

**A PROJECT REPORT  
ON  
STATISTICAL MODELING FOR EVALUATING THE BUSINESS  
ORGANISATION USING MACHINE LEARNING TECHNIQUES**

**Submitted to  
Osmania University  
in partial fulfilment of the requirements for the award of**

**MASTER OF SCIENCE  
IN  
STATISTICS**



**DEPARTMENT OF STATISTICS  
UNIVERSITY COLLEGE OF SCIENCE  
OSMANIA UNIVERSITY  
HYDERABAD – INDIA**

**By**

**Ms M. OM GAYATHRI  
Ms P. GOURI PRIYA  
Ms P. SOWMYA  
Ms G. GAYATHRI  
Ms P. RAMYA  
Ms M. TIRUMALA  
Mr T. CHANDU KUMAR  
Ms AFREEN SULTHANA**

**Roll No.: 1007-17-507-013  
Roll No.: 1007-17-507-012  
Roll No.: 1007-17-507-011  
Roll No.: 1007-17-507-015  
Roll No.: 1007-17-507-010  
Roll No.: 1007-17-507-004  
Roll No.: 1007-17-507-005  
Roll No.: 1007-17-507-020**

**Under the Supervision of  
Dr. M. VENUGOPALA RAO  
2018**

**A PROJECT REPORT  
ON**

**Statistical modelling for Evaluating the Business Organisation  
Using Machine Learning Techniques**

**Submitted to  
Osmania University  
in partial fulfilment of the  
requirements for the award of  
Master of Science in Statistics**



**DEPARTMENT OF STATISTICS  
UNIVERSITY COLLEGE OF SCIENCE  
OSMANIA UNIVERSITY  
HYDERABAD – INDIA**

**By**

**Ms M. OM GAYATHRI  
Ms P. GOURI PRIYA  
Ms P. SOWMYA  
Ms G. GAYATHRI  
Ms P. RAMYA  
Ms M. TIRUMALA  
Mr T. CHANDU KUMAR  
Ms AFREEN SULTHANA**

**Roll No.: 1007-17-507-013  
Roll No.: 1007-17-507-012  
Roll No.: 1007-17-507-011  
Roll No.: 1007-17-507-015  
Roll No.: 1007-17-507-010  
Roll No.: 1007-17-507-004  
Roll No.: 1007-17-507-005  
Roll No.: 1007-17-507-020**

**Under the Supervision of  
Dr. M. VENUGOPALA RAO  
2018**

## CERTIFICATE

|   |                           |
|---|---------------------------|
| This is to certify that Ms M. OM GAYATHRI | Roll No.: 1007-17-507-013 |
| Ms P. GOURI PRIYA                         | Roll No.: 1007-17-507-012 |
| Ms P. SOWMYA                              | Roll No.: 1007-17-507-011 |
| Ms G. GAYATHRI                            | Roll No.: 1007-17-507-015 |
| Ms P. RAMYA                               | Roll No.: 1007-17-507-010 |
| Ms M. TIRUMALA                            | Roll No.: 1007-17-507-004 |
| Mr T. CHANDU KUMAR                        | Roll No.: 1007-17-507-005 |
| Ms AFREEN SULTHAN                         | Roll No.: 1007-17-507-020 |

have submitted the project titled “**Statistical modelling for Evaluating the Business Organisation using Machine Learning Techniques**” in partial fulfilment for the degree of Master of Science in Statistics.

Head

Department of Statistics

Internal Examiner

External Examiner

## **DECLARATION**

The research presented in this project has been carried out in the **Department of Statistics, Osmania University, Hyderabad.** The work is original has not been submitted so far, in part or full, for any other degree or diploma of any university.

**Ms M. OM GAYATHRI**

**Ms P. GOURI PRIYA**

**Ms P. SOWMYA**

**Ms G. GAYATHRI**

**Ms P. RAMYA**

**Ms M. TIRUMALA**

**Mr T. CHANDU KUMAR**

**Ms AFREEN SULTHANA**

Department of Statistics

Osmania University

Hyderabad – 500 007, T.S.

INDIA

## ACKNOWLEDGEMENTS

I deem it a great pleasure to express my deep sense of gratitude and indebtedness to my research supervisor **Dr. M. VENUGOPALA RAO**, Statistics department, College of Science, Osmania University for her valuable guidance, and enlightening discussions throughout the progress of my project work.

I also express my sincere and heartfelt thanks to **Prof. C. JAYALAKSHMI**, Head of the Department, Department of Statistics, Osmania University for providing the necessary support and facilities in the department for completion of this work successfully.

It is indeed with great pleasure I record my thanks to **Dr. G. JAYASREE**, Chairperson, Board of Studies, Department of Statistics, Osmania University for having provided with all the facilities to carry out my work.

I thank **Dr. N. Ch. BHATRACHARYULU**, **Dr. K. VANI**, **Dr. S. A. JYOTHIRANI**, **Dr. G. SIRISHA** and **Mrs. J. L. PADMA SHREE**, for their encouragement and constant help during the research.

I would like to express my deepest gratitude to **T. SANDHYA**, **BALA KARTHIK** for their advice, guidance and involvement at various stages of this work, I would also like to thank them for their understanding and constant encouragement throughout this project.

I thank all Non – Teaching members of the Department of Statistics, who helped me during my Thesis work.

I am thankful to the Osmania University for permitting me to carry out this work.

# CONTENTS

|  | Page No. |
|--|----------|
| 1. INTRODUCTION AND SCOPE OF THE PROBLEM   | 01       |
| 1.1. Need of Statistics for today's Market Environment   | 02       |
| 1.2. Scope of the problem  | 02       |
| 1.3. Questionnaire   | 03 - 05  |
| 1.4. Review of the chapters  | 06       |
| 2. REVIEW OF MACHINE LEARNING PROCESS  | 07       |
| 2.0. Need of Machine Learning  | 08       |
| 2.1. Machine Learning Process  | 08 - 11  |
| 2.2. Types of Machine Learning   | 11 - 14  |
| 2.3. Choosing the algorithm  | 15 - 23  |
| 2.4. Choosing and comparing the models through pipelines   | 24 - 27  |
| 2.5. Model diagnosis with Overfitting and Underfitting   | 28 - 31  |
| 2.6. Overall process of Machine Learning   | 31       |
| 3. MACHINE LEARNING AT WORK  | 32       |
| 3.1. An approach to the problem  | 33 - 34  |
| 3.2. Number of factors extracted   | 35 - 38  |
| 3.3. Cluster Analysis  | 39 - 40  |
| 3.4. Discriminant Analysis   | 40 - 42  |
| 4. CONCLUSIONS   | 43       |
| 4.1. What are the parameters to evaluate the coffee shop   | 44       |
| 4.2. Segments  | 44       |
| 4.3. Whether any Marketing actions need to be taken, given the<br>Customer needs and segment sizes | 44       |
| 4.4. Classification algorithm  | 44       |
| 4.5. Key Attributes  | 45       |
| Bibliography   | 46       |
| 5. APPENDIX  | 47- 55   |
| R- Code, Data  |          |