Walmart Sales Forecasting

Aditya Sharma, Devank Garg, Siddharth Pratap Singh, Panshul

# INTRODUCTION

**Introduction**

ML is a type of AI which allows software’s to predict new results more accurately and easily. ML algorithms use training(historical) data to predict new data(output). Our model helps us to predict future sales. Sales Forecasting is defined as the system through which values of future sales are estimated. In layman terms, which product is selling how much in certain future time period. It encourages and helps the making of well-informed business decisions by forecasting short-term and long-term results. Forecasts might be based on historical sales data, industry comparisons, and economic trends.

Walmart Inc. is an American multinational retail corporation that operates a chain of hypermarkets, discount department stores, and grocery stores from the United States, headquartered in Bentonville, Arkansas. [0,1]

Walmart provides datasets for 98 products across 45 outlets so developers can access information on weekly sales by locations and departments.

The goal with a project of this scope is to make better data-driven decisions in channel optimization and inventory planning.

The decision makers of Walmart should be able to analyse the effect of various factors effecting the sales. The dataset for this project has been taken from Kaggle “Data Analytics Competition”. It contains data of 45 Walmart stores based on location. The land area and type of each store have been mentioned, as well as the holiday weeks. Price markdown data (nearly like discount data) has been provided in addition to these. Our main goal is to measure predict the stores data and Walmart sales data for all the weeks in a year.

We wanted to look at how internal and external influences at one of the country's largest corporations could affect Weekly Number of sales. [1,1]

## Abbreviations and Acronyms

“ML” means machine learning

“AI” means artificial intelligence

## Proposed Methodology

Data Exploration

Exploratory Data Analysis

Data Pre-Processing

Data Manipulation

Feature Extraction

Project Outcomes and Conclusion

*Proposed Algorithm*

Apply the pre-processing steps on the interested datasets

importing necessary libraries for training the dataset and prediction

a) Remove unwanted features if any.

b) Handle Null value or enormous values if any.

c) Use one hot encoding (OHE) or multi hot encoding to labelling, if any.

d) Convert categorical value to numerical value if any.

e) Standardize the data into [0,2].

Step 3: Split the improved Behaviour of the urban traffic of the city of Sao Paulo in Brazil

Data Set after Step 2 into training and testing in the following ratio:

a) 70:30

b) 30:70

Step 4: Train and Test the different regression model using sklearn library function

a) Linear regression

b) Elastic Net

c) Lasso Regression

d) Ridge Regression

e) Random Forest Regressor

Step 5: Evaluate the performance of different regression models on various performance

metrics:

a) Confusion Metric

b) Mean Absolute Error (MAE)

c) Mean squared error (MSE)

d) Root mean squared error (RMSE)

e) R Square / Coefficient of Determination/ Goodness of fit

Step 6: Visualize the results of different regression models into different plots:

School of Computer Science Engineering and Technology

a) regplot ().

b) FacetGrid using lmplot () [2,1]

# MATH

If you are using *Word,* use either the Microsoft Equation Editor or the *MathType* add-on (http://www.mathtype.com) for equations in your paper (Insert | Object | Create New | Microsoft Equation *or* Math Type Equation). “Float over text” should *not* be selected.

## Equations

#### Mean Absolute Error: - MAE = (1/n) \* Σ|yi – xi|

#### Mean squared Error: - E[(X−x^)2|Y=y] =E[(X−g(y))2|Y=y].

#### Root mean squared error: - RMSE = RMSE = √ [ Σ(Pi – Oi)2 / n ]

#### Co-efficient of Determination: - Using Correlation Coefficient = Σ [(X – Xm) \* (Y – Ym)] / √ [Σ (X – Xm)2 \* Σ (Y – Ym)2]

**Dataset Description**

You will be given historical sales data for 45 Walmart shops in various geographies. You are responsible with projecting department-wide sales for each store, which includes a variety of departments.

Walmart.csv

One of the main retail locations in the US, Walmart, might want to precisely anticipate the deals and request. There are sure occasions and occasions which sway deals on every day. There are deals information accessible for 45 stores of Walmart. The business is confronting a test because of unexpected requests and runs unavailable a few times, because of the improper AI calculation. An ideal ML calculation will anticipate request precisely and ingest factors like financial circumstances including CPI, Unemployment Index, and so forth.

This is the chronicled information that covers deals from 2010-02-05 to 2012-11-01, in the record WalmartStoresales. Inside this record you will find the accompanying fields:

Store - the store number

Date - the seven days stretch of deals

Weekly\_Sales - deals for the given store

Holiday\_Flag - whether the week is an exceptional occasion week 1 - Holiday week 0 - non-occasion week

Temperature - Temperature upon the arrival of offer

Fuel\_Price - Cost of fuel in the district

CPI - Prevailing purchaser cost file

Joblessness - Prevailing joblessness rate

Occasion Events

Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13

Workday: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13

Thanksgiving: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13

Christmas: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

Be aware of the different meanings of the homophones “affect” (usually a verb) and “effect” (usually a noun), “complement” and “compliment,” “discreet” and “discrete,” “principal” (e.g., “principal investigator”) and “principle” (e.g., “principle of measurement”). Do not confuse “imply” and “infer.”

Prefixes such as “non,” “sub,” “micro,” “multi,” and “ultra” are not independent words; they should be joined to the words they modify, usually without a hyphen. There is no period after

.

**Graphs**

Chart, bar chart

Description automatically generated

Fig1:-Count plot (Annual Sales)

Chart, bar chart

Description automatically generated

Fig2: - Count plot 2 (Monthly Sales )

A picture containing bar chart

Description automatically generated

Fig3: - Count plot 3 (Daily Sales)

Chart, pie chart

Description automatically generated

Fig 4: - Pie Chart 1 (Annual Sales)

Chart, pie chart

Description automatically generated

Fig 5: - Pie chart 2 (Monthly Sales)

Chart, pie chart

Description automatically generated

Fig6: - Pie chart 3 (Daily Sales)

**Flow Charts**

Diagram

Description automatically generated

Fig 7 :- Flow Diagram

Diagram

Description automatically generated

Fig 8 :- Forecasting Process

# **Conclusion**

## A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Table: -1

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Max\_sales | Min\_sales |
| 1 | Yearly | **2011** | **2012** |
| 2 | Monthly | **April** | **September** |
| 3 | Daily | **Thursday** | **Tuesday** |

Table: -2

|  |  |  |  |
| --- | --- | --- | --- |
| S. NO | Types of Model | Regression Matrices | Values |
| 1. | Linear Regression | Mean Absolute Error  Mean Squared Error  R2- Score | 0.247  0.929  0.922 |
| 2. | Lasso | Mean Absolute Error  Mean Squared Error  R2- Score | 0.246  0.925  0.922 |
| 3. | Ridge | Mean Absolute Error  Mean Squared Error  R2- Score | 0.249  0.932  0.921 |
| 4. | RandomForestRegressor | Mean Absolute Error  Mean Squared Error  R2- Score | 0.233  0.766  0.926 |
| 5. | Elastic Net | Mean Absolute Error  Mean Squared Error  R2- Score | 0.291  0.456  0.085 |

References

[0,1]: Y. Niu, "Walmart Sales Forecasting using XGBoost algorithm and Feature engineering," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2020, pp. 458-461, doi: 10.1109/ICBASE51474.2020.00103.

[1,1]: M YASSER H : [Walmart Dataset | Kaggle](https://www.kaggle.com/datasets/yasserh/walmart-dataset)

[2,1] : M YASSER H: [Walmart Sales Prediction - (Best ML Algorithms) | Kaggle](https://www.kaggle.com/code/yasserh/walmart-sales-prediction-best-ml-algorithms)

[4,2] : A Service-Oriented Solution for Retail Store Network

Planning. Xinxin Bai,Wei Shang,Wenjun Yin, Jin Dong.

[2]. A sales forecasting model for consumer products based

on the influence of online word-of-Mouth.Ching-Chin

Chern,Chih-Ping Wei,Fang-Yi Shen,Yu-Neng Fan.

[3]. Using clustering to improve sales forecasts in retail

merchandising. Mahesh Kumar · Nitin R. Patel.

[4]. Simulation Based Sales Forecasting On Retail stores

LinkedIn Post: -

<https://www.linkedin.com/feed/update/urn:li:ugcPost:6925286006912925696/?updateEntityUrn=urn%3Ali%3Afs_updateV2%3A%28urn%3Ali%3AugcPost%3A6925286006912925696%2CFEED_DETAIL%2CEMPTY%2CDEFAULT%2Cfalse%29>