CMR UNIVERSITY

Private University Established in Karnataka State by Act No. 45 of 2013

SCHOOL OF ENGINEERING AND TECHNOLOGY

Chagalahatti, Bengaluru, Karnataka- 562149

on

"SALES FORECASTING USING MACHINE LEARNING"

Submitted By

MANOJ BAHADUR

(18BBTCS063)

Submitted in partial fulfilment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Under the Supervision of

Dr. Babu Rao K
Professor & HoD
Department of IT,
School of Engineering & Technology
CMR University.

2021 – 2022

CMR UNIVERSITY

Private University Established in Karnataka State by Act No. 45 of 2013

SCHOOL OF ENGINEERING AND TECHNOLOGY

Chagalahatti, Bengaluru, Karnataka- 562149

CERTIFICATE

*This is to certify that the **Capstone Project Build** Work entitled **"Sales Forecasting Using Machine Learning"** has been successfully carried out by **Manoj Bahadur (18BBTCS063)** in partial fulfilment of the requirement for the award of the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING** by **SCHOOL OF ENGINEERING AND TECHNOLOGY, CMR UNIVERSITY** during the academic year 2021-2022. It is certified that all corrections/suggestions indicated for InternalAssessment have been incorporated in the Report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Projectwork prescribed for the said Degree.*

Signature of the project guide

Dr. Babu Rao K,
HoD, Dept. of IT,
School of Engg & Tech,
CMR University.

Signature of the Project Coordinator

Dr.R.Sathiyaraj,
Asst. Prof, Dept. of CSE,
School of Engg & Tech,
CMR University

Signature of the HOD

Dr. P.Rubini,
HOD, Dept. of CSE,
School of Engg & Tech,
CMR University.

Signature of the Dean

Dr. C. Prabhakar Reddy
Dean,
School of Engg & Tech
CMR University.

Examiners

1) _____ _____

2) _____ _____

DECLARATION

I, MANOJ BAHADUR(18BBTCS063), student of Bachelor of Technology, Computer Science and Engineering, CMR University, Bengaluru, hereby declare that the Capstone project work entitled ”Sales Forecasting using Machine Learning” submitted by me, for the award of the Bachelor’s degree in Computer Science and Engineering to CMR University is a record of bonafide work carried out independently by us under the supervision and guidance of Dr. Babu Rao K, Professor and HoD, Dept of IT, CMR University.

I further declare that the work reported in this capstone project work has not been submitted and will not be submitted, either in part or in full, for the award of any other degree in this university or any other institute or University

Place: Bengaluru

Date: 05/05/22

MANOJ BAHADUR

ABSTARCT

Machine Learning is transforming every walk of life and has become a major contributor in real world scenarios. The revolutionary applications of Machine Learning can be seen in every field including healthcare, sales, entertainment, transport and several more, the list is never ending. The traditional approach of sales and marketing goals no longer help the companies, to cope up with the pace of competitive market. Major transformations can be seen in the domain of sales and marketing as a result of Machine Learning advancements. Owning to such advancements, various critical aspects such as consumer's purchase patterns, target audience, and predicting sales for the recent years to come can be easily determined, thus helping the sales team in formulating plans for a boost in their business. The aim of this project is to propose a dimension for predicting the future sales of Walmart keeping in view the sales of previous years. A comprehensive study of sales prediction is done using Machine Learning models such as Linear Regression, K-Neighbors Regressor, and XGBoost Regressor. The prediction includes data parameters such as item store, date, temperature, fuel price, Markdown 1-5, CPI(customer price index), unemployment and holidays.

ACKNOWLEGMENT

The satisfaction that accompanies the successful completion of this project would be incomplete without the mention of the people who made it possible, without whose constant guidance and encouragement would have made efforts go in vain.

We consider ourselves privileged to express gratitude and respect towards all those who guided us through the completion of the project. We express my thanks to our Project Guide **Dr. Babu Rao K, Professor & Head,** Department of Information Technology, School of Engineering and Technology, CMR University for his constant support.

We would like to express my thanks to **Dr. P.Rubini, Professor and Head**, Department of Computer Science and Engineering, School of Engineering and Technology, CMR University, Bangalore, for her encouragement that motivated me for the successful completion of Project work.

We express my heartfelt sincere gratitude to **Dr. C. Prabhakar Reddy, Dean,** School of Engineering and Technology, CMR University for her support.

We would like to express my sincere thanks and gratitude to our Project coordinator **Dr. R. Sathiyaraj** for his support, invaluable guidance and encouragement throughout the tenure of this Capstone Project.

MANOJ BAHADUR (18BBTCS063)

CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

It is a well-known fact that we live in a business dominated world, be it buying a new car or having coffee at your local barista, everything in today's modern world is business-led, and the percentage of newer businesses or startups only grows higher every day. As a matter of fact, the top 10 companies of the world are all business-led. But have you ever wondered how did they manage to get there?

The answer is by using appropriate sales forecasting models and being ready beforehand for higher or lower business demands and being able to keep the pipeline efficient throughout the financial year. The business sector overall contributes 72% of GDP (Gross Domestic Product) in the OECD (Organization for Economic Co-operation and Development), and corporations with more than $1 billion in revenue account for an increasingly large share of that.

Stock analysts depend on accurate forecasts, and shareholders make their buying and selling investment decisions on this type of forecast-based analysis. A missed sales forecast could lead directly to your stock price plummeting, a scenario that is sure to upset your shareholders and members of your board. Simply put, getting your sales forecast wrong at a public company could spell your doom.

But what about at a smaller, private company? Finally, many small companies also have to consider how much incoming cash they should expect, with this revenue forecast not necessarily being the same as your sales forecast. Without a deep-well of customers or the backing of a VC firm, a small company likely depends on incoming cash from customers.

If you expected to bring in $100k in revenue this quarter – of which $50k had been earmarked for general utilities in your office, such as keeping the lights on and the heat running – and you only bring in $70k, that puts your operational planning in a real tough spot. Every cent counts when planning the operations at a small business, so getting both your revenue and sales forecast right absolutely matters.

1.1 BACKGROUND

In uncertain times like these, business leaders would kill to have predictable revenue. Many of them are still grappling with how to forecast revenue for the next year, which is often the starting point for drawing up annual budgets for the organization. With distributed sales teams, businesses are now relying on their ability to forecast, now more than ever, to drive their entire growth strategy. Sales forecasting is both a science and an art. Decision makers rely on these forecasts to plan for business expansion and to determine how to fuel the company's growth. So, in many ways, sales forecasting affects everyone in the organization.

Yet for most companies coming up with an accurate sales forecast is still a major challenge. Thanks to inaccurate forecasting methods based on intuition, companies end up having poor visibility into projected sales. According to Clari, a revenue operations platform, 93 percent of sales leaders are unable to forecast revenue within 5 percent, even with two weeks left in the quarter.

A sales forecast is an expression of expected sales revenue.

A sales forecast estimates how much your company plans to sell within a certain time period (like quarter or year). The best sales forecasts do this with a high degree of accuracy. Sales forecasts differ in where they're getting their inputs – for example, they may rely on sales reps intuition or even artificial intelligence (AI). More on that in the section on tools used to forecast sales revenue. But all sales forecasts answer two key questions:

- **How much**: Each sales opportunity has its own projected amount it'll bring into the business. Whether that's $500 or $5 million, sales teams have to come up with one number representing that new business. To create the number, they take everything they know about the prospect into account.

- **When**: Sales forecasts pinpoint a month, quarter, or year when the sales team expects the revenue to hit.

When a company constantly misses its sales forecast it can have a negative impact on its valuation over the long term. Exceeding your forecasts isn't good news either. When you cannot estimate how much revenue you will generate accurately, you can't hire or invest to keep with the growth and that could lead to several missed opportunities.

Fret not, sales forecasting is no rocket science either. Since you have a ringside view of the business, you probably are in the best place to see where it is going. You just need to put in a repeatable scaleable data-driven process in place.

Let us factor in the important ingredients of who, what, where, why, and how to make their forecasts:

- **Who**: Sales teams make their forecasts based on who their prospects are. Depending on if their prospects are the actual decision makers or just influences, the forecast will be more or less exact.
- **What**: Forecasts should be based on exactly what solutions you plan to sell. In turn, that should be based on problems your prospects have voiced, which your company can uniquely solve.
- **Where**: Where is the buying decision made, and where will the actual products be used? Sales teams see better accuracy when they get closer (at least for a visit) to the center of the action.
- **Why**: Why is the prospect or existing customer considering new services from your company in the first place? Is there a compelling event making them consider it now? Without a forcing function and a clear why, the deal may stall inevitably.
- **How**: How does this prospect tend to make purchasing decisions? If you're not accounting for how they do it now and how they've done it in the past, it may be fuzzy math.

1.2 PROBLEM STATEMENT

The sales forecast indicates as to how much of a particular product is likely to be sold in a specified future period in a specified market at specified price. Sales forecasts can be used to identify benchmarks and determine incremental impacts of new initiatives, plan resources in response to expected demand, and project future budgets.

Sales forecasting, estimation of type, quantity and quality of future sales. Goal for the sales department is decided on the basis of this forecast and these forecasts also help in planning future development of the concern. To develop an application that can forecast and provide accurate results is the task undertaken.

1.3 OBJECTIVE

Holidays can create a huge impact on sales. So, if there is a good prediction on Sales then Walmart can Calculate how much product to order during Holiday time. It will help in predicting which products needs to be purchased during the holiday time. As customers planning to buy something expects the products to be available immediately. And through prediction they can figure out which product will require at what time. Thus soar the trust of Customer on Walmart. This problem can also solve the issue of Marketing Campaigns. As Forecasting is often used to adjust ads and marketing campaigns and can influence the number of sales. Walmart runs several markdown events throughout the year. And these markdown event precede to the prominent holidays. So to solve the issue Walmart can organize such events more efficiently.

Time Series data known to have linearity. A simple mean or median model can't map the linear relationship efficiently or at all. A model like ARIMA (Autoregessive Integrated Moving Average) does not take into account the impact of features like size of store, Type of store, which holiday occur during that time while predicting Sales. Therefore, A Machine Learning model like, a linear regression model can provide robust prediction given the data set satisfies its linearity assumptions. Furthermore, machine learning forecasting is not a black box; the influence of model inputs can be weighed and understood so that the forecast is intuitive and transparent. Machine Learning models can also be updated and become adaptable to the changes in data set. Also, through machine learning help, relation between markdown events and weekly sales can be utilize in correct manner using machine learning model.

Chapter 2

AIM AND SCOPE

2.1 LITERATURE SURVEY

1) Sales Forecasting in Industries.

AUTHORS: Daniel Büttner; Markus Rabe

Sales forecasts are required for planning resources and defining stock levels through the supply chain (SC) because demand is becoming diversified due to higher customer expectations regarding service and higher competitive pressure through products' substitution possibilities. The theoretical foundations of Time Series Analyses (TSA) started in 1927 with the work of Yule and TSA, but it still seems to be only partially established in industry. Meanwhile, the use of machine learning (ML) approaches for forecasting sales volumes has come to the fore in the current big data era due to higher data availability and computing power. In recent years, much experience and new methods have been obtained in this scientific field. Both TSA and ML will continue to play a role. The variety of methods should be tested and compared continuously in a quantitative and qualitative manner to support practical knowledge and advance forecasting. Using data from a company in an electrical industry, this paper compares sales forecasts built from TSA and ML.

2) Business Forecasting using Machine Learning

AUTHORS: Md. Anisur Rahman Mia; Mohammad Abu Yousuf; Rupon Ghosh

Forecasting system can help any business to predict its future sale, profit and loss. So it is very useful while launching a new product or manufacturing existing products. In this work, a business forecasting system is designed and implemented to forecast the amount of future sale of products using Machine Learning (ML). ML algorithms build a pattern from input variables then make decision.

Use of big data technology in researches also increasing. By combining big data and ML together powerful predictive systems can be designed. So, big data processing technology can be used for this work to prepare data for training purpose of the proposed system. Prediction accuracy may range and vary between 99% to 75% for different products and mean absolute percentage error (MAPE) is 7.32%. So this system methodology is very much efficient to predict future sale.

3) Seasonal Sales Prediction and Visualization for Walmart Retail Chain Using Time Series and Regression Analysis

AUTHORS: Rut Vyas; Revathi A

The forecasting of sales forms the foundation of maximizing profits and successful planning operation at the start of the fiscal year for Business to consumer (B2C) models like Walmart retail chain. It is paramount for such companies to understand the needs and buying pattern of their customers. This paper puts forth some crucial sales visualizations that can prove indispensable to a store manager and performs a comprehensive evaluation of the different forecasting models to predict sales of a retail chain. One of the chief goals of the paper is to determine prediction models that can accommodate the seasonality in sales during four holidays.

4) Walmart sales forecasting using XGBoost algorithm and feature engineering

AUTHORS: Yiyang Niu

With the amount of data growing exponentially, the rational use of big data has become the focus of enterprises to serve the future and make better decisions. Using machine learning algorithms to predict the sales of products and commodities has become a hot spot for researchers and companies in recent years. This paper proposes the XGBoost sale prediction model which combines XGBoost algorithm and meticulous feature engineering processing for predicting Walmart's sales problem.

This paper's method can effectively mine attributes of different dimensions to make predictions well. This paper evaluates the XGBoost sale forecast model on the sales data of Walmart supermarkets datasets provided by the Kaggle competition. Experimental results show our method achieves superior performance over the other machine learning approaches.

5) Machine Learning model for Slaes forecasting using XGBoost

AUTHORS: Xie dairu; Zhang Shilong

For modern retail corporations operating a huge chain of businesses, exact sales predication is the key in driving corporations development, even success or failure. Sales forecasting allows corporations to efficiently allocate resources including cash flow, production, and make better informed business plan. In this paper, we propose an efficient and accurate sales forecasting model using machine learning. Initially, feature engineering is conducted for extracting features from historical sales data. Furthermore, we used eXtreme Gradient Boosting (XGBoost) to utilize these features for forecasting the future sales amount. The experiment results on a publicly Walmart retail goods dataset provide by Kaggle competition demonstrate our proposed model performs extremely well for sales prediction with less computing time.

2.2 WHY IS SALES FORECSATING IMPORTANT?

To understand why sales forecasting is so important to business health, think about two example scenarios: one with a car manufacturer and another with an E-commerce shop. In the case of a car manufacturer, cars take a long time to build. The manufacturer has a complex supply chain to ensure every car part is available exactly when they need to build cars, so the number of cars available to purchase will meet demand. When you buy something online, whether that's from a large marketplace or a small boutique, you get a delivery estimate If your delivery comes a day or a week after it's promised, that'll affect your satisfaction with the company and decrease your willingness to want to do business with them again.

Sales forecasting is similar in both cases. Sales forecasts help the entire business plan resources to ship products, pay for marketing, hire employees, and beyond. Accurate sales forecasting yields a well-oiled machine that meets customer demand, both today and in the future. And internally on sales teams, sales revenue that delivers in its estimated time period keeps leaders and collaborators happy, just like a shipment that arrives on time. If forecasts are off, the company faces challenges that affect everything from pricing to product delivery to the end user. Meanwhile, if forecasts are on point, the company can make better investments, perhaps hiring 20 new developers instead of 10, or build a new sales office in a new territory.

2.3 TYPES OF SALES FORECAST

- **Rule-based forecasting**: In rule-based forecasting, predictions are generated using a set of manually developed rules and assumptions based on past data and known trends.
- **Machine Learning forecasting**: Machine learning algorithms would learn the rules that would have to be manually designed in rule-based forecasting. This is done through a process called supervised learning. Supervised learning is the task of learning the relationship between outputs (sales) and inputs (past sales, economic indicator, holiday calendar etc.) Machine Learning algorithms find these relationships by minimizing prediction error – i.e. finding the relationships and parameters that maximize prediction accuracy.

2.4 OBJECTIVES OF SALES FORECAST

The main objective of sales forecasting is to paint an accurate picture of expected sales. Sales teams aim to either hit their expected target or exceed it. When the sales forecast is accurate, operations go smoothly and leaders are pleased with the follow through on the plan. When the forecast is exceeded, the company gets to decide how it'll invest the extra money – and everyone's even happier. At the same time, though, if you're consistently over, you should adjust your forecast for peak precision.

Outside of just aiming for accuracy, though, sales teams hope their forecasts will achieve two simple objectives:

- **Smooth internal operations**: When the forecast is met, the friction inside the organization about all the things revenue funds – melts away. Trade-offs and compromises don't need to be made about things like cutting the workforce, reducing support, or halting product development. Instead, business hums along nicely.
- **Smooth external operations**: Every company wants to delight its customers and partners. When forecasts are met and internal operations are flowing as they should be, your company can continue funding external marketing events, staffing ample customer service touch points, investing in its community, and more. From the outside, it's clear that everything inside is working as it should be.

You'll notice a different (not so great) feeling in the hallways at work when your sales forecast is on the downside – compared to accurate or even on the upside. Your goal is to keep morale and collaboration high with a solid forecast.

2.5 DESIGN A SALES FORECASTING PLAN

Sales forecasting is a muscle, not an item to check off your to-do list. While you should absolutely design a framework for your sales forecasting plan each year, you should also change up your strategies from time to time so new muscles develop.

Craft a sales forecasting plan with your team by focusing on three primary activities:

- **Calculating number and time period**: Your plan should explain how you'll calculate the estimated monetary amount and what the time frames will be.
- **Reviewing and revising**: You should also plan to review the forecast at key milestones and revise it if necessary. Most sales leaders track progress against their forecast daily! But you'll also want to schedule designated check-ins throughout the quarter.
- **Breaking the patterns**: Even the best sales organizations need to shake up their processes once in a while. Breaking your patterns can help you find new ways of crafting even more accurate forecasting. Try skip-level forecasting, ask different questions, have executive sponsorship reviews, and take different angles of the data

2.6 SALES FORECAST IN UNPREDICTABLE TIMES

Unpredictable events have an enormous impact on your sales forecast. Extreme weather, economic crises, global pandemics like COVID-19 – all dramatically change your forecast. What you thought you knew about expected revenue growth can be suddenly flipped on its head. As soon as an extraordinary event hits, sales and finance leaders at your company will quickly want to know:

- **How's our pipeline looking today?**
- **What are the best- and worst-case scenarios?**
- **How has the forecast changed from a week or a month ago?**

Your forecast implicates resourcing, headcount, and more. Although things may be changing quickly, you don't want to give up on your forecast. Your forecast is only as good as the data coming into it from your sales teams. In uncertain times, quick access to sales data and the ability to pivot territories and resource deployment accordingly can make the difference between business continuity and dissolution.

There's no silver bullet to forecast perfectly in a crisis or unforeseen scenario. But vigilantly updating what's in the pipeline and analyzing sales data more frequently than usual will help you see trends and retool your forecast accordingly.

Empathy and care are always fundamental, but this is especially true in these situations. Empathizing with your customers' challenges and caring for your own sales reps should come before anything else. Build trust with internal and external partners. That trust will help you grow again in the future. Learn more about maintaining customer relationships as a sales leader both remotely and in times of crisis.

2.7 EXISTING ML MODELS

In the world of Machine Learning, sales forecasting is a time-series regression problem. A regression is any task concerned with the estimation of a continuous quantity (i.e, sales). Time-series regressions are a particular case of regression, with an additional time dimension.

There are two main types of time series regressions models:

- **Autoregressive models**: These models predict future sales solely based on past sales values. These models include ARIMA, SARIMAX and Exponential Smoothing. They generate predictions by finding trends and seasonality patterns.
- **Multivariate models**: Multivariate models are based on a variety of inputs, including past sales, holiday calendars, or even economic indicators. These models include Linear Regressions, Neural Networks, Decision Tree-based methods and Support Vector Machines.

Chapter 3

METHODOLOGY

3.1 PROPOSED SYSTEM

- Our aim is to develop an application that can be used to predict the sales of any business given the appropriate data set.
- The data set in this case is the Walmart Data set which has been collected from Kaggle.
- The next step is to clean the data and transform it into the desired format by feature selection method.
- The data sets are trained and tested to produce the sophisticated results and henceforth apply the suitable algorithms.
- The evaluation and interpretation of results is performed.
- The results are then integrated and displayed through the Fronted UI.

3.2 ADVANTAGES

- **Accuracy in demand planning** - The economic impact of inaccuracy is huge, and the efficacy of machine learning has already been validated. Machine learning leverages high-performing algorithms for data processing thereby, producing results more effective than traditional demand forecasting.
- **Absorb a tremendous volume of varied data and generate mission** - Machine learning enables the absorption of a great volume of varied data at a startling velocity thereby, improving the forecasting at SKU level comprising of four pillars-Time series member, case ID, transaction amount, and date.
- **Update data in real-time and re-route supply chain processes around it** - Machine learning forecasting leverages superior algorithms to constantly update the 'demand forecasts' in realtime. This helps in aligning the product, customer, pricing, and promotions with great efficiency.

- **Accelerate the data processing mechanism** - The modern machine learning architecture facilitates optimized use of memory storage and propelling the forecasting mechanisms to lightning speed. This enables the decision-makers to make well-informed investments for superior data processing.

3.3 FEATURES

- **Different Models**: This application uses different Machine Learning algorithms to give the most accurate result, taking all the important factors into account for arriving at this conclusion.
- **Forecasting with different time period**: This application helps forecast, periodically based on user input.
- **Features that impact the sales**: This application takes into account, the features like holidays, size and type of store and other factors to predict sales.
- **Data visualisation**: This application uses data visualization tools like bar graphs, line graphs, pie charts, scatter plots, etc, for easier understanding and quick decision making.

3.4 MACHINE LEARNING ALGORITHMS USED

- **Linear Regression**: A regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. The linear regression model provides a sloped straight line representing the relationship between the variables. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.
- **KNN Regression:** K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. It is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighbourhood.

- **XGBOOST Regressor:** XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. It is the most common algorithm used for applied machine learning in competitions and has gained popularity through winning solutions in structured and tabular data. The algorithm's optimization techniques improve performance and thereby provides speed using the least amount of resources.

Model	MAE	RMSE	Accuracy
Linear Regression (Baseline)	14566	21767	8.89
KNNRegressor	8769	14991	56.87
DecisionTreeRegressor	2375	7490	96.02
RandomForestRegressor	1854	5785	96.56
ExtraTreeRegressor	1887	5684	96.42
XGBRegressor	2291	5205	97.23

Fig 3.1 ML Model Comparison

From the above image, it can be observed that the algorithms WMAE/MAE (Weighted Mean Average error) and RMSE (Root Mean Square Error) is inversely proportional to its accuracy. Lesser the value of WMAE and RMSE, greater is the accuracy of that model.

In terms of accuracy,

XGB Regressor > KNN Regressor > Linear Regression

Chapter 4

SYSTEM REQUIREMENT SPECIFICATION

A software requirements specification (SRS) is a description of a software system to be developed. It lays out functional and nonfunctional requirements, and may include a set of use cases that describe user interactions that the software must provide. In order to fully understand one's project, it is very important that they come up with an SRS listing out their requirements, how are they going to meet it and how will they complete the project. SRS also functions as a blueprint for completing a project with as little cost growth as possible.

4.1 FUNCTIONAL REQUIREMENTS

Functional Requirement defines a function of a software system and how the system must behave when presented with specific inputs or conditions. These may include calculations, data manipulation and processing and other specific functionality. Following are the functional requirements on the system:

- Collecting Data sets and data per-processing is performed for that data set.
- The data set will be subjected to various data mining techniques; Clustering will be performed on the given data set.
- The clusters are then subjected to other algorithms like Association rule mining and trend analysis.

4.2 BASIC OPERATIONAL REQUIREMENTS

The customers are those that perform the primary functions of systems engineering, with special emphasis on the operator as the key customer. Operational requirements will define the basic need and, at a minimum, will be related to these following points:

- **Mission profile or scenario**: It describes about the procedures used to accomplish mission objective. It also finds out the effectiveness or efficiency of the system.
- **Performance and related parameters**: It points out the critical system parameters to accomplish the mission.

- **Utilization environments**: It gives a brief outline of system usage. Finds out appropriate environments for effective system operation.
- **Operational life cycle**: It defines the system lifetime.

4.3 HARDWARE SYSTEM CONFIGURATION

Processor: Intel Pentium and above

Ram: 4GB

Hard disk capacity: 50 GB

4.4 SOFTWARE SYSTEM CONFIGURATION

Operating system: Linux, Windows 10, Mac OS

Programming Language: Python, JavaScript, HTML, CSS, XML

Framework: Django

IDE: PyCharm

Libraries: NumPy, Pandas, SciKit, Pickle

- **Python 3**: Python is an interpreted, high-level and general purpose programming language. Its language constructs and object-oriented approach aim to help write clear, logical code for small and large-scale projects.

- **Django**: Django is a Python-based free and open source web framework that follows the model template views (MTV) architectural pattern. It is used to link the front-end and back-end of a program.

- **NumPy**: NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

- **Pandas**: Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

- **SciKit**: Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

- **Pickle**: Python pickle module is used for serializing and de-serializing a Python object structure. Any object in Python can be pickled so that it can be saved on disk. What pickle does is that it "serializes" the object first before writing it to file. Pickling is a way to convert a python object (list, dict, etc.) into a character stream. The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

- **Javascript**: Alongside HTML and CSS, JavaScript is one of the core technologies of the World Wide Web. JavaScript enables interactive web pages and is an essential part of web applications.

- **jQuery**: jQuery is a JavaScript library designed to simplify HTML DOM tree traversal and manipulation, as well as event handling, CSS animation, and Ajax. It is free, open source software using the permissive MIT License. As of May 2019, jQuery is used by 73% of the 10 million most popular websites.

- **Chart.js**: It is a free, open source, JavaScript library for data visualization, which supports 8 chart types: bar, line, pie, area, bubble, radar, polar, and scatter.

Chapter 5

RESULTS AND DISCUSSION

5.1 ARCHITECTURE

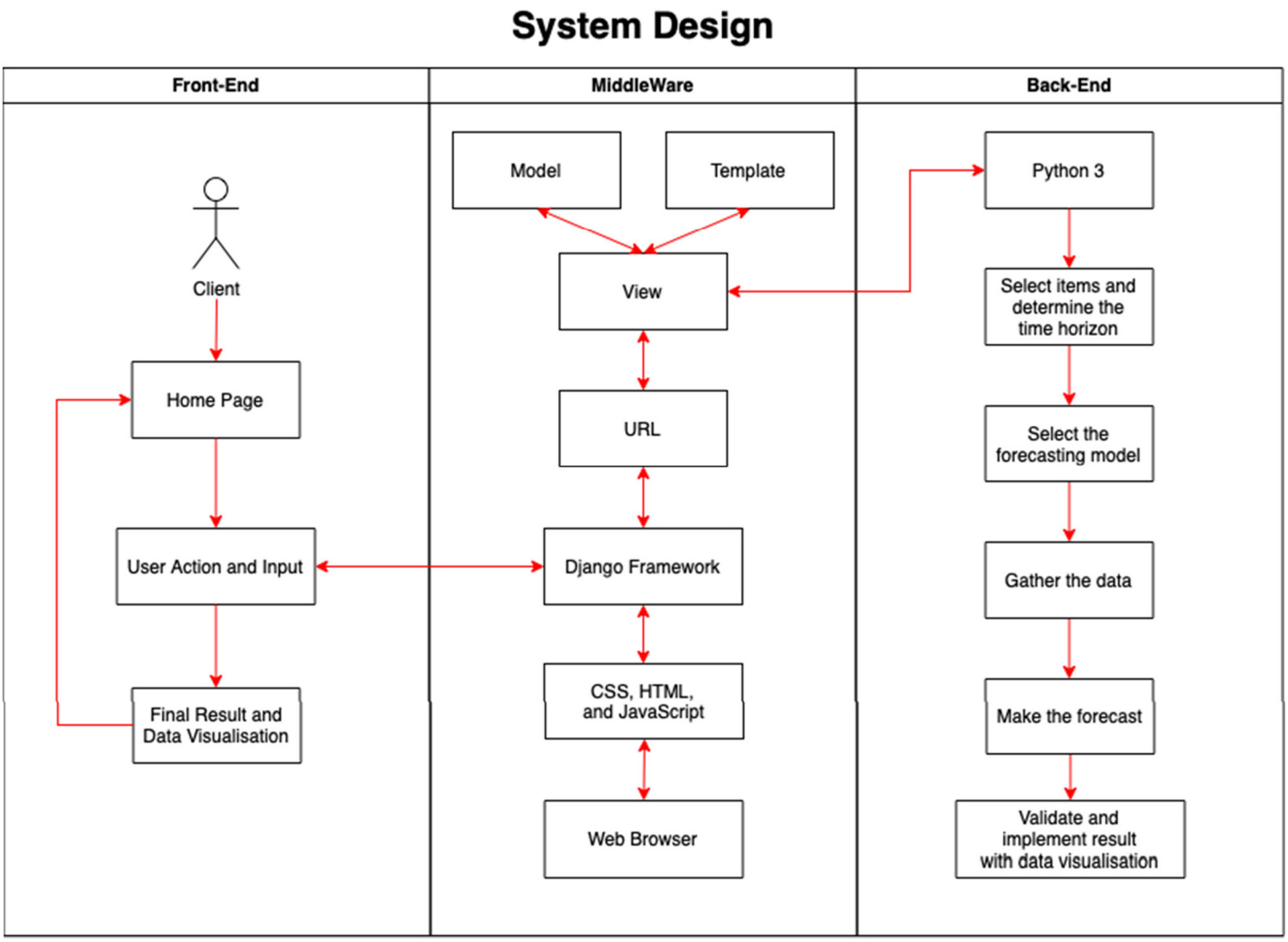

Fig 5.1 System Architecture

5.2 WORKFLOW OF PROJECT

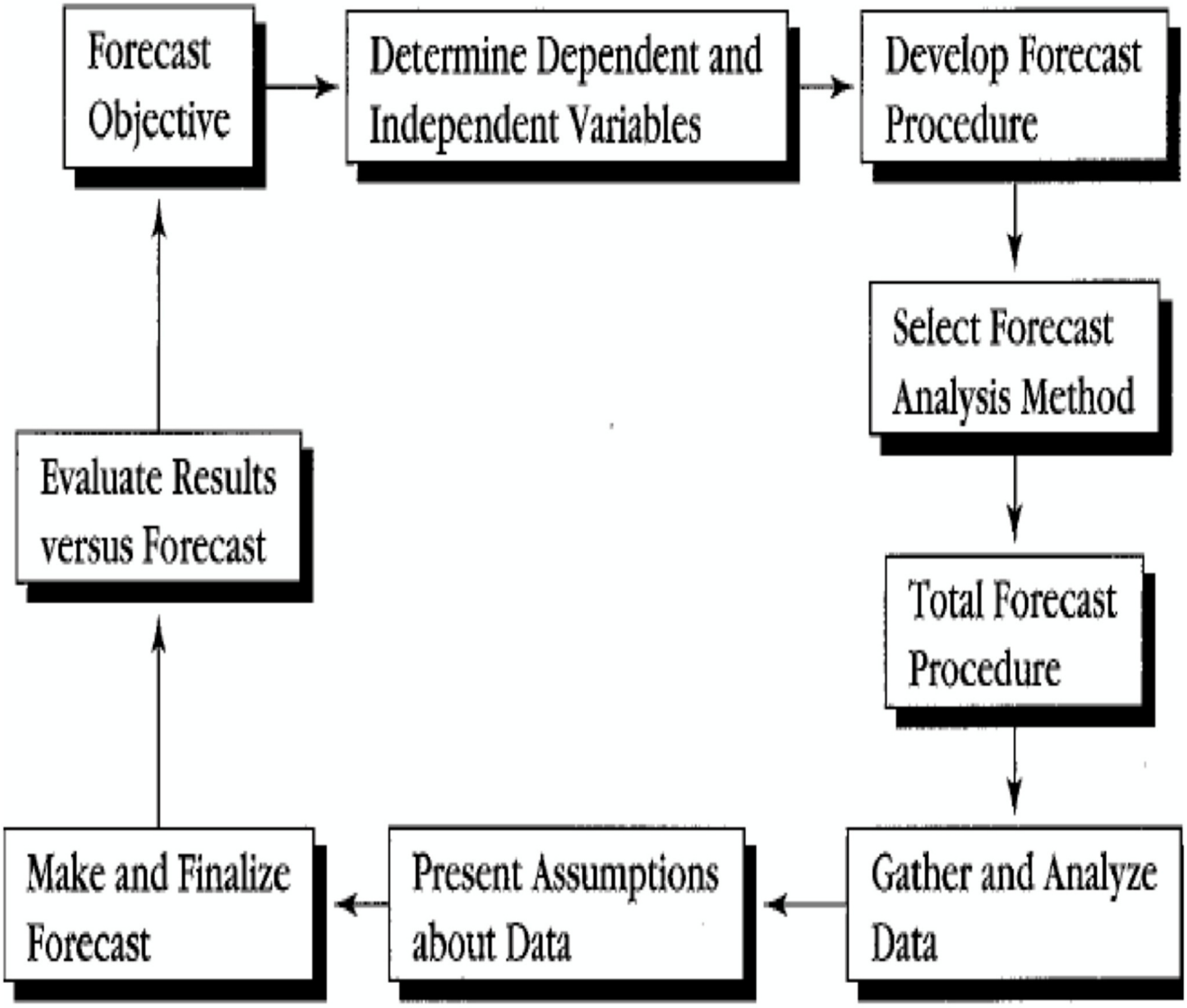

Fig 5.2 Workflow Overview

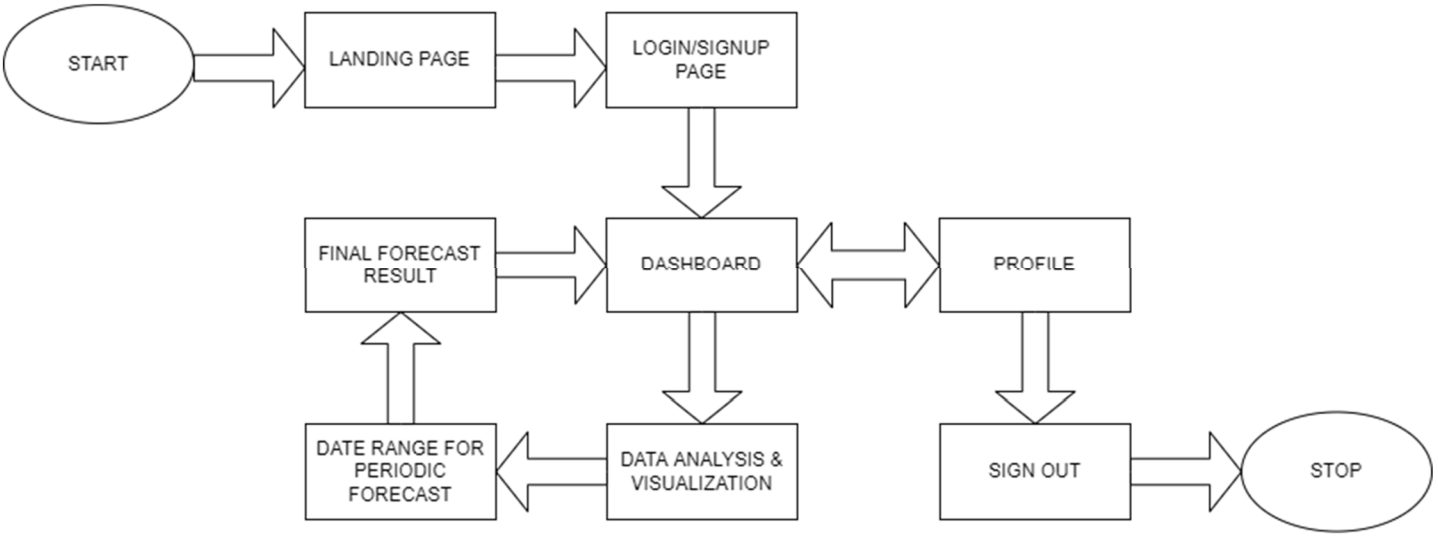

Fig 5.3 Project Workflow

5.3 DATAFLOW DIAGRAM

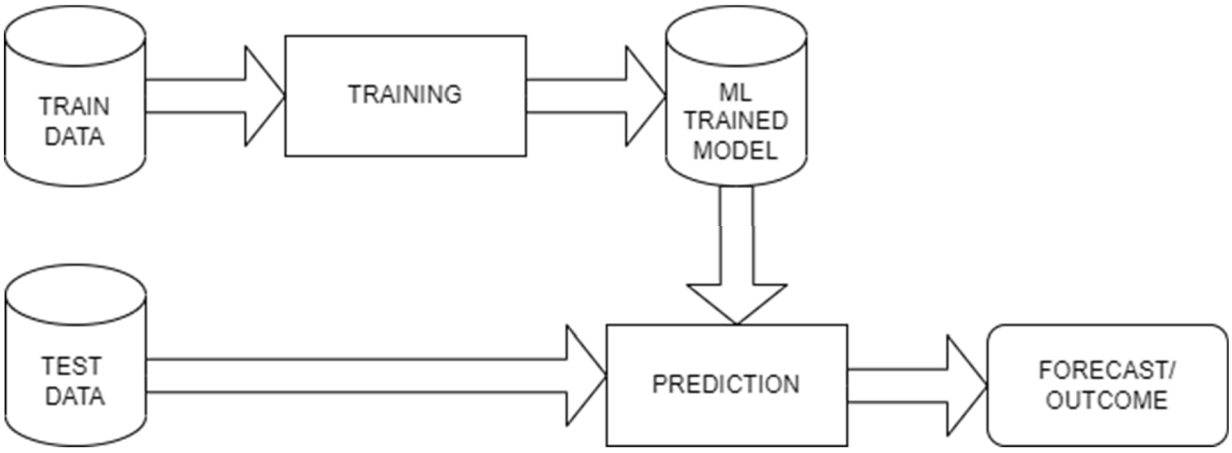

Fig 5.4 Dataflow overview

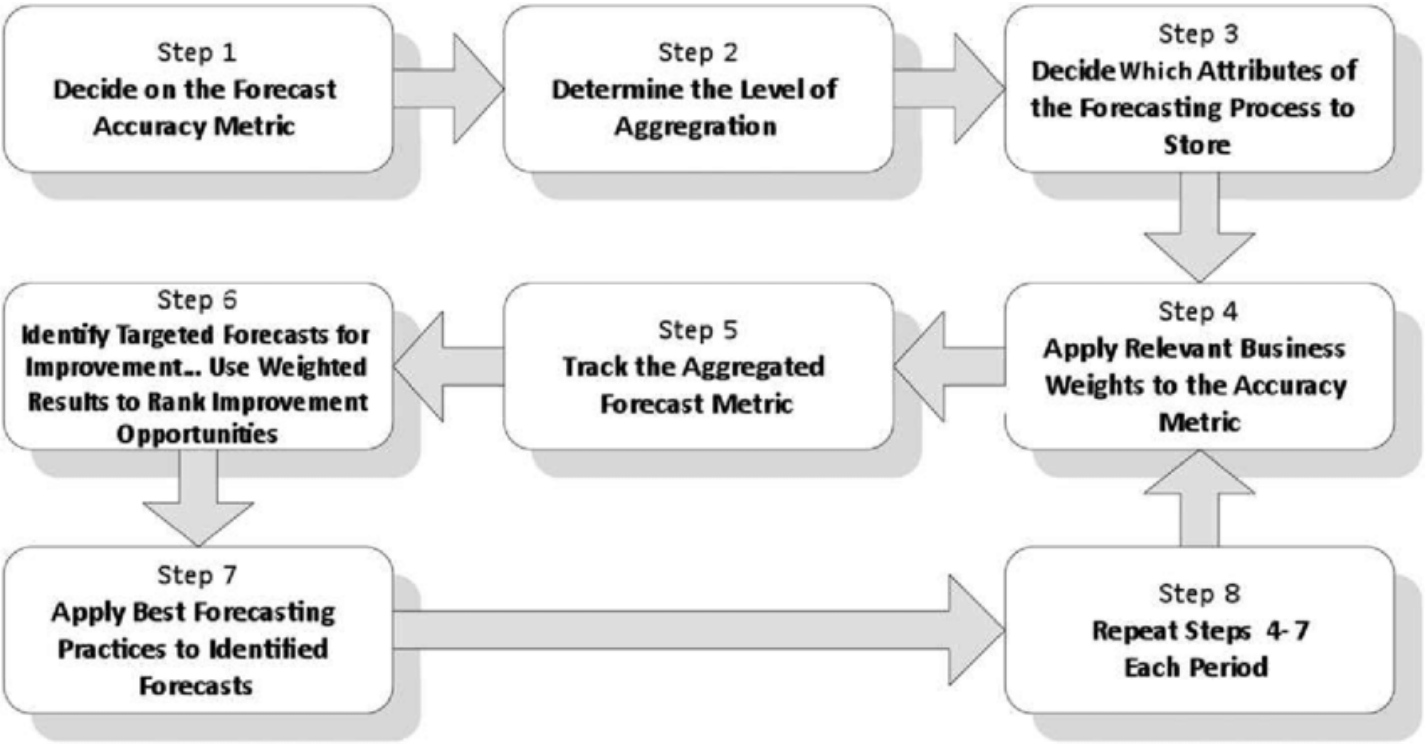

Fig 5.5 Application Dataflow

5.4 IMPLEMENTATION

Let us look at a diagram of the steps for sales forecasting, and from the diagrams such as system design diagram, workflow of the project, data flow diagram, as well as facilities that we will be offering with our application.

Fig 5.6 Sales Forecasting Steps

1. **Determine the use of forecasting**: Forecasting is a technique that uses historical data as inputs to make informed estimates that are predictive in determining the direction of future trends.
2. **Select the items to be forecasted** -The decisions need to be made about what should be forecasted. It is necessary to ask whether forecasts are needed for:
    - Every product line, or group of products?
    - Every sales outlet, or for outlets grouped by region, or only for total sales?
    - weekly data, monthly data or annual data?

3. Determine the time horizon of the forecast - It is important to consider the forecasting horizon. Will forecasts be required for one month in advance, for 6 months, or for ten years? Different types of models will be necessary, depending on what forecast horizon is most important.
4. Select the forecasting models - The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the forecasts are to be used. It is common to compare two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions and usually involves one or more parameters which must be estimated using the known historical data.
5. Gather the Data - We have historical sales data for 45 Walmart stores located in different regions. Each store contains a number of departments, and you are tasked with predicting the department wise sales for each store.

   Given below is a brief overview of the data set being used:
   stores.csv: This file contains anonymized information about the 45 stores, indicating the type and size of store.
   train.csv: This is the historical training data, which covers 2010-02-05 to 2012-11-01. Within this file you will find the following fields:
   - Store - the store number
   - Dept - the department number
   - Date - the week
   - Weekly_Sales - sales for the given department in the given store
   - IsHoliday - whether the week is a special holiday week

   test.csv: This file is identical to train.csv, except we have withheld the weekly sales. You must predict the sales for each triplet of store, department, and date in this file.

features.csv: This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:

- Store - the store number
- Date - the week
- Temperature - average temperature in the region
- Fuel_Price - cost of fuel in the region
- MarkDown1-5 - anonymized data related to promotional markdowns that Walmart is running.
- MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA.
- CPI - the consumer price index
- Unemployment - the unemployment rate
- IsHoliday - whether the week is a special holiday week

6. Make the forecast - We need to consider the different Machine Learning algorithms to give the most accurate result, by taking all the important factors into account before arriving at the conclusion. The model with the least WMAE value, have the highest accuracy.
7. Validate and implement the results - Once a model has been selected and its parameters estimated, the model is used to make forecasts. The performance of the model can only be properly evaluated after the data for the forecast period have become available. Evaluation consists of four steps: testing assumptions, testing data and methods, replicating outputs, and assessing outputs. After validating the methods and outputs, the output is presented.

5.5 OUTPUT SCREENSHOTS

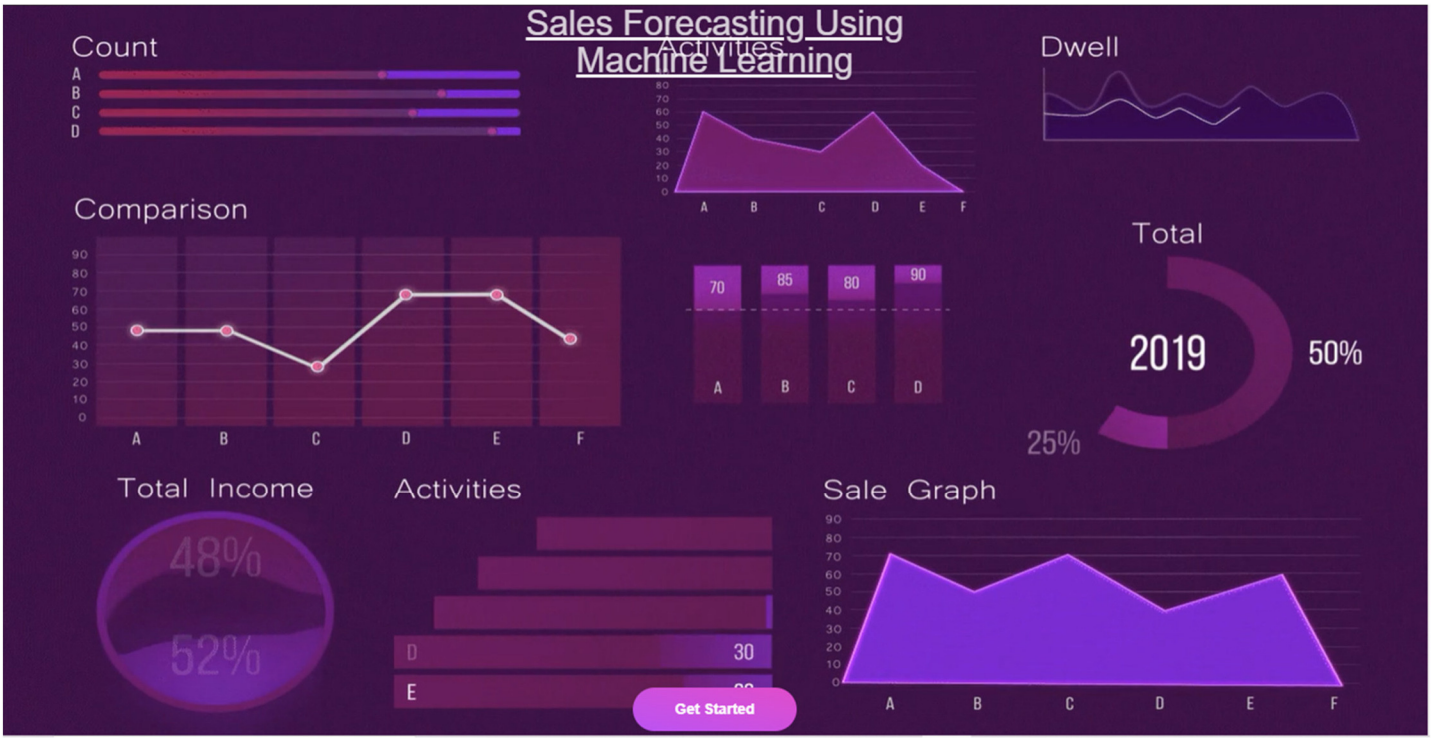

Fig 5.7 Landing Page

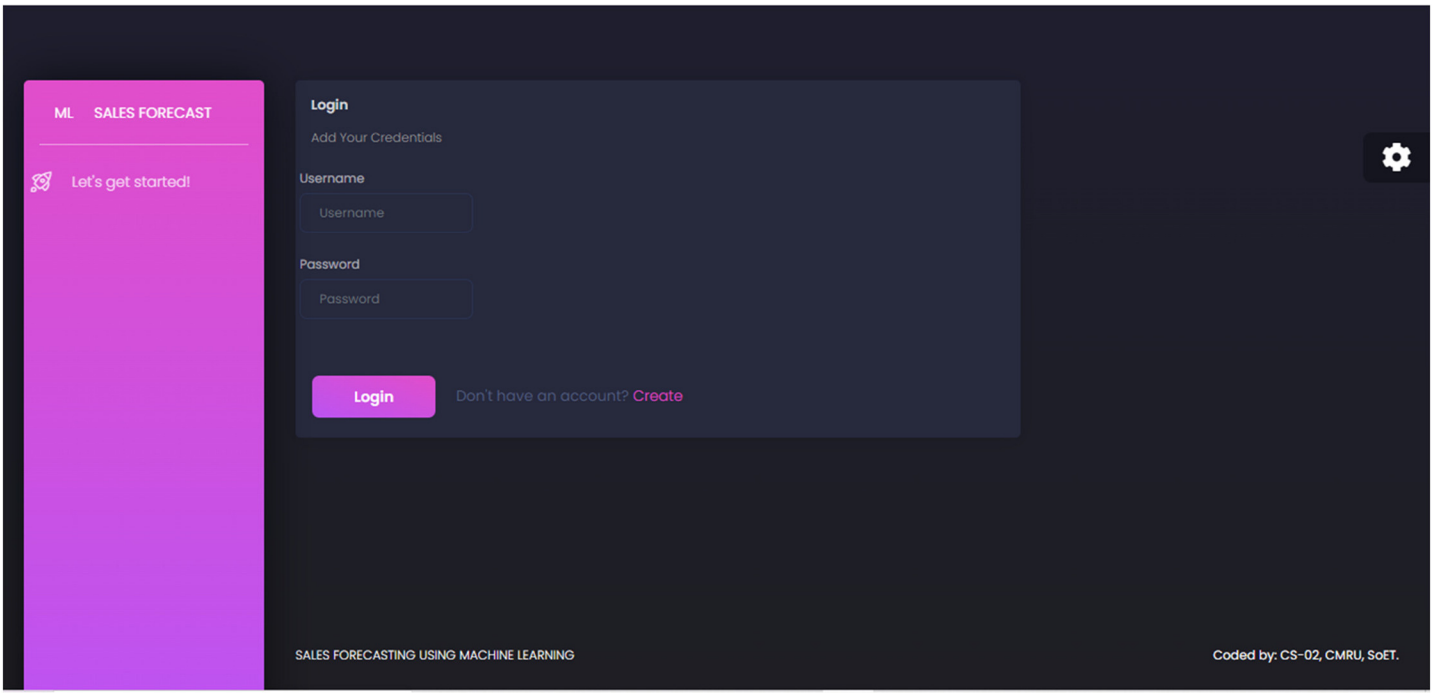

Fig 5.8 Login Page

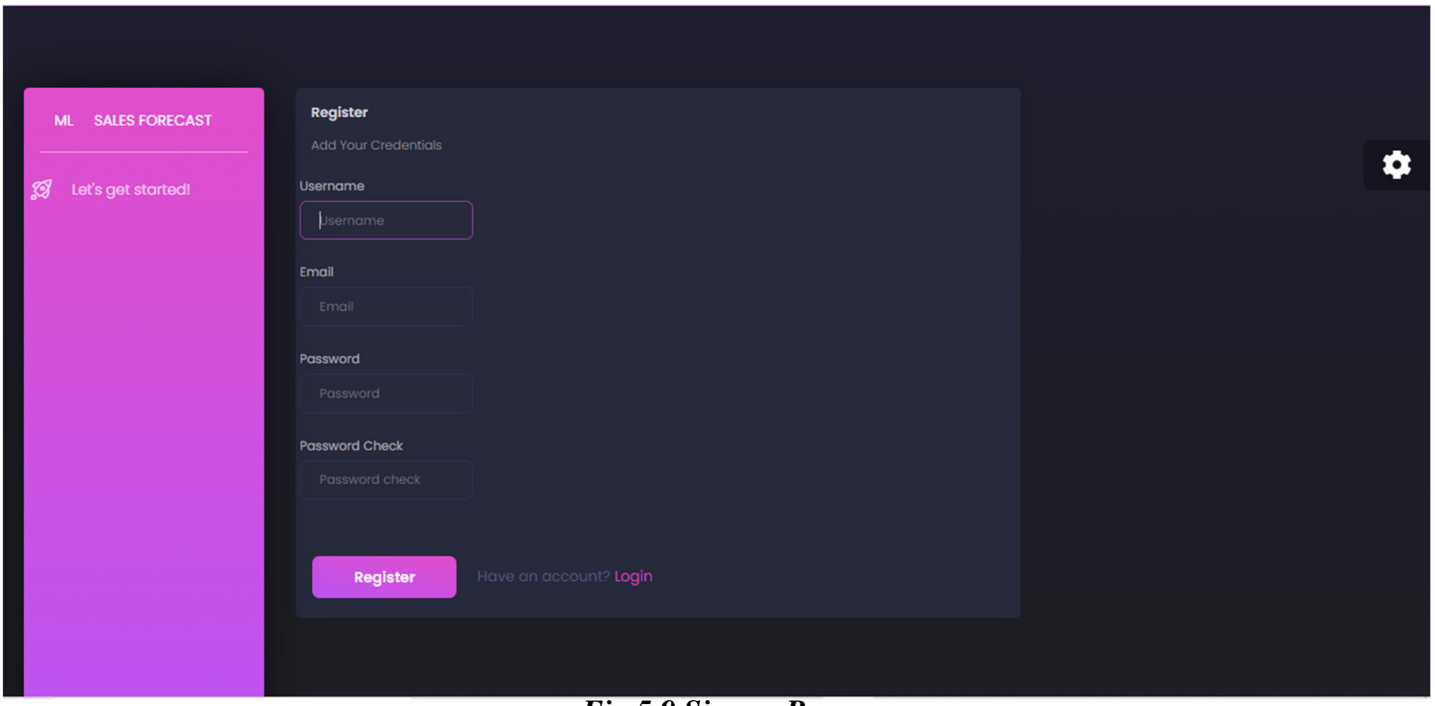

Fig 5.9 Signup Page

Fig 5.10 Dashboard

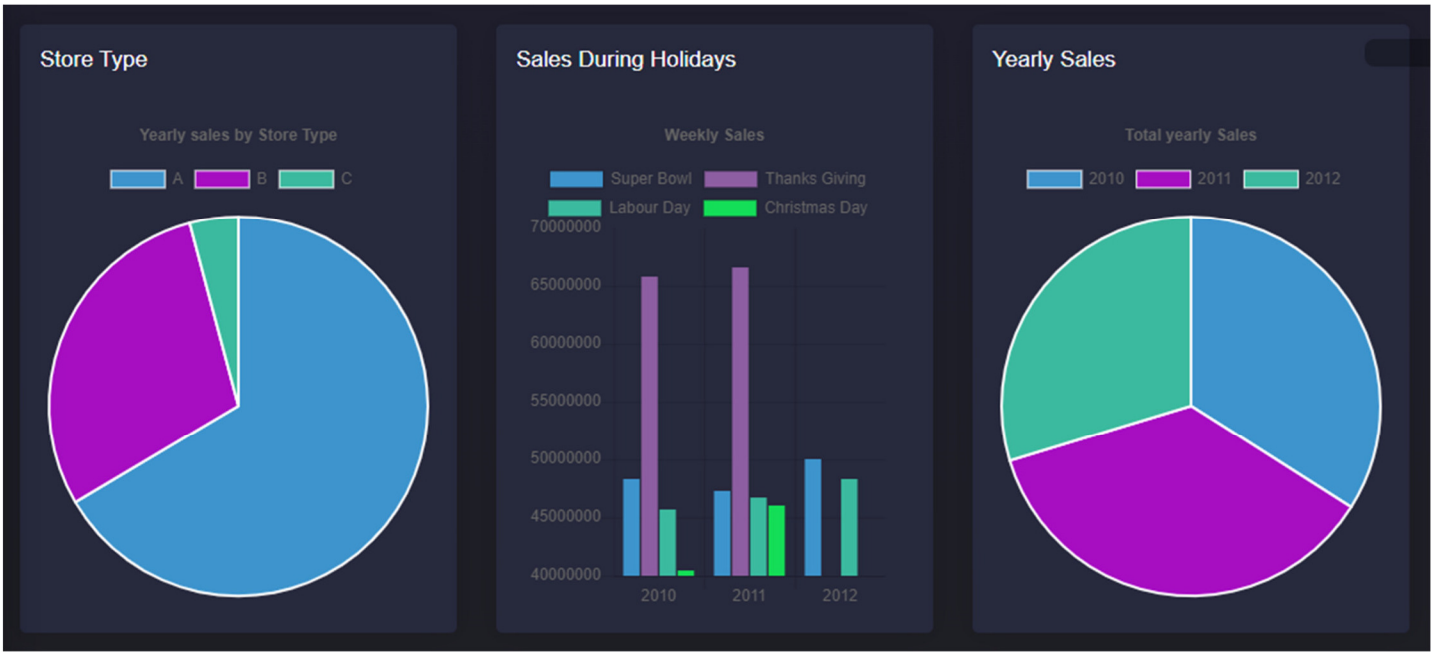

Fig 5.11 Data Visualization

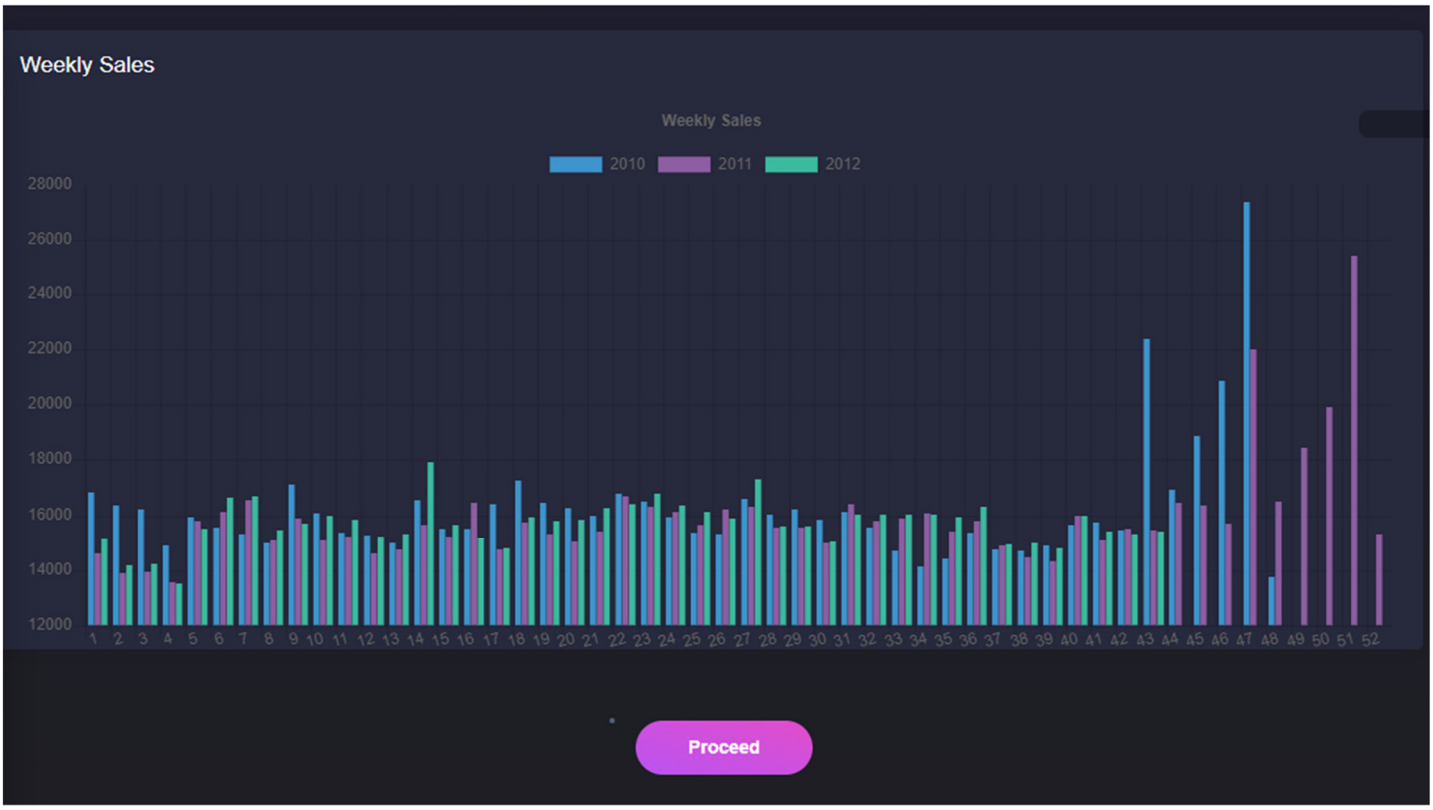

Fig 5.12 Weekly Sales Representation

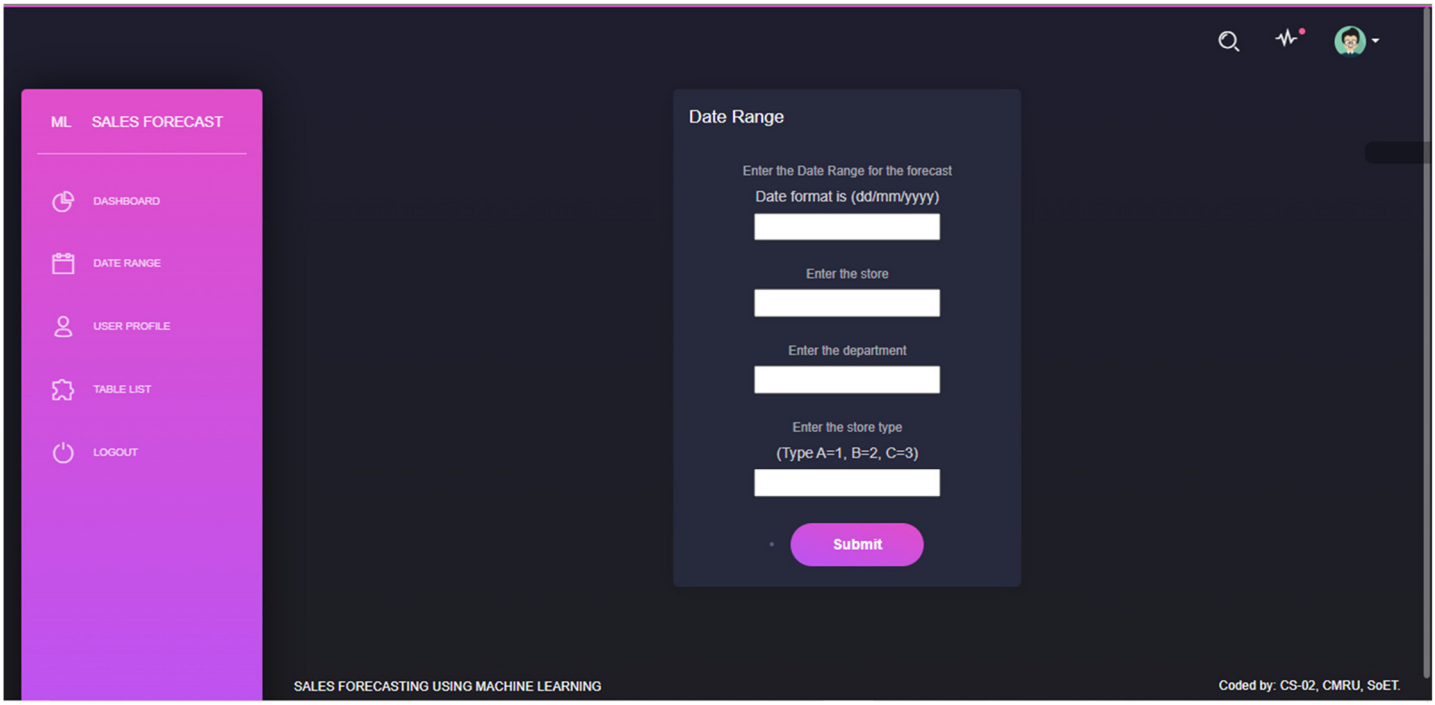

Fig 5.13 Date range to get user input

Fig 5.14 Results from 3 Models

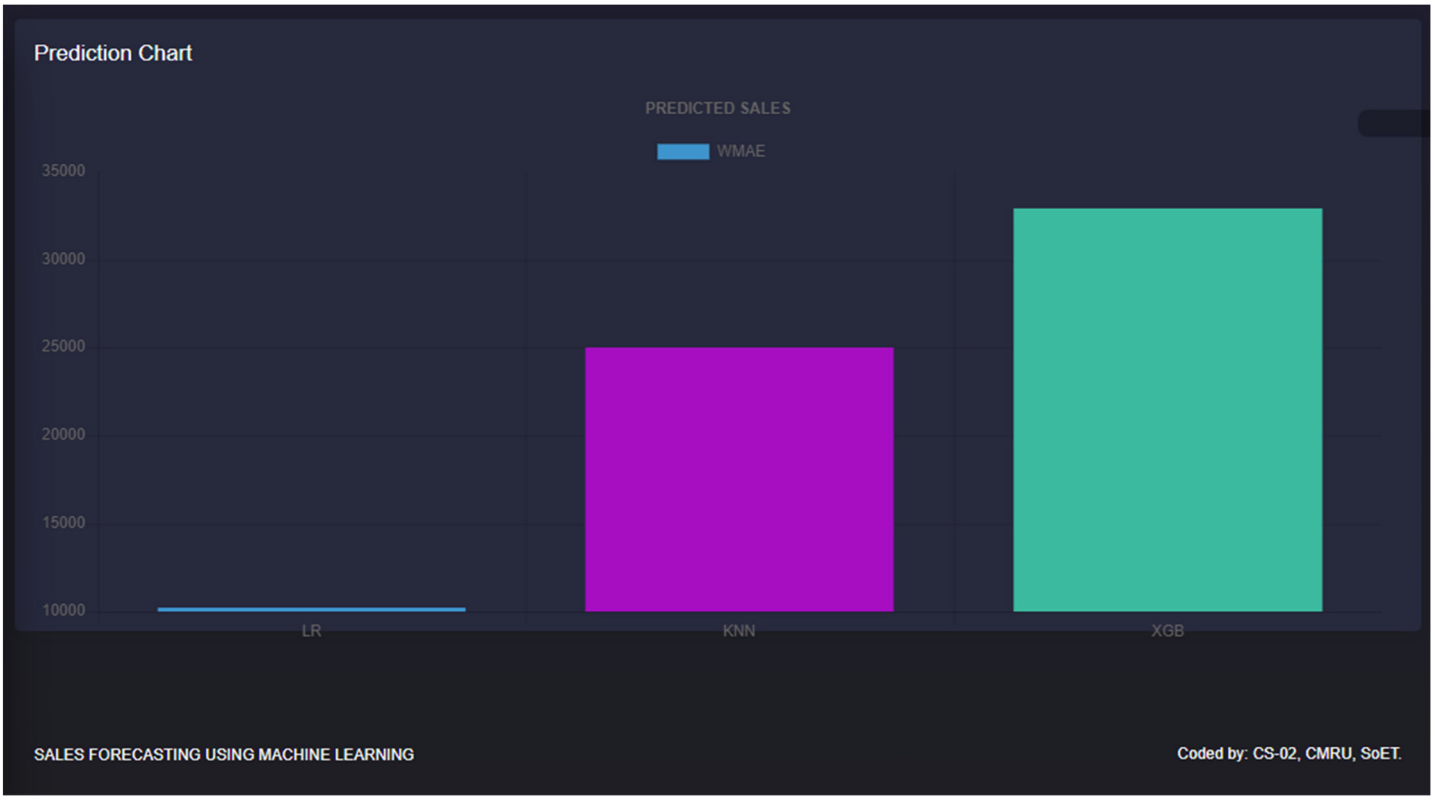

Fig 5.15 Prediction chart

Store	Date	Weekly_Sales	IsHoliday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	05-02-2010	1643690.9	0	42.31	2.572	211.0963582	8.106
1	12-02-2010	1641957.44	1	38.51	2.548	211.2421698	8.106
1	19-02-2010	1611968.17	0	39.93	2.514	211.2891429	8.106
1	26-02-2010	1409727.59	0	46.63	2.561	211.3196429	8.106
1	05-03-2010	1554806.68	0	46.5	2.625	211.3501429	8.106

Fig 5.16 Sample Dataset - Features

Stores CSV File

Store	Type	Size
1	A	151315
2	A	202307
3	B	37392
4	A	205863
5	B	34875

Fig 5.17 Sample Dataset – Stores

Train CSV File

Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	2010-02-05	24924.5	FALSE
1	1	2010-02-12	46039.49	TRUE
1	1	2010-02-19	41595.55	FALSE
1	1	2010-02-26	19403.54	FALSE
1	1	2010-03-05	21827.9	FALSE

Fig 5.18 Train Dataset

Fig 5.19 Test Dataset

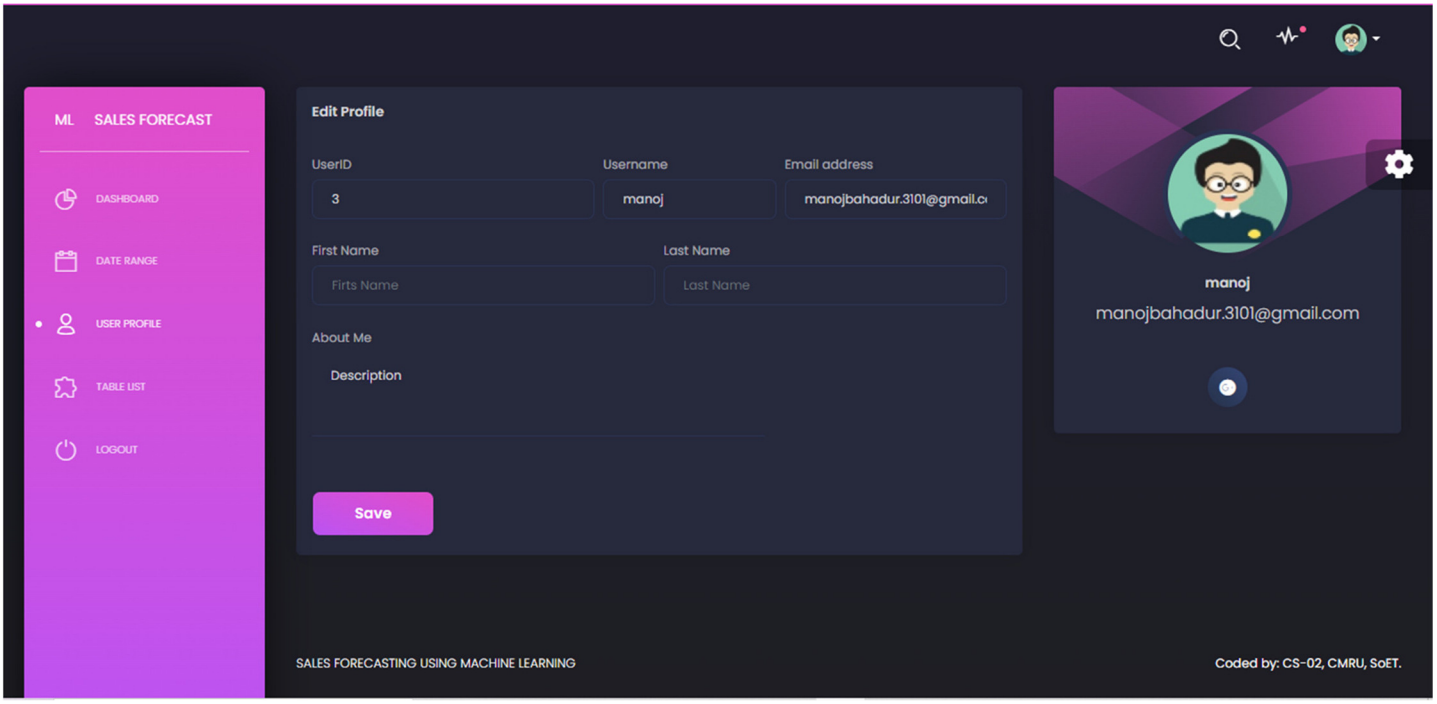

Fig 5.20 Profile Page

Chapter 6

CONCLUSION

Sales forecasting is a pivotal part of the financial planning of business for any organization. It can be said as a self-assessment tool which uses the statistics of the past and the current sales in order to predict future performance. Sales forecasting plays an important role in optimizing the sales process.

Financial and Sales planning with the help of the sales forecasts helps to get the information needed to predict the revenue as well as the profit. Thus, in finding such solution for sales forecasts, machine learning algorithms such as Linear regression, K-Nearest Neighbor (KNN) and XGBOOST Regressor have been evaluated on the Walmart sales data which can forecast the short term sales and help the organization in making the key decisions.

The web application developed for Sale Forecasting, processes the data and provides insights through visualization like bar graphs and pie charts. The weekly sales are predicted using the data from the past and the three models produce the predicted sales as the output. The output is ranked according to its accuracy.

REFERENCES

[1] https://medium.com/analytics-vidhya/walmart-recruiting-store-salesforecasting-kaggle-competition-856c72c9265a

[2] https://www.salesforce.com/resources/articles/building-a-sales-forecast-guide/

[3] https://medium.com/geekculture/walmart-recruiting-store-sales-forecasting-b8b2f4cf19b1

[4] http://www.pnas.org/content/117/30/17656

[5] http://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data

[6] https://ijret.org/volumes/2015v04/i06/IJRET20150406008.pdf

[7] https://www.w3schools.com/

[8] https://stackoverflow.com/

Capstone Build: Monthly Progress Report

Date :	**26-02-2022**
Month:	**FEBRUARY**

Student ID	18BBTCS007	18BBTCS045	18BBTCS063	18BBTCS031

Project Title	**Sales Forecast using Machine Learning.**

Work completed this month	This month is mainly focused on various front-end related UI and functionality such as – landing page, user account creation, user account authentication, along with importing the dataset to the project and showing a sample chart with dummy data			
Deliverables submitted this month	1. Front-end sample UI Design 2. Landing Page 3. Sample Charting with Dummy Data 4. User creation – logic + design 5. User authentication – logic + design			
	18BBTCS007	**18BBTCS045**	**18BBTCS063**	**18BBTCS031**
Individual contribution for the month	Sample Charting with Dummy Data, Importing Dataset, Project Integration	Landing Page creation, front-end CSS file, Design Generation	Defining method and scope for user creation as well as authentication	-
Plan for next month	Data Cleaning for Visualising on Charts appropriately, "Analyse" button to take 'date' input from user and carry out backend business logic, Development of backend Python Scripts, and overall Integration + Testing of Project Modules.			
Progress since last report	-			
Comment by Guide				
Signature of Students with Date	29/08/21	29/08/21	29/08/21	-
Signature of Guide with Date				

Capstone Build: Monthly Progress Report

Date :	**11-04-2022**
Month:	**APRIL**

Student ID	18BBTCS007	18BBTCS045	18BBTCS063	18BBTCS031

Project Title	**Sales Forecast using Machine Learning.**

Work completed this month	This month is mainly focused on the data aspect of the project. This involves data wrangling, data cleaning, data visualisation, exploratory data analysis, as well training the various Machine Learning models as part of the business logic. Moreover, we have also implemented charting with appropriate data			
Deliverables submitted this month	1. Front-end new page with user input and 'analyse' button 2. Notebook with data visualisation 3. Implemented charting data from notebook 4. Sample code and ML model training in notebook 5. Diagrams for application cycle and application workflow			
	18BBTCS007	**18BBTCS045**	**18BBTCS063**	**18BBTCS031**
Individual contribution for the month	Carried out code sampling, exploratory data analysis, and trained ML models.	Created application workflow and application cycle along with documentation and active contribution towards other tasks	Carried out data visualisation and implemented charting with appropriate data set	-
Plan for next month	Implement backend code from notebook. Test the project.			
Progress since last report	Significant progress in terms of the data aspect which includes visualisation and charting, as well as, code sampling, and fitting of dataset to various Machine Learning models			
Comment by Guide				
Signature of Students with Date	11/04/22	11/04/22	11/04/22	-
Signature of Guide with Date				

Capstone Build: Monthly Progress Report

Date :	**02-05-2022**
Month:	**MAY**

Student ID	18BBTCS007	18BBTCS045	18BBTCS063	18BBTCS031

Project Title	**Sales Forecast using Machine Learning.**

Work completed this month	This month is mainly focused on the testing and finalising of the project. Faced an array of issues, errors, and exceptions with respect to the coding environment, which have now been fixed. Faced a massive issue with form and data passing whilst mapping URLs which have also been fixed			
Deliverables submitted this month	1. Front-end page integration with URL mapping 2. Fixed errors and exported entire backend code and business logic 3. Testing entire project and finalising all pages and data flow			
	18BBTCS007	**18BBTCS045**	**18BBTCS063**	**18BBTCS031**
Individual contribution for the month	Fixed errors and exported entire backend code and business logic	Testing entire project and finalising all pages and data flow	Front-end page integration with URL mapping	-
Plan for next month	—			
Progress since last report	Significant progress in terms of the final backend business logic implementation from Jupyter Notebook. Moreover, successfully made required changes for form submission and mapped with internal page URLs. Fixed many errors with respect to code logic, and data passing across different pages.			
Comment by Guide				
Signature of Students with Date	02/05/22	02/05/22	02/05/22	-
Signature of Guide with Date				