Risk Management with Backorder in Supply Chain using Machine Learning Techniques

Anubhav Singh1, Lubna Tharanum2, Mohammed Minaam Qureshi3, Rachana4, Nivedha S5
Assistant Professor2
Department of Computer Science and Engineering
The National institute of Engineering, Mysuru, India

Abstract:
Risk and uncertainty has always been an important issue in supply chain management. Supply chain management is the handling of the entire production flow of goods or services to maximize quantity, the customer experience and profitability. Machine learning prediction approaches are not used in numerous business choice cycles because of the absence of clearness and adaptability. The incorrect information as contributions to the forecast interaction may deliver wrong expectations. We aim to use machine learning models in the area of the business decision process by predicting products backorder while providing flexibility to the decision authority, better clarity of the process, and maintaining higher accuracy. A ML technique is utilized for indicating various degrees of ongoing information which may occur by human mistakes. The tree-based AI is picked for better reasonableness of the model. The backorder of products are anticipated in this examination utilizing Random Forest (RF) and XGBoost Machine (XGBM). We have seen that the exhibitions of the ML models have been improved by 20%. We have utilized a metric to indicate the inventory level, sales level, forecasted sales level, and lead time. We show how this model can be utilized to anticipate the likely backorder items before real deals occur. The referenced strategies in this exploration can be used in other production network cases to forecast backorder.

I. INTRODUCTION
Supply chain management is the handling of the entire production flow of goods or services to maximize quantity, the customer experience and profitability. The experts and researchers accept that powerful supply chain management has become a significant empowering influence to improve association execution and valuable ways of securing competitive advantage. The expanding vulnerability expects them to spend more assets to expect request, supply. At the point when a client arranges an item, which isn’t accessible in the store or impermanent unavailable, and the client chooses to stand by until the item is accessible and vowed to be transported, then, at that point this situation is called backorder of that particular item[1,2]. Besides, the vulnerability in clients requests causes trouble in anticipating the interest which makes the conventional inventory network and the executives frameworks less compelling from multiple points of view, for example, wrong interest gauging or misclassification of delayed purchased items[3,4]. These days, a few organizations anticipate the backorders of items by applying AI expectation cycles to conquer the related unmistakable and theoretical expenses of backorders[5]. Machine learning models may misclassify numerous records if the dataset contains deluding or missing data. This issue is a test to investigate the dataset of this examination. There are exceptionally high negative and positive qualities in a few foreseeing highlights of this dataset. Our dataset contains the quantity of negative records identified with the stock. The negative stock level recommends that the current stock level of the item is under nothing. The circumstances and end results of negative stock are surely known in the production network industry[6]. A stock levelward requesting model examines the connection between stock level and request which reflects what negative stock level can mean for the interest[7]. Recent studies suggest that the correlation factor among the product variety, sales, and inventory level is biasing the inventory level[8,9]. We have played out some theory tests considering backorder situations. The results of the theory’s tests are useful to pick the proper machine learning model for forecast. Random Forest (RF) and XGBoost Machine (XGBM) strategies[10,11] are picked. To determine the imbalanced class issue and to remove outliers, we have used robust scaling on the objective class. We have partitioned our foreseeing highlights in various reaches, and we have passed it to XGBM and RF models for expectation. In addition, the genuine information is taken care of by those models. It very well may be seen that the two models show various qualities in the trial. This examination considers backorder determination utilizing RF and XGBM and reports effectively justifiable plausible backorder choice situations.

II. LITERATURE SURVEY
Thi Thu Ha Nguyen has proposed a paper, Wal-Mart’s successfully integrated supply chain and the need of setting up the Triple-A production network in the 21st century. It is generally perceived that as opposed to zeroing in just on their activities, supply chain capability has become a basic factor to support the seriousness of organizations. As Harrison[12] affirmed, organizations currently consider its stock chains as support the seriousness of organizations. As Harrison[12] affirmed, organizations currently consider its stock chains as support the seriousness of organizations.

IJESC, July 2021

http://ijesc.org/
concerns a simple supply chain system which focuses on only cost effectiveness and speed not on Risk management. Companies like Ericsson (Norrman and Jansson, 2004) and Nokia (Li et al, 2006), have since quite a while ago realized the requirement for effective risk management. The key discoveries indicate that it is necessary to assemble the Triple-A supply chain in the 21st century and in order to achieve an optimal supply chain, firms need to overhaul its supply chain towards an incorporated supply chain with three capacities (agility, adaptability and alignment) in which adjusting incentives can be seen as the most important component to make it risk free. This research confirms that an integrated supply chain will be the most crucial pre-imperative for a firm to improve the supply chain performance and achieve an ideal supply chain.

III. EXISTING SYSTEM

One of the traditional methods is to manufacture inventory randomly without any measures and detailed analysis of customer’s demand. Risk management is important in an organization because without it, a firm cannot possibly define its objectives for the future and will help it achieve its primary objectives. Demand and supply is not regulated properly.

IV. PROPOSED SYSTEM:

In this era everyone is focusing on speed, cost-effective and integrated supply chain. An integrated supply chain will be the first vital prerequisite for a firm to improve the supply chain performance and achieve an ideal supply chain. The main intention is to manage demand risk and it is a process of taking strategic steps to identify, assess and mitigate the risk in our end-to-end supply chain and solve demand risk. Risk is an important issue threatening sustainability and Competitiveness of supply chains. The recurrence, seriousness and assortment of supply chain (SC) chances are speeding up because of expanding globalization, client expectations and more limited item life cycles in SC networks. Efficient Risk management is achieved.

Manufacturing of products is regulated based on demand of customers, thereby, reducing risk and minimizing the loss of the firm.

![Figure 1. Proposed System](image-url)

V. SYSTEM IMPLEMENTATION

1. XG BOOST

XGBoost is an ML algorithm based on decision-trees that uses a gradient boosting framework. It was developed as a research project at the University of Washington, and since then has been credited with winning numerous Kaggle competitions. The two main standouts for XGBoost are System Optimizations and Algorithmic Enhancements.

1. System Optimization:
   • Parallelization: A parallelized implementation is used by XGBoost to approach the process of sequential tree building.

2. Hardware Optimization:
   • This algorithm makes efficient use of hardware resources too. This is accomplished by ‘cache awareness’ which is a concept involving the allocation of internal buffers in each thread to store gradient statistics. Along with this it also supports Sparsity Awareness, Weighted Quantile Sketch, Cross-validation.

   • $F_0(x)$ with which we initialize the boosting algorithm is:
     \[
     F_0(x) = \arg \min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)
     \]
   • The gradient of the loss function is computed iteratively:
     \[
     r_{tm} = \alpha \cdot \left( \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right) |_{F(x_i) = F_{m-1}(x)} \text{ where } \alpha \text{ is the learning rate}
     \]
   • Each $h_m(x)$ is fit on the gradient obtained at each step
   • The multiplicative factor $\gamma_m$ for each terminal node is derived and the boosted model $F(x)$ is defined:
2. RANDOM FOREST

Random forest is a supervised learning algorithm. Random forest comprises countless individual choice trees that work as an ensemble. Every individual tree in the random forest lets out a class forecast and the class with the most votes turns into our model's expectation. The "forest" which is built is nothing but an ensemble of choice trees, generally prepared with the "bagging" strategy. The overall thought of the bagging is that a mix of learning models improve the overall result. In this manner random forest forms numerous decision trees and combines them to get a more precise and stable forecast.

Methodology: Algorithm for Construction of Random Forest is:

**Step 1:** Assume the number of training cases be “n” and let the number of variables included in the classifier be “m”.

**Step 2:** Let the number of input variables used to make a decision at the node of a tree be “p”. Assume p<m.

**Step 3:** Pick a preparation set for the decision tree by picking k occasions with substitution from all "n" available training cases by taking a bootstrap sample.

**Step 4:** For every node of the tree, randomly pick factors on which to look for the best split. New data can be anticipated by considering the majority votes in the tree. Predict data which is not in the bootstrap sample. And compute the aggregate.

**Step 5:** Compute the best split based on these chosen variables in the training set. Base the decision at that node using the best split.

**Step 6:** Pruning is used to cut off the leaf nodes so that the tree can grow further. Here the tree is completely retained.

**Step 7:** The best split is one with the least error i.e. the least deviation from the observed data set.

VI. RESULT DISCUSSION

1. XG Boost

```python
model = XGBClassifier(nthread=-1)
parameters_for_model = {'n_estimators': [10,20,30]}
clf = GridSearchCV(model, parameters_for_model, scoring = 'roc_auc')
model.fit(X_train, y_train)
model.predict(X_test)
```

![ROC AUC Curve for XG Boost](image1)

**Figure 2. ROC AUC Curve for XG Boost**

2. Random Forest

```python
model = RandomForestClassifier(class_weight = 'balanced_subsample')
parameters_for_model= {'n_estimators': [10,50,100], 'max_depth': [1,3,5]}
clf = GridSearchCV(model, parameters_for_model, scoring = 'roc_auc')
fitted_model = clf.fit(X_train, y_train)
model.predict(X_test)
```

![ROC AUC Curve for Random Forest](image2)

**Figure 3. ROC AUC Curve for Random Forest**
For the given dataset XG Boost performs better. However, if we see the overall accuracy of both the models, we can conclude that both the models perform good and these models could be used in improving the overall efficiency of the Supply Chain.

VII. CONCLUSION

We will manage demand risk and it is a process of taking strategic steps and solving demand risk. We have achieved balance between having sufficient inventory levels to meet customer needs without having a risk of loss for the firm due to surplus quantity of goods. We must have to keep track of the customer requirements and study their interest, based on which we will decide the manufacturing of the inventory and regulate the demand by keeping loss at minimum scale. Data pruning can be one of the solutions, but it requires the experts' availability, which is costly more often. In this research, a ranged method is proposed to tackle this issue. The actual data and the ranged data have been utilized in both ML models, and their performances have been analyzed. The correlation result shows that the models' performance increments by around 20% when they are prepared with the ranged data. In this paper, it has been shown that depending on known stock, lead time, sales, and forecasted sales, we can recognize those items that will be backorder items. Nonetheless, the relational factors like local buyout quantity, past due stock, provider' performances, and different product risk flags have not been considered on account of the absence of that data. We may focus on those elements in our future work.

As the vulnerability of demand assumes a crucial part to make the market unstable, the connection between the predicted demand and the predicted backorder may also need attention. As future exploration, the referenced viewpoints can be considered.

VIII. REFERENCES


Overall Accuracies

Table I. F1 scores

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGBoost</td>
<td>92.64763800124254</td>
</tr>
<tr>
<td>Random Forest</td>
<td>89.72965764325762</td>
</tr>
</tbody>
</table>