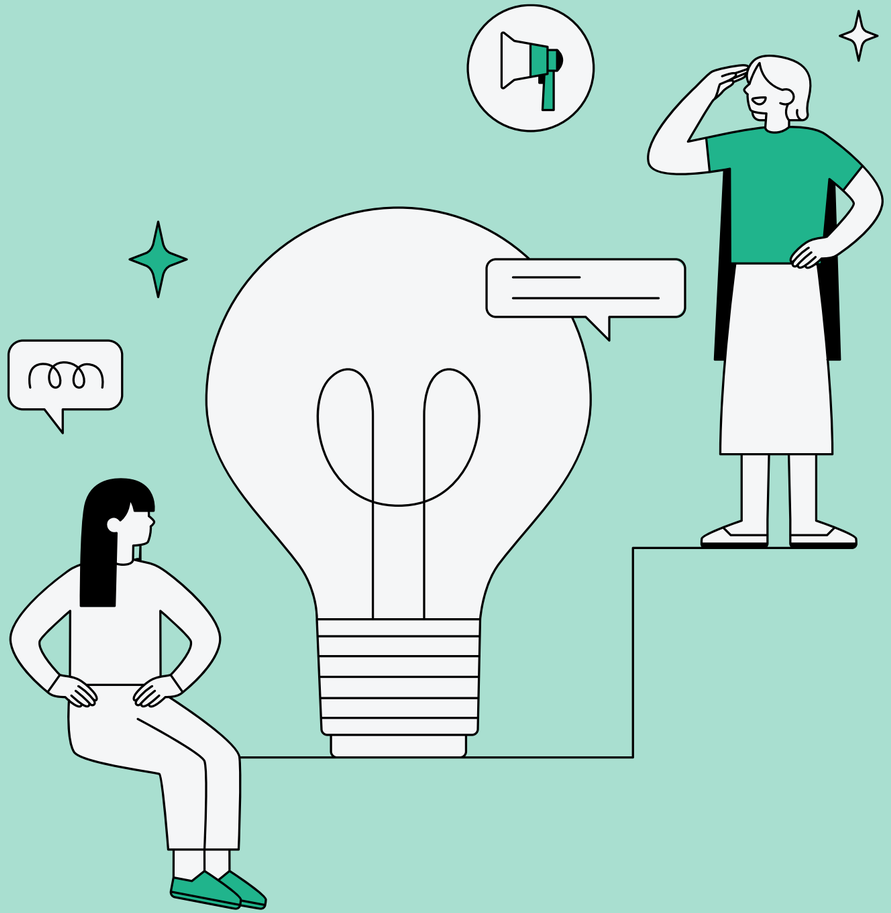
Presented by Satrio Bayu Tirto L (DS24)

BackOrder Detection

For supply chain industry

Objective

Back orders are an important indicator to guarantee the quality and quantity of stock required. So a backorder model is needed that can detect needs that occur in the field quickly so that there are no obstacles in the supply chain.



Main challenges identified

01. Question

- What columns affect BackOrder status?
- What is the condition of the data source?
- Is data modification necessary?

02. Modeling

- Determine the target score you want to achieve?
- What methods are needed to improve model performance

03. Launch

- Is the model used easy to use for user?

Workflow

Data shape :
Row:1.687.861
Collumn:23

Cleaning

Remove missing row,drop forcasting column and impute missing lead time

EDA

Evaluate data, if there weird value change or drop, make analysist

Modeling

Test possible combination for best peformance

Evaluation

Launch

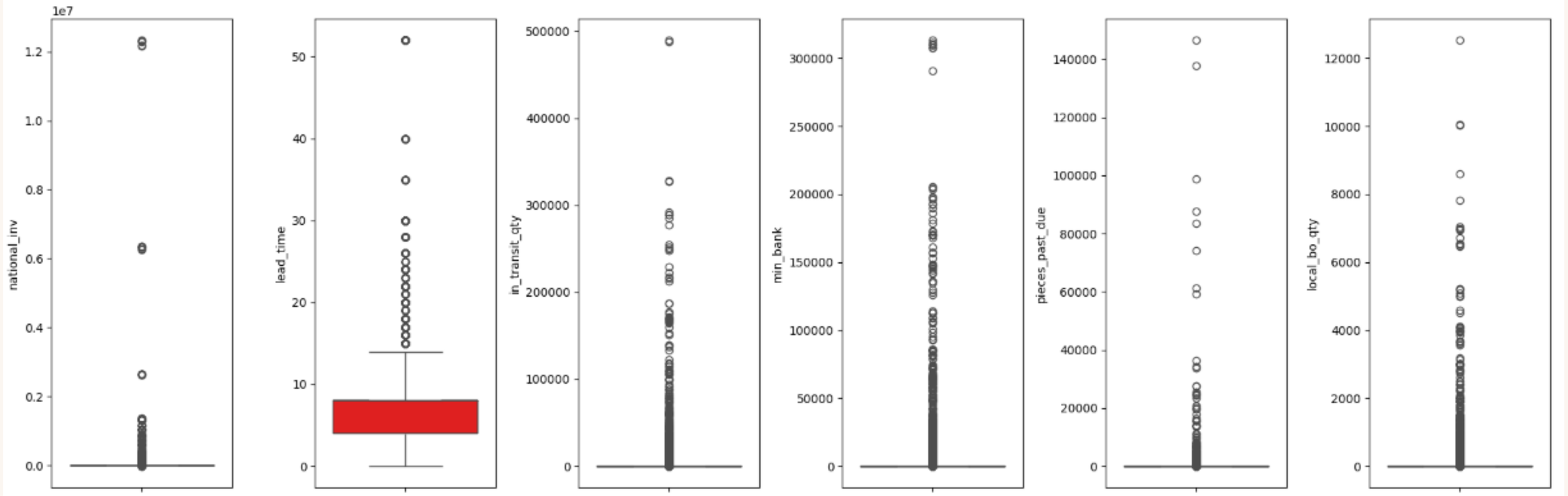
Corelation

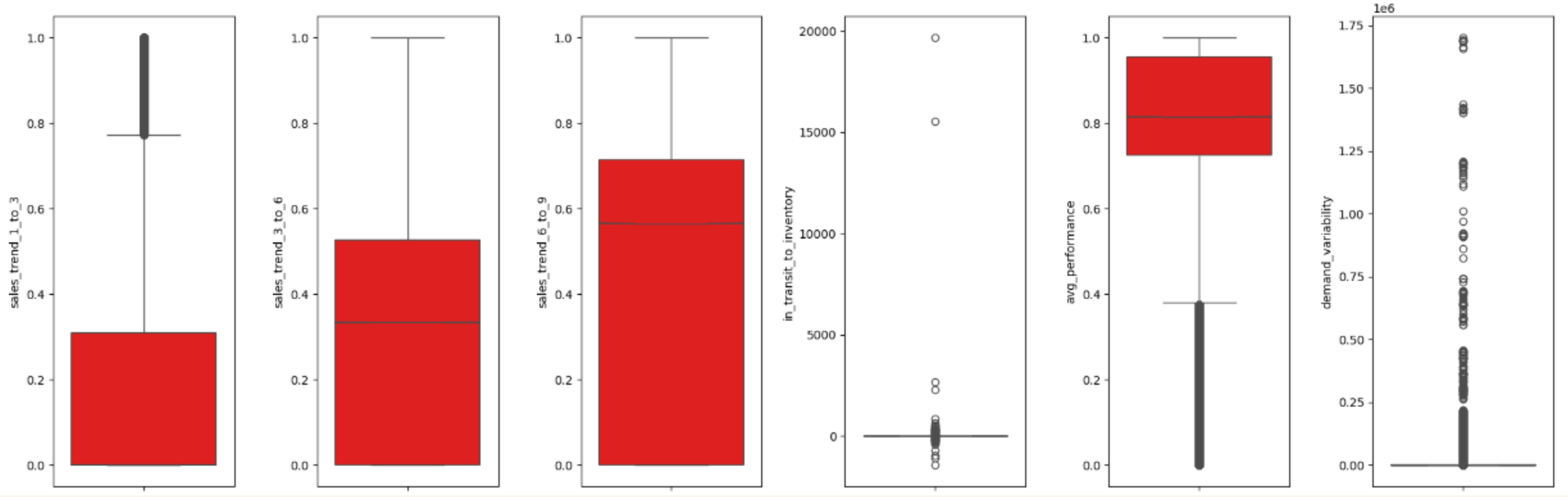
No	feature	vif_score
1	Product actually went on backorder	1.001003
2	Current inventory level for the part	1.541092
3	Transit time for product (if available)	1.074840
4	Amount of product in transit from source	2.646855
5	Sales quantity for the prior_1_month	7.385064
6	Sales quantity for the prior_3_month	48.270187
7	Sales quantity for the prior_6_month	62.490407
8	Sales quantity for the prior_9_month	20.738407
9	Minimum recommend amount to stock	6.552367
10	Parts overdue from source	1.138571
11	Source performance for prior 6 month	16.342734
12	Source performance for prior 12 month	16.338935
13	Amount of stock orders overdue	1.009610
14	Source issue for part identified	1.000607
15	Deck risk(Supply Issue)	1.082952
16	Original Equipment constraint (Production issue)	1.000409
17	Production Part Approval Process risk	1.003414
18	Stop auto buy	1.137085
19	Revision stop	1.003911

Action:

- Treshold = 10
- Combine Sales quantity become 1 feature (demand)
- Make Sales quantity trend
- Combine Performance
- Make rautio in transit with inventory

OUTLIER


1e7
national_inv
lead_time
in_transit_qty
min_bank
pieces_past_due
local_bo_qty


sales_trend_1_to_3
sales_trend_3_to_6
sales_trend_6_to_9
in_transit_to_inventory
avg_performance
1e6
demand_variability

Descriptive Analyst

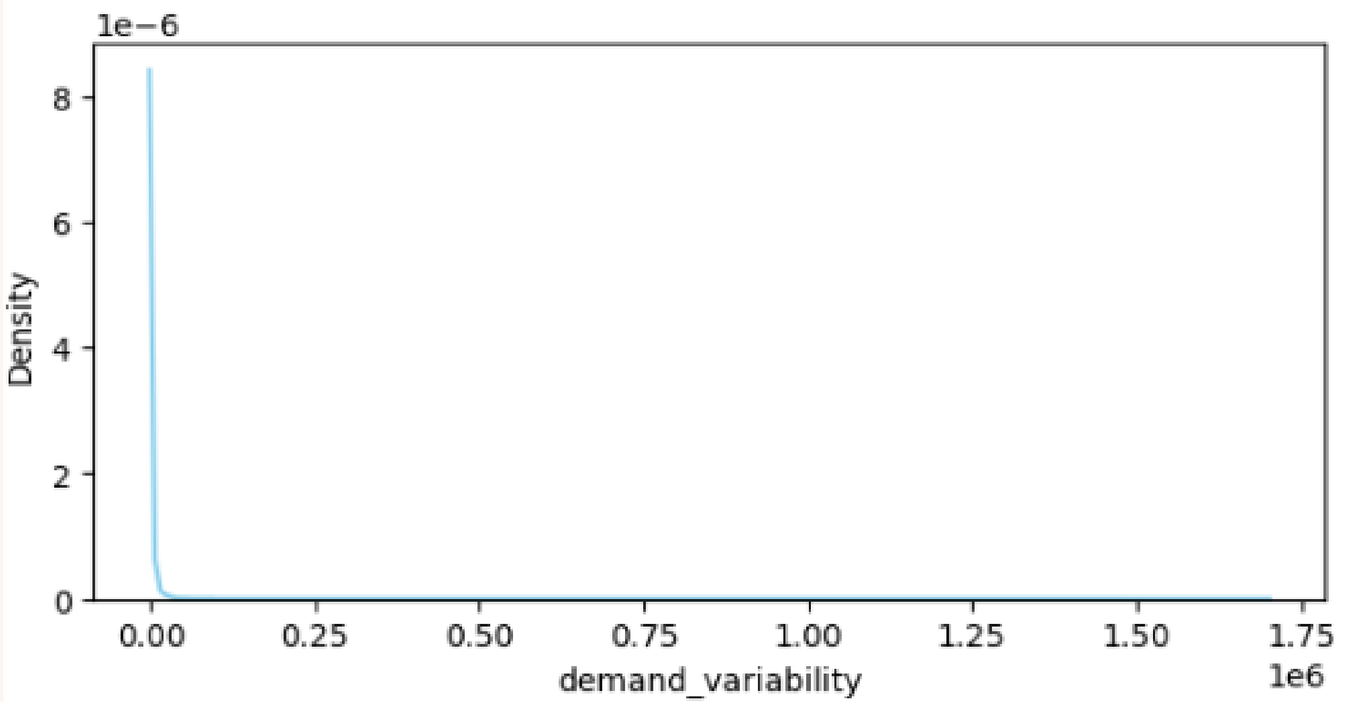
	Current inventory	Transit time	product in transit	sales_1_month	sales_3_month	sales_6_month	sales_9_month	Minimum Stock	Parts overdue from source	perf_6_month_avg	perf_12_month_avg	stock orders overdue
count	1,69E+12	1,59E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12	1,69E+12
mean	4,96E+08	7,87E+06	4,41E+07	5,59E+07	1,75E+08	3,42E+08	5,25E+08	5,28E+07	2,04E+06	-6,87E+06	-6,44E+06	6,26E+05
std	2,96E+10	7,06E+06	1,34E+09	1,93E+09	5,19E+09	9,61E+09	1,48E+10	1,25E+09	2,36E+08	2,66E+07	2,58E+07	3,37E+07
min	-2,73E+10	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	-9,90E+07	-9,90E+07	0,00E+00
25%	4,00E+06	4,00E+06	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	0,00E+00	6,30E+05	6,60E+05	0,00E+00
50%	1,50E+07	8,00E+06	0,00E+00	0,00E+00	1,00E+06	2,00E+06	4,00E+06	0,00E+00	0,00E+00	8,20E+05	8,10E+05	0,00E+00
75%	8,00E+07	9,00E+06	0,00E+00	4,00E+06	1,50E+07	3,10E+07	4,70E+07	3,00E+06	0,00E+00	9,70E+05	9,50E+05	0,00E+00
max	1,23E+13	5,20E+07	4,89E+11	7,42E+11	1,11E+12	2,15E+12	3,21E+12	3,13E+11	1,46E+11	1,00E+06	1,00E+06	1,25E+10

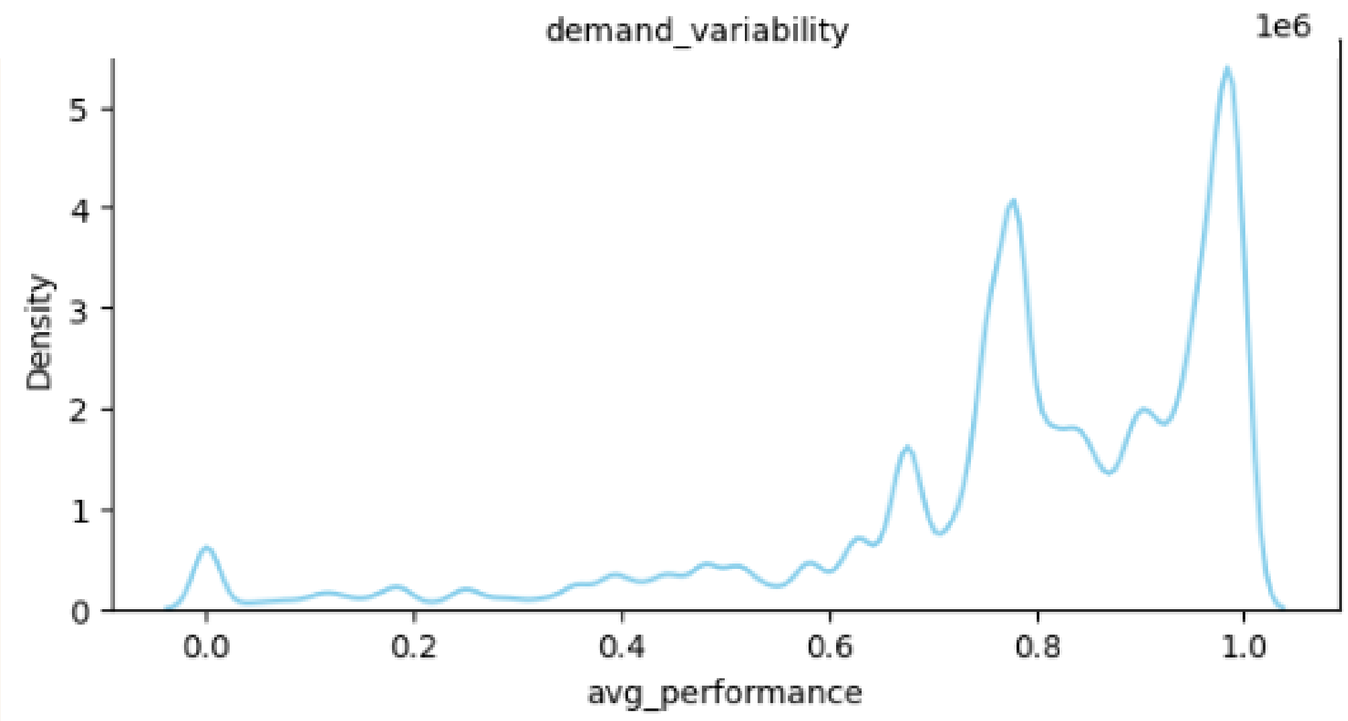
Note

- There is negative value in performance and not reasonable
- Negative value in Curent Inventory(National Invent) reasonable

Distrbution Hlstogram

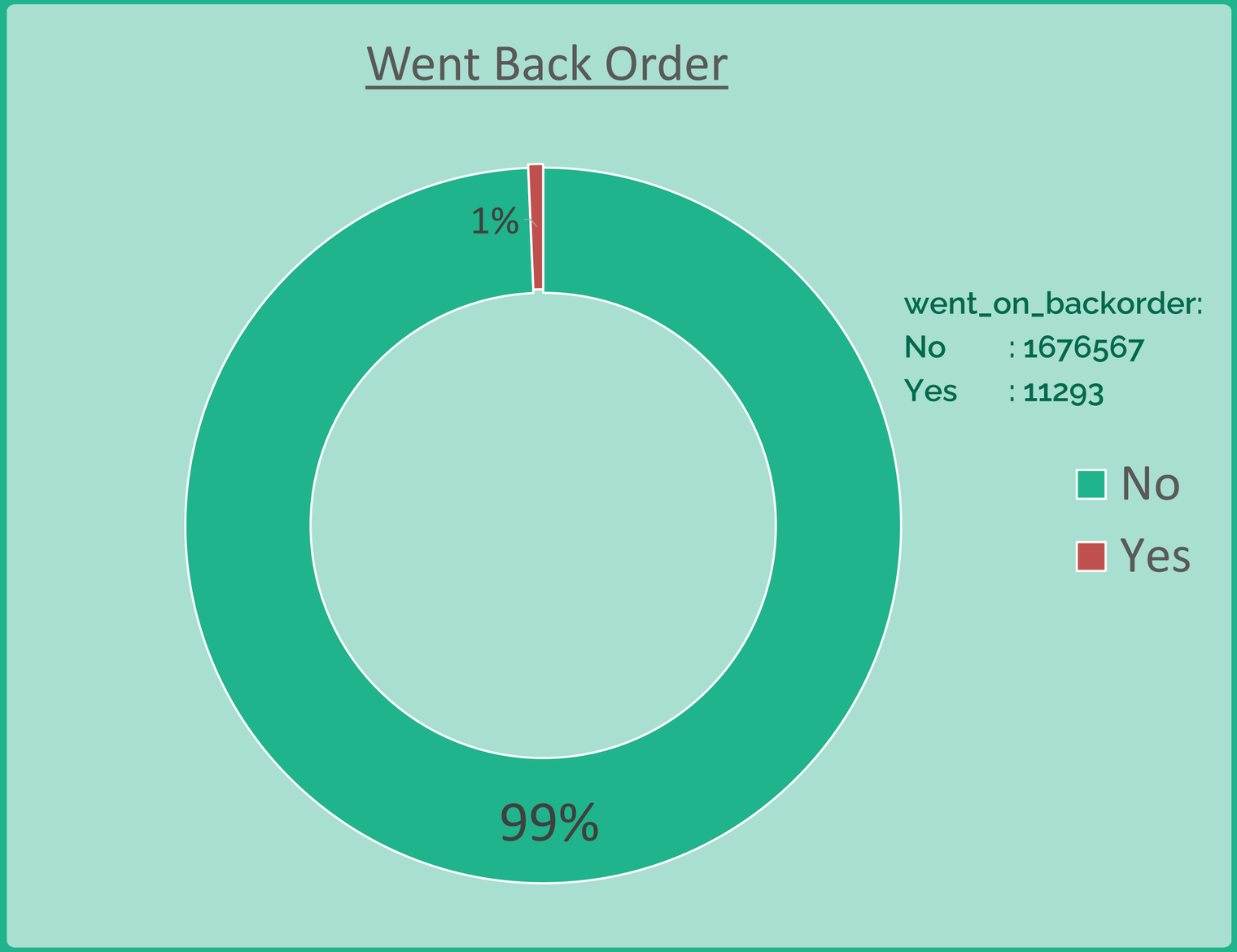
There is very high zero value in demand (sales quantity).
The performance is good





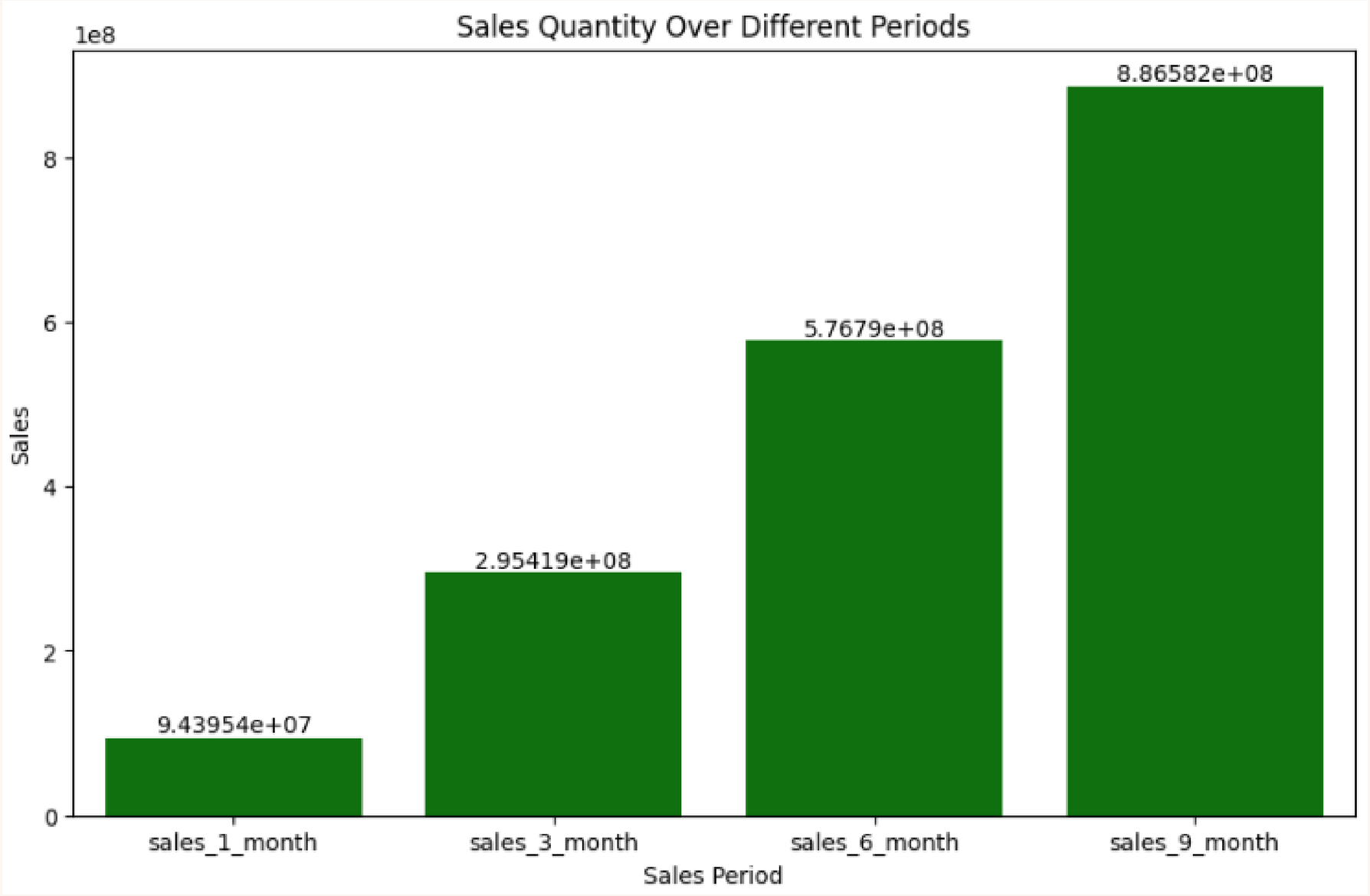
Back Order Analysist

There very high inbalance In target data



Total Sales Quantity

Value from bar chart is value from summing previous month demand with new value



Analysist risk to back order

Back order		No		Yes		Total
variable	value	sum	Percent	sum	Percent	
Deck risk	No	1291000	99.27%	9377	0.72%	1300377
	Yes	385567	99.50%	1916	0.49%	387483
Original Equipment constraint	No	1676330	99.33%	11285	0.66%	1687615
	Yes	237	96.73%	8	3.26%	245
Potential issue	No	1675711	99.33%	11242	0.66%	1686953
	Yes	856	94.37%	51	5.62%	907
Production Part Approval Process risk	No	1474492	99.35%	9534	0.64%	1484026
	Yes	202075	99.13%	1759	0.86%	203834
Revision stop	No	1675836	99.33%	11293	0.66%	1687129
	Yes	731	100%	0	0%	731
Stop auto buy	No	60615	99.22%	471	0.77%	61086
	Yes	1615952	99.33%	10822	0.66%	1626774

Note:
Top 3 issue : Stop auto buy, Deck risk.
Prodcut Part Aproval

Machine Learning

- For model, we will use Randomforest Classification and XGBoost.
- For the evaluation matrix, we will emphasize the recall value (detection ability) rather than precision

Base Model

Random Forest				
	Precision	Recall	F1-Score	SUpport
FALSE	0.99	1	1	502885
TRUE	0.8	0.21	0.33	3473
Accuracy	0.99			506358
Macro avg	0.9	0.6	0.66	506358
Weighted avg	0.99	0.99	0.99	506358

XGBoost				
	Precision	Recall	F1-Score	SUpport
FALSE	0.99	1	1	502885
TRUE	0.68	0.08	0.14	3473
Accuracy	0.99			506358
Macro avg	0.84	0.54	0.57	506358
Weighted avg	0.99	0.99	0.99	506358

Add Feature & Balancing

Random Forest				
	Precision	Recall	F1-Score	SUpport
FALSE	0.99	1	1	502885
TRUE	0	0	0	3473
Accuracy	0.99			506358
Macro avg	0.5	0.5	0.5	506358
Weighted avg	0.99	0.99	0.99	506358

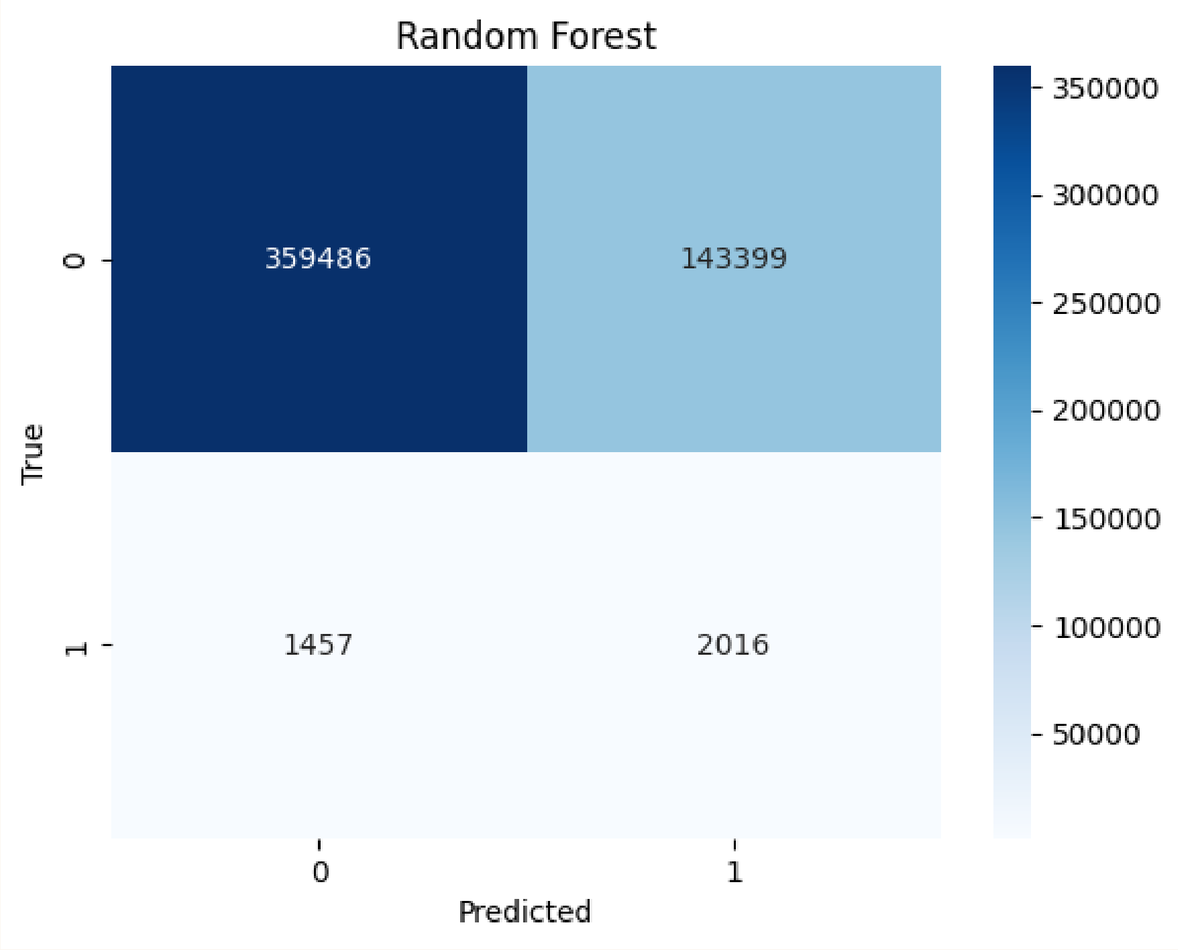
XGBoost				
	Precision	Recall	F1-Score	SUpport
FALSE	0.99	1	1	502885
TRUE	0.04	0.08	0.05	3473
Accuracy	0.98			506358
Macro avg	0.51	0.53	0.52	506358
Weighted avg	0.99	0.98	0.98	506358

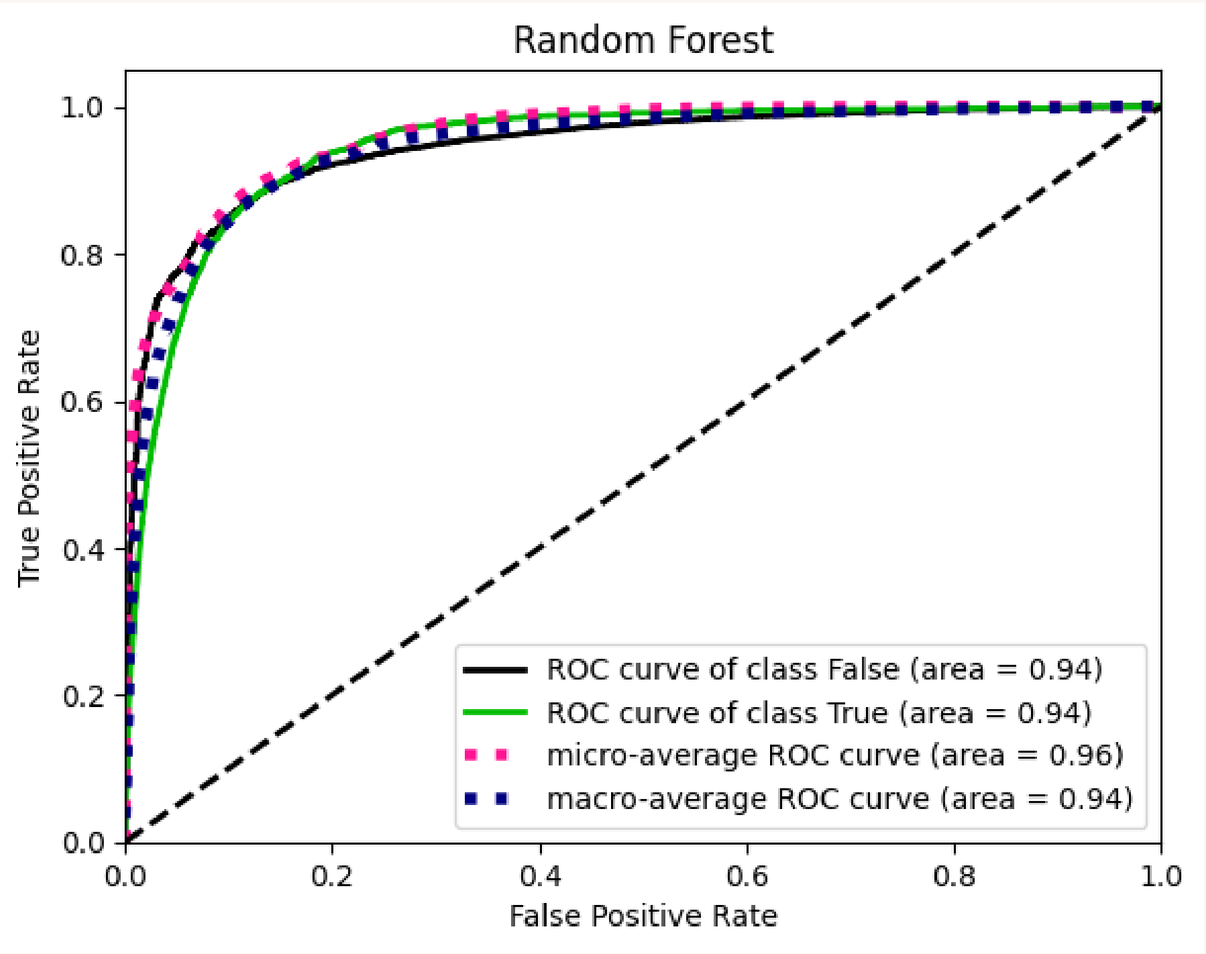
Tunning

Random Forest				
	Precision	Recall	F1-Score	SUpport
FALSE	1	0.71	0.83	502885
TRUE	0.01	0.58	0.03	3473
Accuracy	0.71			506358
Macro avg	0.5	0.65	0.43	506358
Weighted avg	0.99	0.71	0.83	506358

XGBoost				
	Precision	Recall	F1-Score	SUpport
FALSE	1	0.59	0.74	502885
TRUE	0.01	0.64	0.02	3473
Accuracy	0.59			506358
Macro avg	0.5	0.62	0.38	506358
Weighted avg	0.99	0.59	0.74	506358

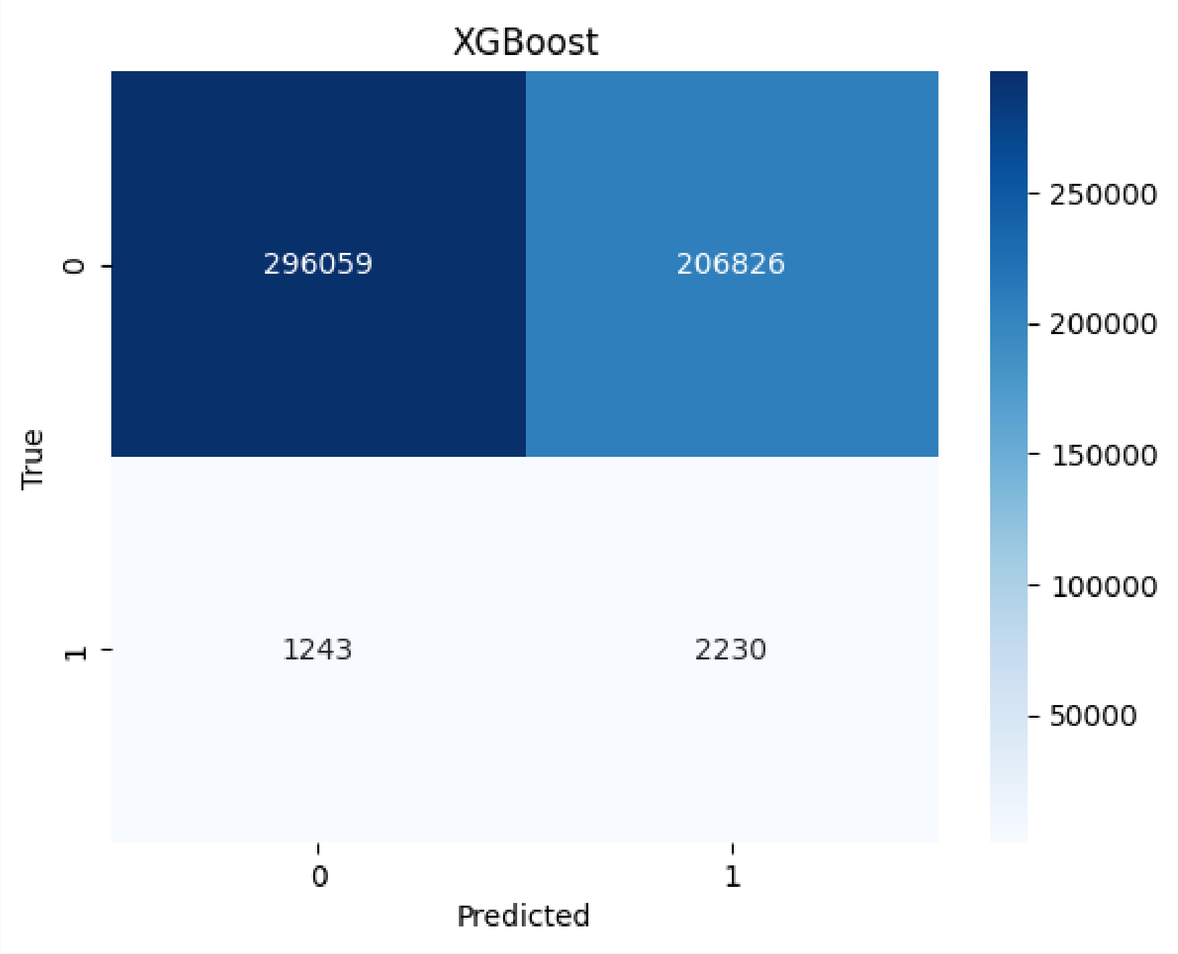
Random Forest Model

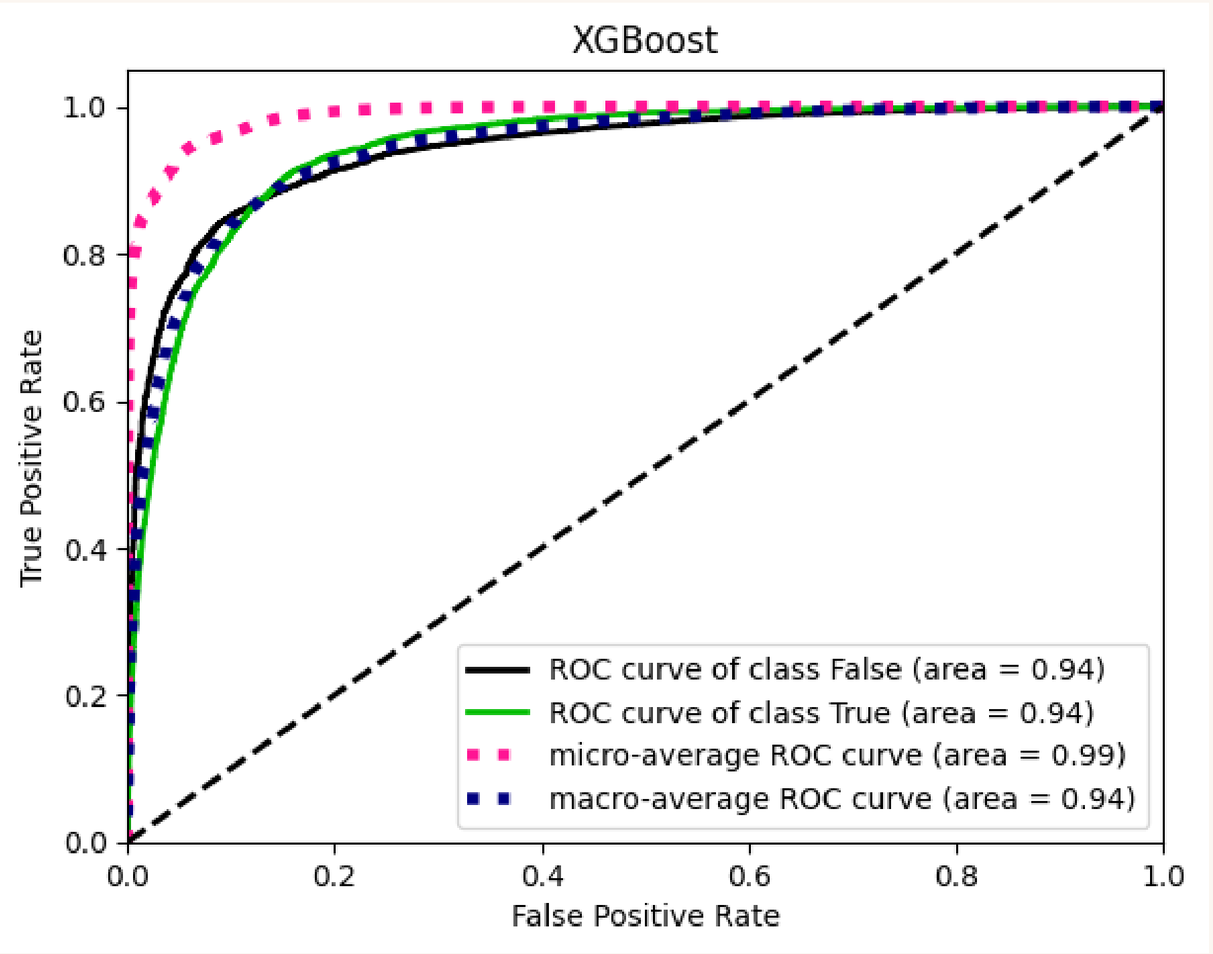




False Negative Rate for Random Forest:
0.4195220270659372

XGBoost Model

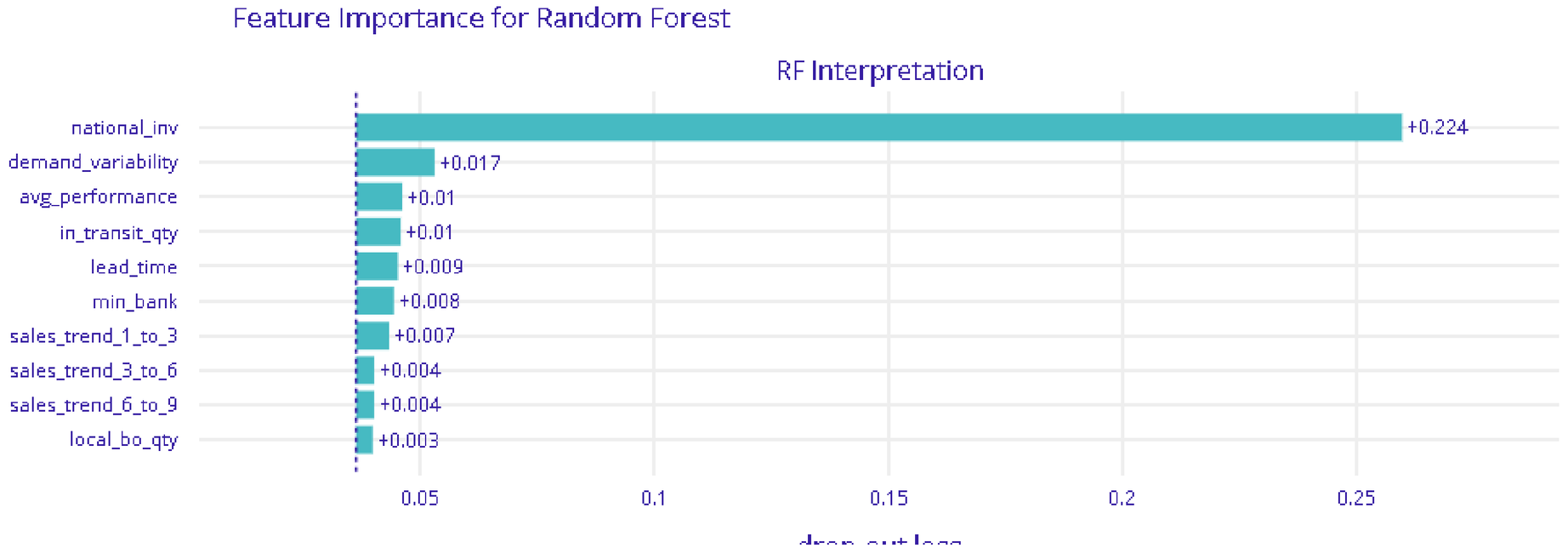


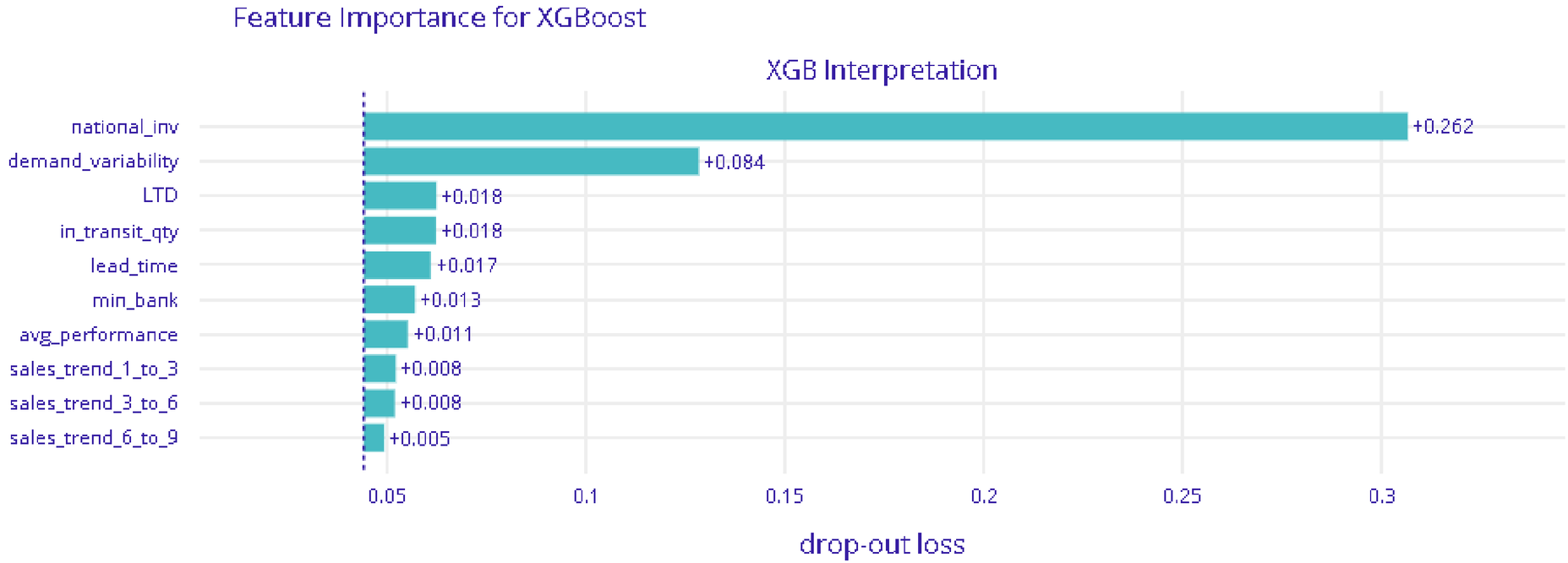


False Negative Rate for XGBoost:
0.35790382954218253

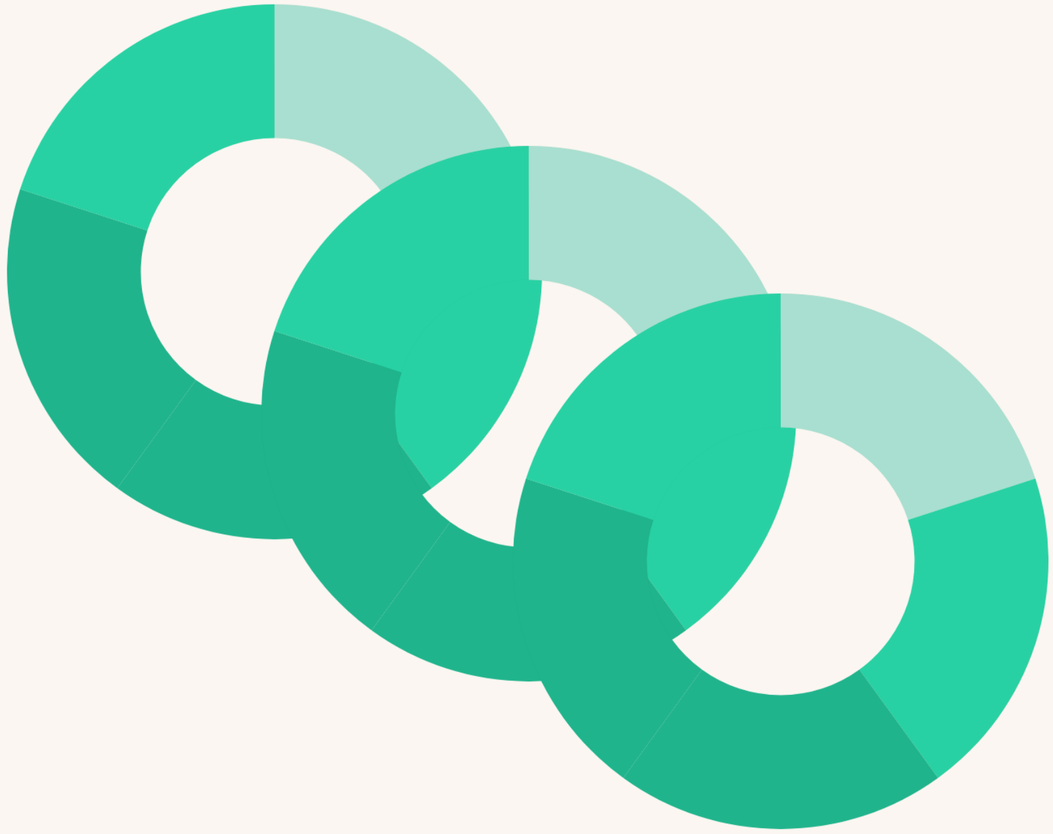
Feature Importance

Feature Importance
- Current inventory level
- Demand
- Amount of product in transit

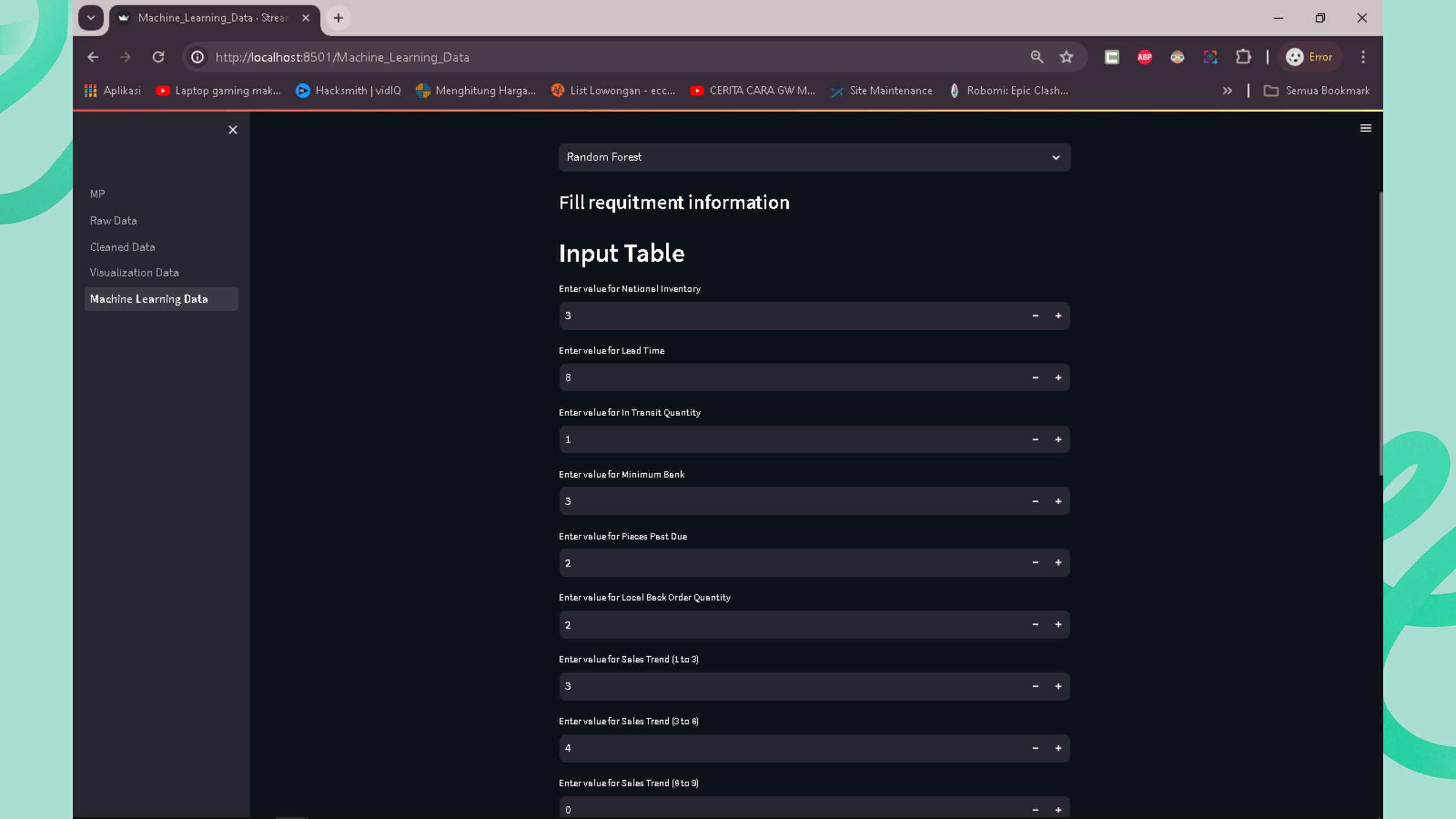


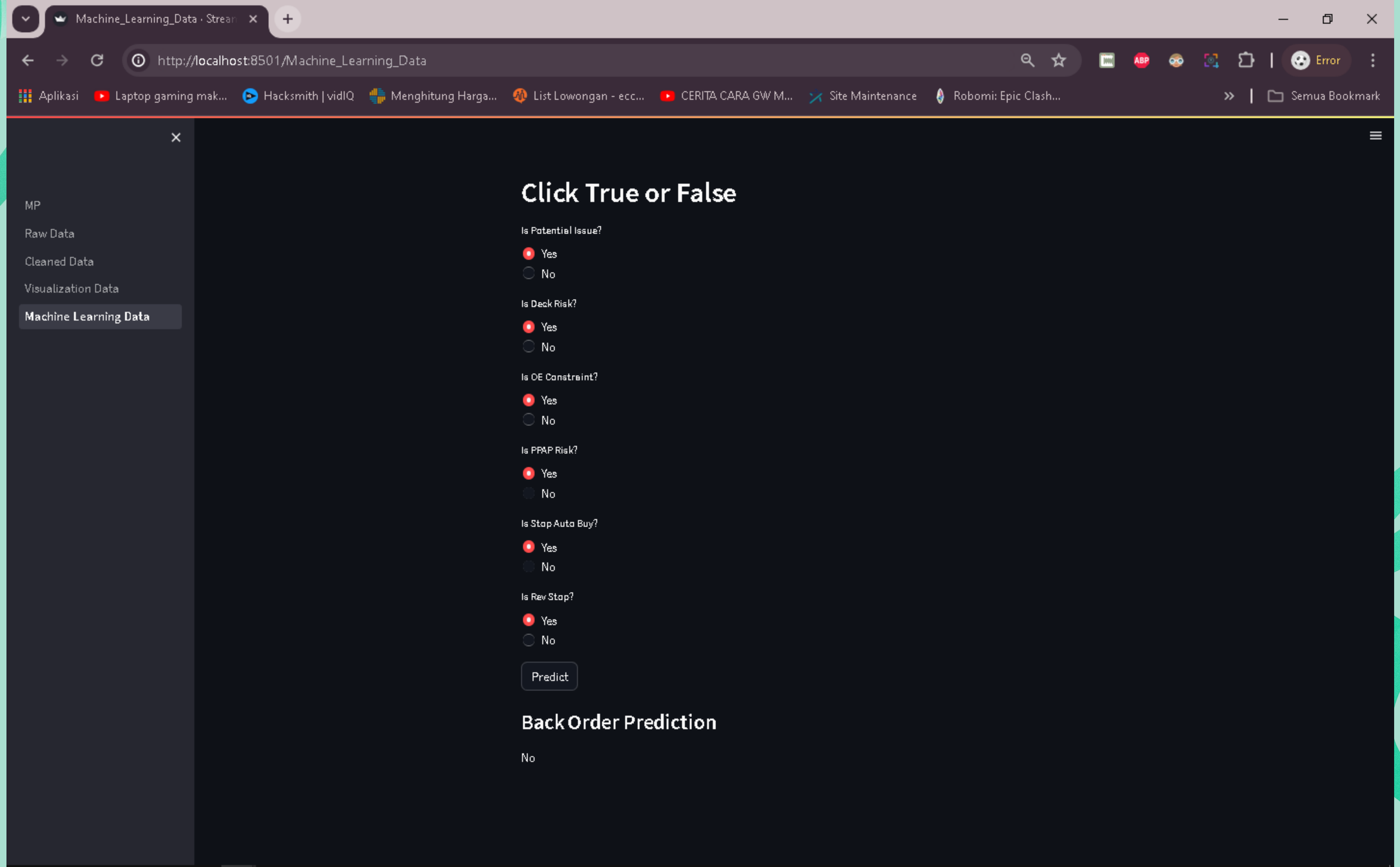


Deploying Test



For testing this model, we will use streamlit.
we will test from the point of view of user for this model,


Machine_Learning_Data · Strean
http://localhost:8501/Machine_Learning_Data
Error
Aplikasi
Laptop gaming mak...
Hacksmith | vidIQ
Menghitung Harga...
List Lowongan - ecc...
CERITA CARA GW M...
Site Maintenance
Robomi: Epic Clash...
Semua Bookmark
MP
Raw Data
Cleaned Data
Visualization Data
Machine Learning Data
Random Forest
Fill requitment information
Input Table
Enter value for National Inventory
3
Enter value for Lead Time
8
Enter value for In Transit Quantity
1
Enter value for Minimum Bank
3
Enter value for Pieces Past Due
2
Enter value for Local Back Order Quantity
2
Enter value for Sales Trend (1 to 3)
3
Enter value for Sales Trend (3 to 6)
4
Enter value for Sales Trend (6 to 9)
0


Machine_Learning_Data · Strean
http://localhost:8501/Machine_Learning_Data
Error
Aplikasi
Laptop gaming mak...
Hacksmith | vidIQ
Menghitung Harga...
List Lowongan - ecc...
CERITA CARA GW M...
Site Maintenance
Robomi: Epic Clash...
Semua Bookmark
MP
Raw Data
Cleaned Data
Visualization Data
Machine Learning Data
Click True or False
Is Potential Issue?
Yes
No
Is Deck Risk?
Yes
No
Is OE Constraint?
Yes
No
Is PPAP Risk?
Yes
No
Is Stop Auto Buy?
Yes
No
Is Rev Stop?
Yes
No
Predict
Back Order Prediction
No

Conclusion

1. Overall, the company's supply chain performance is very good, but we need to investigate further for products that need back orders, what lines are having problems?
2. Check further why many product don't auto buy
3. The import feature makes a lot of sense because it includes stock, but there are less visible potential issues and risks
4. For Model use XGBoost for recall but if need acuration use RandomForest
5. Too many features to include in streamlit. some had to be eliminated or downsized. This is no problem if it is automatically imported.

Thank you very much!

Dataset:https://data.world/amitkishore/can-you-predict-products-back-order
Notebook:https://colab.research.google.com/drive/122-u1mPWUhIT15qvqcKVrKZQjdxF2NIR?usp=drive_link