

**A PROJECT REPORT  
ON  
PREDICTING GROSS REVENUE OF MOVIES**  
Submitted to  
**Osmania University**  
in partial fulfilment of the requirements for the award of  
**Master of Science**  
In  
**Applied Statistics**



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

**By**

<b>Munnur Jyothi</b>	<b>1007-17-508-007</b>
<b>Margie Urmila</b>	<b>1007-17-508-032</b>
<b>Majji Divya</b>	<b>1007-17-508-033</b>
<b>Chittabhathine Sri Lakshmi</b>	<b>1007-17-508-034</b>
<b>Kamatam Gopi Krishna</b>	<b>1007-17-508-035</b>
<b>Godhula Chandra Lekha</b>	<b>1007-17-508-037</b>

**Under the Supervision of  
T. SANDHYA**

**December 2018**

**A PROJECT REPORT  
ON  
PREDICTING GROSS REVENUE OF MOVIES**  
Submitted to  
**Osmania University**  
in partial fulfilment of the requirements for the award of  
**Master of Science**  
In  
**Applied Statistics**



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

**By**

<b>Munnur Jyothi</b>	<b>1007-17-508-007</b>
<b>Margie Urmila</b>	<b>1007-17-508-032</b>
<b>Majji Divya</b>	<b>1007-17-508-033</b>
<b>Chittabhathine Sri Lakshmi</b>	<b>1007-17-508-034</b>
<b>Kamatam Gopi Krishna</b>	<b>1007-17-508-035</b>
<b>Godhula Chandra Lekha</b>	<b>1007-17-508-037</b>

**Under the Supervision of  
T. SANDHYA**

**December 2018**

## **CERTIFICATE**

This is to certify that

<b>Ms. Munnur Jyothi</b>	<b>1007-17-508-007</b>
<b>Ms. Margie Urmila</b>	<b>1007-17-508-032</b>
<b>Ms. Majji Divya</b>	<b>1007-17-508-033</b>
<b>Ms. Chittabathine Sri Lakshmi</b>	<b>1007-17-508-034</b>
<b>Mr. Kamatam Gopi Krishna</b>	<b>1007-17-508-035</b>
<b>Ms. Godhula Chandra Lekha</b>	<b>1007-17-508-037</b>

have submitted the project titled **“Predicting Gross revenue of movies”** in partial fulfilment for the degree of Master of Science in Applied Statistics.

Head

Department of Statistics

Internal Examiner

External Examiner

# **DECLARATION**

The analysis 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 of diploma of any university.

**Munnur Jyothi**

**Margie Urmila**

**Majji Divya**

**Chittabhathine Sri Lakshmi**

**Kamatam Gopi Krishna**

**Godhula Chandra Lekha**

Department of Statistics

Osmania University

Hyderabad – 500 007, T.S.

INDIA

## ACKNOWLEDGEMENTS

We deem it a great pleasure to express our deep sense of gratitude and indebtedness to our research supervisor **T.SANDHYA**, Statistics department, University College of Science, Osmania University for her valuable guidance, and enlightening discussions throughout the progress of our project work.

We also express our 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 we record our thanks to **Dr. G. JAYASREE**, Chairperson, Board of studies, Department of Statistics, Osmania University for having provided with all the facilities to carry out our work.

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

We would like to express our deepest gratitude to **Dr M. VENUGOPALA RAO, BALA KARTHIK** for their advice, guidance and involvement at various stages of this work. I would also thank them for their understanding and constant encouragement throughout this project.

We thank all Non-Teaching staff of Department of Statistics who helped us during our thesis work.

We are thankful to the Osmania University for permitting us to carry out this work.

# CONTENTS

	Page No.
1. INTRODUCTION AND SCOPE OF THE PROBLEM	01 – 04
1.1. Scope of the Problem	02
1.2. Data Description	02 – 03
1.3. Review of the Chapters	04
2. REVIEW OF MACHINE LEARNING TECHNIQUES	05 – 22
2.0. Need of Machine Learning	06
2.1. Machine Learning	06
2.1.1. Business understanding	07
2.1.2. Data understanding	07
2.1.3. Data preparation	07
2.1.4. Modelling	08
2.1.5. Evaluation	08
2.1.6. Deployment	08
2.2. Types of Machine Learning	09 – 11
2.2.1. Supervised learning	09 – 10
2.2.2. Unsupervised learning	10 – 11
2.2.3. Reinforcement learning	11
2.3. Choosing the algorithm	12 – 16
2.3.1. Types of Regression algorithm	12 – 13
2.3.2. Types of Classification algorithm	14 – 15
2.3.3. Types of Unsupervised algorithm	15 – 16
2.4. Choosing and comparing models through pipelines	17 – 18
2.4.1. Model Validation	17
2.5. Model diagnosis with over fitting and under fitting	19 – 21
2.5.1. Bias and Variance	19 – 20
2.5.2. Model performance matrix	20 – 21
2.6. Overall process of Machine learning	22
3. MACHINE LEARNING AT WORK	24 – 34

4. SUMMARY	36
5. APPENDIX	38 – 51
Rcode	38 – 44
Dataset	45 – 48
Bibliography	49