International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Horoscope Analysis and Astrological Prediction using Biased Logistic Regression (BLR)

Rahul Shajan, Gladston Raj S

Abstract*: Majority of people have connected their day to day life with Astrology. Normally the astrological factors would be checked for the prediction of birth details, house warming, marriages etc. The perception of astrology may vary for people. Due to the innumerable instance that provides the influence of astrology in our day to day life, some consider it as divine guidance or a vital source of information that has much to offer apart from solutions to problems in life. People are considering astrology as the traditional science of India. All the popular television channels and newspapers include the astrology programs in their schedules. That is only because of the wide popularity and acceptability of astrology in the minds of common man. Still, modern science is not considering astrology as a proven science. All the astrological calculations and predictions are still performed in the traditional way. Modern scientific methods or inventions are not at all used by the astrologers. In this paper, there is an attempt to connect traditional science with the modern science. Nowadays data mining is a most popular research area. Application of data mining in astrology is a new concept and that type researches are not popular now. If it is possible to prove that the data mining techniques are effective for astrological prediction then it will help to increase the acceptance of astrology in the area of modern science. Thus, there is an effort to increase the credibility of traditional science astrology by connecting it with modern science through the possibilities of Data mining.

Keywords: About four key words or phrases in alphabetical order, separated by commas.

I. INTRODUCTION

Nowadays people are so anxious to know about the things which are going to happen in the future life. Astrology is the only solution for that. Astrology is a traditional science very much popular all around the world, especially in India. About the origin of astrology, a lot of stories are there. Vedas are considered as holy text of Indian culture. Hindu saints consider astrology as a part of Vedas. Astrology is popularly known as "vedanga" means body part of Veda [1]. The life is a mixture of happiness and sadness. No one will get full of happiness or sadness in their life.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Rahul Shajan*, presently doing Phd in School of Computer Sciences, Mahatma Gandhi University, Kottayam
Dr.Gladston Raj Professor & Head, Department of Computer Science @ Government College Nedumangad.

Astrology tells that the good and bad thing happening in the life of a human being is depending on the position of planets at the birth time. Simply astrology is the study of the influence of macrocosm on the microcosm. Astrology never considers the entire planets in the universe. Only the seven planets; Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn and three special positions in the space called Rahu, Kethu, Gulikan are considered. Astrological researchers believe that the influence of these planets will decide the life span of an individual. That's why the "Subhamuhoortham", which means best time, is sought to start the good events.

Astrology has six branches called Jathakam, Golam, Nimitham, Prasnam, Muhoortham, Ganitham[3]. Within these six branches jathakam and prasnam are used to predict the good and bad things that might happen in the life of an individual. Both these branches consider the Astrological chart for the prediction. The astrological chart is a representation of the position of planets in a particular time. It is a circular diagram with twelve regions that is the 360 degree circle and is divided into twelve regions with 30 degrees each [8][2]. Each region is known as Houses, Rasi or Bhavas. These houses represent the twelve regions in the universe. Each house has a name and it is shown in the below table.

Table: I House numbers and names

House number	Name
1	Aries
2	Taurus
3	Gemini
4	Caner
5	Leo
6	Virgo
7	Libra
8	Scorpio
9	Sagittarius
10	Capricorn
11	Aquarius
12	Pisces

All the planets considered in astrology will reside either in one of the houses of astrological chart. The circular astrological chart is tedious to handle by the astrologers so a rectangular representation has been used for the analysis.The representation of the astrological chart is shown below.

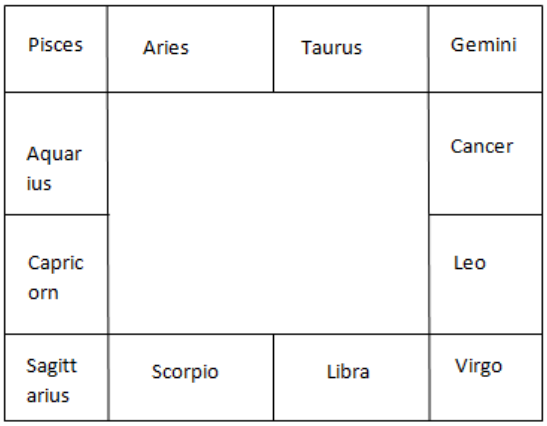

Figure 1: Rectangular representation of Astrological Chart.

When a child is born, the astrology believers will generate the horoscope of that child by consulting an astrologer. Horoscope is a document that describes all the astrological features of a child by analyzing the birth date, birth time and birth place[4]. The predictions in the horoscope are done by the analysis of the astrological chart of a child. As mentioned earlier the past, present, and future of an individual can be predicted by the analysis of the astrological chart. The astrological chart is the basic tool used for astrological prediction. Besides the planets, there is a crucial factor called "Lagnam" is also considered in the Birth chart. Lagnam is the name of the House (Rasi) which is raised at the time of birth [3][4]. The Lagnam may be either one of the twelve houses in the astrological chart. The house which is selected as Lagnam will consider as Bhavam one. For example, if Taurus is the Lagnam then Bhavam one is Taurus and Bhavam two is Gemini and so on.

Example of a Birth chart

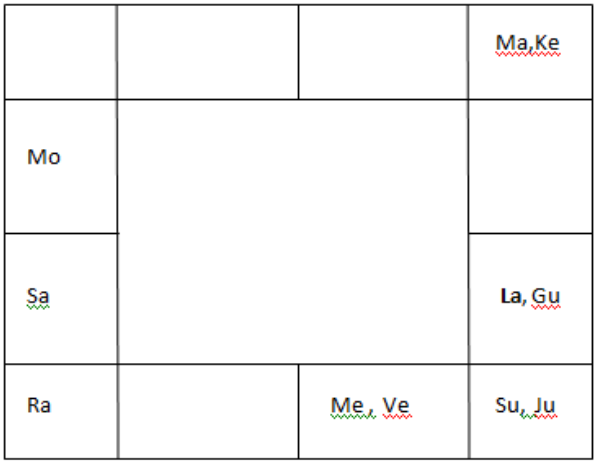

Fig 2: Birth chart of an individual with date of birth 09/10/1992, birth time 4.30 am and birth place Munnar.

In the above birth chart, there are some symbols which is used to represent the planets and Lagnam, the details as follows

Table II: List of planets and its representation symbols.

Planet	Symbol
Lagnam	La
Sun	Su
Moon	Mo
Mars	Ma
Mercury	Me
Jupiter	Ju
Venus	Ve
Saturn	Sa
Rahu	Ra
Kethu	Ke
Gulikan	Gu

From the example, it is clear that the Lagnam or Bhavam one is Leo and the Gulikan is residing in Bhavam one. Virgo is Bhavam 2 and it holds both Sun and Jupiter. In Bhavam three there are Mercury and Venus. Also some Bhavams like Scorpio, Pisces etc does not hold any planets. In astrology, the houses without planet are also considered for the prediction. In this study, this astrological knowledge is used for the classification purpose.

We know that a lot of classification techniques are available in data mining. Some of the existing classification algorithms also show results [5][6]. But in this review, it is identified that the Logistic Regression is the most suitable algorithm for the classification of astrological data. But the basic LR is not giving an expected result.

A. *Logistic Regression*

Logistic regression is a statistical method for analyzing a dataset; that contains one or more independent variables to determine the output [11]. The dependent variable is the class output and the independent variables are the features or attributes. The Logistic curve is a common 'S' shaped sigmoid curve.

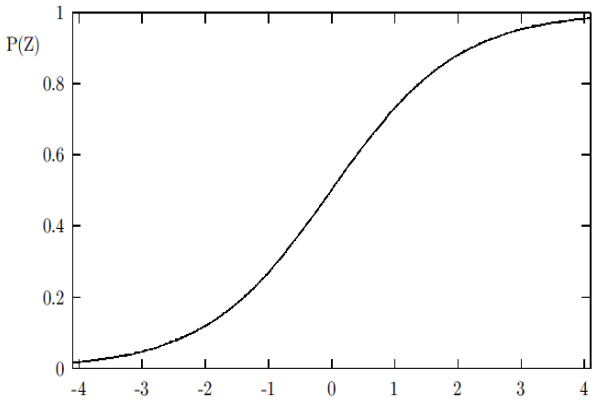

Figure 3: Sigmoid curve.

The sigmoid curve can be represented by the sigmoid function

$$P(Z)=\frac{(Exp\ Z)}{1+Exp\ Z} \qquad (1)$$

P behaves like the distribution function of a symmetrical density, with midpoint zero, as Z moves through the real number axis; P rises monotonically between the bounds of 0 and 1. There is a cost function and it is a single value. It is a measure of "How good" a neural network did with respect to its given training sample and the expected output.

The cost function C (W,B, S^T,E^T)

Where W is the neural network weight, B is Neural network biases, S^T is the input of a single training sample and E^T is the desired output.

Actually, a cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and Y [10]. It is expressed as a difference or distance between the predicted value and the actual value. In larger data set Logistic regression uses the stochastic gradient descent as the classifier function. SGD consider only one random point while changing weights, unlike gradient descent which considers the whole training set [12]. In this paper, there is an attempt to apply the logistic regression on astrological data for classification and prediction.

II. METHODOLOGY

The most challenging phase in this study is to generate the horoscope details of a number of people who are famous in different areas. Every one wishes to know the most suitable job area that he/she can perform well. Astrology tells that the horoscope that has a great influence on the selection of a job by an individual [7][9]. Medical (Nurse), Sports and Literature are the different areas considered in this study. A sample of 200 people's horoscope details are considered, where 50 people are nurses, 50 are sports men's, 50 are writers and 50 belongs to other groups. The details like date of birth, time of birth and the place of birth where collected and based on these details the astrological chart and the birth star was generated by the help of astrology software. The tabulation of these astrological features is an important job, so a new data model was suggested with 134 attributes. Astrological prediction always performed by analyzing the astrological chart. In an astrological chart, there are 12 houses that may or may not contain planets. The houses with and without planets should consider for the analysis, so the presence and absences of planets in houses are equally important. All the possibilities are considered in the proposed data model. The description of the data model is given below:-

Table III: Attributes and its descriptions.

Attribute	Sub division	Description
Id	Id	Numerical values
Star	Star	Numerical (values ranging 1 to 27)
Lagnam	Lagnam house 1 Lagnam house 2 . . Lagnam house 12	Either 1 or 0 depends on the presence or absence of lagnam (total of 12 attributes)
Bhavam1	Bhavam 1 sun Bhavam 1 moon . . Bhavam2 Gulikan	Either 1 or 0 depends on the presence or absence of planets(total of 10 attributes)
Bhavam 2	Bhavam 2 sun Bhavam 2 moon . . Bhavam 2 Gulikan	Either 1 or 0 depends on the presence or absence of planets(total of 10 attributes)
....		
Bhavam 12	Bhavam 12 sun Bhavam 12 moon . . Bhavam12 Gulikan	Either 1 or 0 depends on the presence or absence of planets(total of 10 attributes)
Class	Class	Nurse, Sports, Writer, Other

id	star	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	LagnamHc	Bhavam1s	Bhavam1r	Bhavam1r	Bhavam1r	Bhavam1j	Bhavam1v	Bhavam1s	Bhavam1r
1	19	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	24	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	21	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
7	13	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
8	13	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
9	10	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0
10	7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
11	12	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
13	6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
15	5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
17	10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
18	23	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	20	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
20	21	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0
21	10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
22	5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
23	21	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
24	21	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Bhavam11	Bhavam11	Bhavam11	Bhavam11	Bhavam11	Bhavam11	Bhavam11	Bhavam11	Bhavam11	Bhavam12	Bhavam12	Bhavam12	Bhavam12	Bhavam12	Bhavam12	Bhavam12	Bhavam12	Bhavam12	class
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	other
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	other
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	sports
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	other
0	0	1	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	other
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	writer
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	nurse
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	nurse
0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	sports
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	nurse
0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	other
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	other
0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	other
0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	writer
0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	sports
1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	other
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	nurse
0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	other
0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	other
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	other
0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	other

Figure 4: The proposed data model

III. EXPERIMENTAL ANALYSIS AND RESULTS

Java is used as the language to implement the existing LR algorithm and the proposed Biased Logistic Regression (BLR) algorithm. The BLR algorithm is used to develop a Classifier-Predictor system which can then be used to classify and predict the job of an individual based on an input of astrological data. This is done in three phases,

1. Training and Evaluation phase

In this phase, two Classifiers, one using LR and other using BLR are created in Java using the Apache Mahout Data Analytics. The LR Classifier serves as the benchmark for comparison with BLR Classifier for the same data. For both Classifiers, the input training set is vectored, i.e. converted into a sparse matrix for easier processing. The standard methodology for training LR consists of running trials, in each trial, a fixed number of passes are performed where the input data is fed into the classifier, the weights between the data points modified using SGD, and a cost value calculated at end for each pass. The number of passes is selected based on how quickly the cost value is reduced to a negligible fraction of the order of 10^{-6}. At the end of each trial, a trained classifier is output, which can be used to predict. The trained classifier is then fed with the testing set, for each row, only the astrological data is used as input, while the last column containing the job is used to check the accuracy of the output. For each test data row the classifier predicts a matrix of the form:

[class1: predicted probability1, class2: predicted probability2…]

The class in this case is an integer id of the job, the predicted probability is a value between 1 and 0, and the class with the highest probability value is taken as the accepted prediction for that row. This value is compared with the expected value of job from above and if they match, then that row is marked as a successful prediction. At the end of the trial, the percentage of successful predictions is tallied and an accuracy value is calculated and tabulated.

For each trial, the input data is shuffled so that each trial has a different set of rows for training and testing. This is a common method used in practical data analytics to avoid any processing artifacts that may appear as a result of any data in the input being inconsistent or incorrect with respect to the rest of the data. In a single run, many trials are conducted using the input data and the average of the tabulated accuracy value is used as the accuracy of the LR Classifier for that data set. BLR Classifier uses the same methodology but adds the Bias Table as discussed in detail later.

2. Cherry picked Training phase

In this phase, several runs are conducted on variously modified datasets with the aim of producing a classifier that has the best of the breed prediction behavior. This process is necessary as most real world data may be unpredictably inconsistent or non-deterministic in nature. A cherry picking automated system is used to batch process both the LR and BLR models until both models show consistent behavior for a given dataset within a given timeframe. Classifiers with poor performance are weeded out until only performance units remain.

3. Prediction phase

In the final phase, a Java POJO (Plain Old Java Object) file is output with extension ".pred" which contains the fully trained classifier for prediction. This Classifier-Predictor system can be given an input of astrological data and it will produce a suitable job prediction for that data.

The BLR algorithm was developed to deal with the fundamental limitation of LR in expressing complex relationship type such as the lagnam in astrology. As lagnam is a meta- relation, it is difficult to adequately represent it in a tabular format. Since LR requires all data to have a half open data points, a new strategy was needed to show this lagnam meta-relation. The first step in BLR is to build a Bias table using the input data. The traditional astrologers consider the house called lagnam as the most important feature in the astrological chart. A relation between the lagnam of different individuals is identified based on this knowledge. The individuals with lagnam Aries, Leo, Sagittarius have an internal relation. Similarly (Taurus, Virgo, Capricorn), (Gemini, Libra, and Aquarius) and (Cancer, Scorpio, Pisces) are internally related lagnam groups.

Table IV: Lagnam Group table.

Lagnam group	Houses
G1	Aries, Leo, Sagittarius
G2	Taurus, Virgo, Capricorn
G3	Gemini, Libra, Aquarius
G4	Cancer, Scorpio, Pisces

For creating the bias table, the input matrix is divided into four relations (lagnam) per predictor (job) class, and a bias value is calculated for each [relation x class] group and stored. The bias value is expressed as the total number of items in each [relation x class] divided by number of items in the [relation] and multiplied by a modified value. The modified value is determined through empirical experimentation. The structure of the Bias table is shown below.

Table V: Description of Bias table

Attributes	Description
Class	The predictor value
Matching items	The total number of items with specific class value
Modifier	Experimentally derived constant value
Bias value	It is a numerical value expressed as the total number of items in each [relation x class] divided by number of items in the [relation] and multiplied by a modified value.

This BLR bias table is used in the first phase to introduce bias into the actual matrix of the classifier by intercepting updates between weights in the network and increasing the input value for matching items by the bias value for each pass. This causes the neural network to see these values as a pattern in the input data and consequently modifies the weights (biases the weights) in order to accommodate this. This in turn leads to higher accuracy for the predictions in cases where [relation x class] value is available.

After applying the bias values to lagnam attributes, the BLR gives an accuracy of 71%. The work flow of the LR and BLR algorithms are shown below.

Algorithm : Steps of LR based job prediction using astrological data	
Step 1:	Load the astrological data
Step 2:	Convert the data into sparse matrix
Step 3:	Divide the data into test and training sets (80:20)
Step 4:	Train the training data set by applying Logistic Regression
Step 5:	Test the LR for accuracy
Step 6:	Predict the average accuracy of predicting most matching job by applying LR algorithm on astrological data.

Algorithm : Steps of BLR based job prediction using astrological data	
Step 1:	Load the astrological data
Step 2:	Convert the trained data into sparse matrix
Step 3:	Generate BLR bias table
Step 4:	Generate the new matrix by introducing bias into actual matrix (Lagnam attributes) of each class, which is given in BLR bias table
Step 5:	Produce the output based on new biased matrix from BLR bias table
Step 6:	Determine the predicted probability of each class
Step 7:	If predicted probability of each class is high.
	Accept the prediction
	Else
	Wrong prediction
Step 8:	Measure the performance metric (Accuracy)

IV. CONCLUSION AND FUTURE ENHANEMENTS

Astrology is a most popular traditional science widely used all around the world especially in India. People follow astrology to take the decisions in different situations like the selection of a life partner, selecting a job, starting the construction of a new house, purchase a land etc. So astrology is an unavoidable part of the life. But the astrologers still follow the traditional methods for the prediction. They kept the modern technologies away from the path of astrology. Modern science never accepts astrology as proven science. This study reveals that Data mining can be effectively applied to astrological prediction. The Logistic Regression is an existing classification algorithm that can be used for the classification of astrological data and it gives an accuracy of 60%. It means that the astrological prediction by using LR gives an accuracy of sixty percentages. When the Biased Logistic Regression (BLR) is applied, it is possible to increase the accuracy by 11% than that of LR. That means the BLR gives an accuracy of 71%. The accuracy of the astrological predictions can be increased using the BLR. So the BLR helps to increase the credibility of astrology. Also, the system can decrease the errors that may happen when an astrologer analyses the number of astrological charts. An astrologer has a lot of limitations to identify a common pattern from the large number of astrological data. By using BLR, the similar patterns can easily be identified from the data model and that knowledge can be effectively utilized in the prediction phase. In this study only a lagnam based bias is applied for classification and it gives 71% accuracy. There are a lot of possibilities to increase accuracy by applying different biases or other methods. That type of study can also be included in the future for getting more accuracy.

ACKNOWLEDGMENT

First of all I express my deep and sincere thanks to God Almighty. Without his grace I would not have been able to complete my research work successfully.

I express my immense thanks and respect to my Honorable Guide Dr.Gladston Raj S., Head, Department of Computer Science, Government College Nedumangad, who has helped me a lot to fulfill my work with great success. The valuable guidance and advice besides motivation has led me to the right way of doing my work with perfection. His immeasurable support and boundless co-operation is a rare enchanting experience to me.

I am also deeply remembered my beloved parents, all the members of my family and almost all my dear friends and relatives.

REFERENCES

1. PrasnamargamPoorvardham by punnasseryNeelakandaSarma.
2. PrasnamargamUtharardham , PunnasseryNeelakandasarma.
3. Jyothishavijnanakosham by AnchalMadhavan Pillai and Parameswaranasan.
4. Jyothisha Gurubhoothan by K C Kesava Pillai.
5. Data Mining: Concepts, Models, Methods and Algorithms, Mehmed Kantardzic, ISBN: 0471228524, Wiley- IEEE Press 2002.
6. Data mining techniques by Arun K Pujari.
7. Astrological Prediction for Profession Doctor using Classification Techniques of Artificial Intelligence by Neelam Chaplot, Praveen Dhyani and O. P. Rishi: *International Journal of Computer Applications (0975 – 8887) Volume 122 – No.15, July 2015*
8. Ivan W. Kelly, A Concept of Modern Astrology a Critique. Article in Psychological Reports, 1997
9. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health by Peter C. Austin, Muhammad M. Mamdania, David N. Juurlinka, Janet E. Huxa : Journal of Clinical Epidemiology 59 (2006) 964e969
10. Austin, J. T.,Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. Higher Education: Handbook of Theory and Research, 8, 379–410.
11. Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. Higher Education: Handbook of Theory and Research, Vol. 10, 225–256.
12. Chao-Ying Joanne Peng, Kuk Lida Lee, Gary M. Ingersoll-Indiana University-Bloomington "An Introduction to Logistic Regression Analysis and Reporting"

AUTHORS PROFILE

Rahul Shajan: Completed MCA from Maatma Gandhi niversity and presently doing Phd in School of Computer Sciences, Mahatma Gandhi University, Kottayam as a full time research schoar under the guidance of Dr. Gladston Raj S. Published different research papers in different conferences and Journals.

Dr.Gladston Raj S: He took MSc, MTech and PhD in Computer Science from University of Kerala and having qualified with UGC-NET and PGDCH. Presently serving as Associate Professor & Head, Department of Computer Science @ Government College Nedumangad. Having guide ship at MG University Kottayam, Bharathiar University Coimbatore and University of Kerala. He produced seven PhDs, submitted two more thesis and guiding research scholars of Mahatma Gandhi University, Kottayam. With 20 years of teaching and research experience, he published more than 45 research papers.