Maharshi Karve Stree Shikshan Samstha's

Cummins College of Engineering for Women, Pune

(An autonomous Institute affiliated to Savitribai Phule Pune University)

Department of Computer Engineering

MKSSS's
Cummins
COLLEGE OF ENGINEERING
FOR WOMEN

23PCCE501L

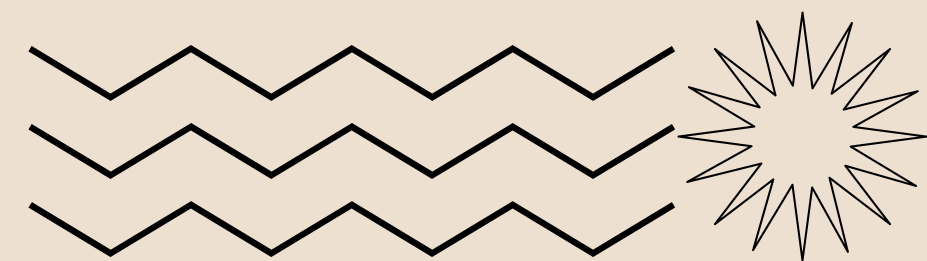**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING LABORATORY**

Smart Retail Analytics

Presented By:

UCE2023408: Anushka Chauhan

UCE2023462: Shruti Shinde

UCE2023474: Vishnupriya Lappasi

Introduction

Problem Statement

- Retailers struggle with stockouts, overstocks, and volatile demand.
- Traditional statistical forecasting is unable to handle non-linear demand, promotions, and seasonality.
- Retail managers lack SKU-level decision support and risk visibility.

Objective

- Build a machine-learning-based inventory optimization system.
- Forecast 30-day SKU demand accurately.
- Predict stockout risk levels (Low/Medium/High).
- Generate actionable reorder recommendations.

Purpose

- Reduce stockouts and overstock losses.
- Improve availability of products for customers.
- Assist retail store managers in data-driven decision-making.

Dataset

Dataset Sources

- M5 Walmart Sales dataset: daily sales per SKU.
- Calendar dataset: events, holidays, SNAP flag.
- Price dataset: sell price per store-SKU.

Dataset Creation Pipeline

- Filtered data for one store (CA_1).
- Merged sales + calendar + price using item_id, store_id, wm_yr_wk.
- Converted 1,941 daily sales columns (d_1...d_1941) into long format.
- Cleaned missing values, converted dates, sorted by item and date.

Final Prepared Dataset

- 40+ features per SKU-day.
- Contains: sales, lags, rolling stats, price, events, forecast_30, current_stock, risk_label.

Features of System

Demand Forecast

Most Critical Input.
Enables the Stockout Predictor
to **look ahead,** driving proactive
ROP and risk alerts.

Stockout risk predictor

Classifies risk
(Low/Medium/High) using
XGBoost and **forecasts** to flag
inventory failure.

Overstock detector

Flags excess inventory by
comparing current stock
against 30-day demand to
reduce holding costs.

Inventory optimization recommender

Generates precise orders
(ROQ) based on calculated
Safety Stock (SS) and **Reorder
Point (ROP)**

Methodology

TECHNIQUES USED

DEMAND FORECASTING

Model	Purpose & Insight
Decision Tree Regressor	Simple, interpretable **baseline model**. Confirmed model ability to capture non-linear relationships.
Random Forest Regressor	**Ensemble Method** that reduces overfitting and stabilizes predictions. Handled high-dimensional feature spaces effectively.
XGBoost Regressor (Best)	**Gradient Boosting** optimized for speed and accuracy. Handles seasonality, patterns, and interactions best. Used to generate the accurate

STOCKOUT RISK CLASSIFICATION

Model	Purpose & Insight
Logistic Regression	Baseline **linear model**. Used for initial binary/ternary risk assessment and coefficient interpretation.
XGBoost Classifier (Best)	**Top performer** for risk prediction. Excels at managing **class imbalance** and capturing complex non-linear relations between forecasted demand, current stock, and event indicators.

Performance Metrics

	precision	recall	f1-score	support
0	0.69	0.59	0.64	1723
1	0.72	0.80	0.75	2247
accuracy			0.71	3970
macro avg	0.70	0.69	0.69	3970
weighted avg	0.70	0.71	0.70	3970

```
Accuracy: 0.9962413261372398

Confusion Matrix:
[[42418    85     0]
 [  178 10030     3]
 [    0     7 19911]]

Classification Report:
              precision    recall  f1-score   support

           0     0.9958    0.9980    0.9969     42503
           1     0.9909    0.9823    0.9866     10211
           2     0.9998    0.9996    0.9997     19918

    accuracy                         0.9962     72632
   macro avg     0.9955    0.9933    0.9944     72632
weighted avg     0.9962    0.9962    0.9962     72632

Stock risk model trained and saved successfully!
```

```
Training XGBoost...
RMSE:  2.1631317206135647
MAE:   0.9976490002633676
R2:    0.653267836655362
MAPE:  57.10480229489171
SMAPE: 52.727432795349074
```

```
Loading data...
Training DecisionTree...
RMSE:  2.7089689033197843
MAE:   1.0878003492379937
R2:    0.4562037143775375
MAPE:  59.86273426254044
SMAPE: 51.35242196629575
```

```
Training RandomForest...
RMSE:  2.131897724435451
MAE:   0.9963421427007951
R2:    0.663208649461525
MAPE:  58.01619827046301
SMAPE: 52.08689648172555
```
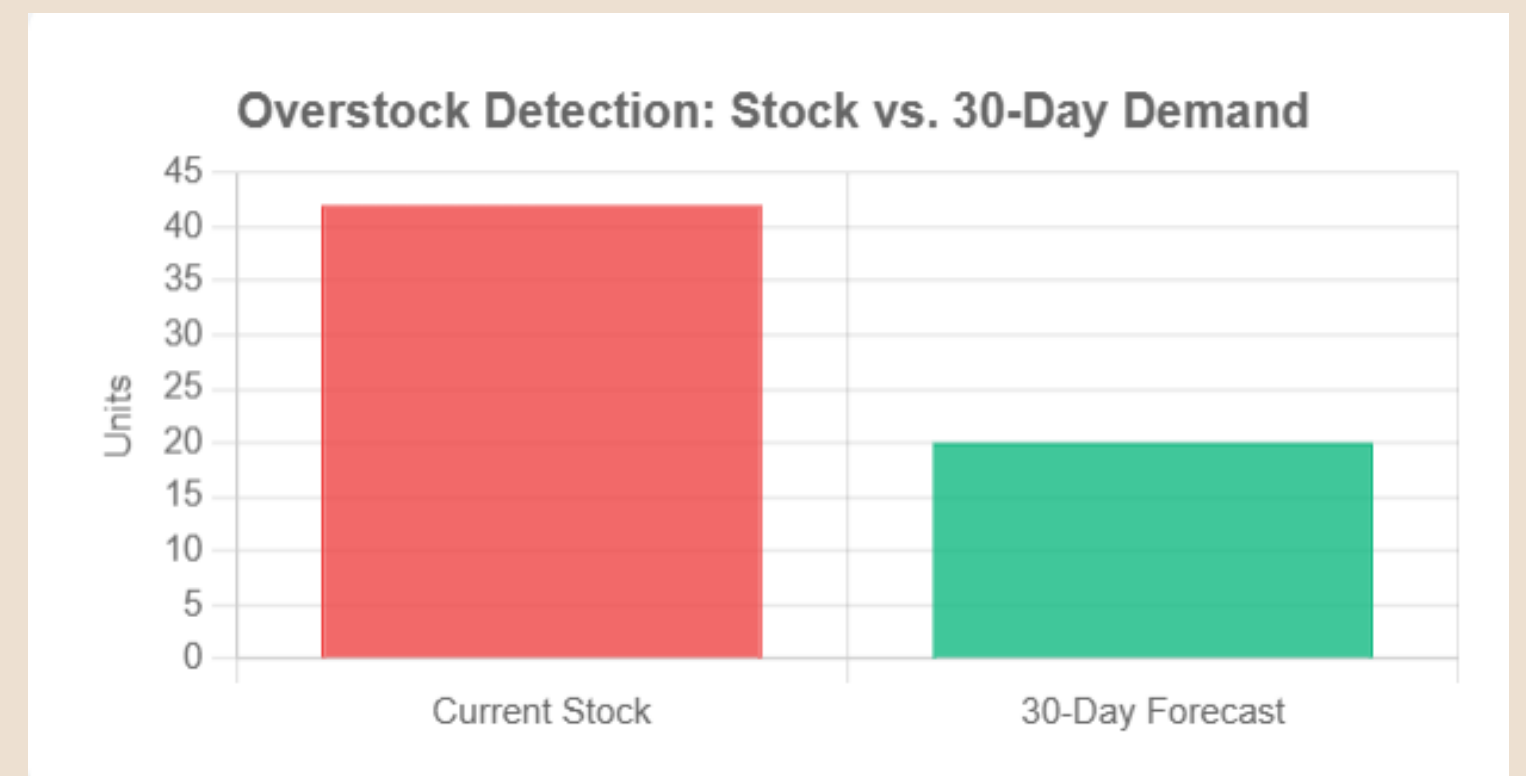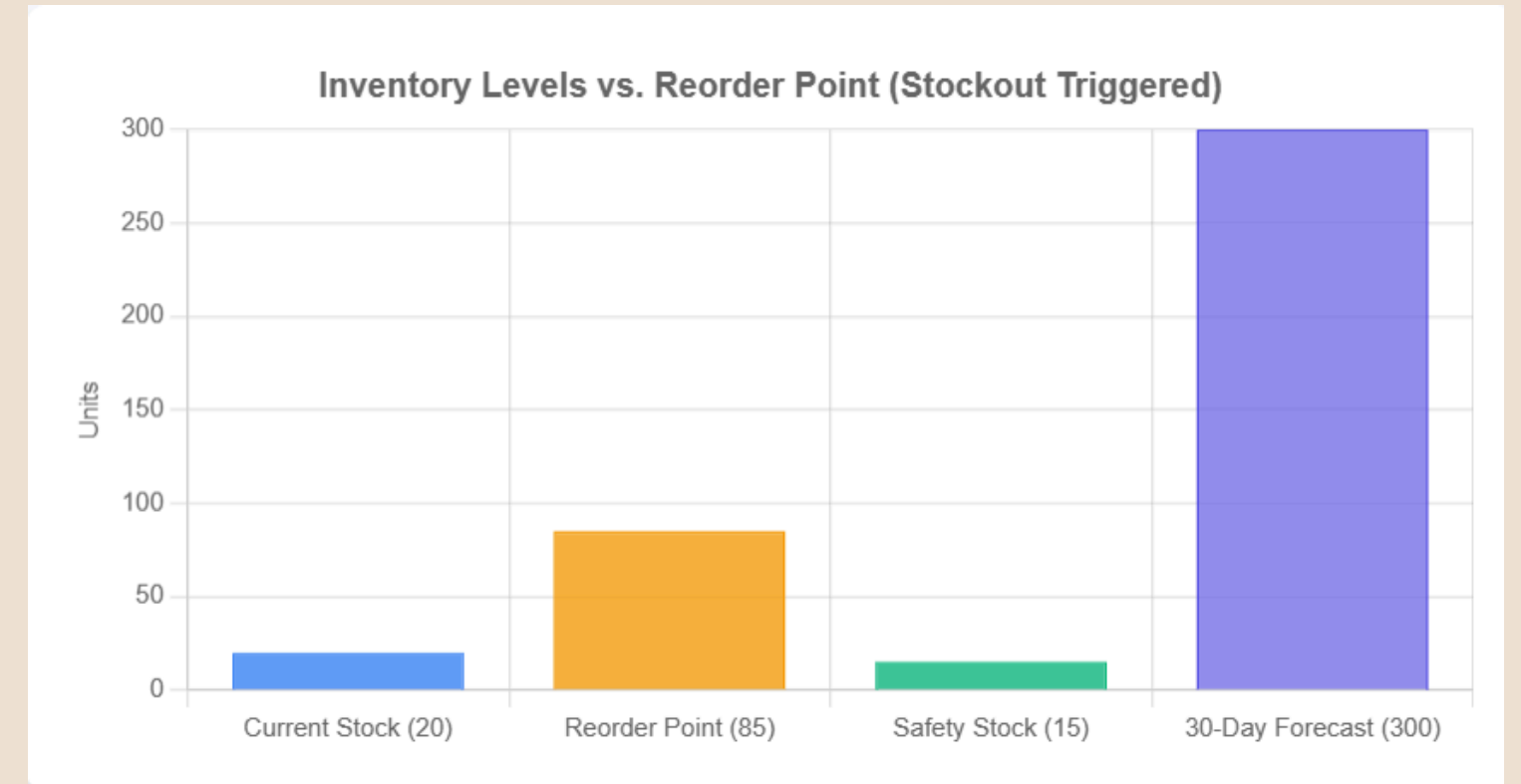
Results & Performance Validation

A. Model Performance Validation

- **Forecasting Accuracy:** XGBoost achieved lower errors (MAE/RMSE) vs. baselines, confirming superior seasonality handling.
- **Classification Accuracy:** Achieved High Accuracy (approx. 98%) and robust separation of risk categories, validating high resilience to noise.

B. Actionable System Output

- **Core Prediction:** Accurate 30-day forecasted demand (forecast_30).
- **Alert & Metric:** Definitive Stockout Risk Level and calculated Reorder Point (ROP).
- **Command:** Generates the Recommended Reorder Quantity for immediate procurement.
- **Visualization:** Displays clear charts for SKU trends and Inventory vs. ROP.

Inventory Levels vs. Reorder Point (Stockout Triggered)

Overstock Detection: Stock vs. 30-Day Demand

Conclusion
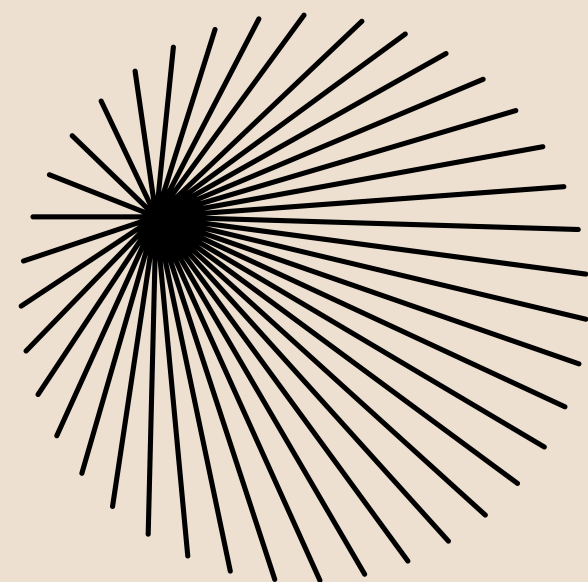
I. Project Impact

- Pipeline Delivered: Fully operational ML-based optimization system.
- Key Result: Achieved high-recall stockout risk classification.
- Actionable Value: System provides prescriptive SKU-level metrics (SS, ROQ).
- Business Benefit: Minimizes stockouts and overstocking for quantifiable savings.

II. Future Strategic Scope

- Dynamic Ordering: Implement Reinforcement Learning (RL) for cost-aware, dynamic ordering policies.
- Live Data Flow: Integrate Deep Learning (LSTM) and live Point-of-Sale (POS) data to capture real-time trends.
- Enterprise Scaling: Deploy system on cloud for multi-store optimization.

Thank You