

# **MACHINE LEARNING**

(Predicting Amount of Purchase)

*Summer Internship Report Submitted in partial fulfillment*

*of the requirement for undergraduate degree of*

**Bachelor of Technology**

In

**Electronics and Communication Engineering**

By

**StudentName**

**StudentRollNumber**

*Under the Guidance of*

**Mr. M. Venkateswarlu**

Assistant Professor



Department Of Electronics and Communication Engineering  
GITAM School of Technology  
GITAM (Deemed to be University)  
Hyderabad-502329  
June 2019

## DECLARATION

I submit this industrial training work entitled “**PREDICTING AMOUNT OF PURCHASE**” to GITAM (Deemed To Be University), Hyderabad in partial fulfillment of the requirements for the award of the degree of “**Bachelor of Technology**” in “**Electronics and Communication Engineering**”. I declare that it was carried out independently by me under the guidance of **Mr. M. Venkateswarlu**, Asst. Professor, GITAM (Deemed To Be University), Hyderabad, India.

The results embodied in this report have not been submitted to any other University or Institute for the award of any degree or diploma.

Place: HYDERABAD

StudentName

Date:

StudentRollNo



### **CERTIFICATE**

This is to certify that the Industrial Training Report entitled **“PREDICTING AMOUNT OF PURCHASE”** is being submitted by StudentName (StudentRollNumber) in partial fulfillment of the requirement for the award of **Bachelor of Technology in Electronics & Communication Engineering** at GITAM (Deemed To Be University), Hyderabad during the academic year 2018-19

It is faithful record work carried out by her at the **Electronics & Communication Engineering Department**, GITAM University Hyderabad Campus under my guidance and supervision.

**Mr. M. Venkateswarlu**

Assistant Professor

Department of ECE

**Dr.K.Manjunathachari**

Professor and HOD

Department of ECE

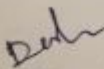


Promize IT Services  
Flat No. 303, Paradise Valley, Road NO.3,  
Madhavapuri Hills, Chanda Nagar,  
Ameenpur village,  
Hyderabad, Telangana, India - 502032

Date: 16<sup>th</sup> June 2019

## CERTIFICATE

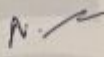
This is to certify that the Internship titled "Predicting amount of Purchase using Multiple Linear Regression" is the bona fide work carried out by [REDACTED] student of the GITAM University, Hyderabad, in partial fulfillment for the award of Bachelor of Technology in Electronics and Communication Engineering during the period 29<sup>th</sup> April 2019 – 15<sup>th</sup> June 2019 at PROMIZE IT SERVICES PRIVATE LIMITED – HYDERABAD. During this period his conduct was found to be very good and he has shown good technical skills.

  
(Dinesh Kumar Jooshetti)

Project Head

Promize IT Services Pvt Ltd.



  
(Suresh Shankar)

Director of Operations

Promize IT Services Pvt Ltd.

## ACKNOWLEDGEMENT

Apart from my effort, the success of this internship largely depends on the encouragement and guidance of many others. I take this opportunity to express my gratitude to the people who have helped me in the successful completion of this internship.

I would like to thank respected **Dr. N. Siva Prasad**, Pro Vice Chancellor, GITAM Hyderabad and **Dr. CH. Sanjay**, Principal, GITAM Hyderabad

I would like to thank respected **Dr. K. Manjunathachari**, Head of the Department of Electronics and Communication Engineering for giving me such a wonderful opportunity to expand my knowledge for my own branch and giving me guidelines to present a internship report. It helped me a lot to realize of what we study for.

I would like to thank the respected faculties **Mr. M. Venkateswarlu** who helped me to make this internship a successful accomplishment.

I would also like to thank my friends who helped me to make my work more organized and well-stacked till the end.

StudentName  
StudentRollNo

## **ABSTRACT**

Machine learning algorithms are used to predict the values from the data set by splitting the data set in to train and test and building Machine learning algorithms models of higher accuracy to predict the values is the primary task to be performed on Cereals data set. My perception of understanding the given data set has been in the view of undertaking a client's requirement of overcoming the stagnant point of sales of the products being manufactured by client.

To get a better understanding and work on a strategical approach for solution of the client, I have adapted the view point of looking at ratings of the products and for further deep understanding of the problem, I have taken the stance of a consumer and reasoned out the various factors of choice of the products and they purchase, and my primary objective of this case study was to look up the factors which were dampening the sale of products and correlate them to ratings of products and draft out an outcome report to client regarding the various aspects of a product manufacturing, marketing and sale point determination.

## **Table of Contents:**

### **LIST OF FIGURESIX**

#### **CHAPTER 1:MACHINE LEARNING1**

1.1 INTRODUCTION.....	1
1.2 IMPORTANCE OF MACHINE LEARNING.....	1
1.3 USES OF MACHINE LEARNING.....	2
1.4 TYPES OF LEARNING ALGORITHMS.....	3
1.4.1 Supervised Learning.....	3
1.4.2 Unsupervised Learning.....	3
1.4.3 Semi Supervised Learning.....	4
1.5 RELATION BETWEEN DATA MINING,MACHINE LEARNING AND DEEP LEARNING.....	5

#### **CHAPTER 2:PYTHON.....6**

2.1 INTRODUCTOIN TO PYTHON.....	6
2.2 HISTORY OF PYTHON.....	6
2.3 FEATURES OF PYTHON.....	6
2.4 HOW TO SETUP PYTHON.....	7
2.4.1 Installation(using python IDLE).....	7
2.4.2 Installation(using Anaconda).....	8
2.5 PYTHON VARIABLE TYPES.....	9
2.5.1 Python Numbers.....	10
2.5.2 Python Strings.....	10
2.5.3 Python Lists.....	11
2.5.4 Python Tuples.....	11
2.5.5 Python Dictionary.....	12
2.6 PYTHON FUNCTION.....	12
2.6.1 Defining a Function.....	12

2.6.2	Calling a	
Function.....	13	
2.7	PYTHON	USING OOP's
CONCEPTS.....	13	
2.7.1		
Class.....	13	
2.7.2 __init__ method in class.....	14	
<b>CHAPTER 3:CASE STUDY.....</b>	<b>15</b>	
3.1		PROBLEM
STATEMENT.....	15	
	vii	
3.2 DATA SET.....	15	
3.3	OBJECTIVE	OF THE CASE
STUDY.....	16	
<b>CHAPTER 4:MODEL BUILDING.....</b>	<b>17</b>	
4.1	PREPROCESSING	OF THE
DATA.....	17	
4.1.1	Getting	the Data
Set.....	17	
4.1.2 Importing the Libraries.....	17	
4.1.3	Importing	the
Data-Set.....	18	
4.1.4	Handling	the Missing
values.....	18	
4.1.5 Categorical Data.....	23	
4.2 TRAINING THE MODEL.....	28	
4.2.1		Method
1.....	28	
4.2.2		Method
2.....	30	
4.3 EVALUATING THE CASE STUDY.....	34	
4.3.1 Building the model(using the statsmodel library).....	39	
4.3.2 Building the model(using splitting).....	41	
<b>CONCLUSION.....</b>	<b>44</b>	



<b>REFERENCES.....</b>	<b>45</b>
------------------------	-----------

## LIST OF FIGURES:

Figure 1 : The Process Flow.....	2
Figure 2 : Unsupervised Learning.....	4
Figure 3 : Semi Supervised Learning.....	5
Figure 4 : Python download.....	8
Figure 5 : Anaconda download.....	9
Figure 6 : Jupyter notebook.....	9
Figure 7 : Defining a Class.....	14

Figure 8 : Importing Libraries.....	17
Figure 9 : Reading the Dataset.....	18
Figure 10 : data before using dropna().....	19
Figure 11 : data after using dropna().....	19
Figure 12 : functioning of fillna(0).....	20
Figure 13 : functioning of interpolate().....	21
Figure 14 : mean imputation.....	22
Figure 15 : median imputation.....	23
Figure 16 : Categorical data.....	24
Figure 17 : dummy set for the above data.....	25
Figure 18 : adding the dummy set to dataframe.....	26
Figure 19 : importing the method.....	27
Figure 20 : handling categorical data.....	27
Figure 21 : importing train_test_split.....	29
Figure 22 : predicting.....	29
Figure 23 : comparing the predicted value with the original one.....	30
Figure 24 : ols and its functioning.....	31
Figure 25 : mathematical operation on input or output.....	32
ix	
Figure 26 : correlation.....	33

Figure 27 : pair plot.....	34
Figure 28 : importing the libraries.....	35
Figure 29 : reading the dataset.....	35
Figure 30 : before handling the missing values.....	36
Figure 31 : after handling the missing values.....	36
Figure 32 : getting dummies for gender column.....	37
Figure 33 : getting dummies for age column.....	37
Figure 34 : getting dummies for city_category column.....	38
Figure 35 : dropping the columns which are not required.....	38
Figure 36 : correlation.....	39
Figure 37 : getting R-squared values using ols.....	40
Figure 38 : improved R-squared value.....	40
Figure 39 : retrieving the input columns.....	41
Figure 40 : retrieving the output columns.....	41
Figure 41 : splitting the data.....	42
Figure 42 : importing the linear regression.....	42
Figure 43 : predicting the output.....	43
Figure 44 : comparing with original data.....	43

x