

I. PROBLEM STATEMENT

There are several big supply chain supermarkets like Walmart, Home Depot, Best Buy and so on, who must maintain an inventory of stock for smooth operation. However, in order to ensure that the inventory is not under-stocked nor overstocked, it is crucial to gain a precise estimate of the sales trends and purchasing needs. This project, *Sales Predictor: Forecasting Sales using Python*, aims to estimate the increase and decrease of sales for several items to predict the sales of the supermarkets in future.

II. PROBLEM ANALYSIS

The problem is broken down into the following phases:

1. **Data collection:** The required dataset for the project is obtained from Kaggle [1]. It is Walmart Weekly Sales dataset, which captures Walmart Sales across different stores in the years 2010, 2011 & 2012.
2. **Data preparation:** The obtained dataset has some missing features and feature engineering needs to be done on some features (date, IsHoliday).
3. **Writing scripts for visualization:** In order to visualize the data in the form of bar graphs, line graphs and scatter plots, scripts are written in python, using standard python libraries.
4. **Writing scripts for prediction:** In order to make predictions, several machine learning models are implemented, and scripts are written to capture the predictions in the form of scatter plots.
5. **Writing results to output file:** The prediction results obtained are written to an output text file, so that it can be accessed later.
6. **Integrating python with php:** An individual, not having knowledge about python, must be able to see visualization and prediction results. For this, the results are hosted over local server, python is integrated with php.
7. **Developing the web interface:** A user interface is developed to group the results in different categories to make these easily accessible and to improve readability.

III. SPECIFICATIONS

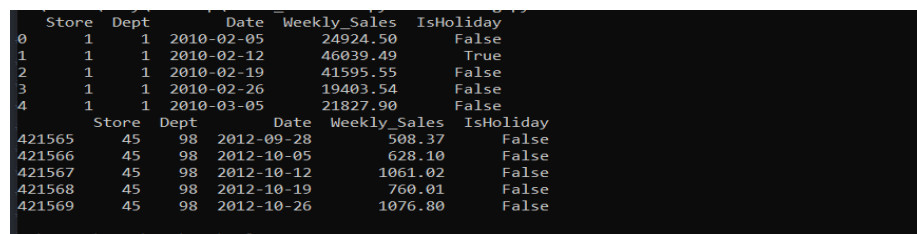
The following specifications are used for project development.

1. DEVELOPMENT ENVIRONMENT

- i. Programming Languages: Python 3.8, php 7.2
- ii. Scripting Languages: HTML5, JavaScript, CSS3
- iii. Algorithms/Models: Linear Regression, K-Nearest Neighbors, Ridge Regression, Random Forest, Decision Trees, XGBRegressor
- iv. Dataset: Walmart Store Sales Forecasting
- v. Online Resources: Kaggle, Github
- vi. Localhost: WAMP Server

2. INPUT SPECIFICATIONS

The input to the project is a Walmart sales forecasting dataset obtained from Kaggle [1]. The dataset is a comma delimited csv file, with 5 columns and 421569 columns with the features as indicated from figure below:



	Store	Dept	Date	Weekly_Sales	IsHoliday
0	1	1	2010-02-05	24924.50	False
1	1	1	2010-02-12	46039.49	True
2	1	1	2010-02-19	41595.55	False
3	1	1	2010-02-26	19403.54	False
4	1	1	2010-03-05	21827.90	False
	Store	Dept	Date	Weekly_Sales	IsHoliday
421565	45	98	2012-09-28	508.37	False
421566	45	98	2012-10-05	628.10	False
421567	45	98	2012-10-12	1061.02	False
421568	45	98	2012-10-19	760.01	False
421569	45	98	2012-10-26	1076.80	False

Figure 1 Walmart Sales Forecasting data

After this, feature engineering is performed on date and IsHoliday feature to obtain a dataset to be used in the project. The below figure shows the details of the prepared dataset.



	Store	Dept	Date	Weekly_Sales	IsHoliday	year	month	week_day	IsHolidayInt
0	1	1	2010-02-05	24924.50	False	2010	2	4	0
1	1	1	2010-02-12	46039.49	True	2010	2	4	1
2	1	1	2010-02-19	41595.55	False	2010	2	4	0
3	1	1	2010-02-26	19403.54	False	2010	2	4	0
4	1	1	2010-03-05	21827.90	False	2010	3	4	0
	Store	Dept	Date	Weekly_Sales	IsHoliday	year	month	week_day	IsHolidayInt
421565	45	98	2012-09-28	508.37	False	2012	9	4	0
421566	45	98	2012-10-05	628.10	False	2012	10	4	0
421567	45	98	2012-10-12	1061.02	False	2012	10	4	0
421568	45	98	2012-10-19	760.01	False	2012	10	4	0
421569	45	98	2012-10-26	1076.80	False	2012	10	4	0

Figure 2 Prepared dataset after feature engineering

3. OUTPUT SPECIFICATIONS

The program outputs data visualization and prediction results in the form of .png images of scatter plots, bar graphs and line graphs, and saves these in a separate folder. It also writes the results of the prediction to a .csv file, which can be accessed any time later.

4. PROGRAM SPECIFICATIONS

1. The program is written in two separate python scripts: one for visualization and other for prediction. The program executes and generates the results in the form of .png images and a .csv file.
2. A web interface is designed to make the visualization process independent of the programming knowledge of the user.
3. Python scripts are integrated with php. The python programs are called from within the web interface when the user clicks a button.

IV. DESIGN

The overall design of the project is described in the following flowchart:

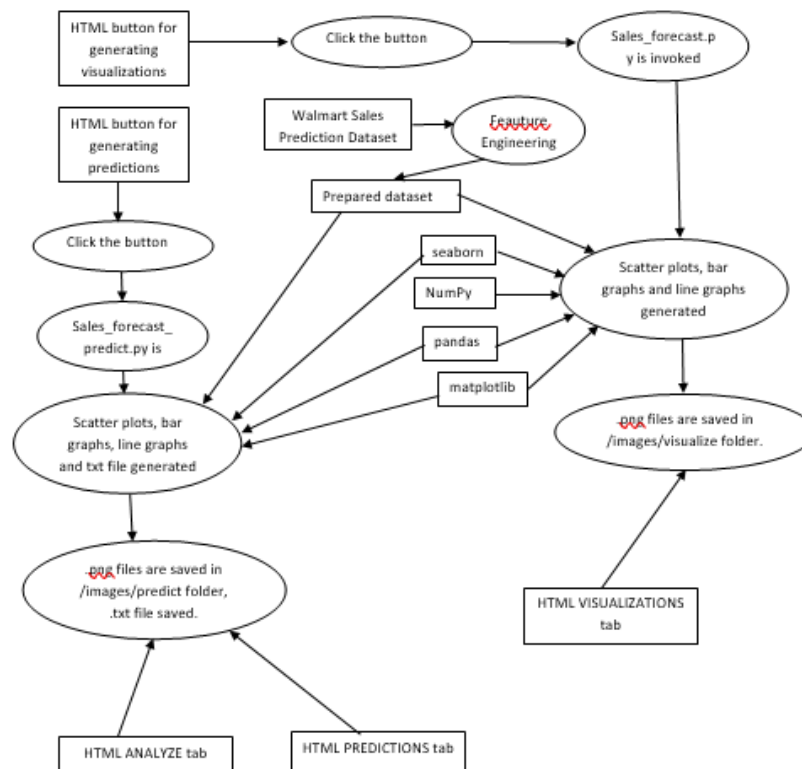


Figure 3 Flowchart of Project

V. TESTING

The website has 5 tabs. The screenshots and utility of each tab is described below:

1. HOME TAB:



Figure 4 Home tab

This tab gives out the overall purpose of the project and describes the various features provided by the website.

2. VISUALIZE TAB

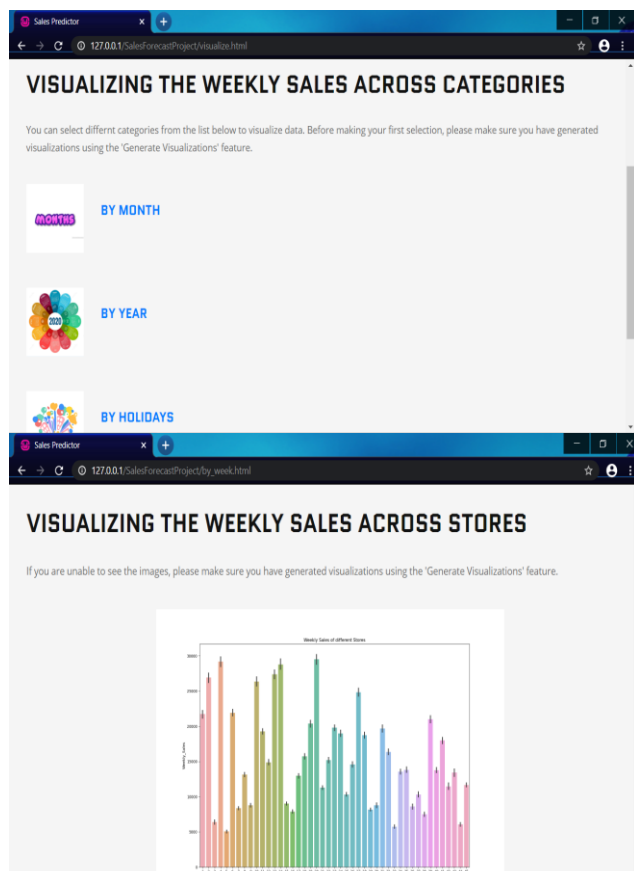


Figure 5 Visualize tab

3. PREDICT TAB

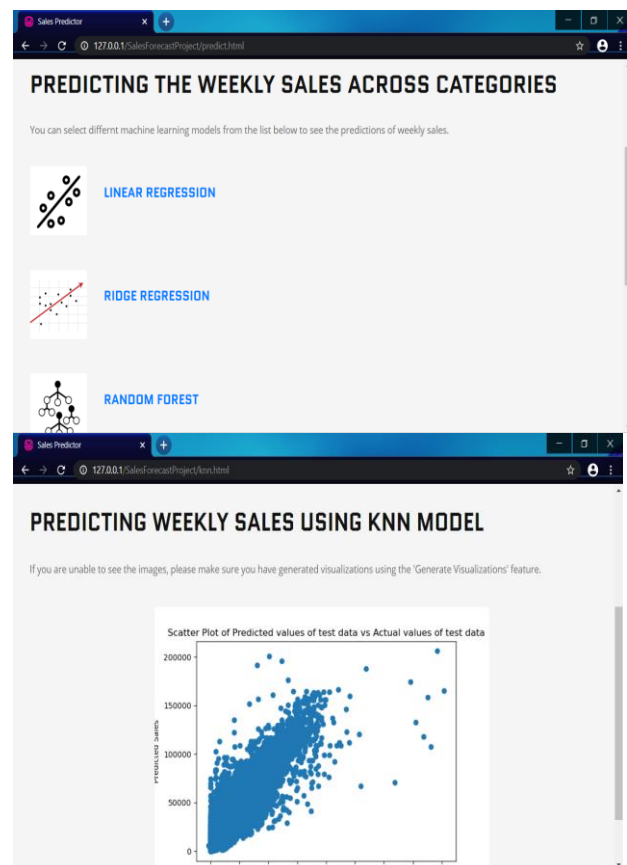


Figure 6 Predict tab

This tab shows the various options for This tab shows the various ML models used

visualization; a user can group by holiday, by week, by month and by year. for prediction. The user can select any model to see the scatter plot.

4. GENERATE VISUALIZATIONS TAB

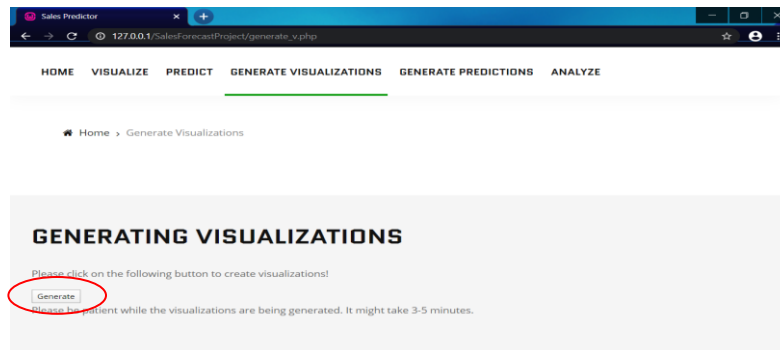


Figure 7 Generate Visualizations tab

There is a 'GENERATE' button on this tab. When the user clicks on the button, a python script is executed, and the .png images generated are saved in /images/visualize folder.

5. GENERATE PREDICTIONS TAB

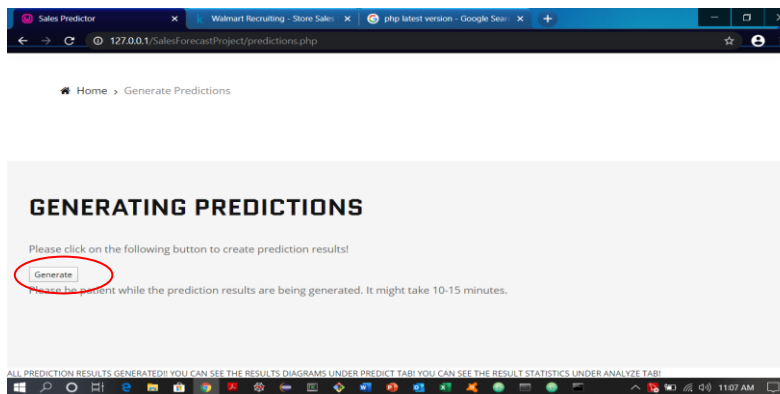


Figure 8 Generate predictions tab

There is a 'GENERATE' button on this tab. When the user clicks on the button, a python script is executed, and the .png images generated are saved in /images/predict folder.

6. ANALYZE TAB

This tab has a hyperlink to the results.txt file. When the user clicks on the 'Results File', it displays the contents of the results.txt file in a new tab.

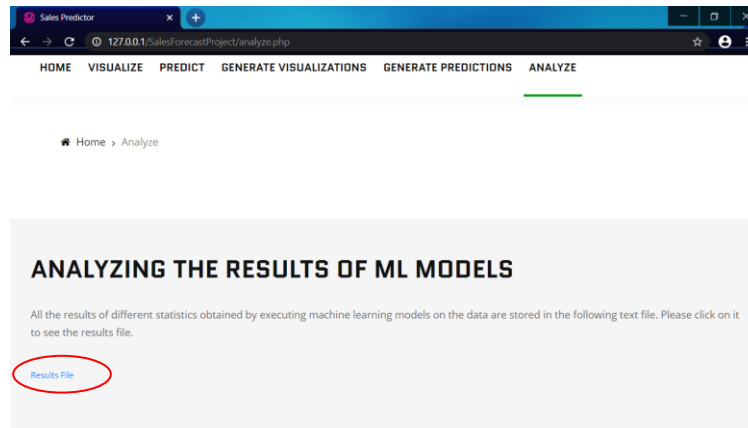


Figure 9 Analyze tab

VI. IMPACT

The project successfully applies ML algorithms to the Sales dataset to generate predictions of future sales. Different models are used, and XGBRegressor predicts the results with the highest accuracy of 96.2%. The web interface is implemented to enable users with no programming knowledge to see the results of visualizations and predictions.

The same program can be implemented using any dataset having the same attributes (Store (int), Dept (int), Date (datetime), IsHoliday (Boolean), Weekly_Sales (float)). Therefore, by changing the dataset, this website can be used by many organizations to make predictions about their future sales.

REFERENCES

[1] Walmart Recruiting- Store Sales Forecasting, Kaggle. Extracted from- <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data>