INVENTORY FORECAST DEMAND

Team Members :

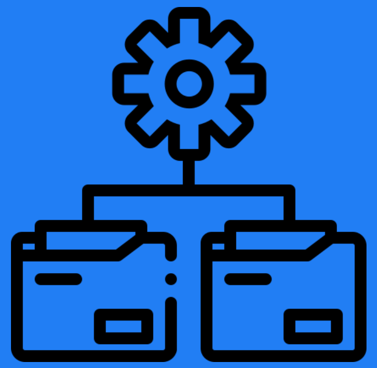
1. Renu B -1RVU22CSE129
2. Karmishtha P-1RVU22CSE077
3. Radhika P-1RVU22CSE126

Overview

Outline

Data preparation

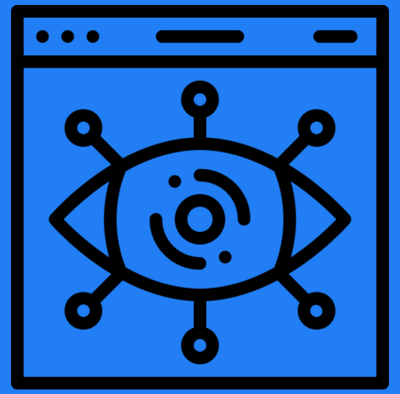
Data Splitting



Model Development

Model Evaluation

Visualisation



Problem Statement

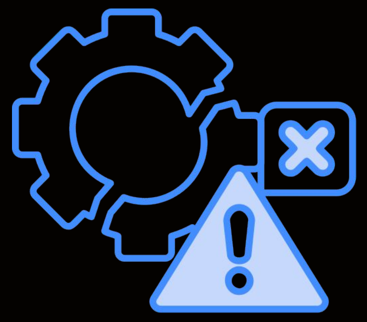
Efficient inventory management is essential for businesses to meet customer demand effectively while minimizing costs. However, accurately predicting inventory needs presents a significant challenge

Stockouts

Running out of products leads to lost sales & unhappy customers.

Obsolete Inventory

Products becoming outdated result in financial losses



Supply Chain Disruptions

Unexpected events like delays or transportation issues disrupt inventory flow, affecting customer satisfaction

Excess Inventory

Holding too much stock ties up money and space, increasing costs.

Inaccurate Forecasts

Poor predictions lead to inefficient ordering and mismatched inventory levels.

Introduction

Welcome to the Retail Sales Prediction presentation. Today, the focus is on analyzing sales data from 2010 to 2012 to forecast demand for Walmart, a major U.S. retail chain. This presentation will cover

Project Overview Understanding the goals.

Modeling Techniques Predicting sales methods.

Data Exploration & Preparation

Best Models and Analysis Evaluating model performance.

Methods & Models

Mount Drive

Import Libraries

Load Dataset

XLS

Print First 5 Rows

EDA

Patterns+anomalies=Hypothesis

Data Pre-Processing

1. Handling Missing values
2. Feature Engineering

Data Visualization

Inference:

1. Best season - summer
2. TOTAL SALES IN EACH YR - 2011
3. TOTAL SALES IN EACH MONTH - july
4. TOTAL SALES IN EACH WEEK - week 51
5. HEAT MAP

Methods & Models

After Data Visualization

Type Casting

Better Analysis

1. Copy code to df1
2. Numerical data to categorical (Explicit)
3. Store, holiday flag & week converted

Data Transformation

Data splitting

1. Numerical Features : ['Temperature', 'Fuel_Price
   'CPI', 'Unemployment']
2. Categorical Features:['Store','Holiday_Flag', 'wee

Methods & Models

After Detecting And Removing Outliers

Detecting & Removing Outliers

Hyperparameter Tuning(GRID SEARCH + RANDOMISED SEARCH)

Cross Validation

1. Z-Score
2. Total outliers: 675 when threshold = 3
3. 3 is standard deviation from mean is a common approach holding about 99.7% of data under a normal distribution

1. Small dataset
2. Few hyperparameters
3. Grid search > Random search for accuracy

Methods & Models

Supervised Learning

Linear Regression	Polynomial Regression	Ridge Regression	Lasso Regression
Global estimator	Non-linear Relationship	Coefficient estimation	Predictive Accuracy
Decision Tree	Random Forest	KNN Regressor	XGB Regressor
Individual predictions	Predictive accuracy	Local estimation	Gradient Boosting

Methods & Models

Unsupervised Learning

K Means Clustering

Grouping unlabeled datasets

Hierachical Clustering

Organizing groups based on similarities

Deep Learning

Multi-Layer Perceptron

Interconnected neural network

Evaluation Metrics

Supervised & Deep

1. Mean Absolute Error (MAE)
2. Mean Squared Error (MSE)
3. Root Mean Squared Error (RMSE)
4. R Squared (R2)

Unsupervised

1. Silhouette score
2. index

Visualization

Distribution Plot

Dataset Details

- Well-documented
- Learning purpose
- covers sales from 2010-02-05 to 2012-11-01
- In the dataset,
  - Store - the store number
  - Date - the week of sales
  - Weekly_Sales - sales for the given store
  - Holiday_Flag - whether the week is a special holiday week 1 – Holiday week 0 – Non-holiday week
  - Temperature - Temperature on the day of sale
  - Fuel_Price - Cost of fuel in the region
  - CPI – Prevailing consumer price index
  - Unemployment - Prevailing unemployment rate
  - Holiday Events
    - Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
    - Labour Day: 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
- One of the leading retail stores in the US, Walmart, would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. There are sales

Experimental Results

	MODEL NAME	DATA	MAE	MSE	RMSE	R2
2	LINEAR REGRESSION	TRAINING	358406.76	204470961002.30	452184.65	31.08 %
3		TESTING	362471.07	210920722575.97	459261.06	31.86 %
4						
5	POLY REGRESSION	TRAINING	58374.79	6886879714.26	82987.23	97.68 %
6		TESTING	69004.3	9318364369.55	96531.68	96.99 %
7						
8	RIDGE REGRESSION GRID SEARCH	TRAINING	358542.43	204480782446.65	452195.51	31.07 %
9		TESTING	362730.83	210893777737.26	459231.73	31.87 %
10						
11	RIDGE REGRESSION RANDOM SEARCH	TRAINING	358484.39	204474451017.80	452188.51	31.08 %
12		TESTING	362622.14	210901978470.13	459240.65	31.86 %
13						
14	LASSO REGRESSION GRID SEARCH	TRAINING	358407.68	204470966370.87	452184.66	31.08 %
15		TESTING	362472.09	210919193605.71	459259.4	31.86 %
16						
17	LASSO REGRESSION RANDOM SEARCH	TRAINING	358407.67	204470966241.68	452184.66	31.08 %
18		TESTING	362472.07	210919212014.84	459259.42	31.86 %
19						
20	DECISION TREE GRID SEARCH	TRAINING	27113.33	2256760434.69	47505.37	99.24 %
21		TESTING	59095.59	8772553461.59	93661.91	97.17 %
22						
23	DECISION TREE RANDOM SEARCH	TRAINING	27000.89	2244776329.23	47379.07	99.24 %
24		TESTING	58738.8	8753530390.98	93560.3	97.17 %

Experimental Results

	A	B	C	D	E	F
1	**MODEL NAME**	**DATA**	**MAE**	**MSE**	**RMSE**	**R2**
26	**RANDOM FOREST GRID SEARCH**	TRAINING	24055.94	1802225821.92	42452.63	99.39 %
27		TESTING	64607.3	13365236840.70	115608.12	95.68 %
28						
29	**RANDOM FOREST RANDOM SEARCH**	TRAINING	40365.92	5119646425.12	71551.7	98.27 %
30		TESTING	65684.06	13874073317.90	117788.26	95.52 %
31						
32		TRAINING	0.0	0.00	0.0	100.0 %
33	**KNN GRID SEARCH**	TESTING	214115.04	79202332295.33	281429.09	74.41 %
34						
35		TRAINING E	246614.87	101506755882.72	318601.25	65.78 %
36		TESTING E	256808.54	110053791468.37	331743.56	64.45 %
37						
38		TRAINING	0.0	0.00	0.0	100.0 %
39	**KNN RANDOM SEARCH**	TESTING	291807.32	133755273840.64	365725.68	56.79 %
40						
41		TRAINING E	346471.61	178306750577.20	422263.84	39.9 %
42		TESTING E	358868.73	192819516586.96	439112.19	37.71 %
43						
44	**XBG REGRESSOR GRID SEARCH**	TRAINING	65424.52	9102237939	95405.65	96.93 %
45		TESTING	75429.65	12495116665.71	111781.56	95.96 %
46						
47	**XBG REGRESSOR RANDOM SEARCH**	TRAINING	94695.14	18516254405.49	136074.44	93.76 %
48		TESTING	103000.97	22844342401.63	151143.45	92.62 %

Experimental Results

	A	B	C	D	E	F
1	**MODEL NAME**	**DATA**	**MAE**	**MSE**	**RMSE**	**R2**
49						
50						
51						
52		INERTIA	12026.7834735127			
53	**K MEANS CLUSTERING**	SILHOUSE SCORE	0.116711518227683			
54						
55		Cophenetic Correlation Coefficient	0.521065897267871			
56	**HIERACHICAL CLUSTERING**	SILHOUSE SCORE	0.159184013358915			
57						
58	**NEURAL NETWORKS - MULTI LAYER PERCEPTRON (MLP)**	TRAINING	391606.57	234613111059.68	484368.78	20.92 %
59		TESTING	395631.04	241326470551.12	491249.91	22.04 %
60						

Conclusion

- Best Supervised Model
  - Decision Tree with Random Search or Random Forest with Grid Search
    - Low MAE, MSE, RMSE
    - High R2 scores on training and testing data
- Best Unsupervised Model
  - Hierarchical clustering
    - Moderate cophenetic correlation
    - Higher silhouette score
- Deep Learning Model (MLP)
  - High errors
  - Low R2 values
- Model Analysis
  - Polynomial Regression, Random Forest, and XGB Regressor outperform others
  - Best scores
  - Generalization capability
  - Avoiding overfitting
  - Polynomial Regression achieves highest score

Future Works

- **Feature Engineering Refinement**

Further exploration and engineering of features to capture more nuanced patterns and relationships in the data

- **Ensemble Methods Exploration**

Investigate the effectiveness of ensemble methods such as stacking or blending to combine the strengths of multiple models and improve overall performance

- **Hyperparameter Fine-Tuning**

Conduct thorough hyperparameter tuning for the selected models to optimize performance further and achieve even better results

- **Incorporate External Data**

Explore the integration of additional external datasets, such as economic indicators or demographic information, to enhance the predictive capabilities of the models.

- **Deep Learning Architectures**

Experiment with more complex deep learning architectures or pre-trained models to leverage the potential of neural networks for improved forecasting accuracy

Thank You