

**DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELHI**



PROJECT REPORT

DATA MINING

(AS PART OF THE COURSE PAPER 404: ADVANCED STATISTICAL COMPUTING AND DATA MINING)

UNDER THE GUIDANCE AND SUPERVISION OF

MR. ABHISHEK K. UMRAWAL
(ASSISTANT PROFESSOR)

SUBMITTED BY:

VISHAL KUMAR
M. SC. STATISTICS, SEMESTER – IV

DATE: APRIL 22, 2016
PLACE: NEW DELHI

ABSTRACT

To understand the computational aspect of Data Mining, a practical approach using statistical software R was introduced in Statistics Department of University of Delhi. The main purpose of this approach is to understand processes involving supervised and unsupervised learning.

This study has been divided into 2 parts:

1. In this part topics of supervised learning were explored like Logistic model, Neural Network and K-Nearest Neighbor approach, and then they were compared on the basis of misclassification they produced in classifying a customer as good credit or bad credit customer.
2. In this part K-mean clustering (unsupervised learning technique) was applied in determining the clothing size of males and females based on heights and weights of some individuals. Then a refinement to female clothing size was recommended.

ACKNOWLEDGEMENTS

I have given my best efforts to make this a good report. However, it would not have been possible without the kind support, guidance and constant supervision of our professor **MR. ABHISHEK K. UMRAWAL** and I would like to express my gratitude towards him for providing necessary help regarding the concepts and techniques used in this case study.

Lecture notes released by **MR. ABHISHEK K. UMRAWAL**, provided the necessary knowledge to tackle the given problem at hand and to understand the basic “why” concept behind any technique and algorithm used in this report.