

1. Project Title:

Predicting Pizza Sales with Machine Learning: A Time Series Forecasting Approach

2. Group Members:

- **Name:** Varun Kumar Medi, **Student ID:** 652174010
- **Name:** Sarath Chandra Balla, **Student ID:** 669161466

3. Problem Statement:

Accurate sales forecasting is essential for pizza stores to optimize inventory, reduce waste, and improve staffing efficiency. Traditional forecasting methods often fail to capture complex patterns influenced by seasonality, weather conditions, and special events. This project aims to develop a robust time series forecasting model that predicts future pizza sales based on historical transaction data.

Key challenges include:

- Handling seasonality, trends, and external factors like weather and holidays.
- Selecting an optimal forecasting model suited for time series data.
- Ensuring the model generalizes well to new sales data while minimizing errors.

By leveraging machine learning, we aim to provide pizza businesses with data-driven insights to optimize sales predictions and decision-making.

4. Dataset:

We will use the **Pizza Sales Dataset** from Kaggle ([link](#)). This dataset contains detailed transactional data from a pizza restaurant, making it ideal for time series forecasting.

Dataset Features:

- **Order Date:** Date of the transaction.
- **Order ID:** Unique identifier for each order.
- **Pizza Name:** Type of pizza sold.
- **Quantity:** Number of pizzas ordered.
- **Unit Price:** Price per pizza.
- **Total Price:** Total revenue for each order.
- **Order Category:** Type of pizza (e.g., classic, gourmet).

- **Pizza Size:** Small, medium, large, or extra-large.
- **Order Type:** Dine-in, delivery, or takeout.

This dataset provides valuable insights into sales trends, seasonality, and the impact of external factors such as promotions and order types, making it well-suited for training and evaluating time series forecasting models.

5. Methodology:

To forecast pizza sales, we will implement and compare the following time series forecasting techniques:

Machine Learning Models:

We will implement and compare the following forecasting techniques:

- **Prophet:** A powerful model designed for time series forecasting, capable of handling seasonality, holidays, and trend shifts.
- **LSTM (Long Short-Term Memory Networks):** A deep learning model designed for sequential time series data, effective in capturing long-term dependencies.
- **SARIMA (Seasonal AutoRegressive Integrated Moving Average):** A statistical model useful for capturing seasonal patterns and trends in the dataset.
- **XGBoost Regression:** A gradient boosting algorithm that excels in capturing complex patterns and interactions in time series data. It will be used to predict future sales based on engineered time-based and external features (e.g., weather, promotions).

Feature Engineering & Data Preprocessing:

- Extracting time-based features (day of the week, month, holiday indicators).
- Analyzing sales trends based on pizza type and size.
- Creating lag features to capture past sales trends.
- Handling missing values and outliers using imputation techniques.
- Normalizing numerical features for better model performance.

Exploratory Data Analysis (EDA):

- **Trend Analysis:** Identifying sales trends over time.
- **Seasonality Detection:** Examining recurring patterns based on weekdays, months, and holidays.
- **Correlation Analysis:** Understanding relationships between sales and external factors using heatmaps and visualizations.

6. Evaluation Metrics:

To assess the forecasting performance of our models, we will use the following metrics:

- **Root Mean Square Error (RMSE):** Measures the average magnitude of forecasting errors.
- **Mean Absolute Error (MAE):** Provides an interpretable measure of prediction accuracy.
- **R² Score:** Evaluates how well the model explains the variance in sales data.

7. Timeline:

- **Week 1:** Data collection, cleaning, and exploratory analysis.
- **Week 2:** Feature engineering and preprocessing.
- **Week 3-4:** Train initial models (Prophet, LSTM, SARIMA, XGBoost) and perform hyperparameter tuning.
- **Week 5:** Model evaluation and selection of the best-performing model.
- **Week 6:** Deployment and testing predictions on unseen data.
- **Week 7:** Finalizing report, presentation, and project submission.

By implementing these methodologies, we aim to develop a reliable and interpretable forecasting model that enhances pizza sales predictions and aids business decision-making.