“BigMart Sale Prediction System”

A Minor Project Report Submitted to
Rajiv Gandhi Proudyogiki Vishwavidyalaya

Towards Partial Fulfillment for the Award
of
Bachelor of Engineering in Computer Science & Information Technology

Submitted by:
Jessica Chouhan (0827CI201087)
Palak Jaiswal (0827CI201126)

Guided by:
Prof. Rachana Bahrawat
Computer Science & Information Technology AITR, Indore

Acropolis Institute of Technology & Research, Indore
July - Dec 2022

EXAMINER APPROVAL

The Project entitled ***"BigMart Sales Prediction System"***

submitted by **Jessica Chouhan (0827CI201087), Palak Jaiswal(0827CI201126)** has been examined and is hereby approved towards partial fulfillment for the award of ***Bachelor of Engineering degree in Computer Science & Information Technology*** discipline, for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the project only for the purpose for which it has been submitted.

(Internal Examiner) **(External Examiner)**

Date: **Date:**

Recommendation

This is to certify that the work embodied in this project entitled “BigMart Sales Prediction System” submitted by **Jessica Chouhan (0827CI201087), Palak Jaiswal(0827CI201126)** is a satisfactory account of the bonafide work done under the supervision of ***Prof. Rachana Bahrawat***, is recommended towards partial fulfillment for the award of the Bachelor of Engineering (Computer Science & Information Technology) degree by Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal.

Prof. Rachana Bahrawat
(Project Guide)

Prof. Nidhi Nigam
(Project Coordinator)

Student Undertaking

This is to certify that project entitled ***"BigMart Sales Prediction System"*** has developed by us under the supervision of ***Prof. Rachana Bahrawat***. The whole responsibility of work done in this project is ours. The sole intension of this work is only for practical learning and research.

We further declare that to the best of our knowledge, this report does not contain any part of any work which has been submitted for the award of any degree either in this University or in any other University / Deemed University without proper citation and if the same work found then we are liable for explanation to this.

Jessica Chouhan (0827CI201087),
Palak Jaiswal (0827CI201126)

Acknowledgement

We thank the almighty Lord for giving me the strength and courage to sail out through the tough and reach on shore safely.

There are number of people without whom this projects work would not have been feasible. Their high academic standards and personal integrity provided me with continuous guidance and support.

We owe a debt of sincere gratitude, deep sense of reverence and respect to our project coordinators **Prof. Nidhi Nigam** and our guide and mentor **Prof. Rachana Bahrawat,** Professor, AITR, Indore for their motivation, sagacious guidance, constant encouragement, vigilant supervision and valuable critical appreciation throughout this project work, which helped us to successfully complete the project on time.

We express profound gratitude and heartfelt thanks to **Dr Shilpa Bhalerao**, HOD IT AITR Indore for her support, suggestion and inspiration for carrying out this project. I am very much thankful to other faculty and staff members of CSIT Dept, AITR Indore for providing me all support, help and advice during the project. We would be failing in our duty if do not acknowledge the support and guidance received from **Dr S C Sharma**, Director, AITR, Indore whenever needed. We take opportunity to convey my regards to the management of Acropolis Institute, Indore for extending academic and administrative support and providing me all necessary facilities for project to achieve our objectives.

We are grateful to **our parent** and **family members** who have always loved and supported us unconditionally. To all of them, we want to say “Thank you”, for being the best family that one could ever have and without whom none of this would have been possible.

Jessica Chouhan (0827CI201087),
Palak Jaiswal (0827CI201126)

Abstract

TITLE OF THE PROJECT

This project is submitted to Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal(MP), India for partial fulfillment of Bachelor of Engineering in Computer Science & Engineering branch under the sagacious guidance and vigilant supervision of ***Prof. Rachana Bahrawat***.

The project is based on predicting the future sales and demand of products in any Bigmart. In the modern era of reaching new lengths of advancement, every company and enterprise are working on their customer demands as well as their inventory management. The models used by them help them predict future demands by understanding the pattern from old sales records. Lately, everyone is abandoning the traditional prediction models for sales forecasting as it takes a prolonged amount of time to get the expected results. Therefore, now the retailers keep track of their sales record in the form of a data set, which comprises price tag, outlet types, outlet location, item visibility, item outlet sales etc.

Key words : Analysis, Big Mart, Data Science, Machine Learning, Prediction, Regression.

List of Figures

List of Tables

List of Abbreviations

SAMPLE EXAMPLE GIVEN

Abbr1: R- CNN- Regional based Covolutional Neural Networks

Abbr2: COCO – Common Objects in context

Abbr3: OpenCV- Open Source Computer Vision

Abbr4: JSON- Java Script Object Notation

Abbr5: CIF- Count In Frame

Abbr6: GPU- Graphical Processing Unit

Abbr7: YOLO- You Only Look Once

Table of Contents

INTRODUCTION

Big Mart is a big supermarket chain, with stores all around the country and its current board set out a challenge to all Data Scientists out there to help them create a model that can predict the sales, per product, for each store to give accurate results. Big Mart has collected sales data from the year 2013, for 1559 products across 10 stores in different cities.

With this information, the corporation hopes we can identify the products and stores which play a key role in their sales and use that information to take the correct measures to ensure the success of their business. Big Mart is a massive network of stores that spans the globe. Big Mart's trends are extremely important, as data scientists analyse them by product and location to identify potential centers. Using a computer to predict Big Mart sales allows data scientists to explore different patterns by shop and product to get the best results. Many businesses rely largely on their information base and require market forecasting. Forecasting involves evaluating data from a wide variety of sources, including consumer trends, buying behaviours, and other considerations. This research would also assist businesses in properly managing their financial means. And that is where machine learning can really be put to good use. In this paper, we employ data mining approaches including discovery, data transformation, feature development, model construction, and testing to forecast sales using various machine learning algorithms. This approach involves pre-processing raw data acquired by a large mart for missing data, abnormalities, and outliers. After that, an algorithm will be trained to create a model depending on the data.
Everyday competitiveness between various shopping centers 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 a period of time, so that each item's volume of sales may be estimated for the organization's stock control, transportation and logistical services.

AIM:

“To find out what role certain properties of an item play and how they affect their sales by understanding Big Mart sales.” In order to help Big Mart achieve this goal, a predictive model can be built to find out for every store, the key factors that can increase their sales and what changes could be made to the product or store’s characteristics.

SPECIFIC OBJECTIVE:

- The objective of this framework is to predict the future sales from given data of the previous year's using Machine Learning Techniques.
- Another objective is to conclude the best model which is more efficient and gives fast and accurate result.
- To find out key factors that can increase their sales and what changes could be made to the product or store's characteristics.

SCOPE:

A sales Prediction is an important component of any business plan. For the sales rep, as well as the entire organization, a sales prediction aims to predict future sales and is used as the basis of planning time and resources. A good forecast should have several objectives, all directed at identifying what you will sell, when you will sell it and to whom.

TECHNOLOGY:

- Python
- Font end – HTML, CSS
- Back end – Flask
- Deployment – Pycharm

Review of Literature

Preliminary Investigations / Current System Available

- Title: - A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression (2018)

1. Author: - Kadam, H., Shevade, R., Ketkar, P. and Rajguru
2. Description: - A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression used Random
3. Forest and Linear Regression for prediction analysis which gives less accuracy. To overcome this we can use XG boost Algorithm which will give more accuracy and will be more efficient.

- Title: - Forecasting methods and applications (2008)

1. Author: - Makridakis, S., Wheelwrigh.S.C., Hyndman. R.J
2. Description: - Forecasting methods and applications contains Lack of Data and short life cycles. So some of the data like historical data, consumer-oriented markets face uncertain demands, can be prediction for accurate result.

- Title: -Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data (2018)

1. Author: - C. M. Wu, P. Patil and S. Gunaseelan
2. Description: - Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data Used Neural Network for comparison of different algorithms. To overcome this Complex models like neural networks is used for comparison between different algorithms which is not efficient so we can use more simpler algorithm for prediction.

- Title: -Prediction of retail sales of footwear using feed forward and recurrent Neural Networks (2018)

1. Author: - Das, P., Chaudhury
2. Description: - Prediction of retail sales of footwear using feed forward and recurrent neural networks used neural networks for prediction of sales. Using neural network for predicting of weekly retail sales, which is not efficient, So XG boost can work efficiently.

Limitations of Current System

Some are great at managing sales leads. Others are focused more on managing relations with existing customers. Some are focused on marketing automation, i.e., automating your follow-on marketing. Others are simple and mostly help you keep track of contacts. Some do better at helping you maximize your social media networking.
Make sure you devote enough time up front to evaluating CRM providers, with a wellorganized and focused approach: ○ Focus on YOUR needs. Don't get dazzled by shiny objects. Write down your two or three primary pain points that you want to fix. ○ Organize your evaluation. Set up a spreadsheet and list each CRM you evaluate, noting your impression of how well it meets your needs. Use live chat and take advantage of demos and free trials -- most providers offer some combination of these. ○ Evaluate price. While isn't everything, in a small business it's very important. In fact, a pricey CRM system often delivers a double whammy, because it may be loaded with bells and whistles that can make it overly complex to implement in a small company

Existing system

This was on k-means Algorithm implementation, Only the two features with the most variance were used to train the model. The model was set to have 2 clusters, 0 being non-fraud and 1 being fraud. We also experimented with different values for the hyper parameters, but they all produced similar results. Changing the dimensionality of the data (reducing it to more dimensions than 2) also made little difference on the final values.

Disadvantages:

The Clustering doesn't produce the less accuracy when compared to Regression methods in scenarios like credit card fraud detection. Comparatively with other algorithms k-means produce less accurate scores in prediction in this kind of scenarios

Proposed System:

Our goal is to implement machine learning model in order to classify, to the highest possible degree of accuracy, credit card fraud from a dataset gathered from Kaggle. After initial data exploration, we knew we would implement a logistic regression model for best accuracy reports.
Logistic regression, as it was a good candidate for binary classification. Python sklearn library was used to implement the project, We used Kaggle datasets for Credit card fraud detection, using pandas to data frame for class ==0 forno fraud and class==1 for fraud, matplotlib for plotting the fraud and non fraud data, train_test_split for data extraction (Split arrays or matrices into random train and test subsets) and used Logistic Regression machine learning algorithm for fraud detection and print predicting score according to the algorithm. Finally Confusion matrix was plotted on true and predicted.

Advantages:

- The results obtained by the linear regression, Random forest algorithm is best compared to any other Algorithms.
- The Accuracy obtained was almost equal to cent percent which proves using of Logistic algorithm gives best results.
- The plots that were plotted according to the proper data that is processed during the implementation
- Requirement Identification overcoming existing gaps
- Create optimized schedules balancing production efficiency and delivery performance
- Maximize output on bottleneck resources to increase revenue
- Synchronize supply with demand to reduce inventories
- Provide company-wide visibility to capacity
- Enable scenario data-driven decision making

System Analysis

The **Systems Development Life Cycle (SDLC)**, or *Software Development Life Cycle* in systems engineering, information systems and software engineering, is the process of creating or altering systems, and the models and methodologies that people use to develop these systems.
In software engineering the SDLC concept underpins many kinds of software development methodologies. These methodologies form the framework for planning and controlling the creation of an information system the software development process.

Project Overview

The aim is to build a predictive model and predict the sales of each product at a particular outlet.

Using this model, BigMart will try to understand the properties of products and outlets which play a key role in increasing sales.

Hardware & Software Requirement

HARDWARE REQUIREMENTS:

- System : Pentium i3 Processor.
- Hard Disk : 500 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 4 GB

SOFTWARE REQUIREMENTS:

- Operating system : Windows 10.
- Coding Language : Python 3.8
- Web Framework : Flask
- Deployment – Pycharm

Functional &Non-functional Requirements

Functional Requirements

- Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provide a permanent copy of the results for later consultation. The various types of outputs in general are:
- External Outputs, whose destination is outside the organization,.
- Internal Outputs whose destination is within organization and they are the
- User's main interface with the computer.
- Operational outputs whose use is purely within the computer department.
- Interface outputs, which involve the user in communicating directly.
- Understanding user's preferences, expertise level and his business requirements through a friendly questionnaire.
- Input data can be in four different forms - Relational DB, text files, .xls and xml files. For testing and demo you can choose data from any domain. UserB can provide business data as input.

Non-Functional Requirements

- Secure access of confidential data (user's details). SSL can be used.
- 24 X 7 availability.

- Better component design to get better performance at peak time
- Flexible service based architecture will be highly desirable for future extension

Data Flow Diagram

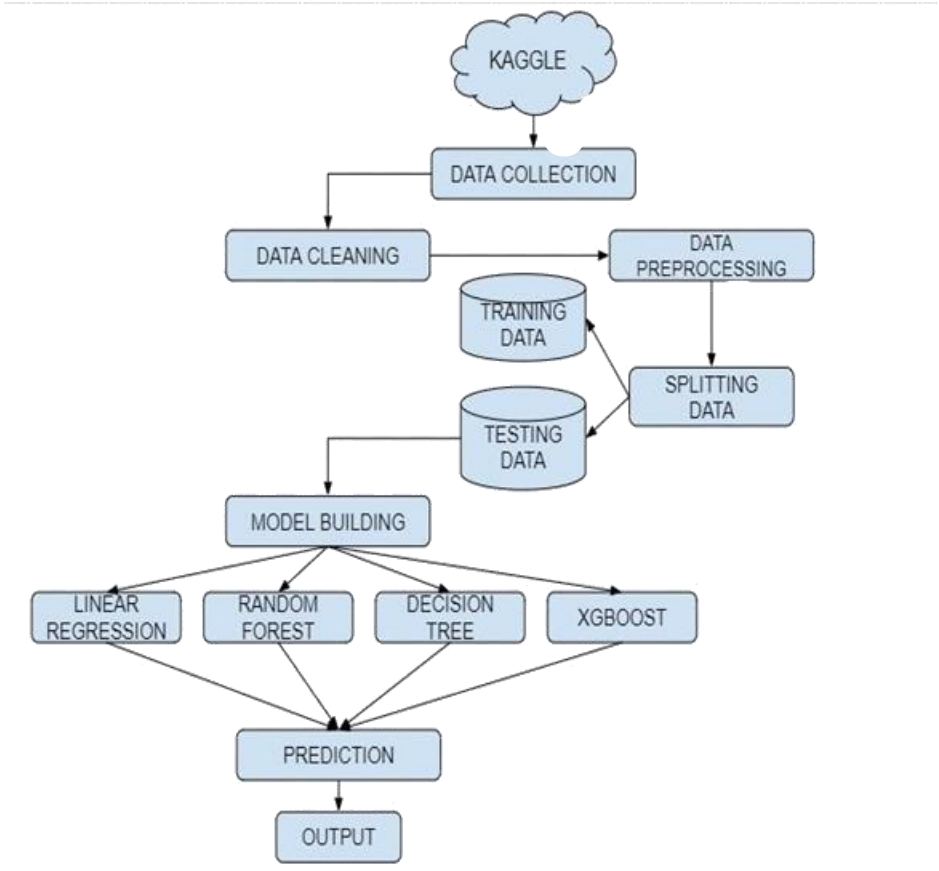
DFDs are the model of the proposed system. They clearly should show the requirements on which the new system should be built. Later during design activity this is taken as the basis for drawing the system's structure charts. The Basic Notation used to create a DFD's are as follows:

Dataflow: Data move in a specific direction from an origin to a destination.

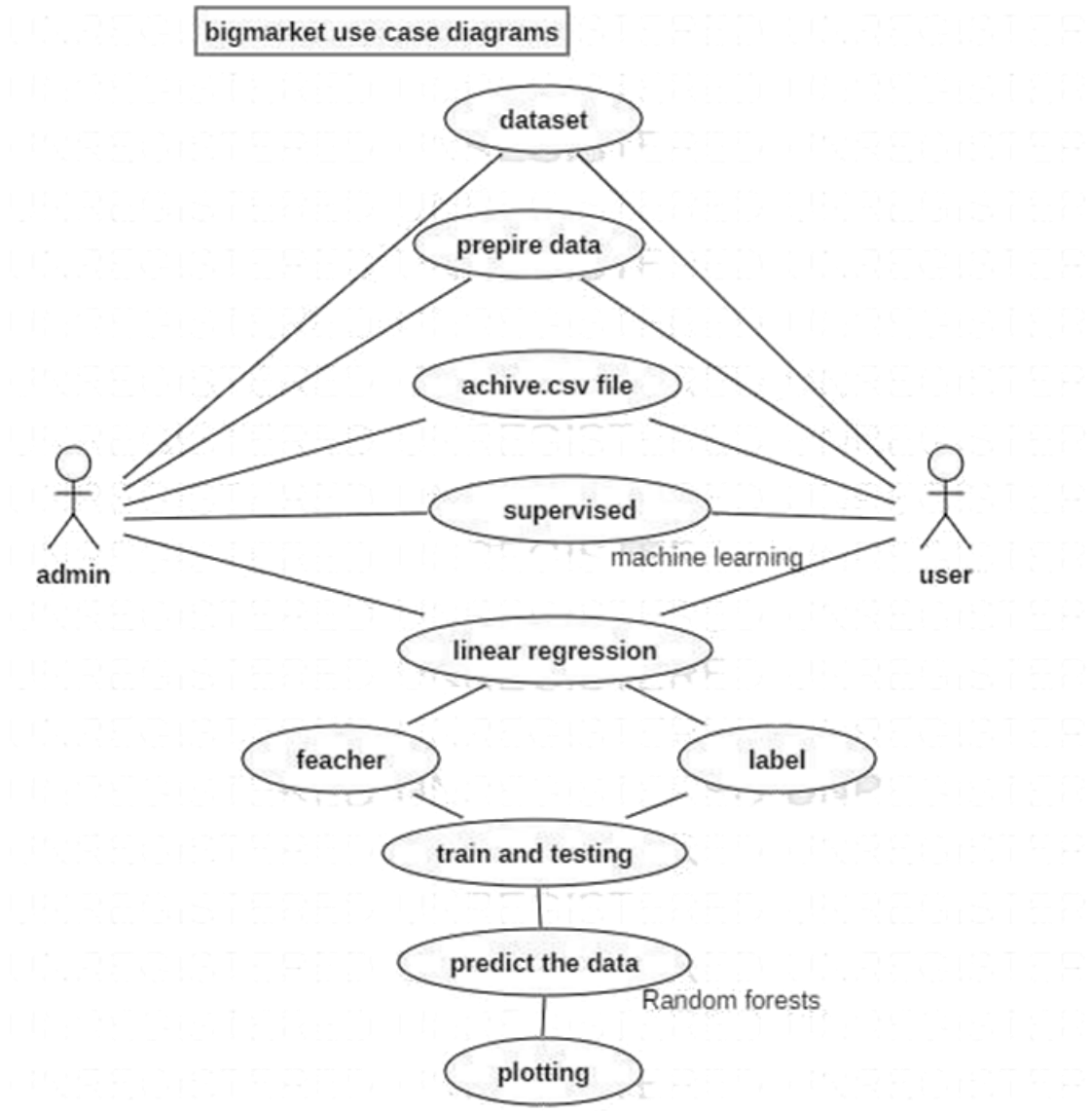
Process: People, procedures, or devices that use or produce (Transform) Data.
The physical component is not identified.

Source: External sources or destination of data, which may be People, programs, organizations or other entities.

Data Store: Here data are stored or referenced by a process in the System.



Use-case Diagram



Implementation

Code is produced from the deliverables of the design phase during implementation, and this is the longest phase of the software development life cycle. For a developer, this is the main focus of the life cycle because this is where the code is produced. Implementation my overlap with both the design and testing phases. Many tools exists (CASE tools) to actually automate the production of code using information gathered and produced during the design phase.

jupyter notebook:

The **Jupyter** Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning.

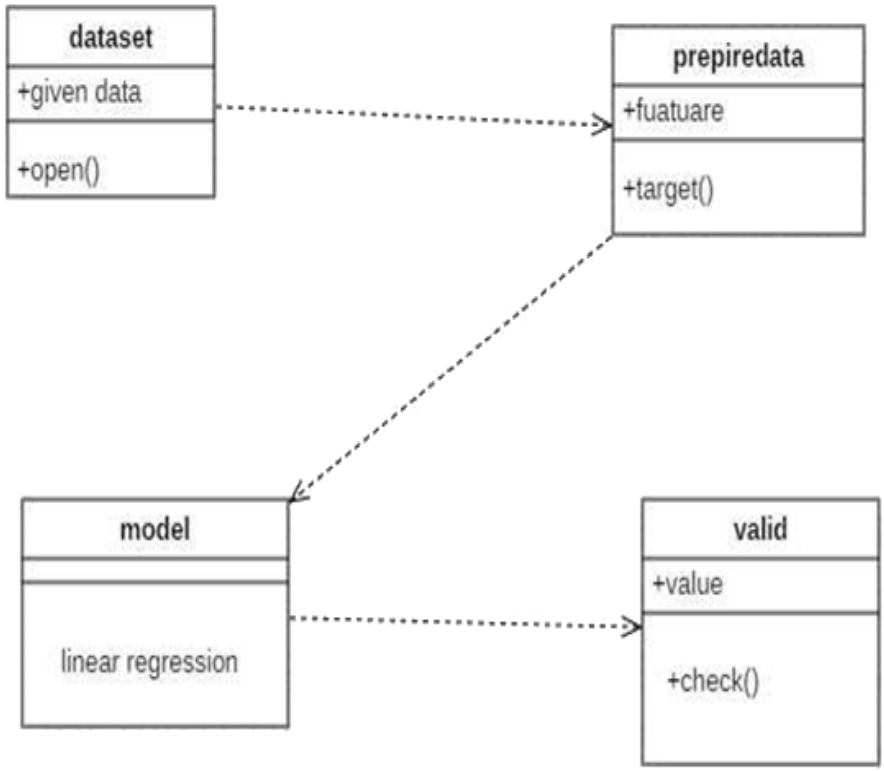
System Design

Software design sits at the technical kernel of the software engineering process and is applied regardless of the development paradigm and area of application. Design is the first step in the development phase for any engineered product or system. The designer's goal is to produce a model or representation of an entity that will later be built. Beginning, once system requirement have been specified and analyzed, system design is the first of the three technical activities design, code and test that is required to build and verify software.

The importance can be stated with a single word "Quality". Design is the place where quality is fostered in software development. Design provides us with representations of software that can assess for quality.

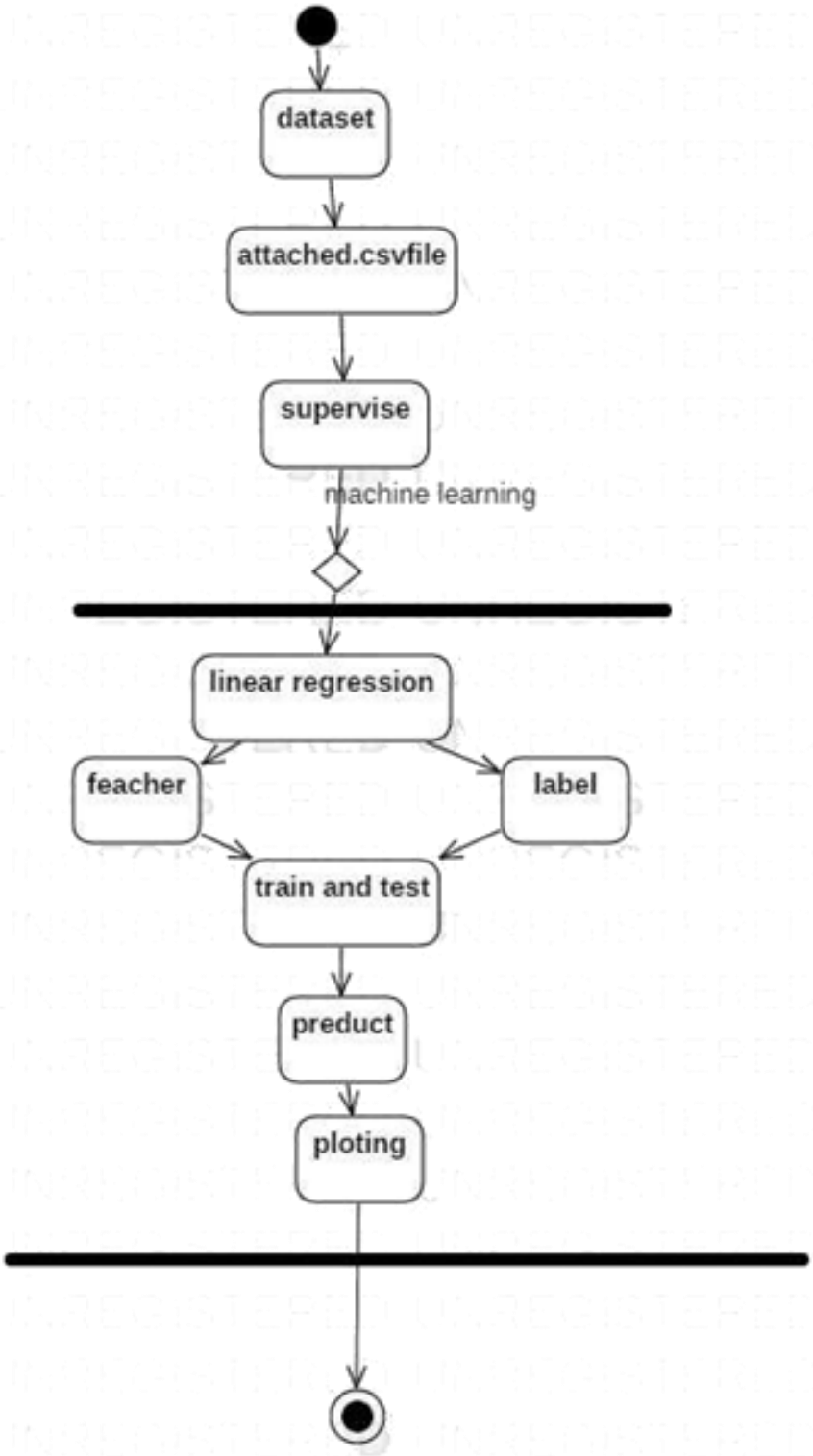
Class Diagram

Class diagrams are the backbone of almost every object-oriented method including UML. They describe the static structure of a system.



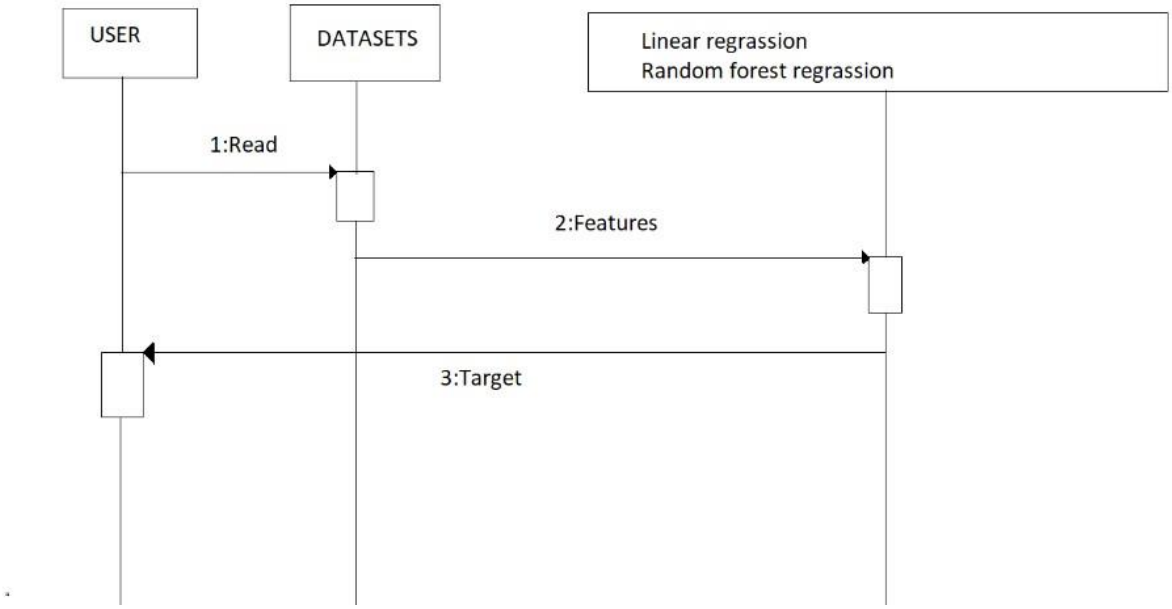
Activity Diagram

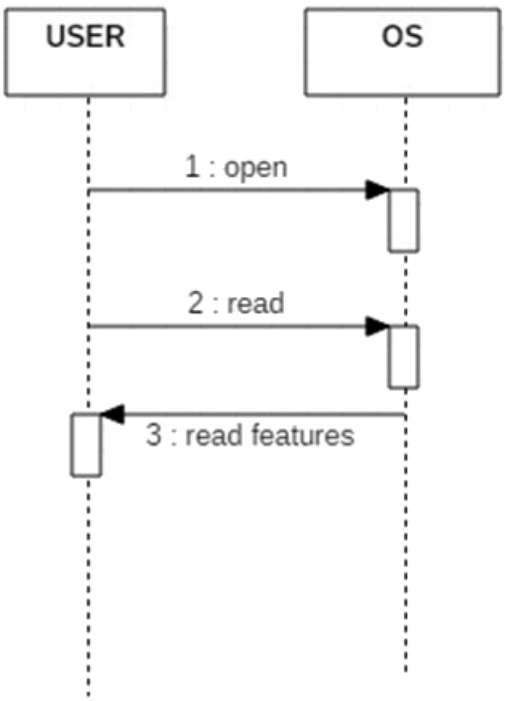
Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



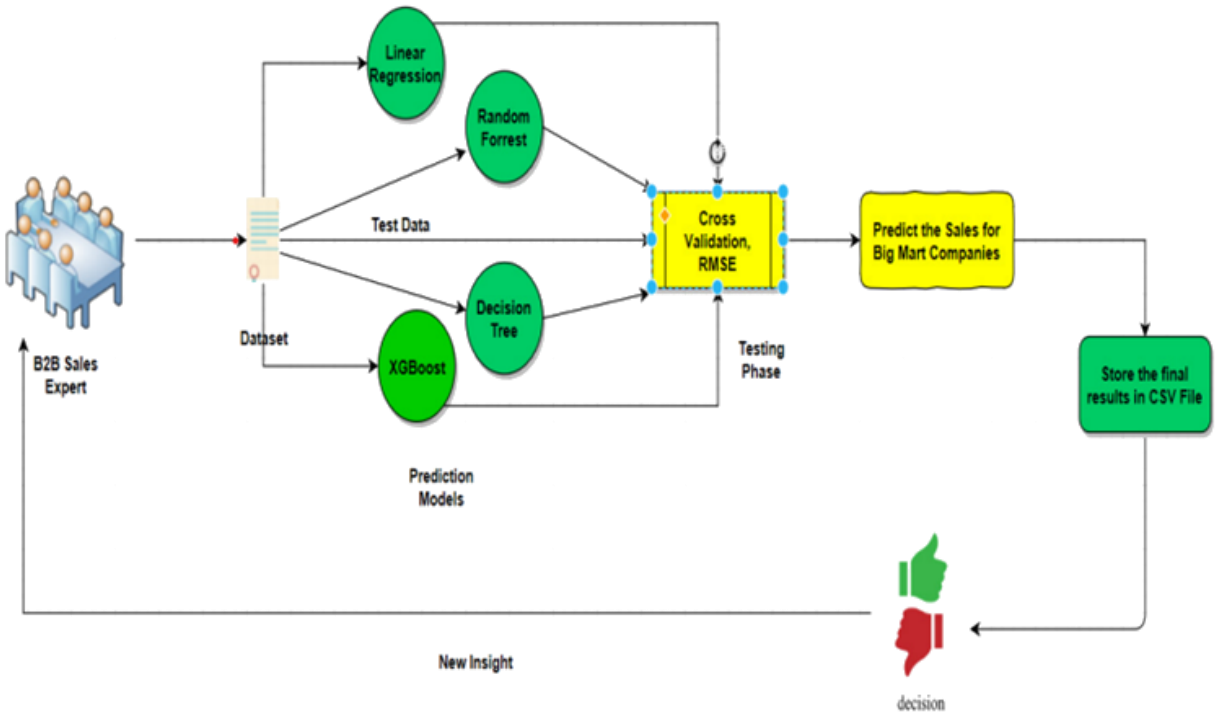
Sequence Diagram

A **sequence diagram** in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows, as parallel vertical lines ("lifelines"), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.

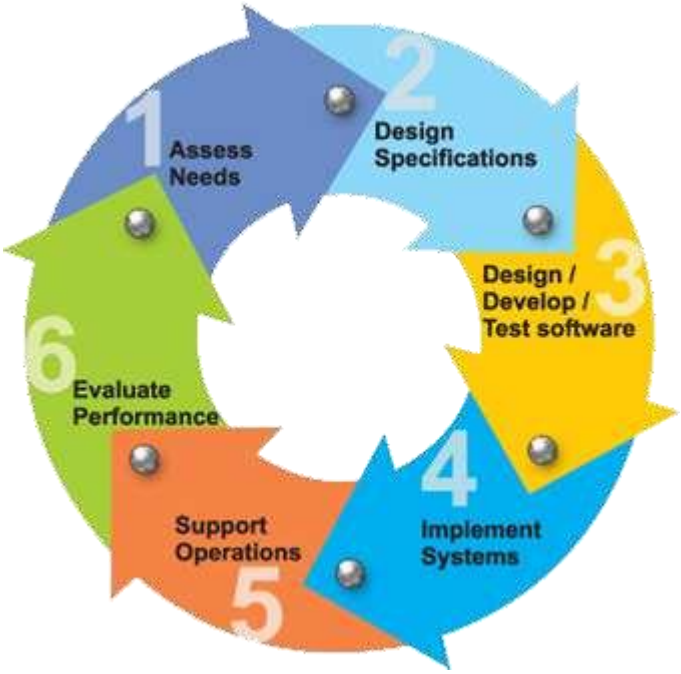




E-R Diagram


Linear Regression
Random Forrest
Test Data
Decision Tree
XGBoost
Dataset
B2B Sales Expert
Prediction Models
Cross Validation, RMSE
Testing Phase
Predict the Sales for Big Mart Companies
Store the final results in CSV File
New Insight
decision

Software Development

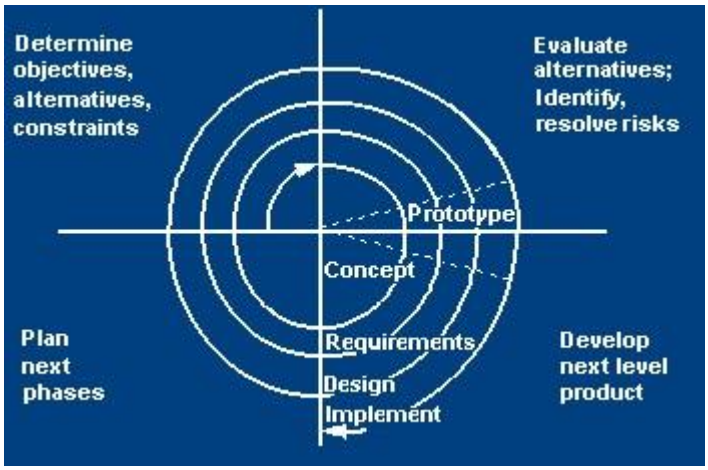


Software Model Used with Justification

SPIRAL MODEL was defined by Barry Boehm in his 1988 article, "A spiral Model of Software Development and Enhancement. This model was not the first model to discuss iterative development, but it was the first model to explain why the iteration models.
As originally envisioned, the iterations were typically 6 months to 2 years long. Each phase starts with a design goal and ends with a client reviewing the progress thus far. Analysis and

engineering efforts are applied at each phase of the project, with an eye toward the end goal of the project.

The following diagram shows how a spiral model acts like:

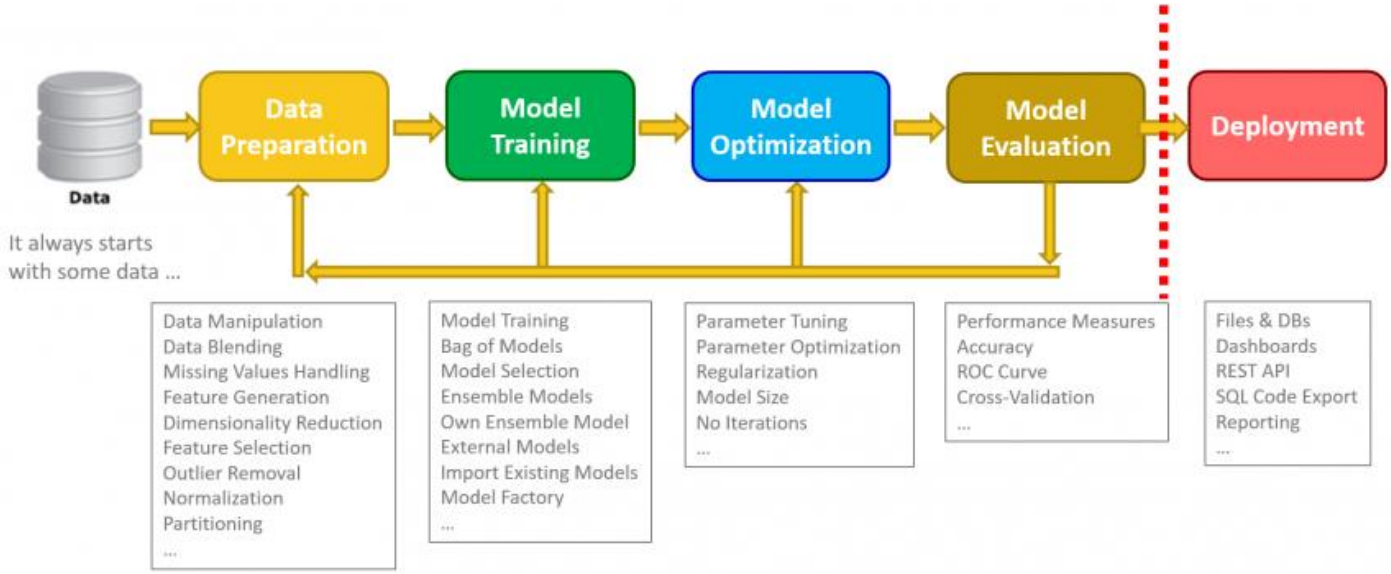


The steps for Spiral Model can be generalized as follows:

- The new system requirements are defined in as much details as possible. This usually involves interviewing a number of users representing all the external or internal users and other aspects of the existing system.
- A preliminary design is created for the new system.

- A first prototype of the new system is constructed from the preliminary design. This is usually a scaled-down system, and represents an approximation of the characteristics of the final product.
- A second prototype is evolved by a fourfold procedure:
- Evaluating the first prototype in terms of its strengths, weakness, and risks.
- Defining the requirements of the second prototype.
- Planning an designing the second prototype.
- Constructing and testing the second prototype.
- At the customer option, the entire project can be aborted if the risk is deemed too great. Risk factors might involved development cost overruns, operating-cost miscalculation, or any other factor that could, in the customer's judgment, result in a less-than-satisfactory final product.
- The existing prototype is evaluated in the same manner as was the previous prototype, and if necessary, another prototype is developed from it according to the fourfold procedure outlined above.
- The preceding steps are iterated until the customer is satisfied that the refined prototype represents the final product desired.
- The final system is constructed, based on the refined prototype.
- The final system is thoroughly evaluated and tested. Routine maintenance is carried on a continuing basis to prevent large scale failures and to minimize down time.

DATA PROCESSING AND METHODOLOGY



A. Data Collection

We have collected the data securely in accordance with an agreed methodology. The rocedure for the collected data may differ from client to client and is dependent on the type, quantity, availability and need of data.

B. Data Cleaning and Preprocessing

The collected data is passed through a 'cleaning' process, so as to make sure that the data is segregated properly and identified gaps in the data are filled with the appropriate information, making data compatible and also fixing errorsin storage systems which can cause data redundancy.

C. Data Modeling

This is primarily a process in which the given dataset and the objects in it are analyzed to get a clear view of the requirements that may help us support our business model. Based on the analysis on patterns present in the data, models are then created on the established flow of the project. This flow offers a better assistance in the utilization of the previously agreed upon semi-formal model that showcases the features of the project. It also provides guidance to follow the relation between the data objects and other objects.

D. Data Prediction

Machine Learning prediction models are trained in this process and then later on evaluated using the data. This will then be applied to the preprocessed dataset. Some of the Models to be used for the prediction are:

- Linear Regression
- Random Forest
- Decision tree
- XG Boost Regressor

E. Data Visualization

Data Analyzed is then further picturized for customers and the admin to reach out conclusions and take effective decisions.

DATA MINING ON WHAT KIND OF DATA?

RELATIONAL DATABASES

A database system, also called a database management system (DBMS), consists of a collection of interrelated data, known as a database, and a set of software programs to manage and access the data A relational database is a collection of tables, each of which is assigned a unique name. Each table consists of a set of attributes (columns or fields) and usually stores a large set of tuples (records or rows)
Each tuple in a relational table represents an object identified by a unique key and described by a set of attribute values A semantic data model, such as an entity-relationship (ER) data model, is often constructed for relational databases. An ER data model represents the database as a set of entities and their relationships Relational data can be accessed by database queries written in a relational query language, such as SQL

DATA WAREHOUSES

A data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site Data warehouses are constructed via a process of data cleaning, data integration, data transformation, data loading, and periodic data refreshing

OBJECT-RELATIONAL DATABASES

Based on an object-relational data model Extends the relational model by providing a rich data type for handling complex objects and object orientation Objects that share a common set of properties can be grouped into an object class. Each object is an instance of its class. Object classes can be organized into class/subclass hierarchies.

ADVANCED DATA AND INFORMATION SYSTEMS

With the progress of database technology, various kinds of advanced data and information systems have emerged and are undergoing development to address the requirements of new

applications handling spatial/temporal data (such as maps) engineering design data (such as the design of buildings, system components, or integrated circuits) hypertext and multimedia data (including text, image, video, and audio data) time-related data (such as historical records or stock exchange data) stream data (such as video surveillance and sensor data, where data flow in and out like streams) the World Wide Web (a huge, widely distributed information repository made available by the Internet).

THE WORLD WIDE WEB

The World Wide Web and its associated distributed information services, such as Yahoo! and Google provide rich, worldwide, on-line information services, where data objects are linked together to facilitate interactive access Capturing user access patterns in such distributed information environments is called Web usage mining (or Weblog mining)

Database or data warehouse server responsible for fetching the relevant data, based on the user's data mining request can be decouples/loose coupled/tightly coupled with the database layer

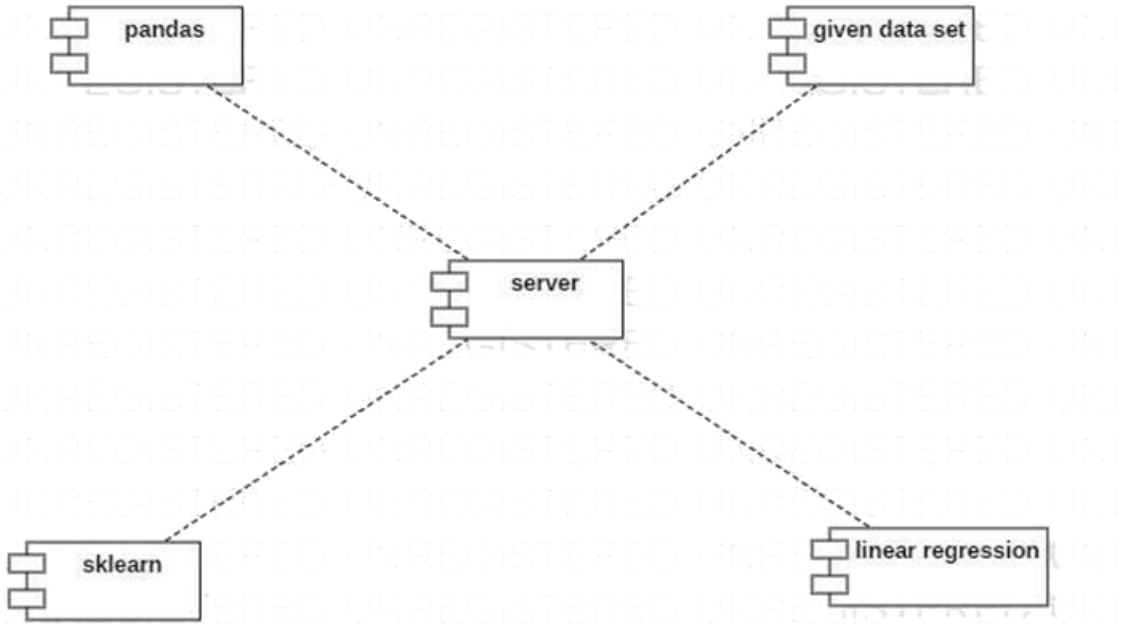
Knowledge base the domain knowledge that is used to guide the search or evaluate the interestingness of resulting patterns interestingness constraints or thresholds, metadata, concept hierarchies, etc.

Data mining engine this is essential to the data mining system and ideally consists of a set of functional modules for tasks such as characterization, association and correlation analysis, classification, prediction, cluster analysis, outlier analysis, and evolution analysis query languages (DMQL) based on mining primitives to access the data

Pattern evaluation module interacts with the data mining modules so as to focus the search toward interesting patterns may use interestingness thresholds to filter out discovered patterns may be integrated with the mining module
User interface communicates between users and the data mining system allows the user to interact with the system by specifying a data mining query or task, providing information to help focus the search, and performing exploratory data mining based on the intermediate data mining results allows the user to browse database and data warehouse schemas or data structures, evaluate mined patterns, and visualize the patterns in different forms.

System Implementation

Component diagram



Python Web Frameworks

A web framework is a code library that makes a developer's life easier when building reliable, scalable and maintainable web applications.

Why are web frameworks useful?

Web frameworks encapsulate what developers have learned over the past twenty years while programming sites and applications for the web.

Frameworks make it easier to reuse code for common HTTP operations and to structure projects so other developers with knowledge of the framework can quickly build and maintain the application.

Common web framework functionality

Frameworks provide functionality in their code or through extensions to perform common operations required to run web applications. These common operations include:

- URL routing
- HTML, XML, JSON, and other output format templating
- Database manipulation
- Security against Cross-site request forgery (CSRF) and other attacks
- Session storage and retrieval

Not all web frameworks include code for all of the above functionality. Frameworks fall on the spectrum from executing a single use case to providing every known web framework feature to every developer. Some frameworks take the "batteries-

included" approach where everything possible comes bundled with the framework while others have a minimal core package that is amenable to extensions provided by other packages.

Comparing web frameworks

There is also a repository called compare-python-web-frameworks where the same web application is being coded with varying Python web frameworks, templating engines and object.

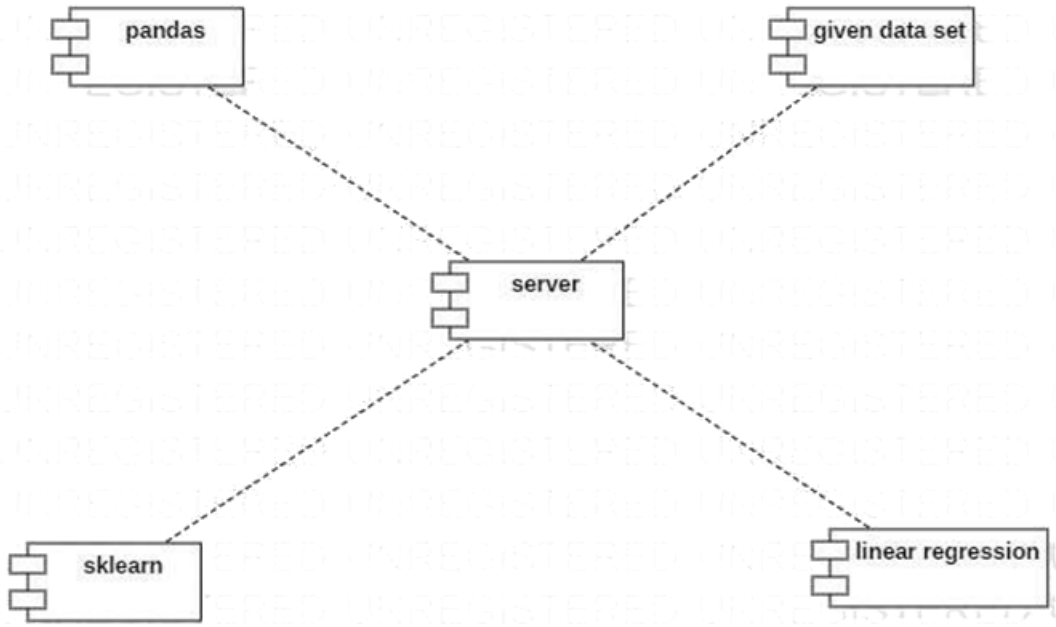
Web frameworks learning checklist

1. Choose a major Python web framework (Django or Flask are recommended) and stick with it. When you're just starting it's best to learn one framework first instead of bouncing around trying to understand every framework.
2. Work through a detailed tutorial found within the resources links on the framework's page.
3. Study open source examples built with your framework of choice so you can take parts of those projects and reuse the code in your application.
4. Build the first simple iteration of your web application then go to the deploymentsection to make it accessible on the web.

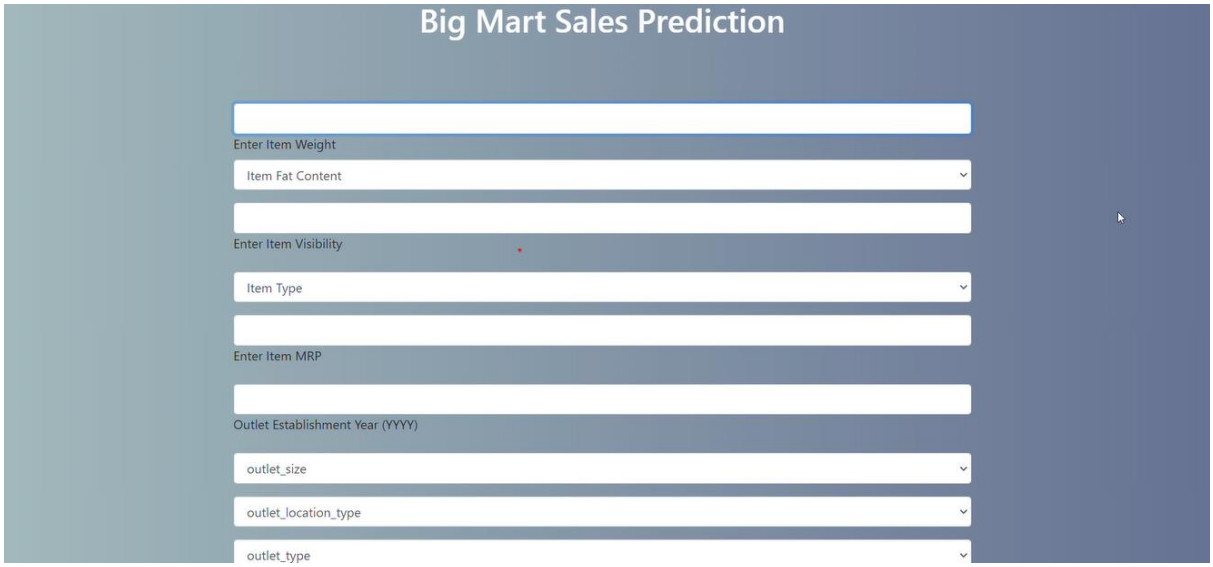
Jupyter notebook:

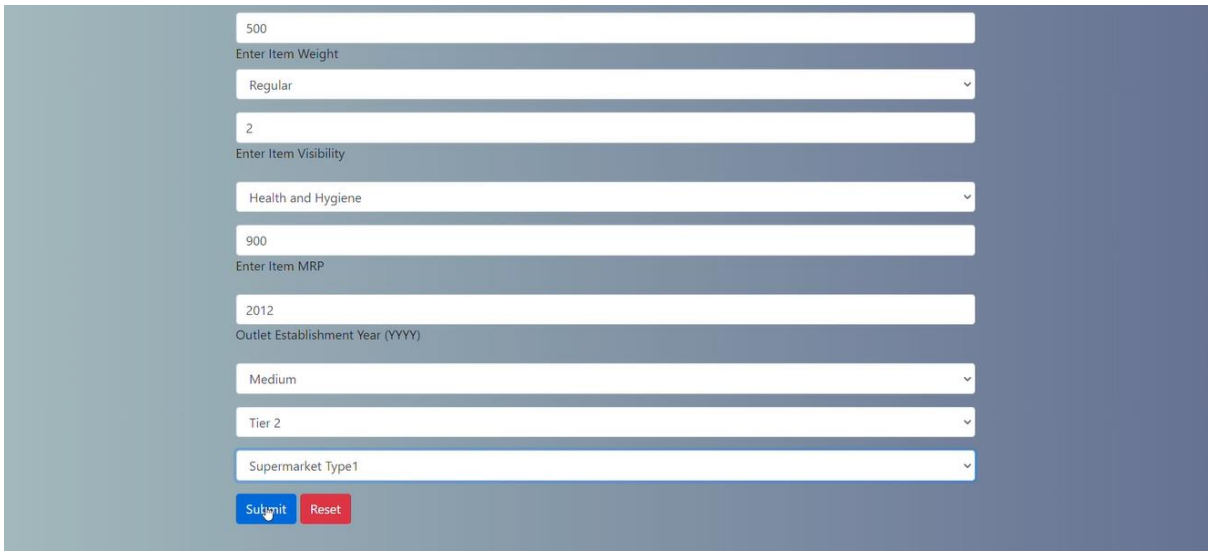
The **Jupyter** Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning

Deployement diagram:



Modules and Snapshots





Result screen:

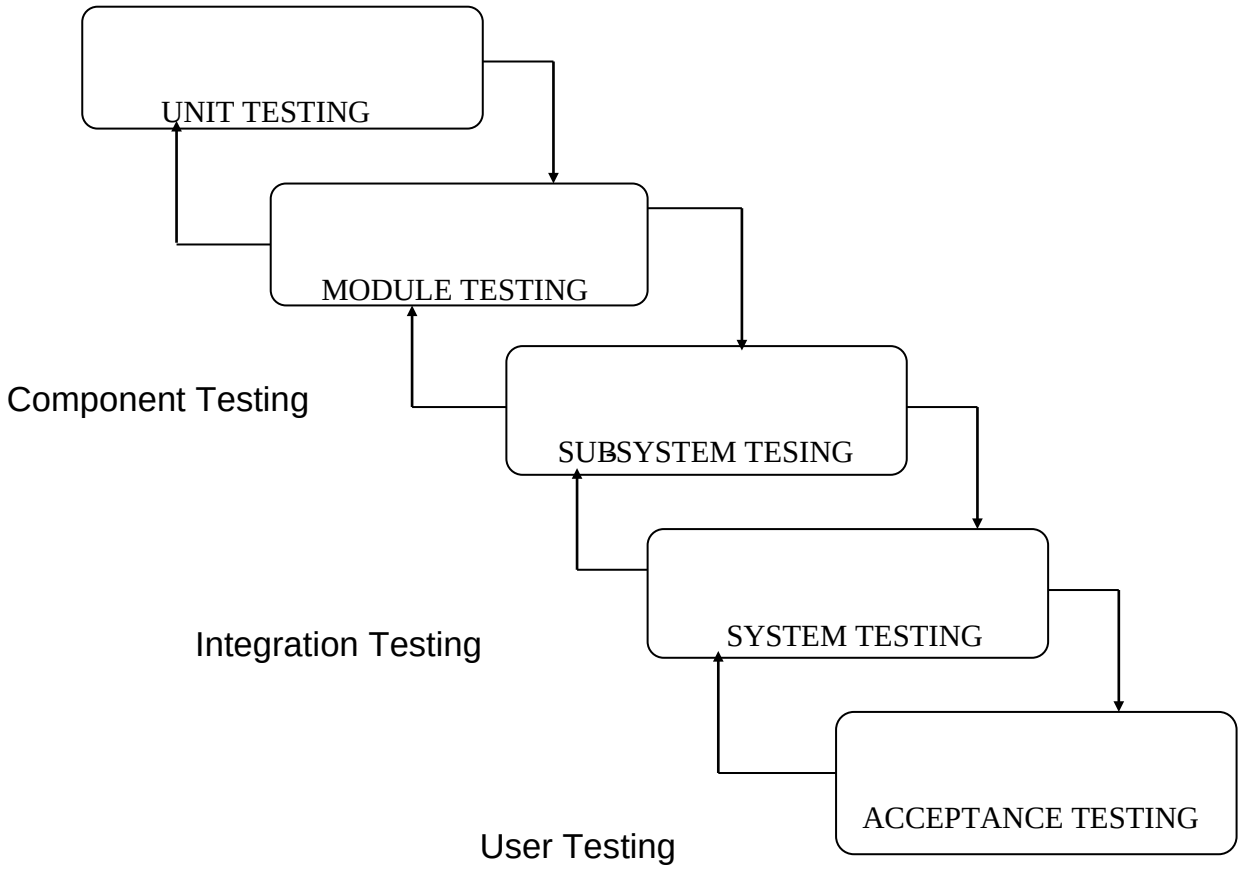
System Testing

INTRODUCTION

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and coding. In fact, testing is the one step in the software engineering process that could be viewed as destructive rather than constructive.

A strategy for software testing integrates software test case design methods into a well-planned series of steps that result in the successful construction of software. Testing is the set of activities that can be planned in advance and conducted systematically. The underlying motivation of program testing is to affirm software quality with methods that can economically and effectively apply to both strategic to both large and small-scale systems.

STRATEGIC APPROACH TO SOFTWARE TESTING



UNIT TESTING

Unit testing focuses verification effort on the smallest unit of software design, the module. The unit testing we have is white box oriented and some modules the steps are conducted in parallel.

1. WHITE BOX TESTING

This type of testing ensures that

- All independent paths have been exercised at least once
- All logical decisions have been exercised on their true and false sides
- All loops are executed at their boundaries and within their operational bounds
- All internal data structures have been exercised to assure their validity.

To follow the concept of white box testing we have tested each form .we have created independently to verify that Data flow is correct, All conditions are exercised to check their validity, All loops are executed on their boundaries.

2. BASIC PATH TESTING

Established technique of flow graph with Cyclomatic complexity was used to derive test cases for all the functions. The main steps in deriving test cases were:

Use the design of the code and draw correspondent flow graph.

Determine the Cyclomatic complexity of resultant flow graph, using formula:

V(G)=E-N+2 or
V(G)=P+1 or
V (G) =Number Of Regions

Where V (G) is Cyclomatic complexity,
E is the number of edges,
N is the number of flow graph nodes,
P is the number of predicate nodes.

Determine the basis of set of linearly independent paths.

3. CONDITIONAL TESTING

In this part of the testing each of the conditions were tested to both true and false aspects. And all the resulting paths were tested. So that each path that may be generate on particular condition is traced to uncover any possible errors.

4. DATA FLOW TESTING

This type of testing selects the path of the program according to the location of definition and use of variables. This kind of testing was used only when some local variable were declared. The *definition-use chain* method was used in this type of testing. These were particularly useful in nested statements.

5. LOOP TESTING

In this type of testing all the loops are tested to all the limits possible.
The following exercise was adopted for all loops:

1. All the loops were tested at their limits, just above them and just below them.

2. All the loops were skipped at least once.

3. For nested loops test the inner most loop first and then work outwards.

4. For concatenated loops the values of dependent loops were set with the help of connected loop.Unstructured loops were resolved into nested loops or concatenated loops and tested as above.

Conclusions

Hence, we propose a software tool for forecasting future sales volume based on the historical sales data. Using this tool, the accuracy of prediction for multiple linear regressions and random forests can be determined. In this report, basics of machine learning and the associated data processing and modeling algorithms have been described, followed by their application for the task of sales prediction in Big Mart shopping centers at different locations. On implementation, the prediction results show the correlation among different attributes considered and how a particular location of medium size recorded the highest sales, suggesting that other shopping locations should follow similar patterns for improved sales.

REFERENCES

1. Rohit Sav, Pratiksha Shinde, Saurabh Gaikwad (2021, June). Big Mart Sales Prediction using Machine Learning.2021 International Journal of Research Thoughts (IJCRT).
2. Inedi. Theresa, Dr. Venkata Reddy Medikonda,K.V.Narasimha Reddy. (2020, March). Prediction of Big Mart Sales using Exploratory Machine Learning Techniques 020 International Journal of Advanced Science and Technology (IJAST).
3. Heramb Kadam, Rahul Shevade, Prof. Deven Ketkar , Mr. Sufiyan Rajguru (2018). A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression. (IJEDR).
4. Gopal Behere, Neeta Nain (2019). Grid Search Optimization (GSO) Based Future Sales Prediction for Big Mart. 2019 International
5. Conference on Signal-Image Technology & Internet-Based Systems (SITIS).
6. Kumari Punam , Rajendra Pamula , Praphula Kumar Jain (2018, September 28-29). A Two-Level Statistical Model for Big Mart Sales Prediction. 2018 International conference on on Computing, Power and Communication Technologies
7. Ranjitha P, Spandana M. (2021). Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms.Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021).
8. Bohdan M. Pavlyshenko (2018, August 25). Rainfall Predictive Approach for La Trinidad, Benguet using Machine Learning Classification. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP).
9. Nikita Malik, Karan Singh. (2020, June). SalesPrediction Model for Big Mart.
10. Gopal Behere, Neeta Nain. (2019, September). A Comparative Study of Big Mart Sales Prediction.
11. Archisha Chandel, Akanksha Dubey, Saurabh Dhawale, Madhuri Ghuge (2019, April). Sales Prediction System using Machine Learning. International Journal of Scientific Research and Engineering Development.
12. https://www.javatpoint.com/machine-learning-random-forestalgorithm
13. https://www.javatpoint.com/machine-learning-decision-treeclassification-algorithm
    A. Chandel, A. Dubey, S. Dhawale and M. Ghuge,;Sales Prediction System using Machine Learning; International Journal of Scientific
14. Research and Engineering Development, vol. 2, no. 2, pp. 1-4, 2019.

[Accessed 27 January 2020].