

Model Specification: SMB Sales & Inventory Forecasting

1. Overview & Business Objective

- **Problem Statement:** Small business owners, such as those in restaurants or consumer goods manufacturing, rely heavily on intuition and manual calculations to predict daily or weekly sales. This often leads to inaccurate demand forecasting, resulting in either costly waste (excess inventory/ingredients) or lost revenue from stockouts.
- **Primary Goal:** To create an automated system that provides small business owners with accurate daily and weekly sales forecasts. The model's output should be directly actionable, enabling smarter decisions on inventory purchasing, production levels, and staff scheduling.
- **Key Stakeholders:**
 - Business Owner / General Manager
 - Kitchen Manager / Production Manager
- **Success Metrics:**
 - **Financial:** Reduce material/ingredient waste costs by 20%. Increase gross profit margin by 5%.
 - **Operational:** Reduce stockout incidents by 50%.
 - **Model Performance:** Achieve a weekly sales forecast with a Mean Absolute Percentage Error (MAPE) of less than 15%.

2. Data Model & Requirements

- **Data Sources:**
 - **Primary Source:** Point of Sale (POS) system (e.g., Square, Toast, Shopify POS).
 - **Secondary Source:** Accounting software (e.g., QuickBooks, Xero).
 - **External Sources (via APIs):**
 - Weather Service (for historical and forecast weather data).
 - Public Holiday Calendars (for national and local holidays).
 - **Geo-aware Event Data:** Using the business's address, query sources like Eventbrite, neighborhood news sites, and city calendars for events within a specific radius (e.g., 1-2 miles) to ensure relevance.
 - **Social Media Platform APIs:** Access to the business's Instagram/Facebook to track metrics like follower counts, engagement, and mentions for quantifying brand awareness.
- **Key Data Entities:**
 - **Sale/Transaction:** A single purchase event.

- **Product:** The individual item being sold.
- **Date:** The day of the sales, enriched with external context.
- **Data Schema (Fields Needed):**
 - **From POS/Sales System:** transaction_id, timestamp, product_id, category, quantity_sold, price_per_unit, total_transaction_value, cost_per_unit.
 - **Generated & External Data:** day_of_week, is_weekend, is_holiday, is_local_event_day, daily_high_temperature, weather_condition.
- **Data Quality & History:**
 - **Ideal Requirement:** A minimum of **1 year** of daily, per-product transaction data from a POS system.
 - **Minimum Requirement / Contingency Plan:** In the absence of long-term POS data, the model will require:
 1. At least **1 year** of total daily sales data (from accounting software).
 2. At least **3 months** of recent, granular per-product POS data to establish a baseline product mix.
 - **Consistency:** Product names must be reasonably consistent.

3. Machine Learning Model

- **Model Type:** A **Time-Series Forecasting** model using a **gradient boosting** algorithm (e.g., LightGBM, XGBoost).
- **Target Variable (The Output):**
 - **Ideal Scenario:** quantity_sold for each key product for each future day.
 - **Minimum Scenario:** total_daily_revenue for the business, which is then broken down into product estimates using the recent product mix.
- **Features (The Inputs):**
 - **Time-based Features:** day_of_week, week_of_year, month, is_weekend.
 - **Hyper-Local Event Features:** is_holiday, is_event_nearby (within a 1-mile radius), nearby_event_count, nearby_event_category (e.g., 'Sports', 'Concert').
 - **Weather Features:** daily_high_temperature, precipitation_inches, weather_condition.
 - **Lag Features (Model's Memory):** sales_from_yesterday, sales_from_same_day_last_week, rolling_7_day_average_sales.
 - **Product-Specific Features (Ideal Scenario only):** product_id, product_category.
- **Evaluation Metrics:**
 - **Primary:** Mean Absolute Percentage Error (MAPE) < 15%.
 - **Secondary:** Mean Absolute Error (MAE).
 - **Methodology:** Time-series cross-validation.

4. Financial Model Integration

- **Automated Daily Decisions:**
 - **Automated Purchasing & Inventory Reports:** Generate suggested daily purchase orders based on sales forecasts and current stock.
 - **Optimized Staff Scheduling:** Recommend the number of staff members needed for a given shift based on revenue forecasts.
 - **Near-Term Cash Flow Projection:** Provide a simple weekly cash flow outlook.
- **Strategic Decision Support: The "Event ROI Calculator"**
 - **Function:** Provides a 360-degree analysis of the potential return on investment for participating in off-site events.
 - **Owner Inputs:** Event Name, Date, Costs, Resource Commitment.
 - **Model Output:**
 1. **Projected Event Profit/Loss:** A simple P&L for the event itself.
 2. **Projected Opportunity Cost:** The simulated impact of reallocating resources away from the main location.
 3. **Projected Visibility & Brand Lift:** An estimate of the increase in social media followers and subsequent store traffic.

5. Feedback & Improvement Loops

- **Core Forecasting Loop (Actuals vs. Forecast):** The model automatically retrains monthly using the latest sales data to adapt to new trends. Performance is continuously monitored to detect and flag anomalies.
- **Inventory & Waste Loop (Recommendations vs. Reality):** The system incorporates a simple waste log, allowing it to learn from over-purchasing and refine its inventory recommendations over time.
- **Human-in-the-Loop for Events (Qualitative Context):** After an event, the system prompts the owner for a one-click qualitative label (e.g., [No Impact], [Very Busy]). This human context enriches the dataset, making future forecasts for similar events more accurate.