



SOEN 6111 WINTER 2024

GROUP - 31

Mid-East Pharmaceutical – Stock Availability

#	Group Members	Student Id
1	Vishnu Rameshbabu	40233562
2	Fiza Altaf Razvi	40236583
3	Kevin Wadera	40261875
4	Khushal Nirmal Jain	40233877

Index

Table of Contents

Overview..... 3

Dataset 3

Research Question..... 4

Model Classes..... 4

Algorithms 4

Overview

Mid-East Canadian Pharmaceutical (MECP) is a medical company specializing in providing cold supply chain solutions for medical equipment to various entities. With an extensive inventory of over 508,000 medical products, they aim to leverage publicly available data from their suppliers to identify in-demand medical products at specific times throughout the year. The goal is to develop a training model using existing data to forecast periods of increased demand for specific medical products. This predictive capability enables them to proactively stock these products ahead of the demand curve, ensuring timely availability and efficient supply chain management.

Dataset

Based on the information provided by the MECP, the dataset needs to be scrapped from internet using a crawler script. From our initial understanding, the dataset contains,

1. **Name(varchar)**: Name of product.
2. **Description (varchar)**: Description of product.
3. **Catalogue Number**: Unique identifier for product in catalogue.
4. **Price(float)**: Price of product.
5. **Availability (varchar)**: Availability status of product.
6. **Manufacturer Number(int)**: Identifier for manufacturer of product.
7. **Invoice Description (varchar)**: Product description on invoice.
8. **Feature Property (varchar)**: Specific features of product.
9. **Sterility(bool)**: Sterility status of product.
10. **Link (varchar)**: URL link for more information about product.
11. **Brand Name (varchar)**: Name of brand associated with product.
12. **Outer Diameter(float)**: Outer diameter measurement of product.
13. **Needle Length(float)**: Length of needle
14. **Volume(float)**: Volume measurement of product.
15. **Color (varchar)**: Color of product.
16. **Inner Diameter(float)**: Inner diameter of product.
17. **Size (varchar)**: Size specification of product.
18. **Age(datetime)**: Age-related information (e.g., expiration date).
19. **Shape (varchar)**: Shape of product.
20. **Catheter Length(float)**: Length of catheter (if applicable).
21. **Composition Ingredient(varchar)**: Composition or ingredients of product.

Research Question

Based upon the requirements gathered from MECP, we have identified the requirement is to create a predictive model that would provide an accurately identify the shortage periods of the medical products.

Model Classes

Based on requirements we aim to use models that adopt supervised learning to train since we need to identify patterns of when a product goes out of stock and during which days in the year.

Algorithms

1. **Random Forest:** Random Forest is used to analyze the factors contributing to stock availability and predict the likelihood of a product going out of stock based on historical data and various product attributes. By training a Random Forest classifier using the columns like product descriptions, prices, manufacturer numbers, etc. it is possible to handle to handle both categorical and numerical features and identify influential factors impacting stock availability.
2. **Time Series Analysis:** Time Series Analysis is a robust method for grasping temporal patterns in stock availability data, aiding in understanding factors leading to product stockouts. Techniques like Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), or Prophet predicts future stock levels based on historical data. These models using forecast analysis and data decomposition capture seasonality, trends, and other temporal patterns, offering insights into demand fluctuations and potential shortages.