

**EFFICIENT CLASSIFICATION MODEL: BY COMPARING LOGISTICS REGRESSION AND ARTIFICIAL NEURAL NETWORK****Ridhima Utreja\* & Namita Rani Behera\*\***

\* Krupanidhi Group of Institutions, Bangalore, Karnataka

\*\* Krupanidhi Degree College, Bangalore, Karnataka



**Cite This Article:** Ridhima Utreja & Namita Rani Behera, “Efficient Classification Model: By Comparing Logistics Regression and Artificial Neural Network”, International Journal of Current Research and Modern Education, Special Issue, March, Page Number 23-26, 2020.

**Abstract:**

In the social sciences, statistical models are employed to understand and to explain many phenomenon's happening around us in the world. The development of such trends must be derived from the notion in order to be scientifically acceptable, rational and practical. In order to do this, techniques might be needed that enable social scientists, even in the lack of strong hypotheses, to efficiently use their field experience. Complex datasets including centuries of variables for the construction of capability model are studied to give an explanation of the phenomena. This article is a gift of a green classification model that shows the findings of the ANN and the logistics regression with the same statistics.

**Introduction:**

The majority of individuals appear to have to wait for the future. It's a grassy hunger for man. The next blackjack card – What an energy to have over the natural world – to recognise in advance about the environment, the inventory market! Laptop technology and statistics technology are currently implementing a factor after many years of study and development, in which predictive models constitute an essential or crucial element of the business environment. Many innovations, like fast-powered computers, less expensive storage, cloud computing, voice popularity, cell devices, artificial intelligence, and powerful application software make them feasible. The world produces extensive statistics. To cite only one example, the quantity of information as a method of leveraging the Internet of Things will diminish what we now call big data. Recent advancements have enabled the democratisation of predictive analytics and trends and revealed smaller companies and clients, avoiding the need to generate and evaluate insights provided by huge software budgets or legions of scientists. Corporate executives in all industries will need to be aware of the scientific truths that shape a future economy, but also to be aware of how extraordinary these coming possibilities are to be used and taken advantage of. We have the strength to change our environment as human beings, not like most other creatures. One of the main ways to achieve this is to evaluate situations based mostly on data and evidence, to prepare and strategy for the fate. One may go so far as to claim that it is an evolutionary survival need or at least enhanced survival probability. It's amazing how our equipment (which is a powerful illustration of statistics) repeats what we need (or assume we want) to remain on record at some point. Think down the wheel, hot air, steel, weaponry, tour, medicine and data production. This is seen as an extension of our herbal survival instincts in predictive computing models. In this globally comprehensive and statistically important predictive way, you wish to learn from everything around us and estimate, examine, or plan what may be in the future. This article is an effort to offer a green classification model that highlights results from data mining i.e. ANN and statistical methods i.e. regression of logistics on the same grounds. This article is an effort to offer a green classification model that highlights results from data mining i.e. ANN and statistical methods i.e. regression of logistics on the same grounds.

**Literature Review:**

In a wide range of domains, data mining and statistical techniques have been utilised to identify particular models within a given population of facts especially for commercial companies' operations. Data mining methods are extremely useful to identify patterns from large amounts of data. These are defined below for each of these techniques:

**Techniques in Statistics:**

Statistical models are usually constructed for two basic reasons: foresight and/or clarity. In social sciences, in most situations statistical mode is used to explain, and the scene and causative factors behind human conduct and/or attitudes are more recognised. This skill is essential for the interchange and enhancement of society. Statistical modelling is usually carried out by use of observed statistics (social studies) and regression analysis to evaluate the linkage between a series of explanatory factors, x and a result, and approximately a theoretically generated causal hypothesis. This involves a research process in which theory constructs are defined, variables are derived, statistical modes along with determined variables can be taken, model results based on effect size and fitness information are compared, conclusions and/or refinement and testing can be reached. This process involves a researchership process. (Shmueli, 2010)

**Techniques for Data Mining:**

If the principle and/or the large, novel or complicated statistics set are not properly specified and contain numerous variables that cannot easily be reduced to a number of testing hypothesis, such statistical modelling can normally only be viable. Social scientists can also forecast hobby phenomena using data mining or system mastery techniques (Hindman, 2015). The fact that social scientists are using social surveys as a standard basis for statistical modelling and fresh record-taking is likely to become increasingly frequent (Mayer-Schönberger V & Cukier Big.K, 2013). This approach is a premise. However, the primary aim of such models remains clarity; predictive modelling evidence should be utilised to improve the basic notion and broaden the hypothesis for future checks. Therefore, it is of vital relevance to separate found data from underlying theoretical notions. If useful, the impacts of information models returned to sociological theory and to the field of researchers and policymakers' knowledge should be connected. Machine knowledge manipulates more and more complicated data sets comprising several variables of

## International Journal of Current Research and Modern Education

Impact Factor 6.725, Special Issue, March 2020 – Conference Proceedings

International Conference on Rise of Disembedded Unilateral Economy: InnoVision in the Era of  
Deglobalization (KRUPACON 2019) On 8<sup>th</sup> & 9<sup>th</sup> November 2019 Organized By

Krupanidhi Group of Institutions, Bangalore, Karnataka

abilities and does nonetheless conduct several exploratory and/or predictive studies to make sure that models are based on the idea and guided by area understanding ( Keim et. Al, 2008; Sedlmair et. Al, 2014)

Kiang was used to take a crucial look at the comparative assessment of techniques (2003). In this observation, Kiang took account of data mining class strategies such as neural networks and decision tree models (LDA), logistical regression analysis and all neighbouring model (kNN) and used artificial statistics to conduct a managed experiment in which the traits of records have been systematically altered in order to introduce imperial character. The observation was concluded by researching how the various class methods were performed when positives on statistical characteristics were broken, and Kiang confirmed that factual characteristics had a considerable influence on technique performance.

Also have a look at the performance of two AI approaches, Neural Network and ID3, through Shavlik et. Al (1991), who introduced them in this line by empirically evaluating the outcomes of three elements. The three variables were school data, poor training examples and the coding of favourable performances. Shavlik et al. confirmed, however, that neural networks did not focus much more the distribution of information instances in a pleasant way, with little scope of education. The impacts of 6 statistic features on the general performance of category technology, ID3 and PLSI were studied in a similar line by Rendell and Cho (1990). (Probabilistic getting to know device). The components examined include the educational size set, number of characteristics, attribute scales, error or noise, the class distribution and the distribution of samples. The study carried out for this thesis aims at adding a brand new dimension to the discovery of these documents by looking at the highest quality sample dimensions that are necessary in order to educate a decision tree, a neural network or a logistic regression model. Conversely, Harrell and Lee (1985), in a simulacrum observation, found that for small samples the usage of actual data in the Johnson and Seshia research (1992) was superior to that of logistic regression (LR), whereas the predictive discriminatory analysis (PDA) functioned better than PDA for small samples. In addition to the overall sampling of accurate classifications, the use of McNemar's contrasting correlated proportions controls may be compared in the two ways and meshbane et al (1996) confirmed that neither theoretical nor informational issues were useful in predicting which of the trends would work better. Thus, it appears also to have a look at the attempt to introduce an efficient classification version which highlights the results of data mining i.e. ANN and logistics regression in identical records.

### Differences between Data Mining / Machine Learning and Statistical Modelling:

The formalisation of links between variables inside the form of mathematical equations called statistical modelling Machine study is a collection of rules which may be studied from records that rely largely on regulatory programming. Although every model does the same work, irrespective of varied outputs, the device acquires knowledge and statistical modelling distinctions, and they can be as follows:

- *Various idea colleges;* the machine is a laptop-based and synthetic intelligence subcampus, which explodes study and manufacture of algorithms, rather than programmed orders.
- Statistical modelling is an arithmetic substratum that focuses on discovering links between variables to anticipate consequences.
- *Data sorts:* Machine-learning algorithms can predict flying facts and be known one by one by billions of observations. Machine research is useful with records that have an enormous variety of characteristics and too many observations.
- *Made assumptions:* statistical modes include a set of hypothesis about the age of the statistics determined and similar data from big populations. Alternatively, records minus checks on models and hypothesis and the forecasting accuracy of fashion can be taken into account.
- *Power and Human Effort:* Machine mastery is the technological know-how of computers being able to behave and thus not human interaction is not specifically designed. It examines the observation and design of algorithms, from which information may be analysed and predicted. Statistical models deal with mathematics and need the model to identify the relationships between the variables earlier than the facts.
- *Birthdays:* In various times, mechanical machinery and statistical modelling was born. For hundreds of years, statistical modelling has been around. In the mid-to-past 1950's and 1990s, however, device studying was more characterised by laptop scientists like Arthur Samuel and Tom Mitchell, who grew from computer technology and evolved via a glimpse of the idea of sample recognition and computational learning in synthetic intelligence.

Although it may seem that statistical modelling and machines that become known are particular predictive modelling branches, there is no straightforward border between two, making them virtually identical.

Over the last decade, limits have started to dissolve quickly, each discipline researching significantly. To continue to succeed in the dynamic world of analytics, scientists must position themselves as equals to learn both methods in order to support their clients or agencies to remain applicable in fast-moving records.

### Objective of the Study:

Distinguishing between Data Mining (ANN) and Statistical Techniques (LR) for type version with equal statistic set. Data Mining, i.e. Artificial neural networks (ANN).

### Methods:

For our research reasons, an extraordinary variant was suggested. One fact is the mining method, for instance the neural network and a statistical approach, for example a technique of logistic regression. For the cause of building the modes the facts mining programme matlab was employed. MATLAB Neural Toolbox appears to be utilised here. For technical computing, MATLAB is a high-performance overall language. Toolboxes consist of a full set of MATLAB (M-documents) functions expanding the MATLAB environment to handle specific problem types. The toolkit is to comprise of signal processing, control structures, neural networks, fluid common sense, wavelets, simulation, etc.

## International Journal of Current Research and Modern Education

Impact Factor 6.725, Special Issue, March 2020 – Conference Proceedings

International Conference on Rise of Disembedded Unilateral Economy: InnoVision in the Era of  
Deglobalization (KRUPACON 2019) On 8<sup>th</sup> & 9<sup>th</sup> November 2019 Organized By

Krupanidhi Group of Institutions, Bangalore, Karnataka

An extensive field survey was carried out by one hundred Credit Officers or Managers. Selection of these banks has been completed to preserve the probability and sensitive information of series of credit officers. The initial interview was done by two sets of dependent surveys from Credit authorities and Branch managers. Thereby, an information base structured on the basis of the respondents' ratings on certain items and afterwards on the basis of their latest rating was established. However, the record collecting was split in a school set and examined with the ANN development approach. Weight is changed with the training situations. For network validation, the control instances are utilised. The diversity of cases required for each class may be calculated with several considerations in mind (Khanna, 1990; Bishop, 1995). The learning process consists of a number of instances, each of which has values for a range of variables input and output. The first decisions must be made: which variables must be used, and the number(s) of instances to be collected.

ANN provides a completely distinct solution to the problem of representation of knowledge, called connectionism (Gallant, 1988). Neural networks attempt to mimic human intuition by replicating the way intuition is entirely built upon physicality, by simulating the process of adaptive organic knowledge building. A neuronal group can potentially react correctly to a certain situation (or to the acceptable reaction, when several reactions are necessary) when the facts are noisy or incomplete. Synthetic neural systems give wonderful benefits over selection guides and expert systems in different terms in unstructured decision situations. This study deals with the premise for unstructured selection problems in neural network software (Hawley et al, 1990).

For the example data set, a basic Perceptron NN is set.

Regression of logistics A logistic regression variant is utilised while the dependent variable, as is the case here, is a categorical variable with continuous or categorical variables of the predictor. This is a semimetric version where multivariate normality and the equivalent requirements of dispersion for the facts are not necessary. The following form uses a logistical function:  $Y = 1 / (1 + e^{-y})$ ,  $y = a + \sum_{i=1}^n \beta_i X_i$ , where  $X_i$  is a set of variables between man and woman,  $\beta_i$  is the  $i$ th coefficient and  $y$  is the chance of a good end outcome.  $Y$  is the random variable of Bernoulli.

For the same facts, the logistics regression model was designed

```
model <- glm(Recommended ~ .-ID, data = dataset, family = binomial)
summary(model)
predict <- predict(model, type = 'response')
```

### Results and Discussion:

The results of analyses performed on the same statistics as the single-species analyses. In the case of ANN, it has long been thought that if more practical units are utilised for network training, the overall performance of the neural network would be significantly advanced. An expert is approached in this sample community. During the destination, it would be possible to make the credit award procedure more difficult and to determine the creditworthiness of the borrowing applicant using new factors. With a significantly lower amount of information, the network shows greater error percentage. As the amount of training data is increased, the error % decreases to a certain number. The decrease in the proportion of errors relies on the range of information taught and no passes utilised during adaptation. Thus banks may utilise more information for educating the community in the actual life scenario and are likely to achieve a decrease of mistakes.

Regression of logistics it is shown that the accuracy of the general prediction for schooling data has become practically identical with the accuracy of test data. The accuracy of the classification has hit the plateau in the size of the model taken for education statistics and cannot be achieved by extending the sample length in terms of prediction accuracy. The fundamental category precision for sample size training data was converted into around 78%. This has not helped tremendously to anticipate the check records more accurately by increasing the pattern dimension.

In the logistic regression version, the category accuracy percentage was almost the same as the ANN for the given sample length. However, if the precision percentage of the ANN category is also increased by increasing the sample size while the regression version itself is no longer real. Therefore, it is advisable to use ANN for the big pattern size in conjunction with the construction of a model category over logistics regress version.

### Acknowledgement:

The authors express gratitude towards the assistance provided by The Management, Krupanidhi Group of Institutions (KGI) and Krupanidhi Research Incubation Centre, KGI in completing the research. We also thank our Research Mentors who guided us throughout the research and helped us in achieving the desired results.

### References:

1. Bishop C (1995), "Neural Networks for Pattern Recognition", Oxford: University Press.
2. Gallant S I (1988), "Connectionist expert systems", Communications of the ACM, 31(2), pp 152–169.
3. Harrell, F. E. Jr. and Lee, K. L. (1985). "A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality", In P. K. Sen (Ed), Biostatistics: Statistics in biomedical, public health and environmental sciences, 333-343.
4. Hawley, D D, Johnson J and Raina D (1990), "Artificial Neural Systems: A new tool for Financial Decision-Making", Financial Analyst Journal, November/December, pp 63-72.
5. Hindman M (2015), Building better models prediction, replication, and machine learning in the social sciences Ann. Am. Acad. Political Soc. Sci., 659 (1) ,2015, pp. 48-62
6. Johnston, B. and Seshia, S. S. (1992). "Discriminant analysis when all variables are ordered", Statistics in Medicine, 11, 1023-1032.
7. Keim, D., Mansmann, F., Schneidewind, J., Thomas, J. and Ziegler, H (2008). Visual analytics: scope and challenges Visual Data Mining, Springer-Verlag , pp. 76-90

**International Journal of Current Research and Modern Education****Impact Factor 6.725, Special Issue, March 2020 – Conference Proceedings****International Conference on Rise of Disembedded Unilateral Economy: InnoVision in the Era of Deglobalization (KRUPACON 2019) On 8<sup>th</sup> & 9<sup>th</sup> November 2019 Organized By Krupanidhi Group of Institutions, Bangalore, Karnataka**

8. Khanna T (1990), "Foundations of Neural Networks", Reading, MA: Addison Wesley.
9. Kiang, M. (2003). "A comparative assessment of classification methods", Decision Support Systems, 35, 441-454.
10. Mayer-Schönberger V and CukierBig.K (2013), Data: A Revolution That will Transform How we Live, Work, and Think, Houghton Mifflin Harcourt.
11. Meshbane, A. and Morris, J. D. (1996). "Predictive discriminant analysis versus logistic regression in two-group classification problems", American Educational Research Association annual meeting, New York.
12. Rendell, L. and Cho, H. (1990). "Empirical learning as a function of concept character", MachineLearning, 5, 267-298.
13. Shavlik, J. W., Mooney, R.J. and Towell, G.G. (1991). "Symbolic and neural learning algorithms: an experimental comparison", Machine Learning, 6, 111-144.
14. Shmueli G (2010), To explain or to predict? Stat. Sci. 25 (3), 289-310.