

DSC412-project

Student: Sharen Zhang

The Proposal

Project Proposal: Machine Learning in Supply Chain Demand Forecasting

1. The Problem: One of the biggest challenges in supply chain management is accurately forecasting product demand. Poor demand forecasting can lead to stockouts, excess inventory, and increased operational costs. The objective of this project is to improve demand forecasting accuracy using machine learning algorithms, which will help businesses optimize inventory levels and reduce costs associated with overstocking or understocking products.
2. The proposed method of solution: Machine learning algorithm and such as regression model like random forest or gradient boosting.
3. Potential Stakeholder:
 - Retailers and Manufacturers: Improved demand forecasting will help them manage inventory more effectively.
 - Logistics Companies: More accurate demand data will enable them to optimize transportation and warehousing resources.
 - Supply Chain Managers: They will benefit from the ability to make better decisions regarding inventory control and order fulfillment.
 - Myself: For the future study of how to forecasting effectively of supply and demand.
4. Potential Obstacles:
 - Data Availability: Finding sufficient and relevant data might be challenging. Some datasets may be inaccessible or hard to find.
 - Data Preprocessing: How to effectively cleaning, transforming, and preparing data for machine learning models.
 - Model and algorithm selection: How to selecting the appropriate machine learning algorithm for the task. Might be encounter a lack of specialized algorithmic knowledge and tuning model accuracy and practicality.

5. What is novel about your problem and your potential approach:
- While demand forecasting models have been used in supply chains, applying machine learning algorithms like advanced ensemble methods offers a novel approach in terms of capturing complex patterns in historical data and external factors. This problem will explore how combining multiple data sources can provide more robust and accurate predictions.

The Plan

- Where do you plan to get your data sets?
 - The data set that I'm looking forward will include historical sales data, product details, and inventory levels from a retailer or manufacturer. Additionally, external data such as economic indicators, seasonality, and weather data will be collected from public APIs or other databases. For Dataset Structure, each data point will include features such as date, product ID, sales volume, price, weather conditions, and economic indicators. I'm planning to get my data set from the Kaggle.
- How do you plan to organize your data?
 - The first step will involve cleaning the raw data by addressing missing values, removing duplicates, and ensuring consistency in the dataset. Categorical variables, such as product categories, will be converted into a machine learning-compatible format. This will prepare the data for effective model training.
 - Then, Creating the new features to enhance the model's performance. Once the data has been cleaned and processed, it will be stored in a local file.
- How do you plan to analyze your data? What are you looking for? What might you find?
 - The initial analysis will focus on identifying trends, seasonality, and correlations in the dataset using tools for visualizing patterns. Based on the data structure and problem statement, initial tests will be conducted with machine learning algorithms like ARIMA and LSTM for time series data, while regression-based models will also be explored for comparative analysis. The goal is to determine which model best captures sales patterns and provides the most accurate demand forecast.

- Based on your problem statement, what type of model do you plan to use? Is it one of the ones we have discussed/will discuss in class?
 - Based on early testing, the model with the highest prediction accuracy and lowest error rate will be selected for further tuning.
 - For improvement, the existing statistical models, such as simple moving averages or exponential smoothing will be used as baselines, and machine learning models will be evaluated based on how much they improve over these traditional methods.
- How will you determine if your model is accurate?
 - For regression models, I will use Root Mean Square Error and Mean Absolute Error to measure the difference between predicted and actual values.
 - For time series models like ARIMA or LSTM, I will also assess the model's performance on out-of-sample data using cross-validation and forecast accuracy metrics, such as Mean Absolute Percentage Error.
 - The model will be tested on a holdout dataset to ensure it generalizes well to unseen data, and any overfitting will be identified through these validation processes.
- Are there similar models/paper/code online that are trying to do what you are doing? If there are, use them as a baseline for improvement.
 - I find out a Kaggle article that covers a similar topic and models to what I'm working on; it focuses on analyzing big data for supply chain demand forecasting.
 - Kaggle article link: <https://www.kaggle.com/code/graisybiswal/demand-forecasting-in-supply-chain-analytics/notebook>