Predicting Farmhouse Pizza Sales

- A Data-Driven Approach for Optimizing Raw Material Procurement

group member

- **amitesh**
- **arpit**
- **abhinav**
- **shidant**

Problem Overview

- Predict the quantity of Farmhouse pizza to be sold the next day across 250 outlets
- Minimize the risk of overproduction (leading to wastage) and underproduction (leading to lost revenue).

CHALLENGES

Overproduction = loss of raw materials

Underproduction = Lost sales and customer dissatisfaction

Objective

Objective of the Solution:

- **DEVELOP A PREDICTIVE MODEL TO FORECAST THE SALES OF FARMHOUSE PIZZA.**
- Ensure accurate demand prediction to optimize raw material procurement, avoiding both overstock and stockouts

Approach Overview

- **DATA COLLECTION:**
- Use historical sales data from 250 outlets for past sales of Farmhouse pizza.
- **DATA PREPROCESSING:**
- Clean data (handle missing values, remove outliers).
- Aggregate data to a daily level for each outlet.
- **FEATURE ENGINEERING:**
- Extract useful features such as day of the week, holidays, promotions, weather, and outlet-specific trends.
- **MODEL SELECTION:**
- Implement time series forecasting models to predict demand.

Data Used for Prediction

Sales Data:

- DAILY SALES DATA OF FARMHOUSE PIZZA PER OUTLET FOR THE PAST X MONTHS.

Additional Features:

- DAY OF THE WEEK: DIFFERENT SALES PATTERNS ON WEEKDAYS VS WEEKENDS.
- WEATHER DATA: CORRELATION BETWEEN WEATHER AND PIZZA SALES.
- PROMOTIONS/EVENTS: SALES SPIKES DUE TO PROMOTIONS, LOCAL EVENTS, ETC.
- Holiday Calendar: Effect of holidays or public events on pizza demand.

- **FEATURE ENGINEERING**
- **TEMPORAL FEATURES:**
- Day of the week
- Month/Season
- Public holidays
- **OUTLET SPECIFIC FEATURES:**
- Location (urban vs rural)
- Historical sales trends
- **EXTERNAL DATA:**
- Weather conditions (e.g., cold weather may increase demand for comfort food).
- Promotion and event data.

- MODEL SELECTION
- MACHINE LEARNING ALGORITHMS CONSIDERED:
- LINEAR REGRESSION: FOR SIMPLER MODELS BASED ON PAST SALES TRENDS.
- DECISION TREES / RANDOM FORESTS: TO CAPTURE NON-LINEAR RELATIONSHIPS.
- TIME SERIES MODELS (ARIMA, SARIMA): TO MODEL TIME-DEPENDENT SALES PATTERNS.
- XGBOOST/GRADIENT BOOSTING: A POWERFUL MODEL FOR HIGH ACCURACY WITH TABULAR DATA.

- **MODEL TRAINING PROCESS**
- **DATA SPLIT:**
- Split data into training (80%) and testing (20%) sets.
- **MODEL TRAINING:**
- Train the model on historical sales data (using Random Forest Regressor).
- Tune hyperparameters using cross-validation.
- **MODEL EVALUATION:**
- Evaluate model accuracy using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

- **MODEL EVALUATION & RESULTS**
- **Key Performance Metrics:**
- **Mean Absolute Error (MAE): Measures average error between predicted and actual sales.**
- **Root Mean Squared Error (RMSE): Measures the spread of forecast errors.**
- **Model Results:**
- RMSE = [X] (indicating how close predictions are to the actual sales).
- MAE = [Y] (indicating the average prediction error).
- **Visualizations:**
- Actual vs Predicted Sales Chart.
- Feature Importance Bar Chart (highlighting most important factors affecting sales).

- **IMPLEMENTATION PLAN**
- **DEPLOYMENT STRATEGY:**
- **REAL-TIME PREDICTIONS: MODEL WILL BE INTEGRATED WITH DAILY OPERATIONS TO PREDICT SALES ON A DAILY BASIS.**
- **RAW MATERIAL PROCUREMENT: AUTOMATED PURCHASING SYSTEM LINKED TO PREDICTED SALES TO ENSURE ADEQUATE SUPPLY.**
- **MONITORING & UPDATES:**
- Regular model retraining with updated sales data every quarter to maintain prediction accuracy.

- **CONCLUSION**
- Accurate demand prediction will help in optimizing raw material procurement.
- Reduces wastage and stockouts, ensuring a better customer experience and financial performance.
- A scalable solution for all 250 outlets across the country.