Robust Supply Chains with Gradient Boosted Trees

Pradeep Kumar Mahato, Apurva Narayan
Department of Computer Science
The University of British Columbia
BC, Canada

Motivation

Forecasting of demands in supply chain is a complex problem. The problem gets complex with larger network where stakes are very high.

Example:

- In 2001, Nike outcasted its $400 million demand forecasting software due to erroneous[1].
- In 2014, Walgreen encountered a $1 billion loss forecasting error[2]

Motivation

Food is a perishable item. In supply chain, food industry are the hardest hit due to inaccurate estimations.

Some major causes for estimation failure :

- Planning issue: The plan itself was inadequate in fulfilling the right amount of items on time. These issues could be and must be avoided as early as possible in the supply chain cycle
- Execution issue: Operational issue due to unexpected circumstances for example, delay in shipment from plant to delivery chain, machine failure, etc
- Configuration issue: Run-time change in system parameters and rules that create the supply chain plans such as safety-stock targets, demand forecasting, master plan, backup plan, etc, contribute to such issues.

Objective

Major goal for our project are :

- Prevent service level failure from happening
- Identifying possible root cause/s for such failure
- Estimate approach magnitude of failure

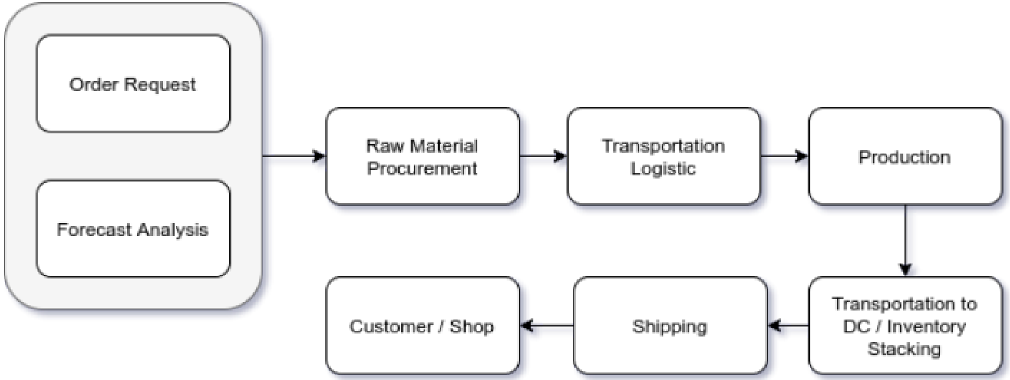


Fig: Supply Chain Network

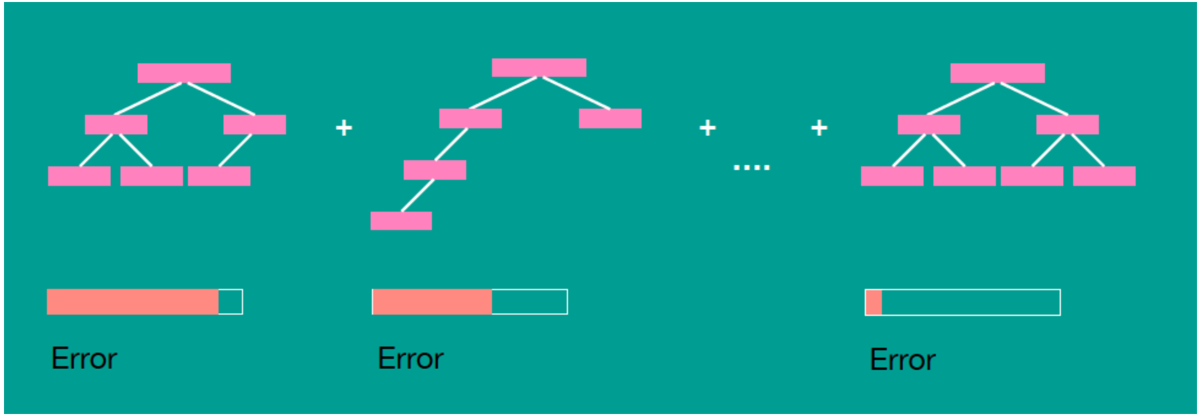
Outline

- Background
- Methodology
- Result
- Conclusion
- Future Work

Background

GBDT : Fundamental

Gradient Boosting descends the gradient by introducing new model while minimising the loss function. In simple term, this means various models are used to improve prediction. It uses a decision tree as its base modelling structure. Decision tree helps in easy explainability for a decision being made.



Source: https://towardsdatascience.com/introduction-to-gradient-boosting-on-decision-trees-with-catboost-d511a9ccbd14

GBDT

Few major reason for choosing decision tree over other algorithm (ex : random forest) :

- Easy to compute and explain why a particular feature has higher importance
- Can be visualized (to a certain extent), easier to explain model implementation
- Simpler model which can be used in real time
- Demerit of decision tree is accuracy which is resolved in GBDT by using Gradient Boosting - using historical evidence.

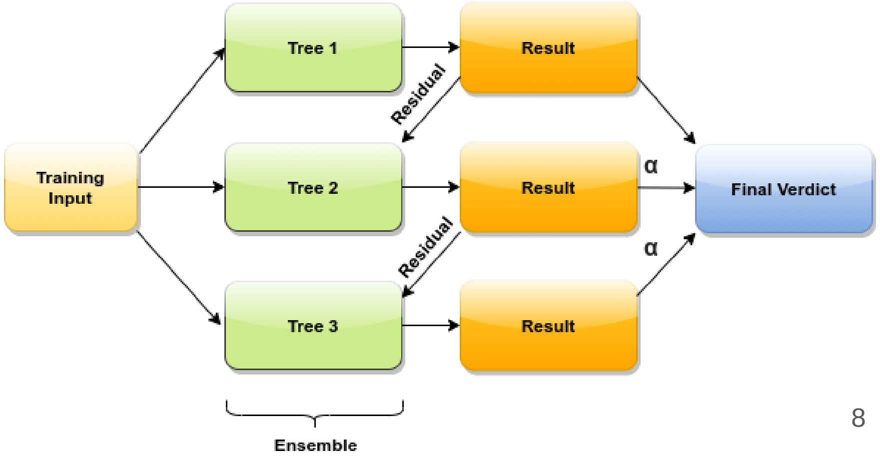


Fig: Sample illustration of GBDT algorithm with 3 trees

Correlation

Correlation (or covariance) function determines the similarity between two entities. Correlation function helps in determining how do features interact among themselves.

Autocorrelation (ACF) and Partial-autocorrelation (PACF) are two predominant techniques often used in time series analysis.

- ACF is a mathematical tool for finding recurrent similarity which is often hidden due to the presence of noise or other factors.
- PACF differs from ACF as it finds the residuals, which survives even after removing earlier correlations before comparing it with the subsequent lags

Shapley Values

Shapley values is a method from coalition game theory which helps to determine the importance of each feature towards the model.

To the best of our knowledge, to date Shapley value is the only mathematical expression which defines the three axioms of interpretability:

- **Dummy player:** If the feature does not have any role towards the model‘s prediction then its contribution should be 0
- **Substitutability:** If the two features add the same marginal value then they possess substitutability.
- **Additivity:** The sum of all contributions in the individual subset should match the overall contribution towards the entire set.

Methodology

Methodology

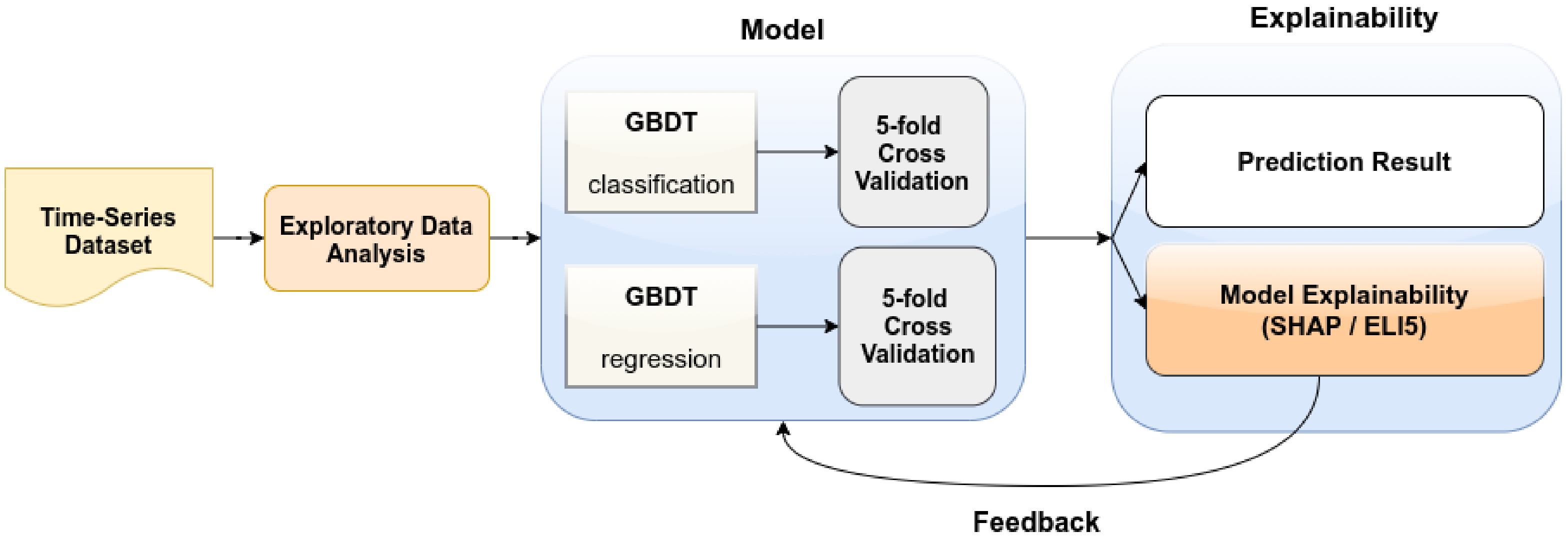


Fig: Model flow diagram for prediction framework

Estimation Algorithm

Algorithm 1: Predict order fulfillment for future time-horizon

Input: $X_{prediction}, model_{classification}, model_{regression}$

Result: Prediction frame with magnitude and probability

foreach *record* $\in$ *dataframe* **do**

  magnitude $\leftarrow$ model$_{regression}$

  prob $\leftarrow$ model$_{classification}$

  class $\leftarrow$ model$_{classification}$

  if *prob* $\leq$ *threshold* & *class* = fullfill **then**

    prob $\leftarrow$ *100 - prob*

    class $\leftarrow$ not-fullfill

  else if *class* = fullfill & *magnitude* $\leq$ $margin$ **then**

    prob $\leftarrow$ $model_{regression}$

    class $\leftarrow$ not-fullfill

  end

Results

General Dataset Overview

Document	Detail
Forecast	Forecast for a material at a plant for the week
Production Ratio	Planned vs Actual for a material at a specific location
Ordered	Quantity of product ordered
Shipped	Shipped quantity
Inventory	Resources in stock per week for specific product per location

Table: General Dataset Overview that applies to majority of supply chain industries

Correlation Analysis

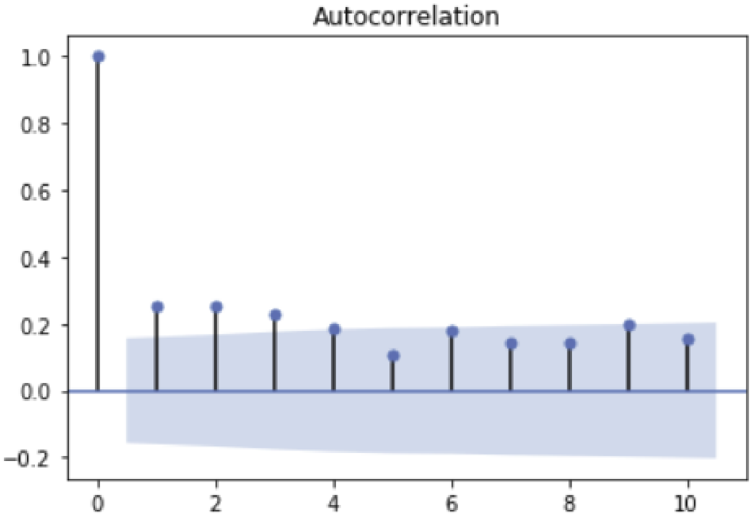


Fig: ACF for Order Quantity

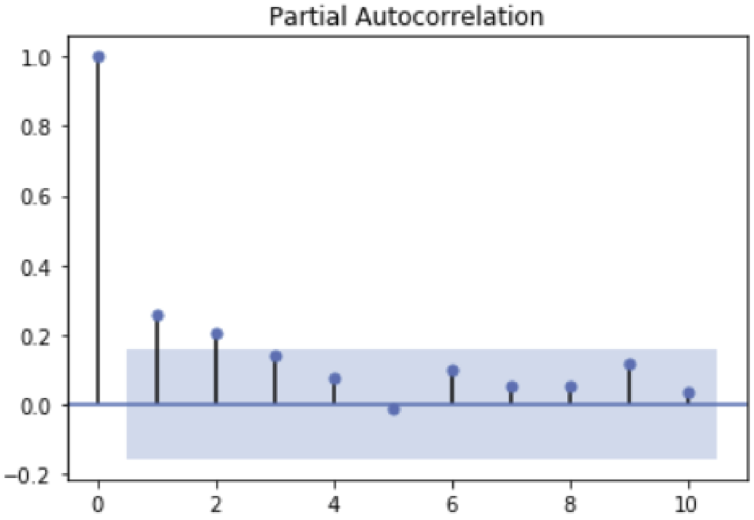


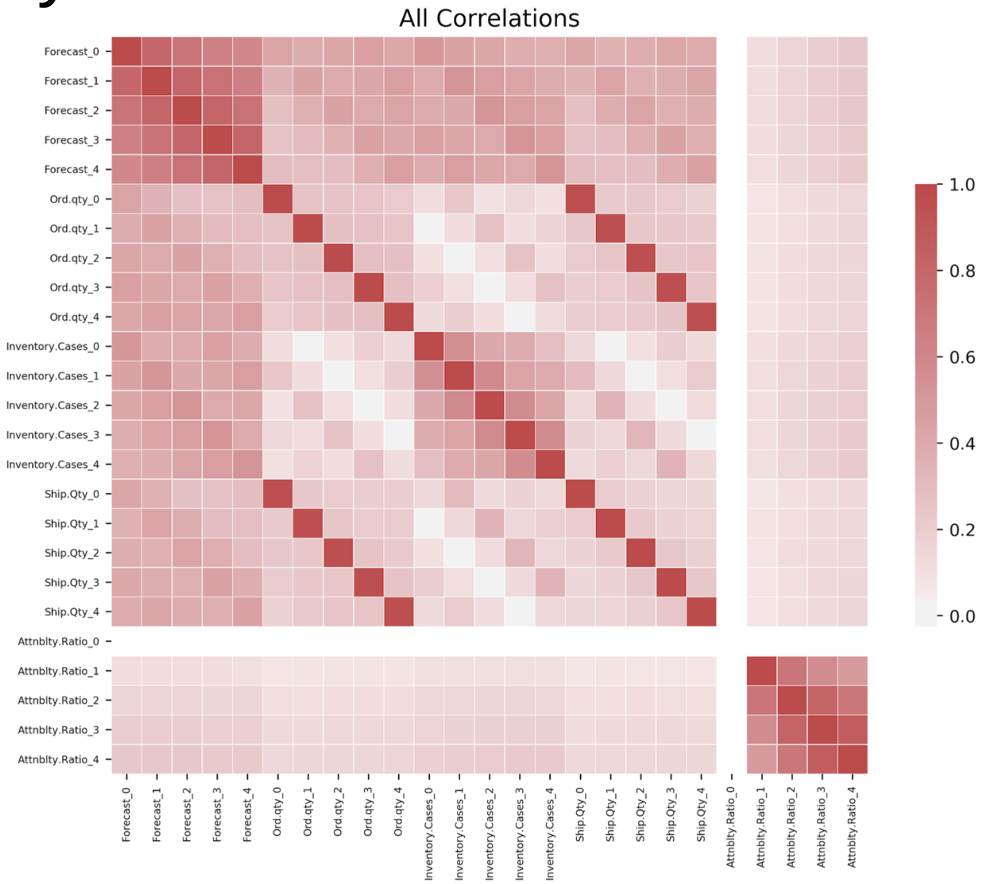
Fig: PACF for Forecast Quantity

Cross-Correlation Analysis

The following chart shows the interaction among the features. Lag of 4 week was considered.

For example, Inventory.Cases_1 (Inventory Case with 1 week lag) has higher interaction effect with Ship.qty_0 (Ship.qty in the present week).

This is apparent since a higher inventory in the previous week would direct affect the present week’s sales.



Impact of individual features

The following explanation for a week in the future time frame which was predicted as a failure case.

As depicted in the figure, due to high Order.qty_0 (for said week) but low Inventory.Cases_1(for previous week), Forecast_1(for previous week) and Attnblty.Ratio_0(for said week), the predicted outcome was a failure.

The show magnitude helps in understanding the impact over the outcome.

[*Note : Please ignore the Plant, <BIAS> and others as there would be parameters which cannot be changed.*]

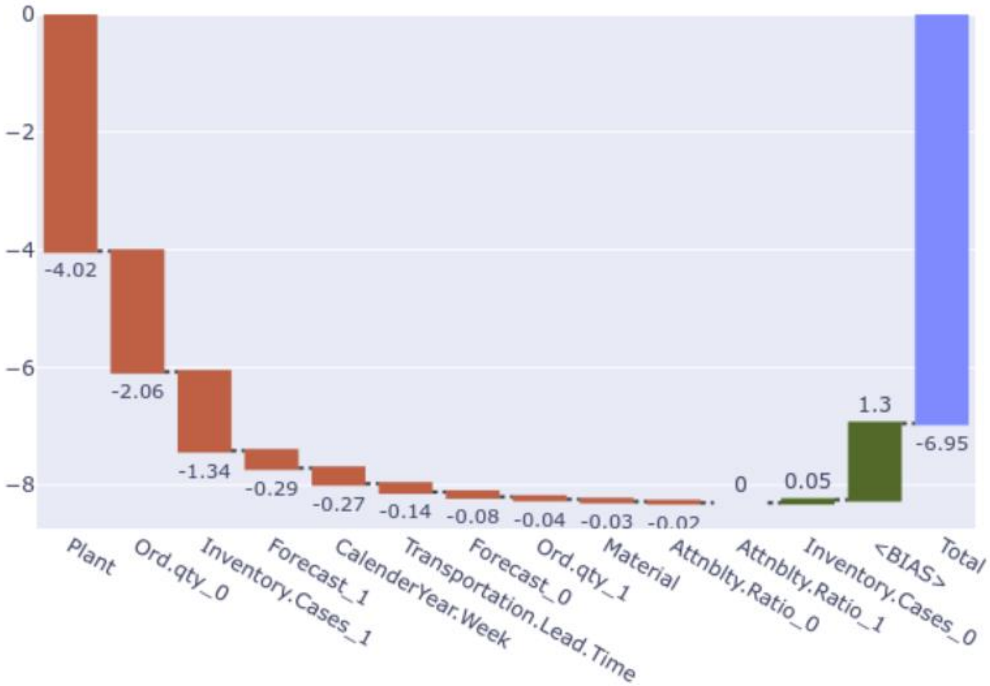


Fig: Explanation for a prediction failure in future

Framework Evaluation

		PREDICTED LABEL					
		LightGBM		XGBoost		CatBoost	
		Fail	Pass	Fail	Pass	Fail	Pass
TRUE	Fail	**4460**	428	4209	679	4345	543
LABEL	Pass	467	**4319**	1007	3779	737	4049

Table 1: Confusion Matrix

	F_1	ROC	Precision	Recall
LightGBM	**0.91**	**0.91**	**0.71**	**0.95**
XGBoost	0.82	0.83	0.70	0.94
CatBoost	0.86	0.87	0.69	0.92

Table 1: F1 and AUC SCORE

Evaluation Model

Confusion Matrix / AUC / F1 Score

	F_1	ROC	Precision	Recall
GBDT	**0.91**	**0.91**	**0.89**	**0.91**
RF	0.68	0.76	0.99	0.51
DART	0.89	0.89	0.88	0.89

Table 3: F1 and AUC SCORE for different decision tree models

Conclusion

Our findings can be stated as below

- LightGBM with GBDT objective are simplest and easiest frameworks to handle data invariance.
- Shapley values, Correlational analysis and Decision tree models offer basic understandable models.
- XGBoost provides best precision however it is very sensitive to parameter tuning
- Simpler and faster models, such as LightGBM with GBDT, helps reduce implementation cost and increases responsiveness for decision makers

Few of the future works could be:

- Improve explainability of the model
- Continuous learning
- Include graph based network modeling

In collaboration with Rich Products Corporation and supported by MITACS grant (IT16503).

References

- C. Koch, “Nike rebounds: How (and why) nike recovered from its supply chain disaster,” 2004
- S. S. Joyjeet Das, “Walgreen cfo’s departure due to $1 bln forecasting error - wsj,” 2014
- Jocelyn D'Souza, A Quick Guide to Boosting in ML, Medium 2018
- Sunil Ray, Quick Introduction to Boosting Algorithms in Machine Learning, Analytics Vidhya, 2015
- Daniel Chepenko, Introduction to gradient boosting on decision trees with Catboost, Medium, 2019
- Cory Maklin, Gradient Boosting Decision Tree Algorithm Explained, Medium, 2019
- S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,”, NIPS 2017
- Y. Zhang, Z. Zhao, and J. Zheng, “Catboost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of northern china,” Journal of hydrology (Amsterdam), vol. 588, p. 125087, 2020.
- T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system.” ACM, 2016, pp. 785–794.
- L. S. Shapley, A VALUE FOR n-PERSON GAMES. Princeton University Press, 2016, p. 307

Thank you