

Capstone Two: Forecasting Perishable Goods Demand Proposal

Problem Statement Worksheet (Hypothesis Formation)

What opportunities exist for grocers or retailers to effectively develop and implement a new Forecasting ML model to better forecast the demand for perishable goods.

H

1 Context

For any grocer or retailer, fresh foods poses a biggest challenge because of their short shelf life, appearance, season dependency, specific transportation and storage requirements and various other factors that can not just impact demand but also led to huge amount of food wastage. Globally, around 2.5 billion tons of food got waste every year, with US tops at first position with 60 million tons - 120 billion pounds food wastage every year. There is an extraordinary need for a better system to resolve this issue at a global level.

2 Criteria for success

Create a machine learning model that accounts for seasonality, weather patterns, and customer behavior to forecast the demand for perishable items like fresh produce or dairy products. A new ML model will be implemented asap.

3 Scope of solution space

A new ML model will be applied on grocery stores and supermarket facilities in order to automate, optimize current operations, better predict the demand for perishable goods, reduce wastage, while keeping up with all customer's expectations.

4 Constraints within solution space

There's a concern regarding the accuracy of data received. Currently, we are relying on the data source from Kaggle website, which we explore as we move forward in the project.

5 Stakeholders to provide key insight

- Mentor - Manish Pathak

In Real- world stakeholders would be

- Grocers
- Supermarkets
- Farmers
- Delivery partner

6 Key data sources

- Kaggle - A data science competition platform and online community for data scientists and machine learning practitioners under Google LLC
- Online web sources including Google

H D E I P

Context: Dataset includes a weekly historical data from 76 different stores of one of the largest retail chains in the world. The data collection ranges from the past three years. Sales and promotional information is also available for each week - product and store wise. The idea is to merge all files together using common keys to extract meaningful insight and create a ML model to better forecast the demand.

There are three CSV files:

File1: Contains columns such as record_ID and units_sold.

File2: Contains columns for week 16/07/13, such as record_ID, week, store_id, sku_id, total_price, base_price, is_featured_sku, is_display_sku.

File3: Contains columns for week 17/01/11, such as record_ID, week, store_id, sku_id, total_price, base_price, is_featured_sku, is_display_sku, units_sold.

Another data source I would like to explore is from the web source, **Food and Agriculture Organization of United nations** while keeping the focus on perishable commodities including fruits, dairy, vegetables etc.

<https://www.fao.org/platform-food-loss-waste/flw-data/en/>

Deliverables - A GitHub repo containing the work you complete for each step of the project, including:

- A slide deck
- A project report