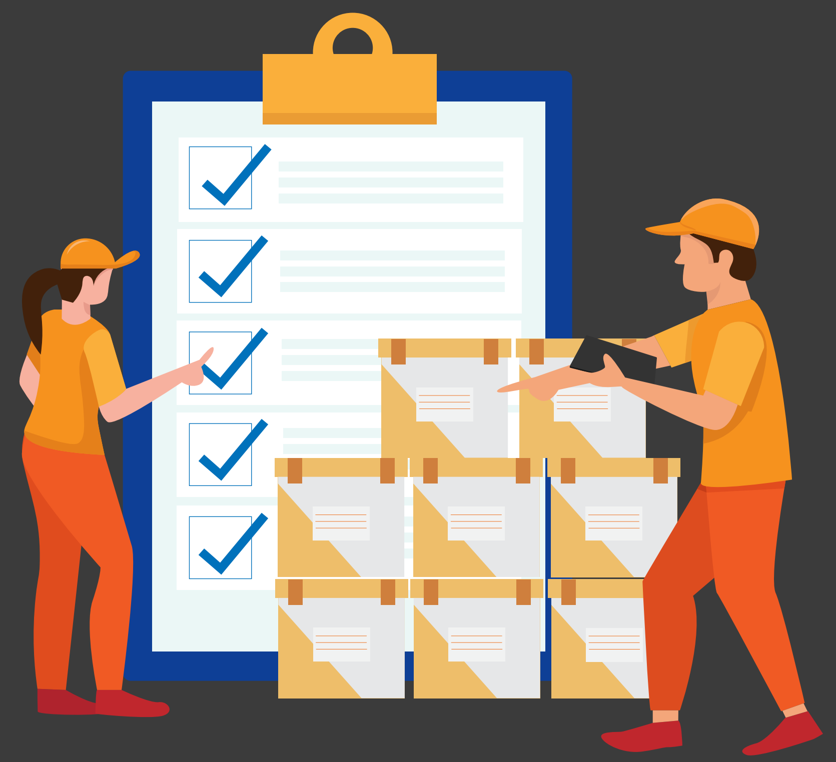
Data Glacier

Your Deep Learning Partner

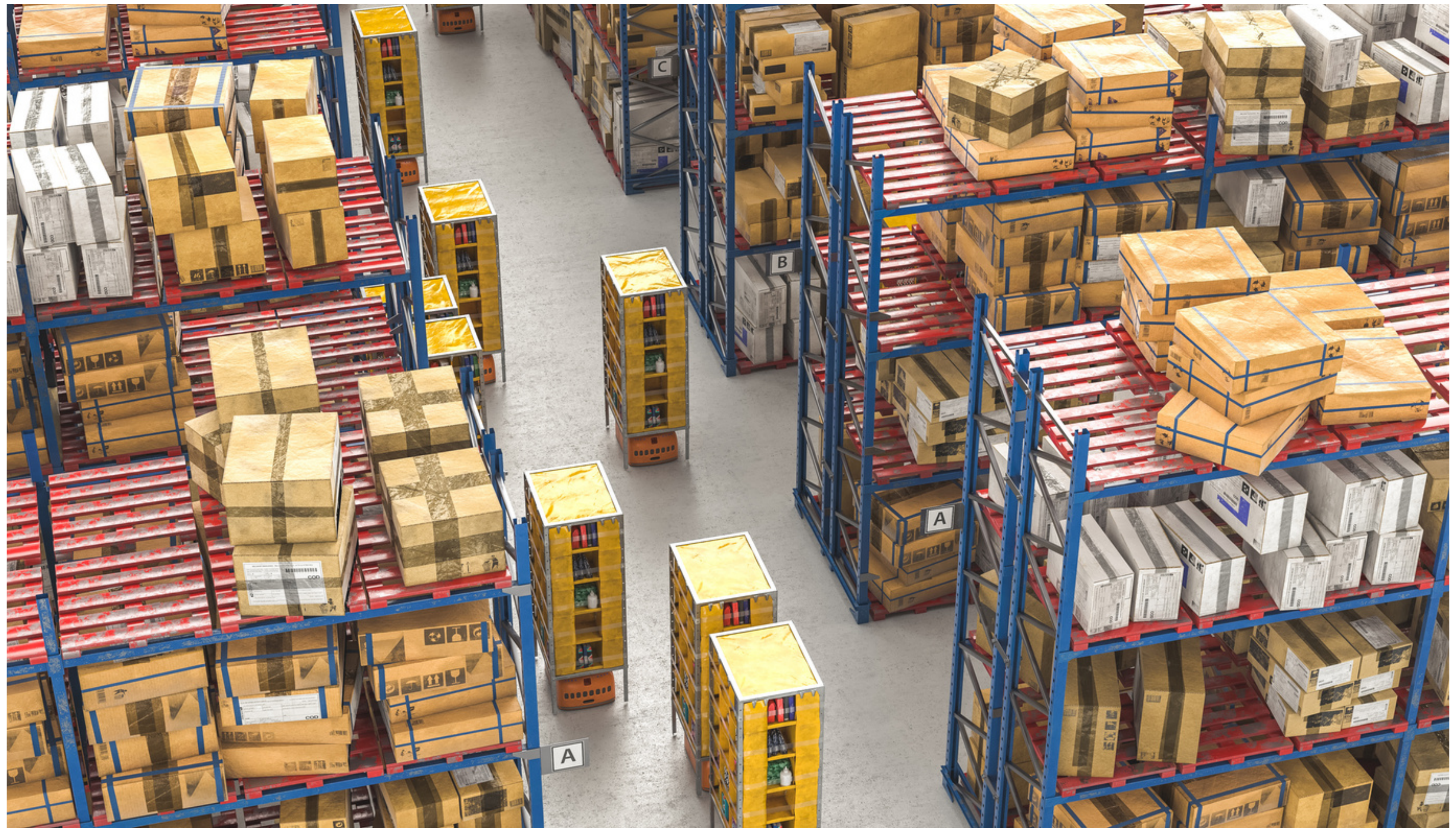
Retail Forecasting for Inventory Management

30-Nov-2023



Data Glacier
Your Deep Learning Partner

Problem Statement



- Our beverage industry client is at a crucial crossroads, focusing on refining their demand forecasting strategies.

- The existing in-house tool has proven unreliable, causing disruptions in inventory management.

- Our mission is to explore AI/ML solutions that promise a more accurate and adaptable forecasting model.

- **The goal is clear:** to elevate operational efficiency and market responsiveness for our valued client.

Data Glacier
Your Deep Learning Partner

Dataset

February 5, 2017 - December 27, 2020

1218 observations - No missing value

- **Product:** Name of the product.
- **Date:** Weekly recording date for sales data.
- **Sales:** Weekly unit sales.
- **Price Discount (%):** Percentage discount applied to the product's price.
- **In-Store Promo:** Presence of in-store promotions (1 for yes, 0 for no) during the week.
- **Catalogue Promo:** Presence of catalogue promotions (1 for yes, 0 for no) during the week.
- **Store End Promo:** Presence of store end promotions (1 for yes, 0 for no) during the week.
- **Google_Mobility:** Data indicating the impact of Google Mobility on sales.
- **Covid_Flag:** Flag representing the influence of COVID-19 on sales.
- **V_DAY, EASTER, CHRISTMAS:** Indicators of specific holidays/events and their impact on weekly sales.

Data Glacier
Your Deep Learning Partner

Summary of EDA

- **Sales by product:** SKU3 has the highest sales among the 6 products, whereas SKU2 has the lowest.
- **Weekly sales:** For products other than SKU6, there is no recorded data for the last 6 weeks within the date range of the dataset. Therefore, these weeks have been excluded from the dataset.
- **Discounted sales:** The highest sales were achieved during the discounts ranging from 40% to 50%.
- **Promotions:** Three different types of promotions have been applied to the products, with in-store promotions being the most frequently implemented.
- **Pandemic effect:** 22.7% of the sales in the dataset occurred during the COVID-19 period.

Modelling

- In this study, the total number of variables increased from 12 to **88** after applying feature engineering.
- These variables include date, lag, rolling_mean, and exponentially_weighted_mean features. Therefore, some variables contain NaN values.
- Accordingly, four different machine learning models have been selected: **LightGBM**, **XGBoost**, **CatBoost**, and **Histogram Gradient Boosting**.
- The calculated SMAPE values after running the models are sorted as follows:
  - **Histogram Gradient Boosting:** 33.622
  - **LightGBM:** 33.659
  - **XGBoost:** 36.480
  - **CatBoost:** 39.238
- The **Histogram Gradient Boosting model** seems to be the **most suitable** for the given task, based on the evaluated SMAPE metric.

Thank you.

Data Glacier

Your Deep Learning Partner