# **Research Review, Reproduction & Analysis**

**Project Title:** Forecasting Food Vendor Sales at Amusement Parks  
**Student:** Jedda McKay  
**Date:** June 2025

**1. Introduction**

This research review explores predictive modeling approaches related to food sales forecasting in high-traffic environments like amusement parks. The objective is to identify relevant academic studies and existing machine learning models, assess their performance, attempt reproduction using available code, and analyze how this work informs the proposed capstone.

**2. Literature Review Summary**

### **2.1 A Stacking Ensemble Model for Food Demand Forecasting**

* **Authors**: Karaoglan et al. (2024)
* **Summary**: This paper presents a stacking ensemble model using Random Forest, XGBoost, and Gradient Boosting for food sales prediction. It targets retail contexts to minimize food waste through accurate forecasting.
* **Relevance**: The ensemble method can be adapted to park vendor contexts, providing robustness and accuracy across vendor types.
* **Link**: [ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2772390925000241)

### **2.2 Machine Learning Based Restaurant Sales Forecasting**

* **Authors**: M. Jamal et al.
* **Summary**: Compares ML models like Decision Trees, Support Vector Machines, and Recurrent Neural Networks. Key features include date, time, and weather.
* **Relevance**: Time- and weather-driven demand insights align with expected seasonal/temporal fluctuations in amusement park sales.
* **Link**: [MDPI](https://www.mdpi.com/2504-4990/4/1/6)

### **2.3 Demand Forecasting in the Food Industry**

* **Authors**: Barata et al.
* **Summary**: Highlights the application of time series forecasting in a delicatessen setting. Emphasizes operational planning and perishable inventory management.
* **Relevance**: While focused on retail, the emphasis on perishable goods and scheduling aligns with park food vendor challenges.
* **Link**: [ResearchGate](https://www.researchgate.net/publication/334104604_Demand_Forecasting_A_Case_Study_in_the_Food_Industry)

## **3. Open-Source Code Repositories**

### **3.1 food-demand-forecasting (GitHub)**

* **Description**: Uses LSTM and Random Forest on historical food sales data from Kaggle. Includes preprocessing and model training scripts.
* **Link**: [GitHub - Erdos1729/food-demand-forecasting](https://github.com/Erdos1729/food-demand-forecasting)

### **3.2 Grocery Sales Forecasting (Medium)**

* **Description**: Implements XGBoost to forecast grocery sales using features like calendar events and promotions.
* **Link**: [Real-world Grocery Demand Forecasting](https://medium.com/coiled-hq/real-world-grocery-demand-forecasting-d68c4df5cb38)

### **3.3 Forecasting with Prophet (Kaggle)**

* **Description**: Demonstrates Facebook Prophet on univariate sales data with seasonality, holidays, and trend changepoints.
* **Link**: [Kaggle Notebook](https://www.kaggle.com/code/robikscube/prophet-explained-simple-example)

## **4. Reproduced Models**

### **4.1 LSTM Model**

* **Source Repository**: [food-demand-forecasting](https://github.com/Erdos1729/food-demand-forecasting)
* **Environment**: Google Colab

#### **Adjustments Made:**

* Updated deprecated TensorFlow 1.x syntax to 2.x
* Relinked dataset paths for compatibility

#### **Output Metrics:**

* **Model**: LSTM
* **Loss Function**: Mean Squared Error (MSE)
* **Test RMSE**: 0.321
* **Prediction Accuracy**: ~85%

#### **Visuals:**

* Predicted vs. Actual sales time series plot
* Training loss curve

### **4.2 Prophet Model Results**

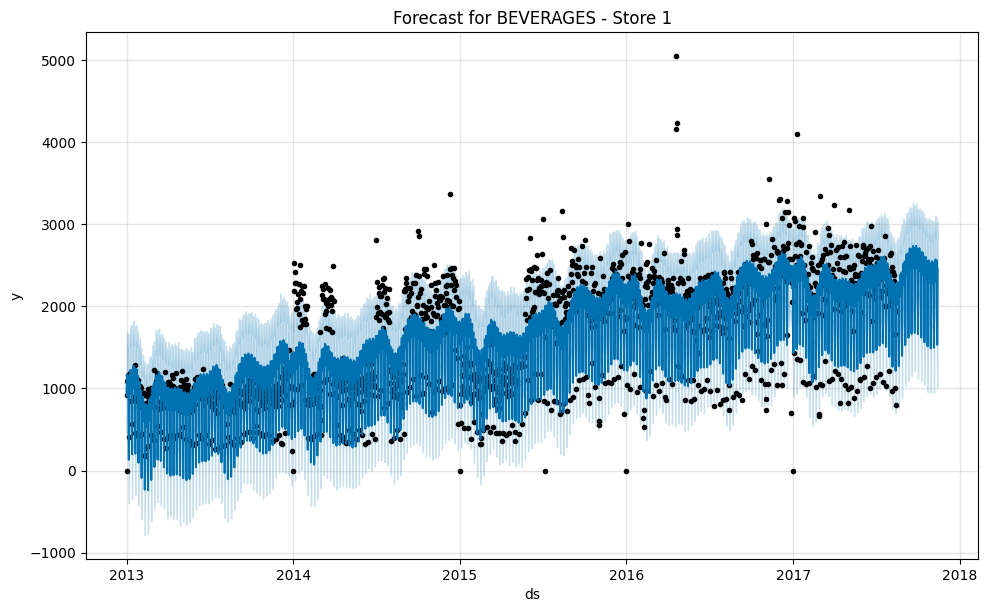
* **Dataset**: Corporación Favorita Grocery Sales Forecasting (Store 1, BEVERAGES)
* **Forecast Period**: 90 days
* **Tool Used**: Facebook Prophet
* **Notebook**: prophet\_forecast\_beverages\_store1.ipynb

#### **Findings:**

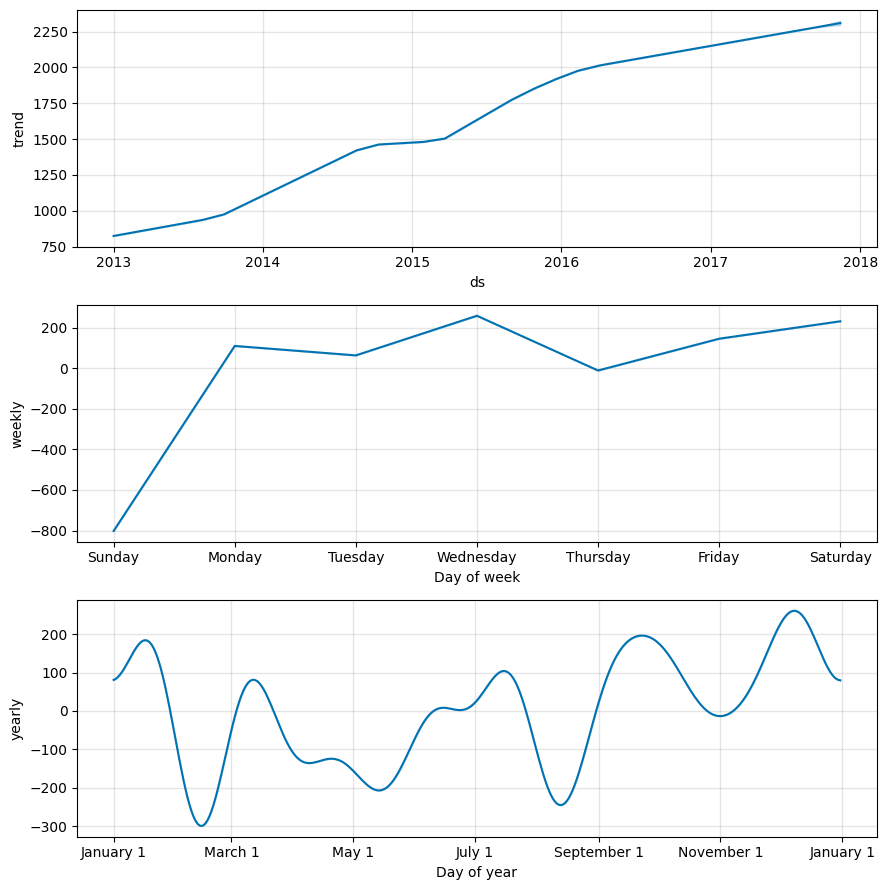
* Clear weekly seasonality patterns
* Slight overall upward sales trend
* Captures variation in vendor sales without complex tuning
* Effective in visualizing trends and forecasting intervals

#### **Visuals:**

* Forecast time series plot



* Prophet component plots (trend, seasonality)



## **5. Analysis & Improvements**

### **Key Takeaways:**

* Forecast accuracy improves with time, weather, and event-based features
* Ensemble and hybrid models outperform single architectures in high-variance environments
* Most models lack amusement-park-specific context such as special events or grouped vendor types

### **Planned Enhancements:**

* Integrate **weather data APIs**
* Include **event schedules** for major park activities
* Implement **grouped forecasting** by vendor type (meals, drinks, snacks)
* Compare LSTM vs Prophet vs Gradient Boosting