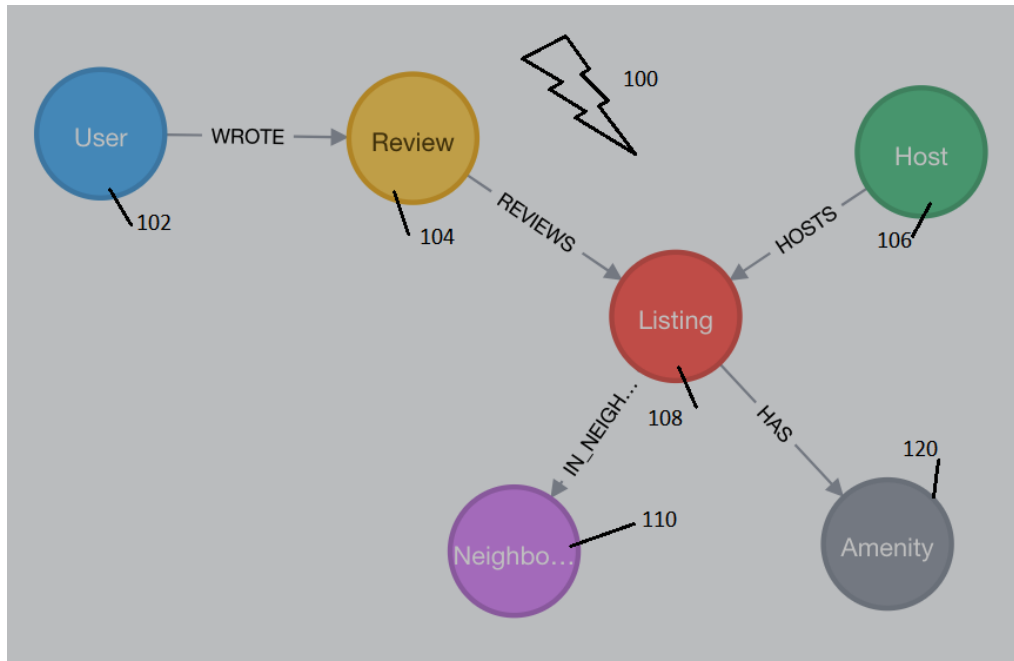
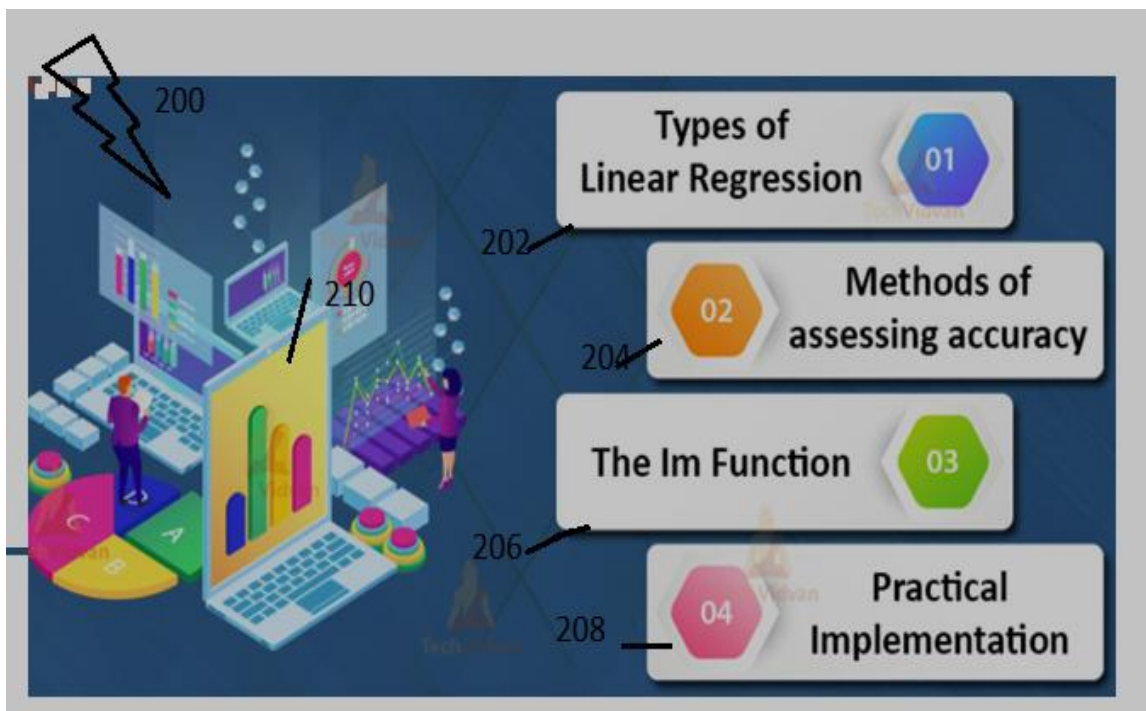


FIG. 1: Types of Regression in Data Science.



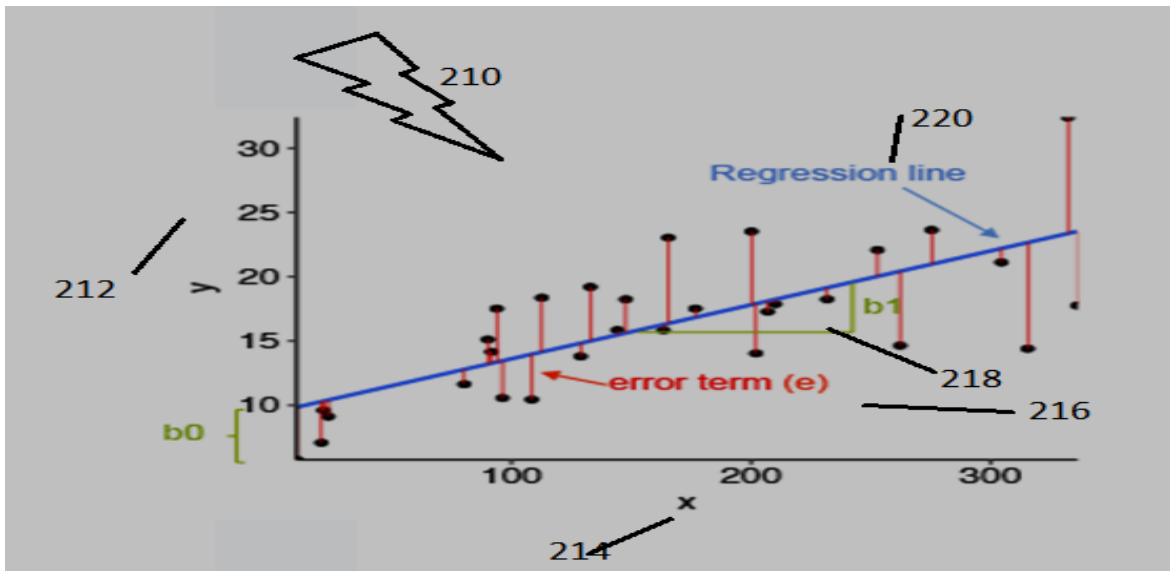


FIG. 2: Linear Regression in R using lm () Function.

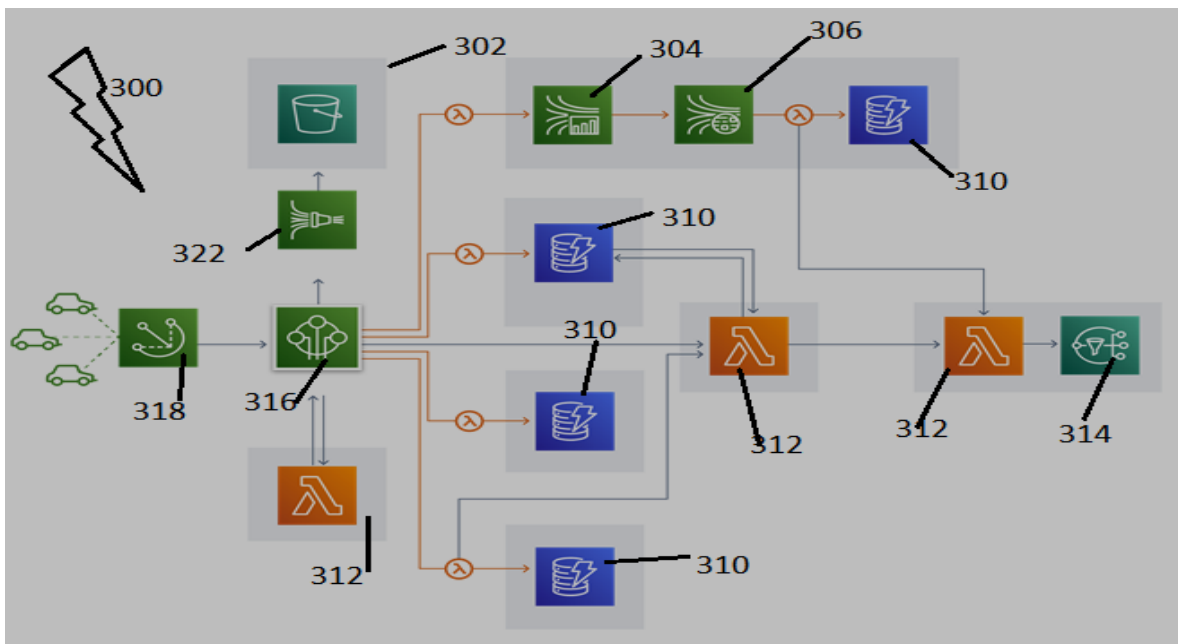


FIG.3: AWS block diagram of a core architecture.

### RESEARCH DESCRIPTION

The following elaborated description, varied specific details area unit set forth to supply an intensive understanding the invention. However, it'll be understood by those consummate within the art that the invention is also practiced while not these specific details. In some instances, well-known strategies, procedures, elements and circuits haven't been delineate well therefore as to not obscure this revelation.



It is supposed that disclosed herein is enforced on an individual basis or combined in any method and/or combination. Aspects delineate with relevance be incorporated in numerous though not specifically delineate relative to that. That is, all unit and options of any embodiments is combined in any method and/or combination.

As used herein, a “service” includes, however isn't restricted to, a software system and hardware service, like cloud services within which software system, platforms, and infrastructure area unit provided remotely through, for instance, the web. A service is also provided exploitation software system as a Service (SaaS), Platform as a Service (PaaS), and/or Infrastructure as a Service (IaaS) delivery models. within the SaaS model, customers usually access software system residing within the cloud employing a skinny shopper, like a browser, for instance. within the PaaS model: The client generally creates and deploys the software system within the cloud typically exploitation tools, libraries, and routines provided through the cloud service supplier. The cloud service supplier could offer the network, servers, storage, and different tools accustomed host the customer's application(s). within the IaaS model, the cloud service supplier provides physical and/or virtual machines beside hypervisor(s).

The client installs package pictures beside application software system on the physical and/or virtual infrastructure provided by the cloud service supplier.

FIG. 1:

Regression analysis may be a sort of prophetic modelling technique that investigates the connection between a dependent (target) and variable quantity (s) (predictor).

This technique is employed for prognostication, statistic modelling and finding the causative result relationship between the variables. as an example, relationship between rash driving and variety of road accidents by a driver is best studied through regression. multivariate analysis is a very important tool for modelling and analysing knowledge. Here, we tend to match a curve / line to the information points, in such a fashion that the variations between the distances of information points from the curve or line is reduced. I'll make a case for this in additional details in returning sections.

We should somehow quantify a relationship between the unknown variable (price) and therefore the legendary variables (bedrooms, bathrooms, and variety of guests accommodated). several worth predictors within the housing market use total variety of rooms to predict housing worth. They assume that the link between worth and variety of rooms is linear. we'll extend this idea to the short-term rental market. regression toward the mean is one tool want to analyse the relationships between variables, and here it is wont to quantify the link between rental worth per night and therefore the total variety of rooms within the listing.

As mentioned on top of, multivariate analysis estimates the link between 2 or additional variables. Let's perceive this with a straightforward example:

Let's say, you wish to estimate growth in sales of an organization supported current economic conditions. you've got the recent company knowledge that indicates that the expansion in sales is around 2 and 0.5 times the expansion within the economy. mistreatment this insight, we are able to predict future sales of the corporate supported current & past info.

There are multiple advantages of mistreatment multivariate analysis. they're as follows:

1. It indicates the numerous relationships between variable and experimental variable.
2. It indicates the strength of impact of multiple independent variables on a variable.

Regression analysis conjointly permits US to check the results of variables measured on totally different scales, like the result of worth changes and therefore the variety of promotional activities. These advantages facilitate market researcher's / knowledge analyst's / knowledge scientists to eliminate and appraise the most effective set of variables to be used for building prophetic unit.

There are a unit numerous varieties of regression techniques obtainable to form predictions. These techniques area unit largely driven by 3 metrics (number of freelance variables, kind of dependent variables and form of regression line). We'll discuss them well within the following sections. For the artistic ones, you'll even cook up new regressions, if you're feeling the necessity to use a mix of the parameters on top of, which individuals haven't used before. however, before you begin that, allow us to perceive the foremost normally used regressions.

### **Linear Regression**

It is one among the foremost wide better-known modeling technique. regression toward the mean is sometimes among the primary few topics which individuals decide whereas learning prognosticative modeling. during this technique, the variable is continuous, freelance variable(s) are often continuous or separate, and nature of curve 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).

1. There must be linear relationship between independent and dependent variables
2. Multiple regression suffers from multicollinearity, autocorrelation, heteroscedasticity.
3. statistical regression is extremely sensitive to Outliers. It will very have an effect on the curve and eventually the forecasted values.
4. multiple regression will increase the variance of the constant estimates and build the estimates terribly sensitive to minor changes within the model. The result's that the constant estimates are unstable.



5. 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

FIG:2: In the generalized linear models' tutorial, we tend to learned regarding numerous GLM's like rectilinear regression, supplying regression, etc. during this tutorial of the TechVidvan's R tutorial series, we tend to ar progressing to check up on rectilinear regression in R well. we are going to learn what's R rectilinear regression and the way to implement it in R. we are going to check up on the smallest amount sq. estimation technique and can conjointly find out how to visualize the accuracy of the model.

Linear regression in R could be a technique want to predict the worth of a variable mistreatment the value(s) of 1 or additional input predictor variables. The goal of rectilinear regression is to ascertain a linear relationship between the specified output variable and therefore the input predictors.

To model an eternal variable Y as a operate of 1 or additional input predictor variables Xi, so the operate may be wont to predict the worth of Y once solely the values of Xi are proverbial. the final variety of such a linear relationship is:

$$Y = \beta_0 + \beta_1 X$$

FIG:3: design diagrams are a good thanks to communicate your style, deployment, and topology. On this page you'll realize an officer assortment of AWS design Icons (formerly straightforward Icons) that contain AWS product icons, resources, and alternative tools to assist you build diagrams.

Customers and partners are allowable by AWS to use the resources below to make design diagrams. The icons are designed to be easy in order that you'll simply incorporate them in your diagrams and place them in your whitepapers, shows, datasheets, posters, or any technical material.

Browse AWS reference design diagrams and learn the way to creator additional expeditiously and effectively on AWS with our professional steering and best practices within the AWS design Center.

The enterprise framework core design explains the input-output relationship between the core functionalities. The line illustrates the association to the four functions indicates the ensue store to method, method to predict, and at last visualize. the shop perform interacts with the method, predict and visualize functions through the enterprise framework and contrariwise.

The fictitious the framework contains adapters for the software system engineers to set up the large information hub, whereby these hardware/software engineers will simply share, outlined store, method and predict distinctive functionalities with the set of knowledge scientists and program (UI) developers.

## **RESEARCH CLAIMS**

1. Our Invention point to point regression analysis big data/AWS processing using machine learning is a stepwise mapped regression analysis-based global big data /AWS processing Technology. The invention is a technological advance creates many difficulties for R&D practitioners in analyzing novel technologies and also the information analysis is an effective tool in this situation. The invented technology is

- a AWS offers the broadest and deepest set of machine learning unique services and multiple supporting AWS/cloud infrastructure, putting machine learning in the hands of all developer, data scientist and expert practitioner. Big data/AWS is helping more than one hundred thousand customers accelerate their machine learning journey. The invention is collected operation defined parameters as dependent variables (tx, ty, tz), using other operation fixed parameters as independent variables and writing single order linear equations, wherein the parameters are in a linear relationship. The invention is a method and system for providing big data/ AWS analytics framework for predictive and qualitative predictive analysis for social application developers data scientists and system unique engineers without having technology specific programming experience.
2. According to claim1# the invention is to a point-to-point regression analysis-based big data processing using machine learning is a stepwise complex regression analysis-based global big data processing method and Big data/AWS is helping more than one hundred thousand customers accelerate their machine learning journey.
  3. According to claim1,2# the invention is to an also using some of the collected operation defined parameters as dependent variables (tx, ty, tz), using other operation fixed parameters as independent variables and writing linear equations, wherein the parameters are in a linear relationship.
  4. According to claim1,2,3# the invention is to a method and system for providing big data analytics framework for predictive and qualitative predictive analysis for social application developers data scientists and system unique engineers without having technology specific programming experience.

## Reference:

1. [https://aws.amazon.com/free/machine-learning/?trk=ps\\_a134p000006gEZhAAM&trkCampaign=acq\\_paid\\_search\\_brand&sc\\_channel=PS&sc\\_campaign=acquisition\\_IN&sc\\_publisher=Google&sc\\_category=Machine%20Learning&sc\\_country=IN&sc\\_geo=APAC&sc\\_outcome=acq&sc\\_detail=%2Baws%20%2Bmachine%20%2Blearning&sc\\_content=Machine%20Learning\\_bmm&sc\\_matchtype=b&sc\\_segment=477000499331&sc\\_medium=ACQ-P|PS-GO|Brand|Desktop|SU|Machine%20Learning|Solution|IN|EN|Text&s\\_kwid=AL!4422!3!477000499331!b!!g!!%2Baws%20%2Bmachine%20%2Blearning&ef\\_id=Cj0KCQiA15yNBhDTARIsAGnwe0WkmlNdhubunxavJCeeCJ-Dtd85joENglP\\_zmYh6zhlyeo8M\\_-K6kaAm3fEALw\\_wcB:G:s&s\\_kwid=AL!4422!3!477000499331!b!!g!!%2Baws%20%2Bmachine%20%2Blearning](https://aws.amazon.com/free/machine-learning/?trk=ps_a134p000006gEZhAAM&trkCampaign=acq_paid_search_brand&sc_channel=PS&sc_campaign=acquisition_IN&sc_publisher=Google&sc_category=Machine%20Learning&sc_country=IN&sc_geo=APAC&sc_outcome=acq&sc_detail=%2Baws%20%2Bmachine%20%2Blearning&sc_content=Machine%20Learning_bmm&sc_matchtype=b&sc_segment=477000499331&sc_medium=ACQ-P|PS-GO|Brand|Desktop|SU|Machine%20Learning|Solution|IN|EN|Text&s_kwid=AL!4422!3!477000499331!b!!g!!%2Baws%20%2Bmachine%20%2Blearning&ef_id=Cj0KCQiA15yNBhDTARIsAGnwe0WkmlNdhubunxavJCeeCJ-Dtd85joENglP_zmYh6zhlyeo8M_-K6kaAm3fEALw_wcB:G:s&s_kwid=AL!4422!3!477000499331!b!!g!!%2Baws%20%2Bmachine%20%2Blearning).
2. <https://aws.amazon.com/blogs/machine-learning/building-machine-learning-workflows-with-aws-data-exchange-and-amazon-sagemaker/>
3. <https://docs.aws.amazon.com/machine-learning/latest/dg/machinelearning-dg.pdf>
4. <https://docs.aws.amazon.com/machine-learning/latest/dg/machinelearning-dg.pdf>