

A Project Report on

# **A Two Level Statistical Model for Big Mart Sales Prediction**

Submitted in partial fulfillment of the requirements for the award  
of the degree of

**Bachelor of Engineering**

in

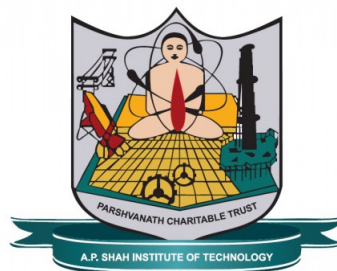
**Computer Engineering**

by

**Rudresh Lagwankar(17202008)**  
**Shubham Mahamunkar(17202007)**  
**Bilal Sayyed(17202005)**

Under the Guidance of

**Prof. Sukhada Aloni**



**Department of Computer Engineering**  
A.P. Shah Institute of Technology  
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615  
UNIVERSITY OF MUMBAI  
**Academic Year 2019-2020**

## Approval Sheet

This Project Report entitled “*A Two Level Statistical Model For Big Mart Sales Prediction*” Submitted by “*Rudresh Lagwankar*”(17202008), “*Shubham Mahamunkar*” (17202007), “*Bilal Sayyed*”(17202005) is approved for the partial fulfillment of the requirement for the award of the degree of *Bachelor of Engineering* in *Computer Engineering* from *University of Mumbai*.

(Prof. Sukhada Aloni)  
Guide

(Prof. Sachin Malave)  
Head Department of Computer Engineering

Place: A.P. Shah Institute of Technology, Thane

Date:

## CERTIFICATE

This is to certify that the project entitled “*A Two Level Statistical Model for Big Mart Sales Prediction*” submitted by “*Rudresh Lagwankar*” (17202008), “*Shubham Mahamunkar*” (17202007), “*Bilal Sayyed*” (17202005) for the partial fulfillment of the requirement for award of a degree *Bachelor of Engineering* in *Computer Engineering*, to the University of Mumbai, is a bonafide work carried out during academic year 2019-2020.

Prof. Sukhada Aloni  
Guide

Prof. Sachin Malave  
Head Department of Computer Engineering

Dr. Uttam D.Kolekar  
Principal

External Examiner(s)

1.

2.

Place: A.P. Shah Institute of Technology, Thane

Date:

## Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

---

(Signature)

---

(Rudresh Lagwankar (17202008))  
(Shubham Mahamunkar (17202007))  
(Bilal Sayyed (17202005))

Date:

## **Abstract**

Sales forecasting is an important aspect of different companies engaged in retailing, logistics, manufacturing, marketing and wholesaling. It allows companies to efficiently allocate resources, to estimate achievable sales revenue and to plan a better strategy for future growth of the company. In this paper, prediction of sales of a product from a particular outlet is performed via a two-level approach that produces better predictive performance compared to any of the popular single model predictive learning algorithms. The approach is performed on Big Mart Sales data of the year 2013. Data exploration, data transformation and feature engineering play a vital role in predicting accurate results. The result demonstrated that the two-level statistical approach performed better than a single model approach as the former provided more information that leads to better prediction.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Objective . . . . .	1
1.3	Problem Definition . . . . .	1
1.4	Scope . . . . .	1
1.5	Technology Stack . . . . .	2
1.6	Benefits for environment Society . . . . .	2
<b>2</b>	<b>Literature Review</b>	<b>3</b>
<b>3</b>	<b>Project Design</b>	<b>5</b>
3.1	Proposed System . . . . .	5
3.2	Use Case Diagram . . . . .	6
3.3	Activity Diagram . . . . .	7
<b>4</b>	<b>Modules</b>	<b>8</b>
4.1	Module-1 . . . . .	8
4.2	Module-2 . . . . .	8
4.3	Module-3 . . . . .	8
<b>5</b>	<b>Result</b>	<b>9</b>
<b>6</b>	<b>Conclusions and Future Scope</b>	<b>10</b>
	<b>Bibliography</b>	<b>11</b>
	<b>Publication</b>	<b>14</b>

# List of Abbreviations

MAE:	Mean Absolute Error
RI:	Rule Induction
k-NN:	k-Nearest Neighbor
MRP:	Maximum Retail Price

# Chapter 1

## Introduction

### 1.1 Introduction

Sales is a lifeblood of each and every company and sales forecasting Plays a vital role in conducting any business. Good forecasting helps to develop and improve business strategies by increasing the knowledge. Sales forecasting is sales prediction that is based on the available resources from the past

### 1.2 Objective

To plan for timely procurement of raw material of right quality to be available at the right time. Raw material management can save company a lot of money by maintaining sufficient stocks to produce stocks as per sales prediction.

### 1.3 Problem Definition

Accurate sales forecast is an important tool for companies to have. It helps CEOs gauge the demand for their products. It helps companies better manage inventory. Sales forecasting allows companies to see into the future and strategically plan their moves to increase growth. we will develop web.

### 1.4 Scope

1. In this paper, prediction of sales of a product from a particular outlet is performed via a two-level approach that produces better predictive performance compared to any of the popular single model predictive learning algorithms. 2. The approach is performed on Big Mart Sales data of the year 2013. Data exploration, data transformation and feature engineering play a vital role in predicting accurate results. 3. The result demonstrated that the two-level statistical approach performed better than a Single model approach as the former provided more information that leads to better prediction.



## 1.5 Technology Stack

### 1. Google Colab:

Google Colab is a free cloud service and now it supports free GPU! You can: improve your Python programming language coding skills.

### 2. Python Flask:

Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.

## 1.6 Benefits for environment Society

1. Two-Level Statistical Prediction Model systems help the big mart Manager to get sales prediction for future, helps big mart manager to take correct decisions and redefine the sales management

2. A Two-Level Statistical Model provide prediction helps to manage business.

# Chapter 2

## Literature Review

1).Business data mining - A machine learning perspective Bose, Indranil, and Radha K. Mahapatra, Information management, (3), pp.Vol. 39211-225, February, 2001.

The objective of this paper is to inform the information systems (IS) manager and business analyst about the role of machinelearning techniques in business data mining. Data mining is a fast growing application area in business. Machine learningtechniques are used for data analysis and pattern discovery and thus can play a key role in the development of data miningapplications. Understanding the strengths and weaknesses of these techniques in the context of business is useful in selectingan appropriate method for a speci@c application.

2).Is Combining Classifiers with Stacking Better Than Selecting the Best One? Deroski, Saso and Bernard Enko.”, Machine learning, Vol. 54(3), pp. 255-273, March.

We empirically evaluate several state-of-theart methods for constructing ensembles of heterogeneous classifiers with stacking and show that they perform (at best) comparably to selecting the best classifier from the ensemble by cross validation. We then propose a new method for stacking, that uses multi-response model trees at the meta-level,and show that it clearly outperforms existing stacking approaches and selecting the best classifier by cross validation.

**3).A Survey on Retail Sales Forecasting and Prediction in Fashion Markets**  
**Beheshti-Kashi and Samaneh, Systems Science Control Engineering, Vol. 3, pp.**  
**154.**

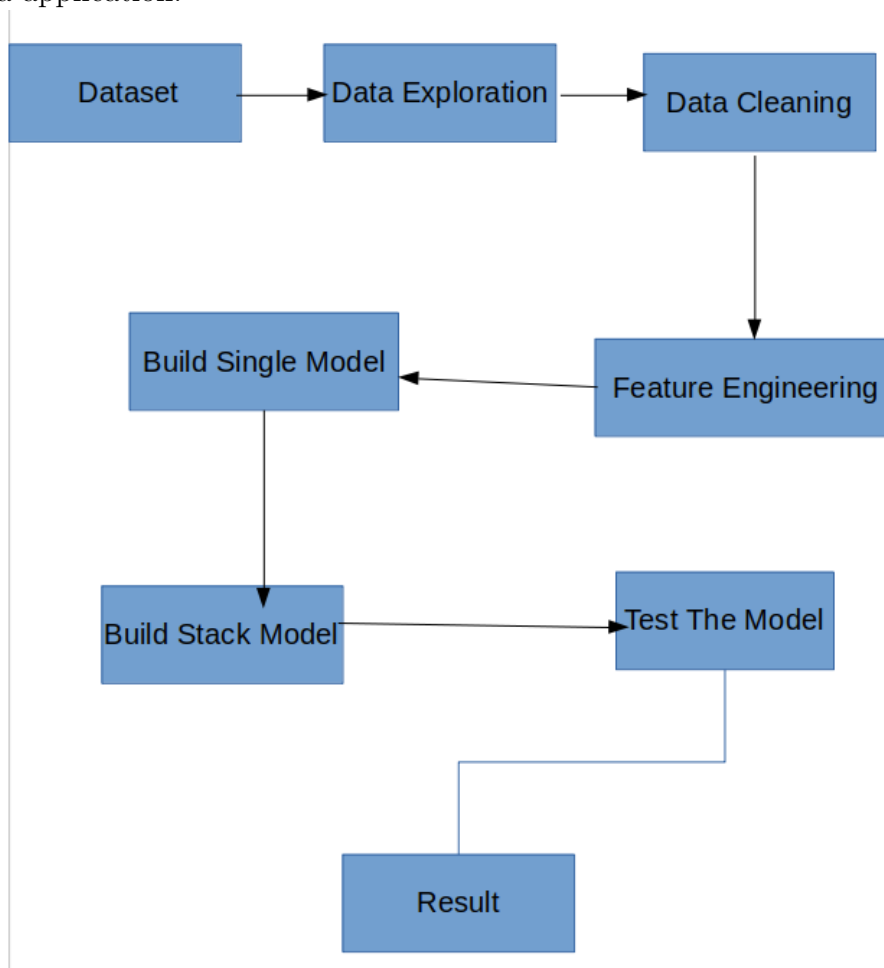
Sales forecasting is an essential task in retailing. In particular, consumer-oriented markets such as fashion and electronics face uncertain demands, short life cycles and a lack of historical sales data which strengthen the challenges of producing accurate forecasts. This survey paper presents state-of-the-art methods in the sales forecasting research with a focus on fashion and new product forecasting. This study also reviews different strategies to the predictive value of user-generated content and search queries.

# Chapter 3

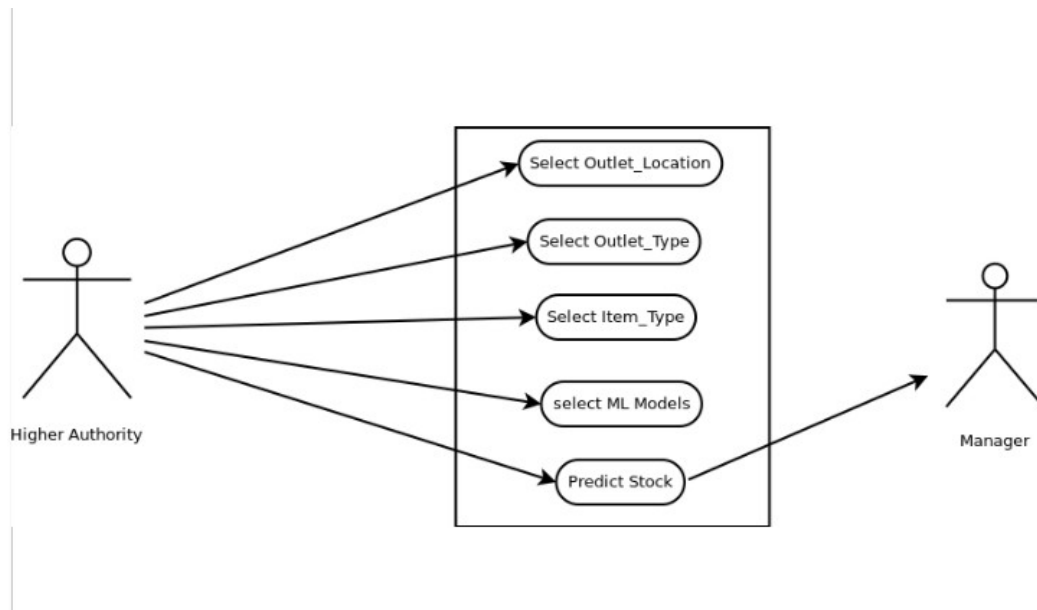
## Project Design

### 3.1 Proposed System

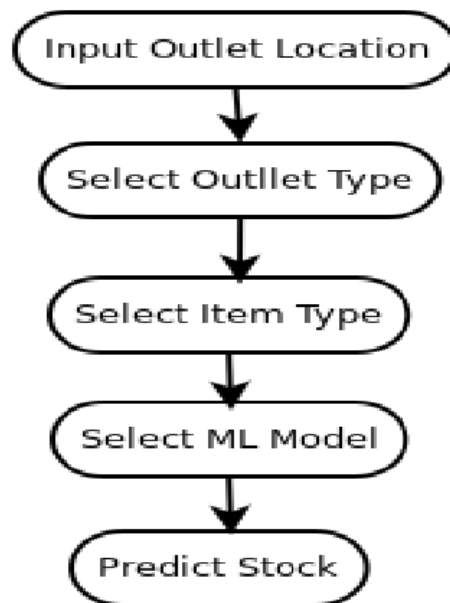
The given system will be an application that will be simply intergrated with a predicting sales of product. The process of predicting sales becomes much more simpler with the help of a application.



## 3.2 Use Case Diagram



### 3.3 Activity Diagram



# Chapter 4

## Modules

### 4.1 Module-1

Dataset Generation : We have studied and done research on big mart sales and we had downloaded the dataset from [www.kaggle.com](http://www.kaggle.com).

### 4.2 Module-2

User-interface: We will create UI in which we can perform operations like selection Of Models,selection of State and Items,analyse the Better Results.

### 4.3 Module-3

Machine learning Module. We are using two models:

- 1)Linear Regression.
- 2)Support Vector Regression.

# Chapter 5

## Result

We have learned how to use linear regression and Support Vector regression. We have also learned to use Overleaf and Google collab.



# Chapter 6

## Conclusions and Future Scope

Each and every company desires to know the demand of the customer in any season beforehand to avoid the shortage of products. As time passes by, the demand of the companies to be more accurate about the predictions will increase exponentially. So, huge research is going on in this sector to make accurate predictions of sales. Better predictions are directly proportional to the profit made by the company. We Will implement two way statistical model by combining ml models Like linear regression, support vector regression to predict sales of Perticular product in perticular region.

# Bibliography

- [1] Business data mining - A machine learning perspective Bose, Indranil, and Radha K.Mahapatra, Information management, (3), pp.Vol. 39211-225, February, 2001.
- [2] Is Combining Classifiers with Stacking Better Than Selecting the Best One? Deroski,Saso and Bernard Enko.”, Machine learning, Vol. 54(3), pp. 255-273, March.
- [3] A Survey on Retail Sales Forecasting and Prediction in Fashion Markets Beheshti- Kashi and Samaneh,Systems Science Control Engineering, Vol. 3, pp. 154

# Appendices

## 1. Installing Python on Windows

There are quite a few implementations of Python to choose from in Windows, but for purposes of this guide, I'll be using the .msi installer found on [python.org](https://python.org).

Begin by downloading the installer from the Python download page. Run the installer - you can accept all the default settings. To see if your install is working as expected, open a command prompt and run `python -V`.

## Acknowledgement

We have great pleasure in presenting the report on **A Two-Level Statistical Model for Big Mart Sales Prediction**. We take this opportunity to express our sincere thanks towards our guide **Prof. Sukhada Aloni** Department of Computer, APSIT Thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of project.

We thank **Prof. Sachin Malave** Head of Department, Computer, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof. Amol Kalugade** BE project co-ordinator, Department of COMPUTER, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

**Student Name1: Rudresh Lagwankar**  
**Student ID1: 17202008**

**Student Name2: Shubham Mahamunkar**  
**Student ID2: 17202007**

**Student Name3: Bilal Sayyed**  
**Student ID3: 17202005**

# Publication

Paper entitled “**A Two Level Statistical Model for Big Mart Sales Prediction**” is presented at “**International Conference on Computing, Power and Communication Technologies (GUCON)**” by “**Kumari Punam**”.