

Presentation

Y. Fan
06.05.2020

Presentation

01

Motivation

Why did I come to this topic?

02

Preparation

Where did the data come from?
Which tools were used?

03

Data Processing

How did I process the data to get valuable
information?

04

Data Modelling

How did I develop a data model to solve the
problem?

05

Evaluation

How is the performance of the data model?

06

Conclusion

What has been achieved? What still needs
to be improved?

01 Motivation

Backorder is a common problem and big challenge for companies.

- Canceled customer orders
- Decreased customer loyalty
- Extra time effort and cost

→ Backorder Predict Modell

02 Preparation

● Tool

Python (Libraries: Pandas, numpy, sklearn, seaborn...)

Machine Learning Technique: Classification (Logistic Regression)

Jupyter notebook & Github

● Data Source

Open source supply chain dataset:

<https://data.world/amitkishore/can-you-predict-products-back-order>

● Data Information

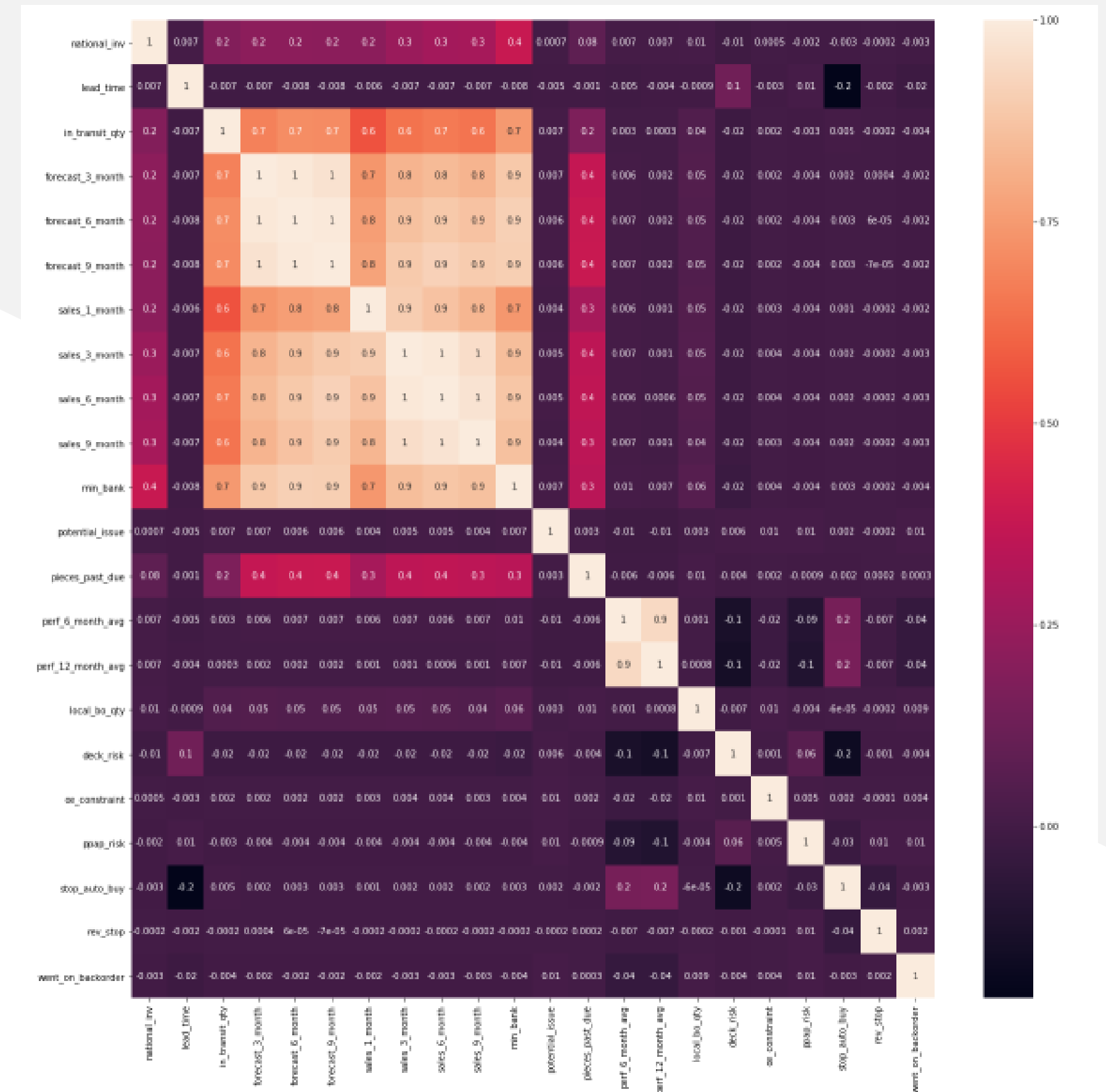
Volume: 1.929.937 rows, 23 columns

The dataset contains information such as: Part ID, inventory level for the part, average transit time for the part, sales forecast, sales history, potential issue, back order indicator etc.

sku	object
national_inv	float64
lead_time	float64
in_transit_qty	float64
forecast_3_month	float64
forecast_6_month	float64
forecast_9_month	float64
sales_1_month	float64
sales_3_month	float64
sales_6_month	float64
sales_9_month	float64
min_bank	float64
potential_issue	object
pieces_past_due	float64
perf_6_month_avg	float64
perf_12_month_avg	float64
local_bo_qty	float64
deck_risk	object
oe_constraint	object
ppap_risk	object
stop_auto_buy	object
rev_stop	object
went_on_backorder	object
dtype:	object

03 Data Processing

- Remove or merge duplicates
- Evaluate and handle missing data
- Convert categorical attributes to binary variables
- Data normalization
- Dimensionality Selection
- Data balancing



04 Data Modelling

Machine Learning Technique: Classification

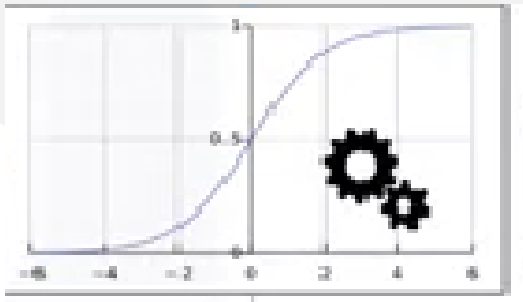
80% Training Set

	Feature 1	Feature 2	Feature 3	Feature 4	...	Feature N	Target
1	28	69	33	Yes	...	12	No
2	29	50	11	No	...	9	No
3	23	90	12	Yes	...	5	Yes
4	19	25	30	Yes	...	16	No
5	20	45	22	Yes	...	26	Yes

Modelling

	Feature 1	Feature 2	Feature 3	Feature 4	...	Feature N	Target
6	29	34	11	No	...	19	?

20% Testing Set

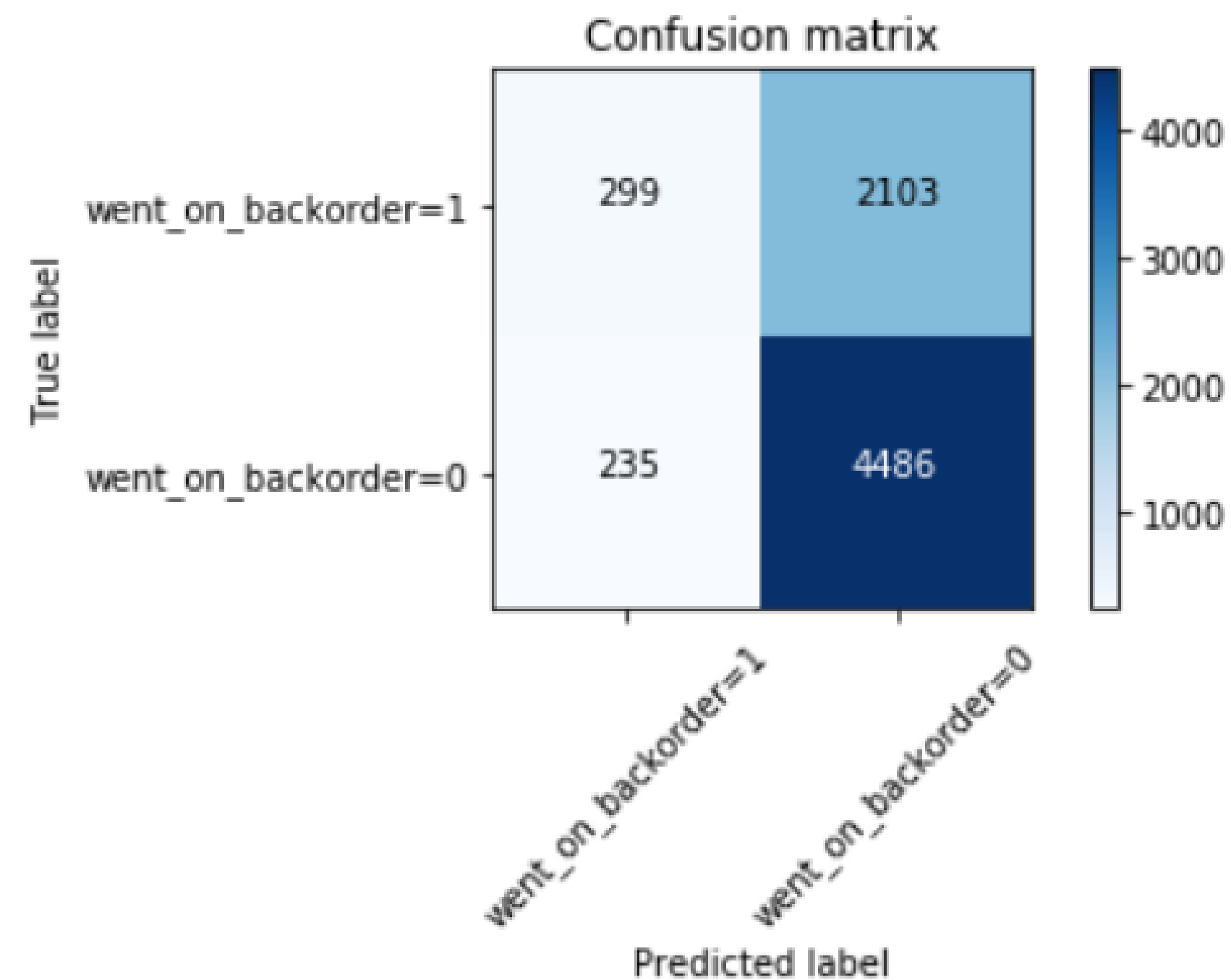


Classifier

Predicting

05 Evaluation

Data Model 1.0: include 14 Features



4486 samples:

Predict = Actual = No
Nothing will happen

299 samples:

Predict = Actual = Yes
Action can be taken to avoid backorder situation
→ Avoid order cancellation and extra cost

2103 samples:

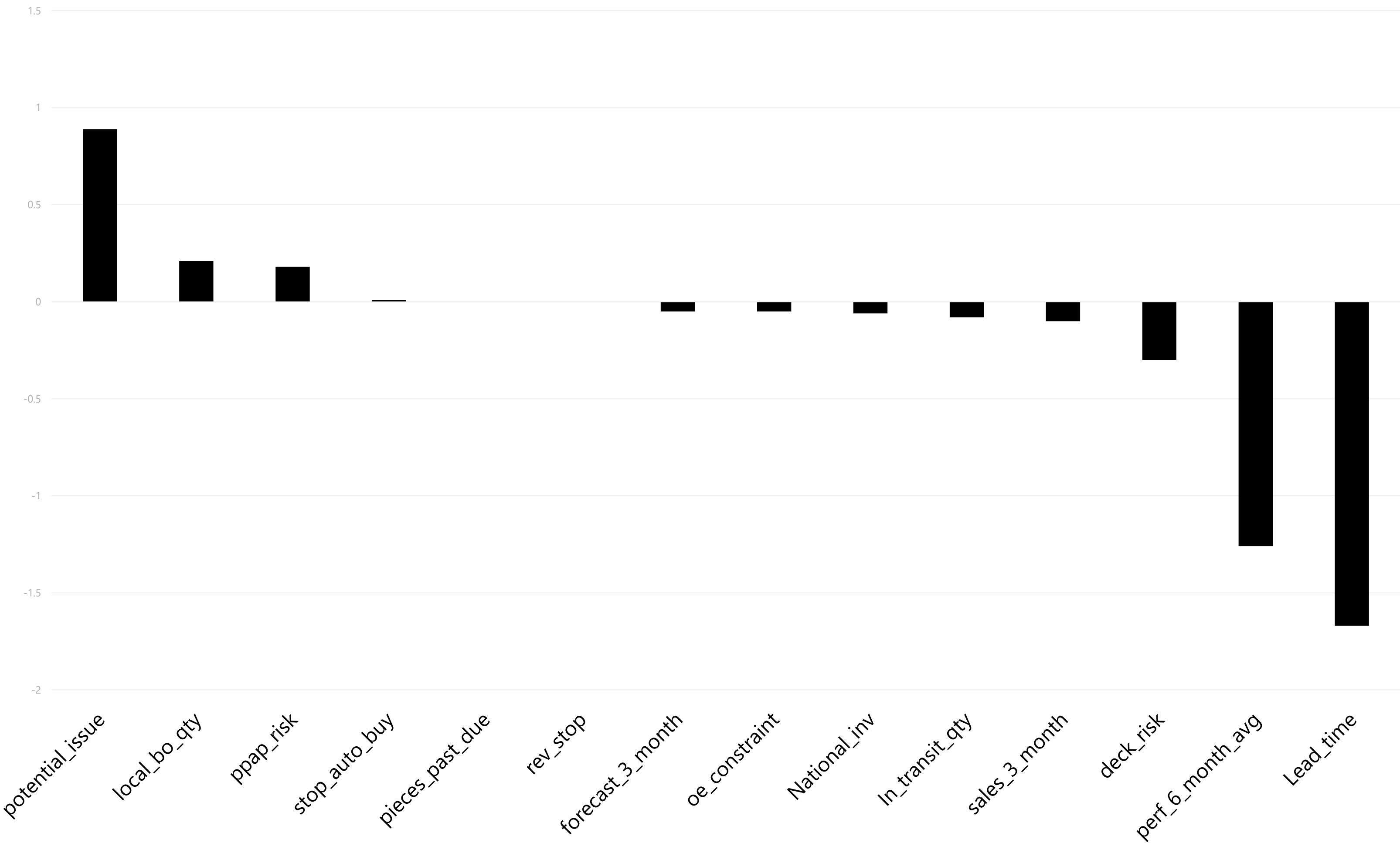
Predict = No
Actual = Yes
The situation has not been improved

235 samples:

Predict = Yes
Actual = No
Action can be taken to avoid backorder situation
→ Cause extra inventory cost

05 Evaluation

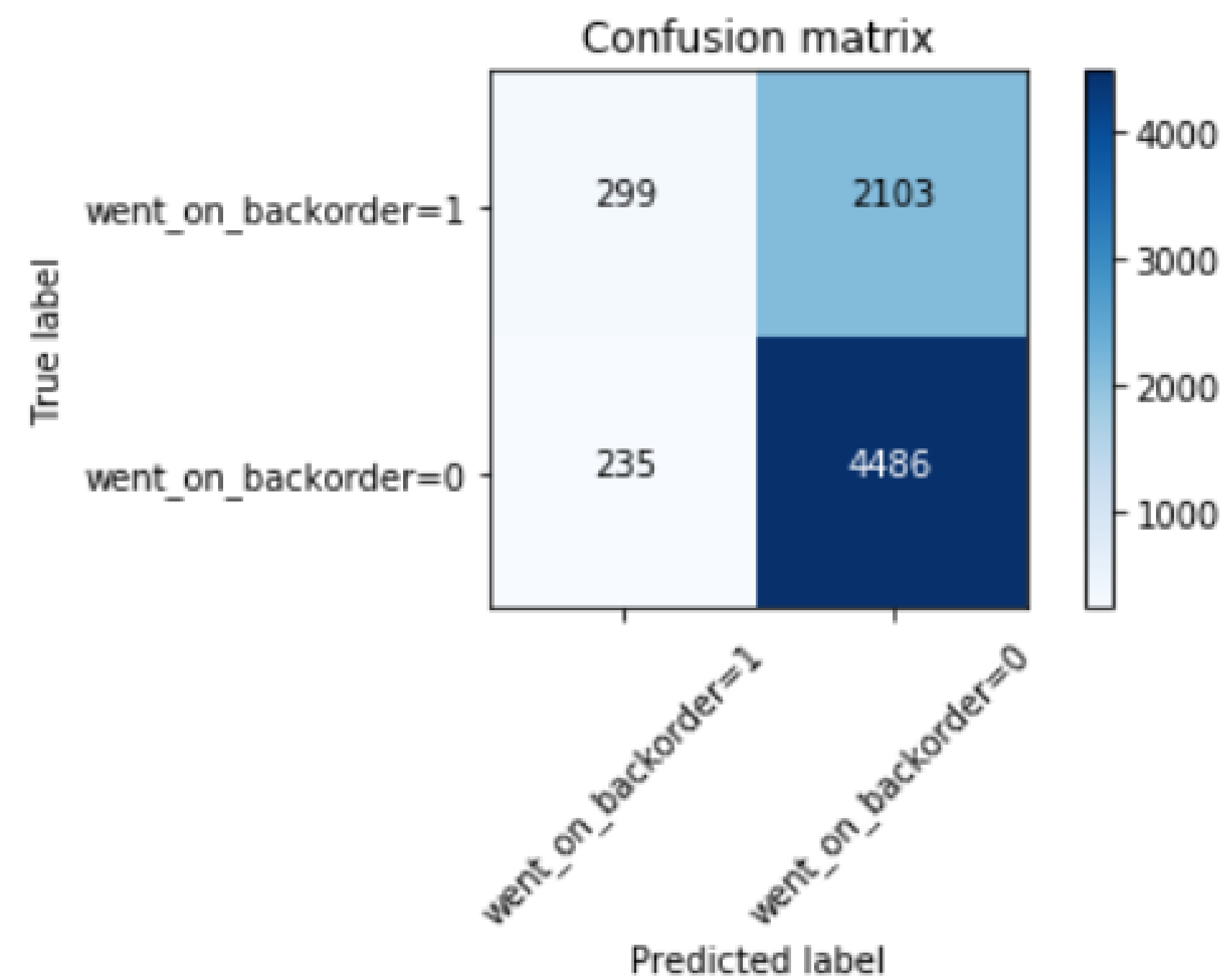
Select features according to importance and Rebuild Model



Top 7 Features
Lead_time
Perf_6_month_avg
Potential_issue
Deck_risk
Local_bo_qty
Ppap_risk
Sales_3_month

05 Evaluation

Data Model 2.0: include 7 Features



06 Conclusion

Pros

“Went on backorder” situation can be predicted to a certain degree to save time and cost in advance.

Cons

The model performances not that well in predicting the case, in which the “went-on-order” is “Yes”.

Reasons:

Insufficient data information due to the data imbalance.

Thank you!