PREDICTIVE ANALYSIS OF BIGMART SALES USING MACHINE LEARNING

A Project Report submitted in the partial fulfilment of the requirements for the award of the degree

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted by

B.Sriram (19471A05405)

Ch.Chandra Sekhar (19471A0508)

G.Jashuva (19471A0524)

Under the esteemed guidance of

D.Venkat Reddy **M.Tech., (Ph.D)**

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NARASARAOPETA ENGINEEING COLLEGE

(AUTONOMOUS)

Accredited by NAAC with A+ Grade and NBA under Cycle-I

NIRF rank in the band of 251-320 and an ISO 9001:2015 Certified

Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada

KOTAPPAKONDA ROAD, YALAMANDA VILLAGE, NARASARAOPET- 522601

2022-2023

NARASARAOPETA ENGINEEING COLLEGE
(AUTONOMOUS)
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled "PREDICTIVE ANALYSIS OF BIGMART SALESUSING MACHINE LEARNING" is a bonafide Work done by **"B.Sriram (19471A0505), Ch.ChandraSekhar (19471A0508), G.Jashuva (19471A0524)"** in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in the Department of **COMPUTER SCIENCE AND ENGINEERING** during 2022-2023

PROJECT GUIDE
Mr.D.Venkat Reddy M.Tech., (Ph.D)
Asst. Professor

PROJECT CO-ORDINATOR
Dr. M. Sireesha M.Tech., Ph.D
Assoc. Professor

HEAD OF THE DEPARTMENT
Dr. S. N. Tirumala Rao, M.Tech., Ph.D

EXTERNAL EXAMINER

DECLARATION

We declare that this project work titled “PREDICTIVE ANALYSIS OF BIGMART SALES USING MACHINE LEARNING” is composed by ourself that the work contain here is ourown except where explicitly stated otherwise in the text and that this work has been submitted for any other degree or professional qualification except as specified

B.Sriram (19471A0505)
Ch.Chandra Sekhar (19471A0508)
G.Jashuva (19471A0524)

ACKNOWLEDGEMENT

We wish to express our thanks to carious personalities who are responsible for the completion of the project. We are extremely thankful to our beloved chairman sri **M.V.Koteswara Rao,B.Sc.,** who took keen interest in us in every effort throughout this course. We owe out sincere gratitude to our beloved principal **Dr.M.Sreenivasa Kumar, M.Tech., Ph.D., MISTE., FIE(I).,** for showing his kind attention and valuable guidance throughout the course.

We express our deep felt gratitude towards **Dr.S.N.Tirumala Rao,M.Tech.,Ph.D.,** HOD of CSE department and to our guide **Mr.D.Venkat Reddy ,M.Tech.,(Ph.D.),**of CSE department whose valuable guidance and unstinting encouragement enable us to accomplish our project successfully in time.

We extend our sincere thanks towards **Dr. M.Sireesha, M.Tech., Ph.D.,** Associate professor & Project coordinator of the project for extending her encouragement. Their profound knowledge and willingness have been a constant source of inspiration for us throughout this project work.

We extend our sincere thanks to all other teaching and non-teaching staff to department for their cooperation and encouragement during our B.Tech degree.

We have no words to acknowledge the warm affection, constant inspiration and encouragement that we received from our parents.

We affectionately acknowledge the encouragement received from our friends and those who involved in giving valuable suggestions had clarifying out doubts which had really helped us in successfully completing our project.

By

B.Sriram (19471A0505)
Ch.Chandra Sekhar (19471A0508)
G.Jashuva (19471A0524)

ABSTRACT

- Currently, supermarket run-centres, Big Marts keep track of each individual item's sales data in order to anticipate potential consumer demand and update inventory management.

- Anomalies and general trends are often discovered by mining the data warehouse's data store. For retailers like Big Mart, the resulting data can be used to forecast future sales volume using various machine learning techniques like big mart.

- A predictive model was developed using Xgboost, Linear regression, Polynomial regression, and Ridge regression techniques for forecasting the sales of a business such as Big -Mart, and it was discovered that the model outperforms existing models.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

INSTITUTE VISION AND MISSION

INSTITUTION VISION

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community.

INSTITUTION MISSION

M1: Provide the best class infra-structure to explore the field of engineering and research

M2: Build a passionate and a determined team of faculty with student centric teaching, imbibing experiential, innovative skills

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION OF THE DEPARTMENT

To become a center of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

MISSION OF THE DEPARTMENT

The department of Computer Science and Engineering is committed to

M1: Mold the students to become Software Professionals, Researchers and Entrepreneurs by providing advanced laboratories.

M2: Impart high quality professional training to get expertise in modern software tools and technologies to cater to the real time requirements of the Industry.

M3: Inculcate team work and lifelong learning among students with a sense of societal and ethical responsibilities.

NARASARAOPETA
NEC ENGINEERING COLLEGE
(AUTONOMOUS)

Program Specific Outcomes (PSO's)

PSO1: Apply mathematical and scientific skills in numerous areas of Computer Science and Engineering to design and develop software-based systems.

PSO2: Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering

PSO3: Promote novel applications that meet the needs of entrepreneur, environmental and social issues.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Educational Objectives (PEO's)

The graduates of the programme are able to:

PEO1: Apply the knowledge of Mathematics, Science and Engineering fundamentals to identify and solve Computer Science and Engineering problems.

PEO2: Use various software tools and technologies to solve problems related to academia, industry and society.

PEO3: Work with ethical and moral values in the multi-disciplinary teams and can communicate effectively among team members with continuous learning.

PEO4: Pursue higher studies and develop their career in software industry

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Outcomes

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Project Course Outcomes (CO'S)

CO425.1: Analyze the System of Examinations and identify the problem.

CO425.2: Identify and classify the requirements.

CO425.3: Review the Related Literature

CO425.4: Design and Modularize the project

CO425.5: Construct, Integrate, Test and Implement the Project.

CO425.6: Prepare the project Documentation and present the Report using appropriate method.

Course Outcomes – Program Outcomes mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1		√											√		
C425.2	√		√		√								√		
C425.3				√		√	√	√					√		
C425.4			√			√	√	√					√	√	
C425.5					√	√	√	√	√	√	√	√	√	√	√
C425.6									√	√	√		√	√	

Course Outcomes – Program Outcome correlation

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1	2	3											2		
C425.2			2		3								2		
C425.3				2		2	3	3					2		
C425.4			2			1	1	2					3	2	
C425.5					3	3	3	2	3	2	2	1	3	2	1
C425.6									3	2	1		2	3	

Note: The values in the above table represent the level of correlation between CO's and PO's:

1. **Low level**
2. **Medium level**
3. **High level**

Project mapping with various courses of Curriculum with Attained PO's:

Name of the course from which principles are applied in this project	Description of the device	Attained PO
C3.2.4, C3.2.5	Gathering the requirements and defining the problem, plan to develop a **smart bottle for health care using sensors.**	PO1, PO3
CC4.2.5	Each and every requirement is critically analyzed, the process model is identified and divided into **five modules**	PO2, PO3
CC4.2.5	Logical design is done by using the unified modelling language which involves individual team work	PO3, PO5, PO9
CC4.2.5	Each and every module is tested, integrated, and evaluated in our project	PO1, PO5
CC4.2.5	Documentation is done by all our four members in the form of a group	PO10
CC4.2.5	Each and every phase of the work in group is presented periodically	PO10, PO11
CC4.2.5	Implementation is done and the project will be handled by the **hospital management and in future updates in our project can be done based on air bubbles occurring in liquid in saline.**	PO4, PO7
CC4.2.8 CC4.2.	**The physical design includes hardware components like sensors, gsm module, software and Arduino.**	PO5, PO6

INDEX

List of figures

1. INTRODUCTION

1.1 Introduction

- Everyday competitiveness between various shopping centres as and as huge marts is becoming higher intense, violent just because of the quick development of global malls also online shopping.

- Each market seeks to offer personalized and limited-time deals to attract many clients relying on period of time, so that each item's volume of sales may be estimated for the organization's stock control, transportation and logistical services.

- The current machine learning algorithm is very advanced and provides methods for predicting or forecasting sales any kind of organization, extremely beneficial to overcome low – priced used for prediction.

- Always better prediction is helpful, both in developing and improving marketing strategies for the marketplace, which is also particularly helpful.

1.2 Existing System

In existing system they used, Artificial neural organizations for nearby income estimations. It implements k-means algorithm. To develop a few deals of forecast standards regression, Auto regressive integrated moving average(ARIMA), Autoregressive moving average(ARMA) are used

Advantages:

1. It keep track of each item sales data in order to forecast implicit consumer demand and update force operation.
2. It aims to provide offers and limited time deals to attract numerous guests over time

Dis-advantages:

1. K-means algorithm produce less accurate scores in prediction.
2. ARIMA has poor performance for long term forecasts and It is computationally expensive

1.3 Proposed System

The proposed system gives most effective analytics solution for sales forecasting which focuses on colourful algorithm operations to the dataset.

The system is more effective to handle massive datasets due to the inclusion of Ridge regression and linear regression models

Advantages:

1.The accurate result is available in the form of graph and pie chart for easy understanding.
2. It also helps retailer to get how to improve his sales and fulfill the demands of customer

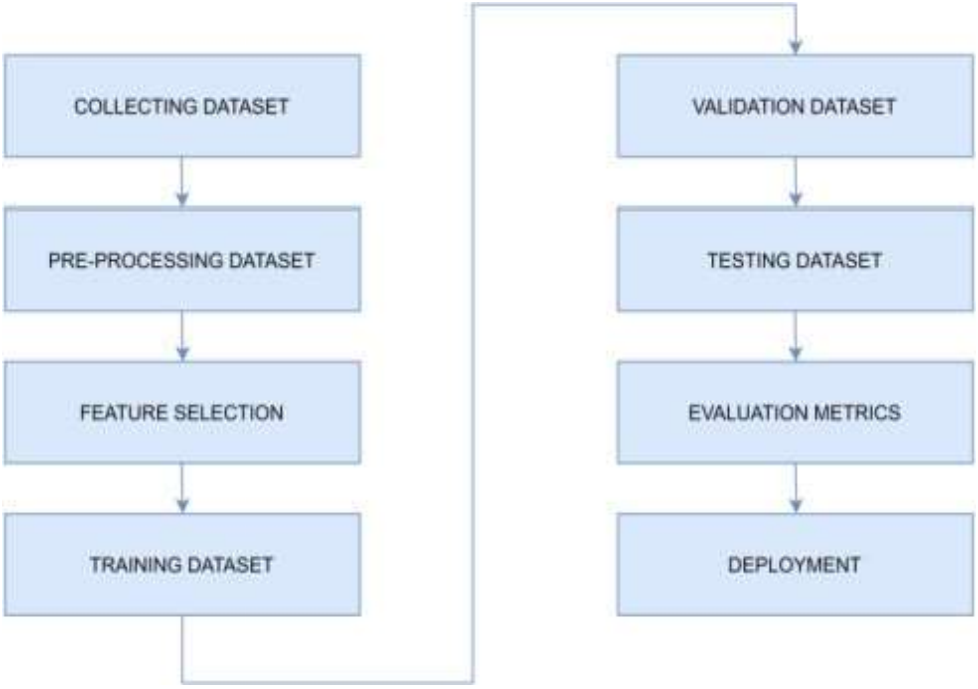


Fig: Proposed System

1) Input: The input is a dataset in which the dataset contains different types of Open, Close, Low, High, Volume, Adjacent Close and the Date. The input variables are the company stock price for the day. The Open column represent about the stock at which cost it has opened. The Close column is used to tell the closing price of the stock in a day. The Low column will tells about the stock price how the price has been dropped in that particular day. The High column is opposite to the low column in the dataset. The Volume column is used to show about how many stock of a company are in market and how many are sold and how many are in buy. The Adjacent close tells about the volume of stock of the company.

2) Pre-processing: Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. The Pre-Processing will be used for removal of noisy data from the dataset. The dataset is checked for the noisy (or) missing values in it, if there is a noisy data present in it, then it will be removed and replaced in the dataset. By that the data prediction is accurate and the prediction will be done. Due to present of the noisy data the prediction may inaccurate.

3) Feature extraction: Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process. Feature extraction is the name for methods that select and /or combine variables into features, effectively reducing the amount of data that must be processed, while still accurately and completely describing the original data set. The feature extraction will be done on the large dataset for the dimensionality Reduction in thedataset.

4) Training Dataset: The deep learning method Neural Network (NN) was used for the clustering and also classify the data. The Neural Networks are used in the dataset for the prediction propose in the dataset. The dataset is classified into training dataset and the test dataset which are used for the prediction propose. The Neural Network plays a major to classify the data into train and test dataset. The train dataset will give training in such a way that will used to test the test dataset. The training dataset is about 70% to 80% in size.
5) Validation dataset: The validation dataset is used for the validation propose of the dataset. The data will be used for the test dataset which will be more accurate. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation dataset is between the both train and test dataset.

6) Testing Dataset: The test dataset will be the last dataset which was divided. The size of the dataset is up to 20% to 30% for the prediction propose for the dataset. The test dataset will be used for the prediction and it uses the train dataset for it. The test dataset is another subset of original data, which is independent of the training dataset.

7) Evaluation Metrics: The Evaluation Metrics is about the accuracy of the models which is used for the given metrics in the dataset. The dataset will be taken and perform the different types of the models to find the accuracy. By finding the accuracy for different models gives an idea about the model and its accuracy.

8) Deployment: The deployment is a method showing output to the given dataset it also known as the Visualization of the dataset. By deploying the dataset there will idea about the different types of the output which makes a different types of the attributes and also the prediction of the dataset in the output screen which is deployment of the dataset.

1.4 System Requirements

Hardware Requirements:

- Processor : Intel(R) Core™2 i5-5500U CPU @ 2.50GHZ
- RAM : 8GB(gigabyte)
- System Type : 64- bit operating system, x64-based processor

Software Requirements:

- Browser : Any Latest browser like Chrome
- Operating System : Windows 10
- Language : Python
- Platform : Jupyter Notebook

2. LITERATURE

2.1 Literature Survey

Several studies have been conducted in the field of predictive analytics for retail sales. Here are some relevant literature and research articles related to our project:

Predictive analytics for sales forecasting and strategic planning in the retail industry" by Barreto and Ferreira. This study explores the use of predictive analytics for sales forecasting in the retail industry. The authors found that predictive analytics can improve the accuracy of sales forecasting and help retailers make better decisions. Predictive modeling in retail: A study on sales forecasting and customer lifetime value" by Gaurav et al. This study examines the use of predictive modeling for sales forecasting and customer lifetime value in the retail industry. The authors found that predictive modeling can help retailers make more accurate sales forecasts and identify high-value customers.Sales forecasting using machine learning algorithms: A case study of a retail chain" by Mohd et al. This study uses machine learning algorithms to predict sales in a retail chain. The authors found that the random forest algorithm provided the best results for sales forecasting.Predictive modeling of retail sales using machine learning algorithms" by Anwar et al. This study explores the use of machine learning algorithms for predictive modeling of retail sales. The authors found that the gradient boosting algorithm provided the best results for sales forecasting.

Overall, the literature suggests that predictive analytics can be a powerful tool for sales forecasting and strategic planning in the retail industry. Machine learning algorithms such as random forest and gradient boosting have been found to be effective for predictive modeling of retail sales. Our project aims to build on these previous studies and provide insights into sales prediction for BigMart.

2.2 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

2.3 Machine Learning Methods

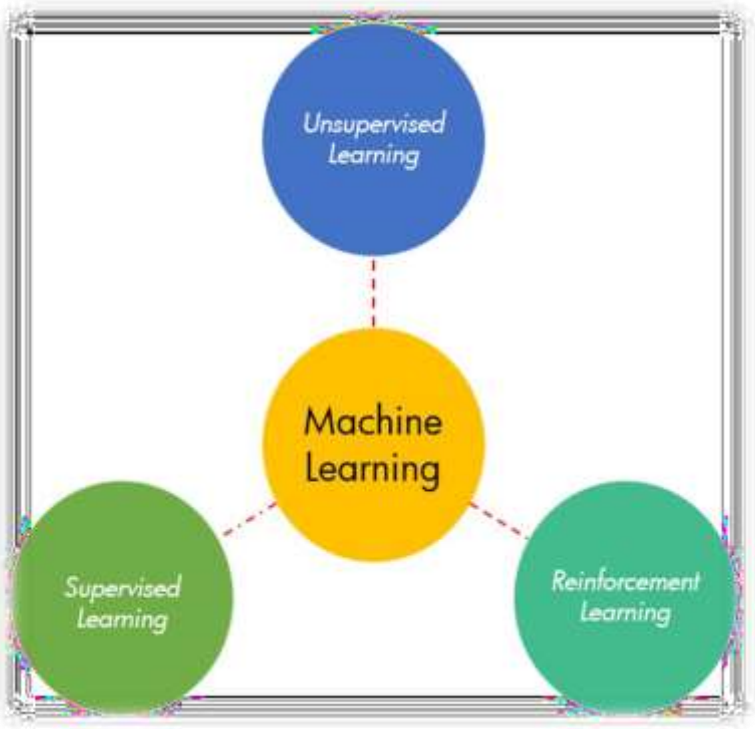


Fig : 2.1 Machine Learning Methods

Supervised machine learning algorithms: Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

Unsupervised machine learning algorithms: Unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

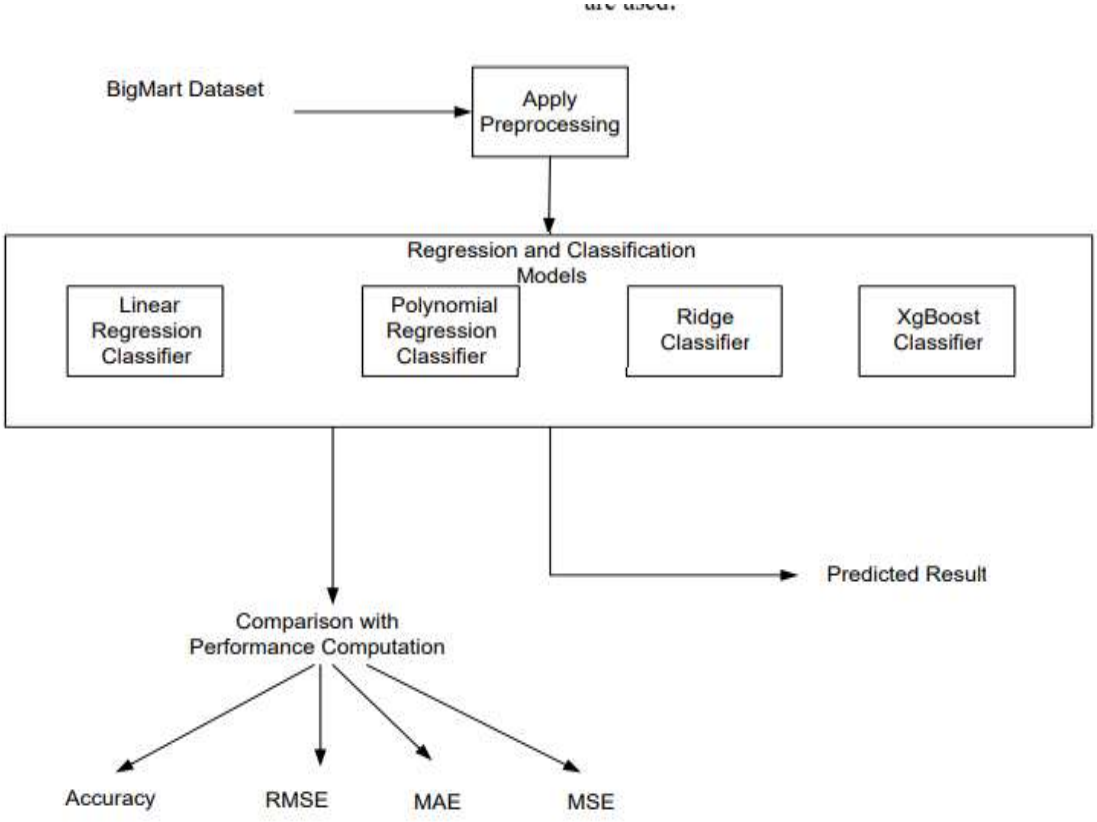
Reinforcement machine learning algorithms: Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best. This is known as the reinforcement signal.

2.4 Applications of Machine Learning

1. Virtual Personal Assistants
2. Predictions while Commuting
3. Videos Surveillance
4. Social Media Services
5. Email Spam and Malware Filtering
6. Online Customer Support
7. Search Engine Result Refining
8. Product Recommendations
9. Online Fraud Detection

3.SYSTEM ANALYSIS

3.1 METHODOLOGY



- The architecture Diagram of the proposed model where they focus on the different algorithm application to the dataset. Where we are calculating the Accuracy, MAE, MSE, RMSE and final concluding the best yield algorithm. Here are the following Algorithm are used

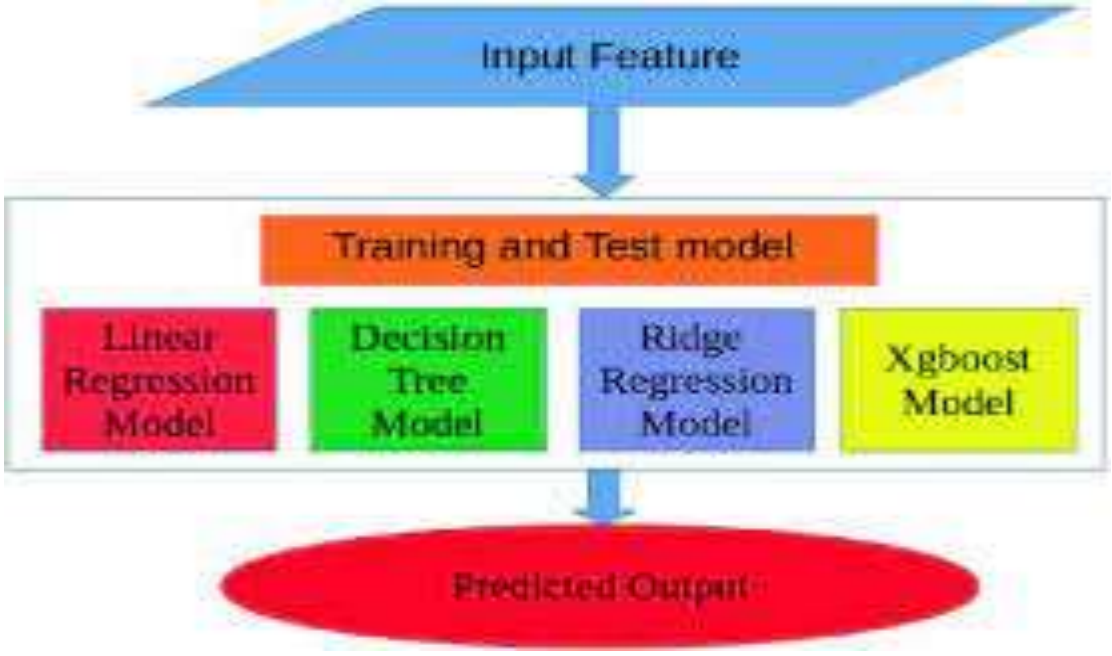


Fig:3.1.1 Architecture

3.2 Implementation of machine learning using Python

Python is a popular programming language. It was created in 1991 by Guido van Rossum.

It is used for:

1. Web development (server-side)
2. Software development
3. Mathematics
4. System scripting

The most recent major version of Python is Python 3. However, Python 2, although not being updated with anything other than security updates, is still quite popular.It is possible to write Python

in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse, Anaconda which are particularly useful when managing larger collections of Python files.Python was designed for its readability. Python uses new lines to complete a command, asopposed to other programming languages which often use semicolons or parentheses.

Python relies on indentation, using whitespace, to define scope; such as the scope of loops,functions and classes. Other programming languages often use curly-brackets for this purpose.

In the older days, people used to perform Machine Learning tasks manually by coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that used in Machine Learning are:

1. Numpy
2. Scipy
3. Scikit-learn
4. Pandas
5. Matplotlib

NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra,

Fourier transform, and random number capabilities. High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

SciPy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.

Skikit-learn is one of the most popular Machine Learning libraries for classical Machine Learning algorithms. It is built on top of two basic Python libraries, NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with Machine Learning.

Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

Matpoltlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, histogram, error charts, bar chats, etc.

3.3. Scope of the Project

The scope of this system is to maintain sales details in datasets, train the model using thelarge quantity of data present in datasets and predict whether increases or decreases of sales on new data during testing.

3.4. Analysis

The dataset contains 12 attributes which are used to predictive analysis ofthe Bigmart sales using Machine learning such as

TABLE 1: Attributes Information

Attribute	Description
Item_Identifer	It is the unique product Id number.
Item Weight	It will include the product's weight.
Item_Fat_Content	It will mean whether the item is low in fat or not.
Item -Visibility	The percentage of the overall viewing area assigned to the particular item from all items in the shop.
Item -Type	To which group does the commodity belong
Item-MRP	The product's price list
Outlet-Identifier	a distinct slot number
Outlet-Establishment Year	The year that the shop first opened its doors.
Outlet-Size	The sum of total area occupied by a supermarket.
Outlet-Location	The kind of town where the store is situated.
Outlet-Type	The shop is merely a supermarket or a grocery store.
Item-Outlet-Sales	The item's sales in the original shop

DataSet:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Item_Iden	Item_Wei	Item_Fat_	Item_Visi	Item_Type	Item_MRP	Outlet_Id	Outlet_Es	Outlet_Si	Outlet_Lo	Outlet_Ty	Item_Outlet_Sales	
2	FDA15	9.3	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermar	3735.138	
3	DRC01	5.92	Regular	0.019278	Soft Drink	48.2692	OUT018	2009	Medium	Tier 3	Supermar	443.4228	
4	FDN15	17.5	Low Fat	0.01676	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermar	2097.27	
5	FDX07	19.2	Regular	0	Fruits and	182.095	OUT010	1998		Tier 3	Grocery St	732.38	
6	NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermar	994.7052	
7	FDP36	10.395	Regular	0	Baking Go	51.4008	OUT018	2009	Medium	Tier 3	Supermar	556.6088	
8	FDO10	13.65	Regular	0.012741	Snack Foo	57.6588	OUT013	1987	High	Tier 3	Supermar	343.5528	
9	FDP10		Low Fat	0.12747	Snack Foo	107.7622	OUT027	1985	Medium	Tier 3	Supermar	4022.764	
10	FDH17	16.2	Regular	0.016687	Frozen Fo	96.9726	OUT045	2002		Tier 2	Supermar	1076.599	
11	FDU28	19.2	Regular	0.09445	Frozen Fo	187.8214	OUT017	2007		Tier 2	Supermar	4710.535	
12	FDY07	11.8	Low Fat	0	Fruits and	45.5402	OUT049	1999	Medium	Tier 1	Supermar	1516.027	
13	FDA03	18.5	Regular	0.045464	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermar	2187.153	
14	FDX32	15.1	Regular	0.100014	Fruits and	145.4786	OUT049	1999	Medium	Tier 1	Supermar	1589.265	
15	FDS46	17.6	Regular	0.047257	Snack Foo	119.6782	OUT046	1997	Small	Tier 1	Supermar	2145.208	
16	FDF32	16.35	Low Fat	0.068024	Fruits and	196.4426	OUT013	1987	High	Tier 3	Supermar	1977.426	
17	FDP49	9	Regular	0.069089	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermar	1547.319	
18	NCB42	11.8	Low Fat	0.008596	Health an	115.3492	OUT018	2009	Medium	Tier 3	Supermar	1621.889	
19	FDP49	9	Regular	0.069196	Breakfast	54.3614	OUT049	1999	Medium	Tier 1	Supermar	718.3982	
20	DRI11		Low Fat	0.034238	Hard Drink	113.2834	OUT027	1985	Medium	Tier 3	Supermar	2303.668	
21	FDU02	13.35	Low Fat	0.102492	Dairy	230.5352	OUT035	2004	Small	Tier 2	Supermar	2748.422	
22	FDN22	18.85	Regular	0.13819	Snack Foo	250.8724	OUT013	1987	High	Tier 3	Supermar	3775.086	

Fig: 3.4.1 Dataset

Fig 3.4 is the data set of predictive Analysis of bigmart sales using machine learning models which contains attributes Item_idebtifier, Item_weight, Item_fat_content, Item_visibility, Item_type, Item_MRP, Outlet_Identifier, Outlet_Establishment_year, Outlet_Location, Outlet_Size, Outlet_Type, Item_Outlet_Sales.

. Data Pre-processing

Before feeding data to an algorithm we have to apply transformations to our data which is referred as pre-processing. By performing pre-processing the raw data which is not feasible for analysis is converted into clean data. In-order to achieve better results using a model in Machine Learning, data format has to be in a proper manner. The data should be in a particular format for different algorithms. For example, if we consider Random Forestalgorithm it does not support null values.

So that those null values have to be managed using raw data.

Data Pre-processing:

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

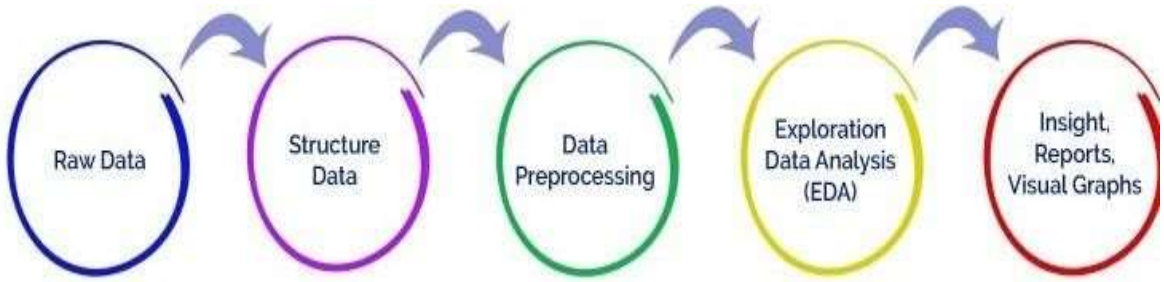


Fig:3.4.2 Data Preprocessing

Need of Data Preprocessing: For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format. For example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set. Another aspect isthat data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

3.5.1 Missing values

Filling missing values is one of the pre-processing techniques. The missing values in the dataset is represented as '?' but it a non-standard missing value and it has to be converted into a standard missing value NaN. So that pandas can detect the missing values. In my collected dataset there are no missing values.

3.5.2 Correlation coefficient method

We can find dependency between two attributes p and q using Correlation coefficient method using the formula.

rp, q= ∑(pi-p)(qi-q)/nσpσq =∑(pi qi)-np q/ nσpσq

n is the total number of patterns, pi and qi are respective values of p and q attributes in patterns i, p and q are respective mean values of p and q attributes, σp , σq are respective standard deviations values of p and q attributes. Generally, -1≤ rp,q ≤ +1. If rp,q < 0, thenp and q are negatively correlated. If rp,q =0, then p and q are independent attributes and there is no correlation between them. If rp,q > 0, then p and q are positively correlated. Wecan drop the attributes that are having correlation coefficient value as 0 as it indicates that the variables are independent with respect to the prediction attribute. Fig:3.8.2 is the correlation matrix. There are no correlated features in the dataset

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	O
Item_Identifier	1.000000	0.044518	-0.114660	-0.025482	-0.017973	0.012853	-0.008602	-0.012772	
Item_Weight	0.044518	1.000000	-0.021157	-0.012049	0.028015	0.024756	-0.007975	-0.008301	
em_Fat_Content	-0.114660	-0.021157	1.000000	0.047314	-0.139434	0.006063	0.000764	0.003151	
Item_Visibility	-0.025482	-0.012049	0.047314	1.000000	-0.035249	-0.001315	-0.100439	-0.074834	
Item_Type	-0.017973	0.028015	-0.139434	-0.035249	1.000000	0.032651	0.001656	0.004970	
Item_MRP	0.012853	0.024756	0.006063	-0.001315	0.032651	1.000000	0.003319	0.005020	
Outlet_Identifier	-0.008602	-0.007975	0.000764	-0.100439	0.001656	0.003319	1.000000	0.079035	
_Establishment_Year	-0.012772	-0.008301	0.003151	-0.074834	0.004970	0.005020	0.079035	1.000000	
Outlet_Size	0.000855	-0.015096	-0.001365	0.087796	0.000827	-0.002473	0.053224	0.470343	
tlet_Location_Type	0.003656	0.004068	-0.001598	-0.028099	0.003084	0.000232	-0.716176	-0.089216	
Outlet_Type	-0.001178	-0.000566	0.002199	-0.173468	0.003053	-0.001975	0.099873	-0.122304	
em_Outlet_Sales	0.002869	0.011550	0.018719	-0.128625	0.017048	0.567574	0.162325	-0.049135	

.

Fig:3.5.2.1 Correlation

3.5.3 Cross Validation:

Cross-validation is a technique in which we train our model using the subset of the data- set and then evaluate using the complementary subset of the data-set. The three steps involved in cross-validation are as follows :

– Reserve some portion of sample data-set.

– Using the rest data-set train the model.

– Test the model using the reserve portion of the data-set

3.5.4 Information Gain:

Information gain is a preprocessing technique, which is used to calculate the reduction in entropy. It is commonly used in the construction of decision trees from a training dataset, by evaluating the information gain for each variable, and selecting the variable that maximizes the information gain, which in turn minimizes the entropy and best splits the dataset into groups for effective classification.

3.5.5. Data Visualization:

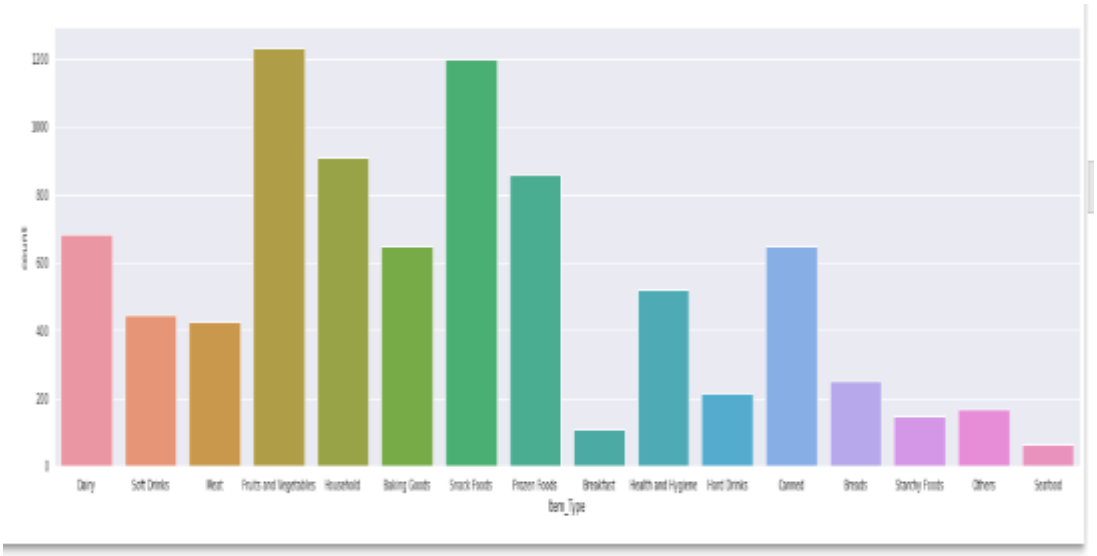


Fig: 3.5.5.1 Data Visualization

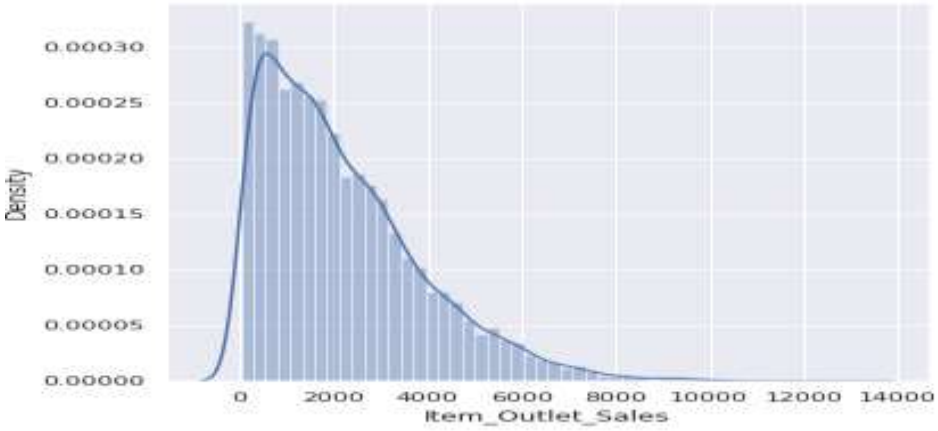


Fig:3.5.5.2 Item Outlet Sales

3.6 Classification

- It is a process of categorising data into given classes. Its primary goal is toidentify the class of our new data

.

3.6.1 Machine Learning algorithm for classification

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models orto draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Linear Regression, Polynomial Regression, Ridge Regression, Xgboost Regression, etc.

1.Linear Regression

Build a fragmented plot.1) a linear or non-linear pattern of data and 2) a variance (outliers). Consider a transformation if the marking isn't linear. If this is the case, outsiders, it can suggest only eliminating them if there is a non-statistical justification.Link .The data to the least squares line and confirm the model assumptions using the residual plot (for the constant standard deviation assumption) and the normal probability plot (for the normal probability assumption) A transformation might be necessary if the assumptions made do not appear to be met. If required, convert the data to theleast square using the transformed data, construct a regression line

When a "good-fit" classic is defined, write the least-square regression line equation. Consist of normal estimation, estimation, and Rsquared errors.

- Linear regression formulas look like this:

Y=o1x1+ o2x2+…… onxn

R-Square: Defines the difference in X (depending variable) explains the total variance in Y (dependent variable) (independent variable). This can be expressed mathematically as

$$R - Square = 1 - \frac{\sum(Y_{actual} - Y_{predicted})^2}{\sum(Y_{actual} - Y_{mean})^2}$$

2.Polynominal Regression :

- Polynomial Regression is a relapse calculation that modules the relationship here among dependent(y) and the autonomous variable(x) in light of the fact that as most extreme limit polynomial. The condition for polynomial relapse is given beneath: y= b0+b1x1+ b2x1 2+ b2x1 3+..... bnx1n
- It is regularly alluded to as the exceptional instance of various straight relapse in ML. Since we apply some polynomial terms to the numerous straight relapse condition to change it to polynomial relapse adjustment to improve accuracy.
- The informational collection utilized for preparing in polynomial relapse is of a non-straight nature
- It uses a linear regression model to fit complex and non-linear functions and dataset

3. Ridge Regression :

Ridge regression is a model tuning tool used to evaluate any data that suffers from multicollinearity. This method performs the L2 regularization procedure. When multicollinearity issues arise, the least squares are unbiased and the variances are high, resulting in the expected values being far removed from the actual values.

The cost function for ridge regression:

$$\text{Min}(||Y - X(\text{theta})||^2 + \lambda||\text{theta}||^2)$$

4. XGBoost Regression :

"Extreme Gradient Boosting" is same but much more effective to the gradient boosting system. It has both a linear model solver and a tree algorithm Which permits "xgboost" in any event multiple times quicker than current slope boosting executions. It underpins various target capacities, including relapse, order and rating. As "xgboost" is extremely high in prescient force however generally delayed with organization, it is appropriate for some rivalries. It likewise has extra usefulness for cross-approval and finding significant factors.

4.IMPLEMENTATION CODE

db.py:

```
#importing packages

import numpy as np
import pandas as pd
import pickle
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import LabelEncoder
from xgboost import XGBRegressor
from sklearn.model_selection iexmport train_test_split
from sklearn import metrics
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import Ridge
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestRegressor



big_mart_data = pd.read_csv('Train.csv')

# find missing values
big_mart_data.isnull().sum()

big_mart_data['Item_Weight'].mean()

big_mart_data['Item_Weight'].fillna(big_mart_data['Item_Weight'].mean(), inplace=True)
mode_of_Outlet_size=big_mart_data.pivot_table(values='Outlet_Size',
columns='Outlet_Type', aggfunc=(lambda x: x.mode()[0]))

miss_values = big_mart_data['Outlet_Size'].isnull()
```

```
big_mart_data.head()
big_mart_data['Item_Fat_Content'].value_counts()

big_mart_data.replace({'Item_Fat_Content': {'low fat': 'Low Fat', 'LF': 'Low Fat', 'reg':
'Regular'}}, inplace=True)

big_mart_data['Item_Fat_Content'].value_counts()

encoder = LabelEncoder()

big_mart_data['Item_Identifier']=encoder.fit_transform(big_mart_data['Item_Identifier'])

big_mart_data['Item_Fat_Content']=encoder.fit_transform(big_mart_data['Item_Fat_Content'])

big_mart_data['Item_Type'] = encoder.fit_transform(big_mart_data['Item_Type'])

big_mart_data['Outlet_Identifier'] = encoder.fit_transform(big_mart_data['Outlet_Identifier'])

big_mart_data['Outlet_Size'] = encoder.fit_transform(big_mart_data['Outlet_Size'])

big_mart_data['Outlet_Location_Type'] =
encoder.fit_transform(big_mart_data['Outlet_Location_Type'])

big_mart_data['Outlet_Type'] = encoder.fit_transform(big_mart_data['Outlet_Type'])

X=big_mart_data.drop(columns=['Item_Identifier','Outlet_Identifier','Item_Outlet_Sales'],
axis=1)

Y = big_mart_data['Item_Outlet_Sales']



X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)

regressor = XGBRegressor(n_estimators= 100,random_state=2)
regressor.fit(X_train, Y_train)
 pickle.dump(regressor, open('model.pkl', 'wb'))
```

app.py:

```
import os
import pickle
import numpy as np
import pandas as pd
from flask import Flask, request, json, render_template, jsonify


app = Flask(__name__)

model = pickle.load(open('model.pkl', 'rb'))


@app.route("/")
def home():
    return render_template("big.html")
  @app.route('/predict', methods=['POST', 'GET'])
def predict():
    item_weight = float(request.form['item_weight'])
    item_fat_content = str(request.form['item_fat_content'])
    item_visibility = float(request.form['item_visibility'])
    item_type = str(request.form['item_type'])
    item_mrp = float(request.form['item_mrp'])
    outlet_establishment_year = int(request.form['outlet_establishment_year'])
    outlet_size = str(request.form['outlet_size'])
    outlet_location_type = str(request.form['outlet_location_type'])
    outlet_type = str(request.form['outlet_type'])

    X = np.array([[item_weight, item_fat_content, item_visibility, item_type, item_mrp,
                outlet_establishment_year, outlet_size, outlet_location_type, outlet_type]])

    my_prediction = model.predict(X)
    r=my_prediction[0]
      return render_template("result.html", **locals())
if __name__ == "__main__":
    app.run(debug=True,port=7895)
```

big.html:

```
 <!DOCTYPE html>
<html lang="en" dir="ltr">

<head>
   <title> Sriram BigMart Sale Project</title>
   <meta charset="utf-8">
   <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.1.3/css/bootstrap.min.css"
      integrity="sha384-
MCw98/SFnGE8fJT3GXwEOngsV7Zt27NXFoaoApmYm81iuXoPkFOJwJ8ERdknLPMO"
crossorigin="anonymous">
   <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='css/Style.css')}}">

   <style>
      body {
         background: #ffffff;
         background: linear-gradient(to left, #0b0000 0%, #ffffff 100%);

      }
      h1{
         text-align: center;
         font-size: 60px;
         color: #ef0303;
         font-style: italic;
         }

      footer a {
         color: #ef0303;
      }

      footer a:hover {
         color: #fff;
      }

      .centerdiv {
         height: 15vh;
         display: flex;
         justify-content: center;
         align-items: center;
      }
         centerdiv a {
         height: 30px;
```

```
            width: 30px;
            background-color: #f5f6fa;
            border-radius: 50px;
            text-align: center;
            margin: 5px;
            line-height: 30px;
            box-shadow: 1px 4px 2px 2px #dcdde1;
            position: relative;
            overflow: hidden;
        }

        .centerdiv a i {
            transition: all 0.3s linear;
        }

        .centerdiv a:hover i {
            transform: scale(1.5);
            color: #f5f6fa;
        }

        .centerdiv a:before {
            content: "";
            width: 120%;
            height: 120%;
            position: absolute;
            top: 90%;
            left: -50%;
            transform: rotate(60deg);
        }

        .centerdiv a:hover:before {
            animation: socialicons 0.8s 1;
            animation-fill-mode: forwards;
        }

        function myFunction() {
        alert("Hello! I am an alert box!");
       }

      .header {
  padding: 10px;
  text-align: center;
 background: rgb(54, 170, 48);
}
nav ul{
```

```
flex:1;
text-align: right;
padding-right: 30px;
}
nav ul li{

   display: inline-block;
   list-style: none;
   margin: 1px 30px;
   font-size: 20px;
   }
   nav ul li a
{
font-size: large;
   font-style: italic;
   color:red
}
     </style>
   <script>
     function myFunction() {
     alert("Reset values in form");
     }
     </script>

</head>

<body>
<div class="header">
    <nav>

          <ul>
             <li><a href="big.html"><b>Home</b></a></li>
             <li><a href="about.html"><b>About</b></a></li>
             <li><a href="contact.html"><b>Contact</b></a></li>
          </ul>
    </nav>

</div>
       <h1><b><u>Big Mart Sales Prediction</u></b></h1>
<div class="wrapper">
       <div class="container my-5">
          <div class="row">
             <div class="col-md-10 col-sm-6 mx-auto">
                <form class="" action="/predict" method="post">
                   <div class="form-group">
                      <input type="text" name="item_id" id="item_ID" class="form-control" required >
                      <label for="item_id" class="ph-area">Item ID</label> </div>
```

```
        <div class="form-group">
            <input type="number" name="item_weight" id="item_weight"
class="form-control" required min="4.00" max="21.00">
            <label for="item_weight" class="ph-area">Enter Item Weight</label>
        </div>

            <div class="form-group">
                <select class="form-control" name="item_fat_content" required="required">
                    <option value="" selected>Item Fat Content</option>
                    <option value="0">Low Fat</option>
                    <option value="1">Regular</option>

                </select>
            </div>
            <div class="form-group">
             <input type="float" name="item_visibility" id="item_visibility" class="form-
        control" min="0.000" max="0.300"
                    required>
                <label for="item_visibility" class="ph-area">Enter Item Visibility</label>
            </div>


            <div class="form-group">
                <select class="form-control chosenn" name="item_type" required="required">

                    <option value="" selected>Item Type</option>
                    <option value="0">Baking Goods</option>
                    <option value="1">Breads</option>
                    <option value="2">Breakfast</option>
                    <option value="3">Canned</option>
                    <option value="4">Dairy</option>
                    <option value="5">Frozen Foods</option>
                    <option value="6">Fruits and Vegetables</option>
                    <option value="7">Hard Drinks</option>
                    <option value="8">Health and Hygiene</option>
                    <option value="9">Household</option>
                    <option value="10">Meat</option>
                    <option value="11">Others</option>
                    <option value="12">Seafood</option>
                    <option value="13">Snack Foods</option>
                    <option value="14">Soft Drinks</option>
                    <option value="15">Starchy Foods</option>

                </select>
            </div>
```

```
            <div class="form-group">
                                <input type="number" name="item_mrp" id="item_mrp"
        min="31.00" max="267.00" class="form-control" required >
                <label for="item_mrp" class="ph-area">Enter Item MRP</label>
            </div>
            <div class="form-group">
              <select class="form-control chosenn" name="item_type" required="required">
                <option value="" selected>Outlet_Identifier</option>
                <option>OUT010</option>
                <option>OUT013</option>
                <option>OUT017</option>
                <option>OUT018</option>
                <option>OUT019</option>
                <option>OUT027</option>
                <option>OUT035</option>
                <option>OUT045</option>
                <option>OUT046</option>
                <option>OUT049</option>
                <option>Other</option>
              </select>

            </div>

            <div class="form-group">
                    <input type="number" name="outlet_establishment_year"
id="outlet_establishment_year"
                class="form-control" required min="1985" max="2009">
                    <label for="outlet_establishment_year" class="ph-area">Outlet
Establishment Year
                (YYYY)</label>
        </div>

            <div class="form-group">
              <select class="form-control chosenn" name="outlet_size" required="required">
                <option value="" selected>outlet_size</option>
                <option value="0">High</option>
                <option value="1">Medium</option>
                <option value="2">Small</option>
              </select>
            </div>




            <div class="form-group">
```

```
                    <select class="form-control chosenn" name="outlet_location_type"
                required="required">
                        <option value="" selected>outlet_location_type</option>
                        <option value="0">Tier 1</option>
                        <option value="1">Tier 2</option>
                        <option value="2">Tier 3</option>
                    </select>
                </div>

                <div class="form-group">
                    <select class="form-control chosenn" name="outlet_type" required="required">
                        <option value="" selected>outlet_type</option>
                        <option value="0">Grocery Store</option>
                        <option value="1">Supermarket Type1</option>
                        <option value="2">Supermarket Type2</option>
                        <option value="3">Supermarket Type3</option>
                    </select>
                </div>
                <div class="form-group">
                    <input type="submit" class="btn btn-primary" value="Submit">
                   <input type="reset" class="btn btn-danger"  onclick="myFunction()"
                value="Reset">
                </div>
            </form>
        </div>
    </div>
</div>
<script src="https://code.jquery.com/jquery-3.3.1.slim.min.js"
    integrity="sha384-
q8i/X+965DzO0rT7abK41JStQIAqVgRVzpbzo5smXKp4YfRvH+8abtTE1Pi6jizo"
    crossorigin="anonymous"></script>
<script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.14.3/umd/popper.min.js"
    integrity="sha384-
ZMP7rVo3mIykV+2+9J3UJ46jBk0WLaUAdn689aCwoqbBJiSnjAK/l8WvCWPIPm49"
    crossorigin="anonymous"></script>
<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.1.3/js/bootstrap.min.js"
    integrity="sha384-
ChfqqxuZUCnJSK3+MXmPNIyE6ZbWh2IMqE241rYiqJxyMiZ6OW/JmZQ5stwEULTy"
    crossorigin="anonymous"></script>
</body></html>
```

Result.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>BIg Mart Sales</title>
   <style>
   body{
   background: black;
   }

         h1{

         text-align: center;
         font-size: 50px;
         color: green;
         margin-top: 5%;

       }
       a{
        font-size: 30px;
        }
        legend{
        color:white;
        font-size: 50px;
        }
</style>

</head>
<body>
<fieldset>
   <legend> Result page</legend>

   <h1>Prediction value is:{{r}}</h1>
   <h1><a href="/">Back and Again Predict!!!</a></h1>
   </fieldset>
</body>
</html>
```

About.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>about page</title>
```

```
   <style>
    fieldset{
       width:80%;
       height:50%;

       margin-top: 80px;
       }
       body{
       background: #ebf8e1;
       }
       legend{
       font-size:40px;
       color:red
       }
       pre{
       color:green;
       font-size:20px
       }

   </style>
</head>
<body>
<fieldset>
   <legend> About Us</legend>
   <pre>Everyday competitiveness between various shopping centres as and as huge marts is becoming
       higher intense,violent just because of the quick development of global malls also online shopping.
       Each market seeks to offer personalized and limited-time deals to attract many clients relying on
       period of time, so that each item's volume of sales may be estimated for the organization's stock
control,
       transportation and logistical services.

       The current machine learning algorithm is very advanced and provides methods for predicting or f

forecasting
       sales any kind of organization, extremely beneficial to overcome low – priced used for prediction.
Always
       better prediction is helpful, both in developing and improving marketing strategies for the
marketplace,
       which is also particularly helpful.

</pre>
   </fieldset>

</body>
</html>
```

Contact.html:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>Hello Worl!</title>
   <style>
      h1{
      color:orange;

       text-align:center
      }
      fieldset{
      width:80%;
      height:50%;
       margin-left: 70px;

      margin-top: 80px;
      }
      body{
      background: #a7ea73;;
      }
      legend{
      font-size:40px;
      color:red}
   </style>
</head>
<body>
<fieldset>
   <legend> Contact Us</legend>

   <h1> sriramburri5@gmail.com</h1>
   <h1>chandraSekharchanna81@gmail.com</h1>
   <h1>jashuva123@gmail.com</h1>
</fieldset>
</body>
</html>
```

5. RESULT ANALYSIS

Algorithms	Accuracy
Linear Regression	**50.57**
Polynomial Regression	**57.79**
Rigid Regression	**50.57**
XgBoost	**86.12**

Fig:5.1 Accuracy Table

The above table shows the accuraces of different models which are created by using the mentioned machine learning algorithms. Among all above models ,the model which is created by using XgBoost algorithm got good accurance .So we consider it as the final model

6.OUTPUT SCREENS

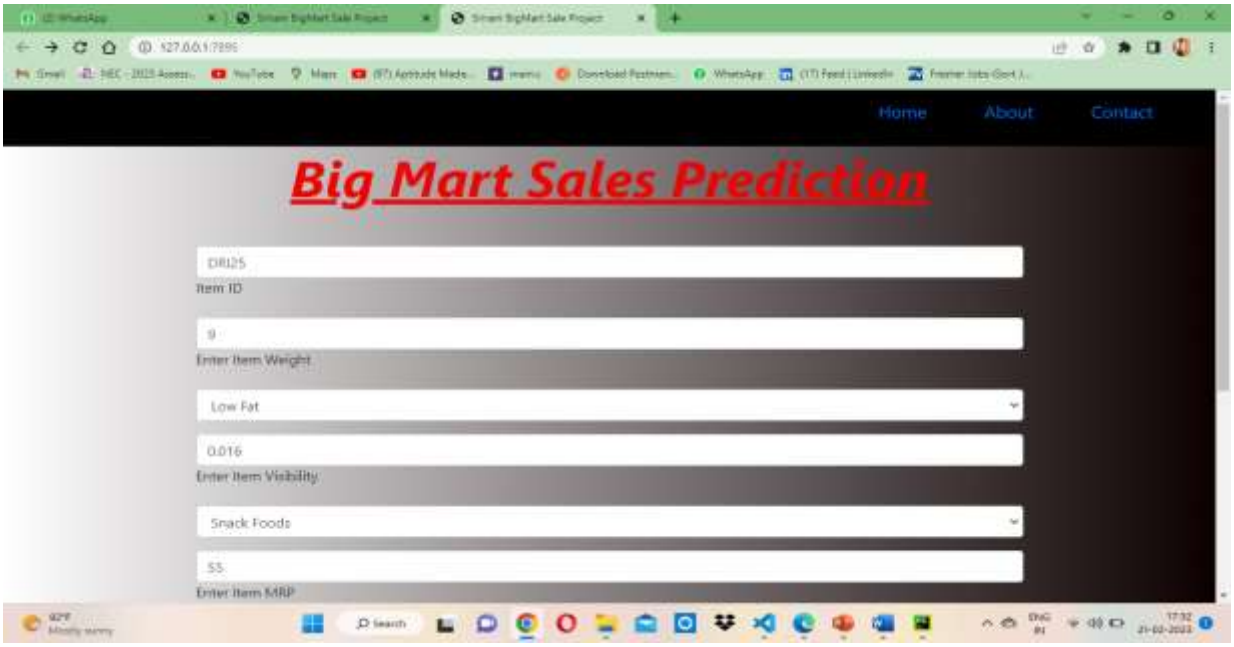


Fig :6.1 Home

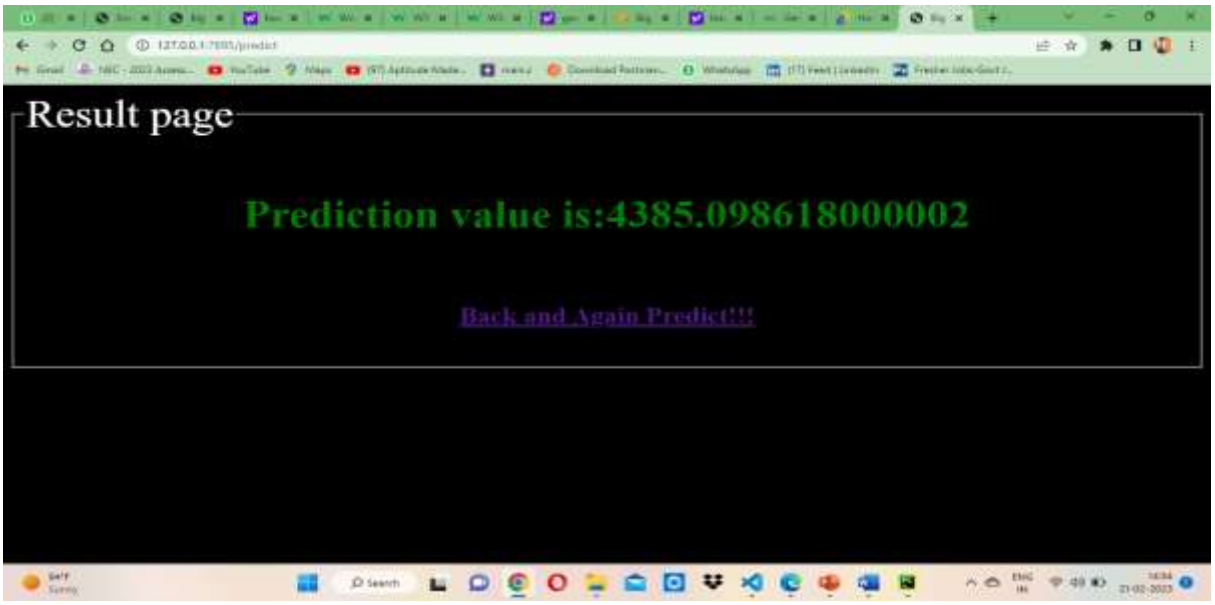


Fig: 6.2 Result

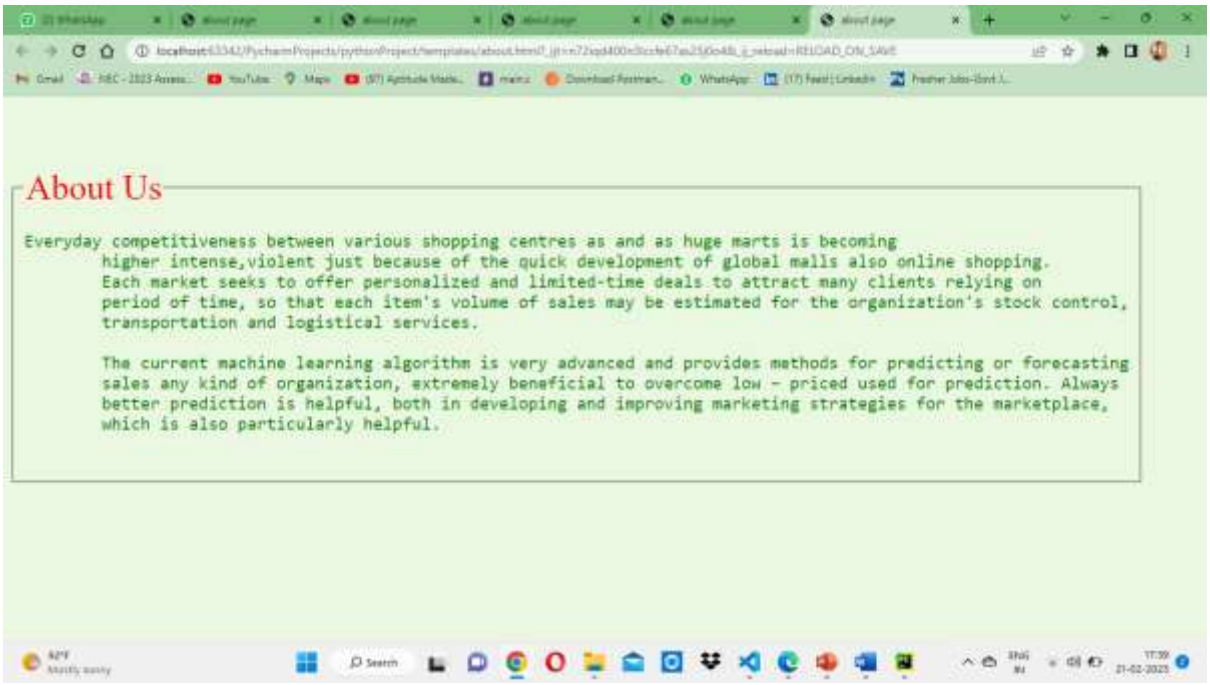


Fig: 6.3 About Us

Fig:6.4 Contact Us

7.TEST CASES

Test case1: please enter Id feild

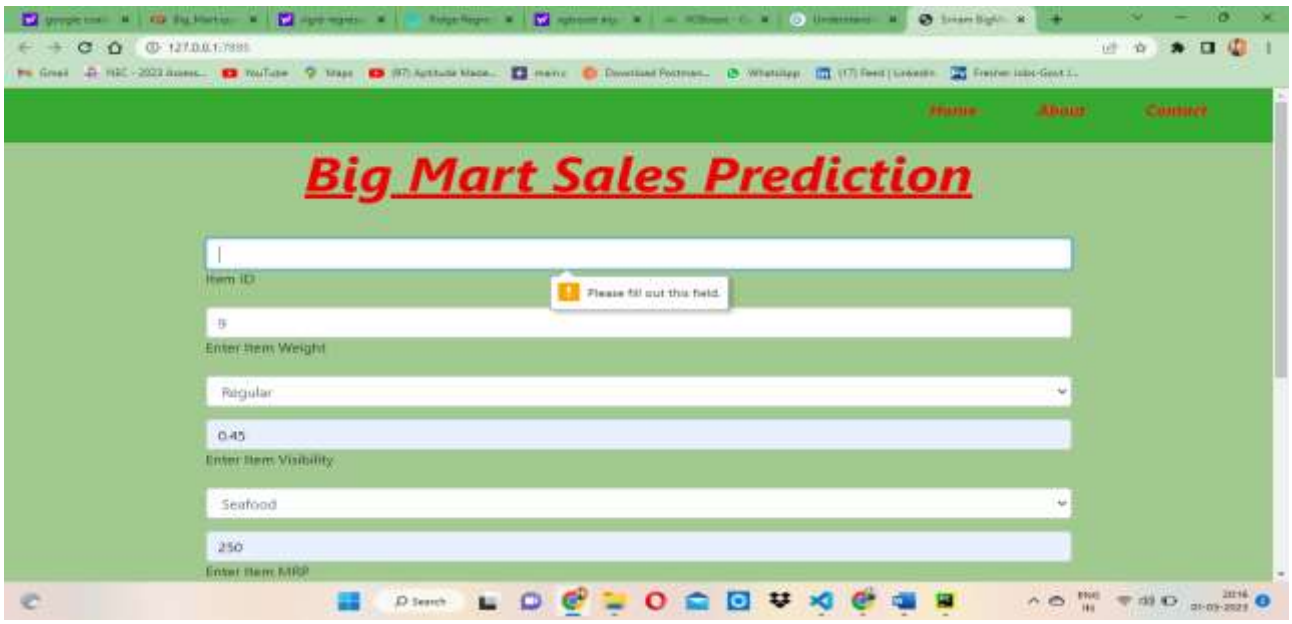


Fig:7.1 Testcase1

Test Case 2: Establishment Year Value must be greater than or equal to 1985

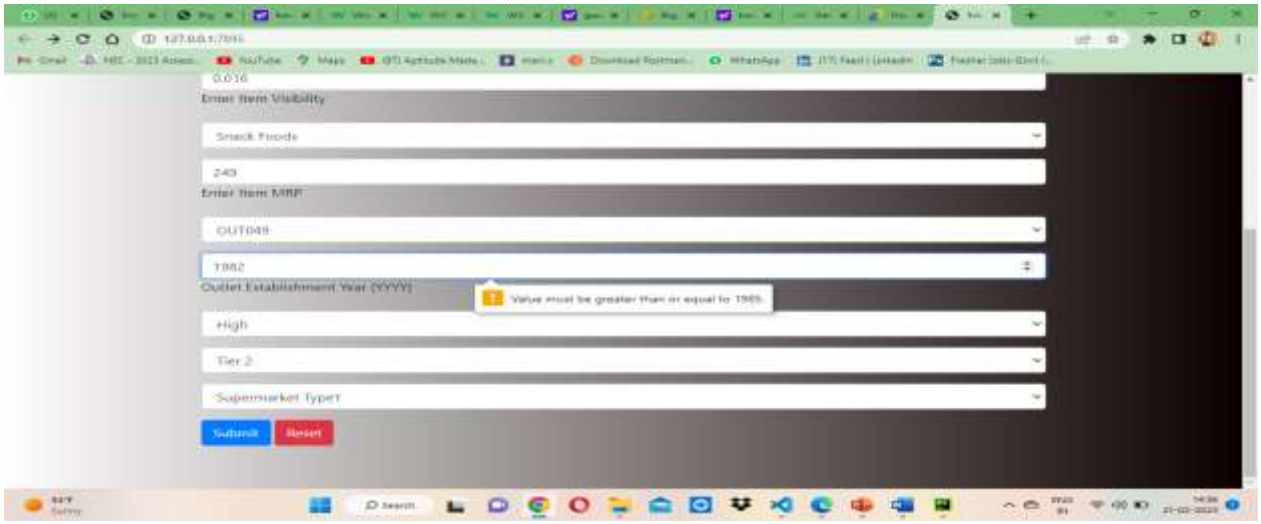


Fig:7.2 Testcase2

Test Case 3: Reset values in form successfully

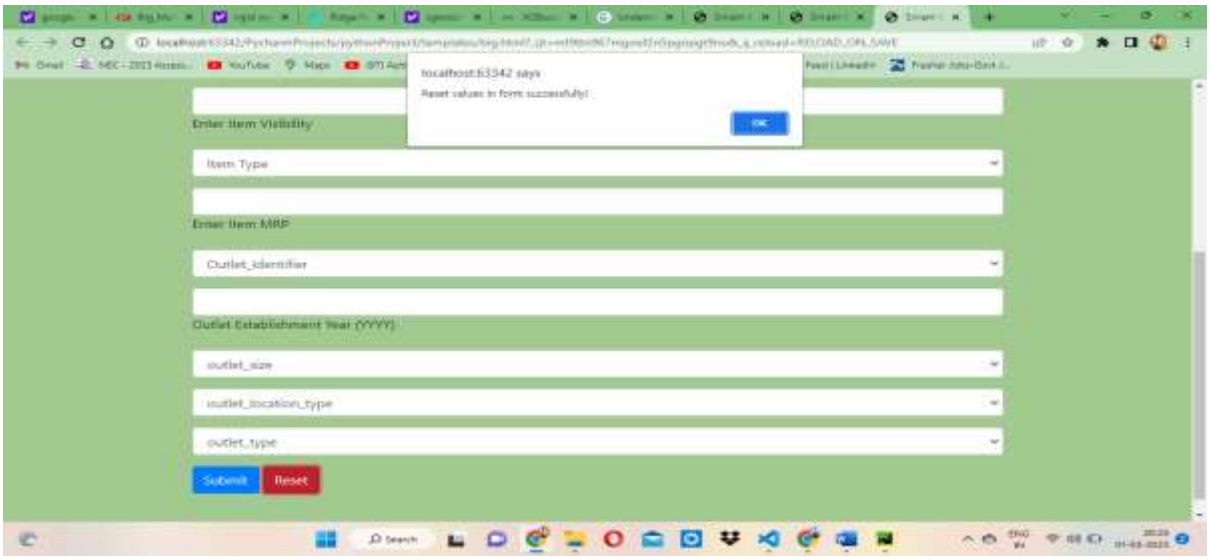


Fig:7.3 Testcase 3

Test Case 4: MRP values must be greater than or equals to 31

Fig: 7.4 Testcase 4

8.CONCLUSION

In this work, the effectiveness of various algorithms on the data on revenue and review of, best performance-algorithm, here propose a software to using regression approach for predicting the sales centered on sales data from the past the accuracy of linear regression prediction can be enhanced with this method, polynomial regression, Ridge regression, and Xgboost regression can be determined. So, we can conclude ridge and Xgboost regression gives the better prediction with respect to Accuracy, MAE and RMSE than the Linear and polynomial regression approaches.

FUTURE SCOPE

In future, the forecasting sales and building a sales plan can help to avoid unforeseen cash flow and manage production, staff and financing needs more effectively.In future work we can also consider with the ARIMA model which shows the time series graph.

9.BIBLIOGRAPHY

[1] Ching Wu Chu and Guoqiang Peter Zhang, “A comparative study of linear and nonlinear models for aggregate retails sales forecasting”, Int. Journal Production Economics, vol. 86, pp. 217- 231, 2003.

[2] Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.- 2. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of D

[3] Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101- 110

[4] Giuseppe Nunnari, Valeria Nunnari, “Forecasting Monthly Sales Retail Time Series: A Case Study”, Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

[5]https://halobi.com/blog/sales-forecasting-five-uses/. [Accessed: Oct. 3, 2018]

[6] Zone-Ching Lin, Wen-Jang Wu, “Multiple LinearRegression Analysis of the Overlay Accuracy Model Zone”, IEEE Trans. on Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[7] O. Ajao Isaac, A. bdullahi Adedeji, I. Raji Ismail, “Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis”, Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.

[8] C. Saunders, A. Gammerman and V. Vovk, “Ridge Regression Learning Algorithm in Dual Variables”, Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY 2010 3561.