**LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH**

**ABSTRACT:**

**Abstract**

Decision taking is attained by probabilistic and predictive approaches developed by various machine learning algorithms. This paper discusses about logistic regression ad its mathematical representation. This paper adheres to logistic regression as a machine learning tool in order to actualize the predictive and probabilistic approaches to a given problem of loan approval prediction. Using logistic regression as a tool , this paper specifically delineates about whether or not loan for a set of records of an applicant will be approved.

**1. INTRODUCTION**

**1.1 INTRODUCTION:**

The technical world is advancing toward complete automation. In order to attain automation various concepts are being developed and put to use , as can be observed from the numerous developments being made and symposiums being held. One of the most striking features that excites scientists and technologists , in regards to the development of automation , is Artificial Intelligence. Artificial Intelligence is the concept of simulate human like intelligence in a computer[6]. To make a machine think exactly like a human is intriguing to the scientists and developers , and they strive to achieve this goal by putting Artificial Intelligence to use. The idea is not to overpower the human society but to work with the man so that the combined intelligence can lead to many more revelations in this technological era.

Artificial Intelligence dates back to the time of advent of computers and since then it has diversified into numerous field[6]. Over the years, technologists have acquired a great understanding in this field which has lead to development of defined models and further application of these models to real world problems. Various domains in Artificial Intelligence domain include machine learning, neural networks, fuzzy logic , natural language processing , expert systems[6]. These concepts are deployed according to the specificity of the desired requirements. In regards with this paper , one of the concepts in the domain of machine learning is exploited and also applied to a real world application. Machine Learning is a tool which facilitates development of analytical models without explicit programming[4]. Various machine learning algorithms are developed to tailor to the problem requirements. All the leading edge industries are now utilizing the capabilities of machine learning to gain higher sales growth and statistics have shown that they are getting positive results. With institutions generating more and more data, exploitation of data manually becomes difficult , hence machine learning, having the capability of analytical modelling is sought to, as a solution.

Distribution of the loans is the core business part of almost every banks. The main portion the bank’s assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approves loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it emphasize different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through this system. Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight .A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Paper is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing. Result against particular Loan Id can be send to various department of banks so that they can take appropriate action on application. This helps all others department to carried out other formalities.

**2. LITERATURE SURVEY**

**An introduction to logistic Regression Analysis and Reporting**

.

The purpose of this article is to provide researchers, editors, and readers with a set of guidelines for what to expect in an article using logistic regression techniques. Tables, figures, and charts that should be included to comprehensively assess the results and assumptions to be verified are discussed. This article demonstrates the preferred pattern for the application of logistic methods with an illustration of logistic regression applied to a data set in testing a research hypothesis. Recommendations are also offered for appropriate reporting formats of logistic regression results and the minimum observation-to-predictor ratio. The authors evaluated the use and interpretation of logistic regression presented in 8 articles published in The Journal of Educational Research between 1990 and 2000. They found that all 8 studies met or exceeded recommended criteria.

**An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain Park**

The purpose of this article is twofold: 1) introducing logistic regression (LR), a multivariable method for modeling the relationship between multiple independent variables and a categorical dependent variable, and 2) examining use and reporting of LR in the nursing literature. Methods: Text books on LR and research articles employing LR as main statistical analysis were reviewed. Twenty-three articles published between 2010 and 2011 in the Journal of Korean Academy of Nursing were analyzed for proper use and reporting of LR models. Results: Logistic regression from basic concepts such as odds, odds ratio, logit transformation and logistic curve, assumption, fitting, reporting and interpreting to cautions were presented. Substantial shortcomings were found in both use of LR and reporting of results. For many studies, sample size was not sufficiently large to call into question the accuracy of the regression model. Additionally, only one study reported validation analysis. Conclusion: Nursing researchers need to pay greater attention to guidelines concerning the use and reporting of LR models.

**A comprehensive introduction to machine learning that uses probabilistic models and inference as a unifying approach.**

Today's Web-enabled deluge of electronic data calls for automated methods of data analysis. Machine learning provides these, developing methods that can automatically detect patterns in data and then use the uncovered patterns to predict future data. This textbook offers a comprehensive and self-contained introduction to the field of machine learning, based on a unified, probabilistic approach.

The coverage combines breadth and depth, offering necessary background material on such topics as probability, optimization, and linear algebra as well as discussion of recent developments in the field, including conditional random fields, L1 regularization, and deep learning. The book is written in an informal, accessible style, complete with pseudo-code for the most important algorithms. All topics are copiously illustrated with color images and worked examples drawn from such application domains as biology, text processing, computer vision, and robotics. Rather than providing a cookbook of different heuristic methods, the book stresses a principled model-based approach, often using the language of graphical models to specify models in a concise and intuitive way. Almost all the models described have been implemented in a MATLAB software package—PMTK (probabilistic modeling toolkit)—that is freely available online. The book is suitable for upper-level undergraduates with an introductory-level college math background and beginning graduate students.

**3. SYSTEM ANALYSIS**

The Systems Development Life Cycle (SDLC), or Software Development Life Cycle in [systems engineering](http://en.wikipedia.org/wiki/Systems_engineering), [information systems](http://en.wikipedia.org/wiki/Information_systems) and [software engineering](http://en.wikipedia.org/wiki/Software_engineering), is the process of creating or altering systems, and the models and [methodologies](http://en.wikipedia.org/wiki/Methodologies) that people use to develop these systems. In software engineering the SDLC concept underpins many kinds of [software development methodologies](http://en.wikipedia.org/wiki/Software_development_methodologies).

**3.1 EXISTING SYSTEM**:

Distribution of the loans is the core business part of almost every banks. The main portion the bank’s assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approves loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants

**DISADVANTAGES:**

Existing works mostly depend on manual verification process.

It may take months or weeks to get loan approval.

**3.3 PROPOSED SYSTEM:**

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight.

**ADVANTAGES:**

Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique.

#### SOFTWARE REQUIREMENTS:

#### 

#### Operating system : Windows XP/7/10

* Coding Language : Html, JavaScript,
* Development Kit : Flask Framework
* IDE : Anaconda prompt

#### HARDWARE REQUIREMENTS:

#### 

#### System : Pentium IV 2.4 GHz.

#### Hard Disk : 100 GB.

#### Monitor : 15 VGA Color.

#### Mouse : Logitech.

#### RAM : 1 GB.