BU7155
Business Data Mining

Data-Driven Optimisation of Supply Chain Management

Group 9 - Zihan Zheng, Xi Zhang, Chen Wang, Ning Xia and Yantong Xiang

November 2024

Business Problem

Key Objective

- Optimising demand forecasting capability
- Determine the key factor involved in this process
- Reducing holding cost while improving efficiency

Stakeholder

- Customer
- Supply Chain Partner
- Internal team within company's logistics and supply chain process

Opportunities vs Challenges

- Opportunities including enhancing efficiency and cost reduction in the supply chain and drive digital transformation initiatives
- Challenges are higher cost generated during the implementation and financial risks

Data Mining Problem

Goal: To optimise supply chain management by reducing order delays and enhancing supply chain capacity.

01

Analyse the factors that will result in an order delay

- **Type of Problem:** Binary Classification - Supervised, Explanatory
- **Model:** Random Forest
- **Outcome Variable:** Label (Delayed/On-time)

02

How can we predict the future order demand in order to optimise supply chain capacity

- **Type of Problem**: Time Series Forecasting - Supervised, Predictive
- **Model**: ARIMA (AutoRegressive Integrated Moving Average)
- **Outcome Variable**: Product Quantity

About the Data

- **Data source and size**: Logistics Supply Chain Real World Data from [Kaggle], 15549 rows, 41 columns. The dataset contains information related to the orders, customers, products, and transactions.
- **Output:** label - Delayed/On-time (-1 early arrival, 0 on time, 1 delayed)
- **Chosen features**: label, shipping_mode, order_region, category_name, order_item_total_amount, customer_state, customer_segment, department_name, payment_type
- **Data partitions**: Training (40%), Validation (30%), Test (30%)

shipping_mode	order_region	category_name	order_item_total_amount	customer_segment	customer_state	department_name	order_status	payment_type	label
Standard Class	Western Europe	Cardio Equipment	84.99157	Consumer	PR	Footwear	COMPLETE	DEBIT	-1
Standard Class	South America	Water Sports	181.99	Consumer	CA	Fan Shop	PENDING	TRANSFER	-1
Second Class	Western Europe	Indoor/Outdoor Games	93.81015	Consumer	PR	Fan Shop	COMPLETE	DEBIT	1
Second Class	Central America	Cleats	99.8906	Consumer	PR	Apparel	PROCESSING	TRANSFER	0
Standard Class	Central America	Water Sports	171.07587	Consumer	CA	Fan Shop	COMPLETE	DEBIT	1
Standard Class	East of USA	Electronics	145.46329	Consumer	PR	Footwear	CLOSED	CASH	1
Standard Class	West of USA	Indoor/Outdoor Games	167.99	Corporate	PR	Fan Shop	COMPLETE	DEBIT	1

Methods

RANDOM FOREST

- Suitable for dealing with complex multi-factor problems
- Effectively reduce the risk of overfitting
- Effectively handle noise, not easily affected by outliers
- High prediction accuracy

TIME SERIES ARIMA Model

- Flexible : Automatic adjustment various patterns
- High interpretability : easy to understand and explain the model behavior
- Simplicity: Model construction and parameter estimation are relatively simple

Evaluation

Random Forest

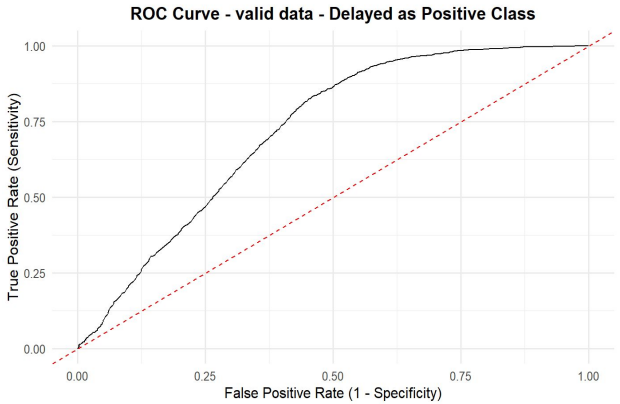
Validation

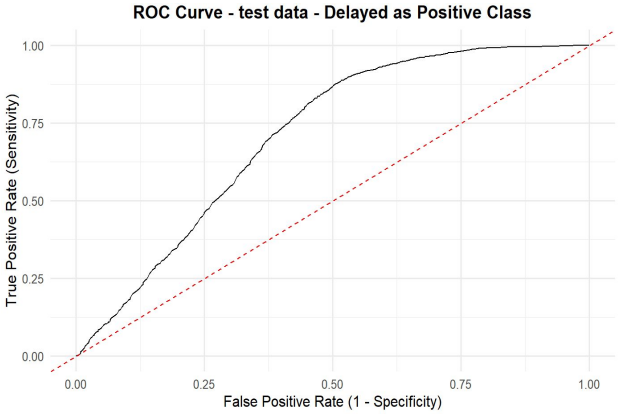
Prediction	Reference: Delayed	Reference: Not Delayed
Delayed	1531	403
Not Delayed	1162	1569

Test

Prediction	Reference: Delayed	Reference: Not Delayed
Delayed	1529	439
Not Delayed	1163	1532

Set	Accuracy	Precision	Recall	F1 Score	AUC
Validation	0.6645	0.7916	0.5685	0.6618	0.7197
Test	0.6564	0.7769	0.5680	0.6562	0.7103



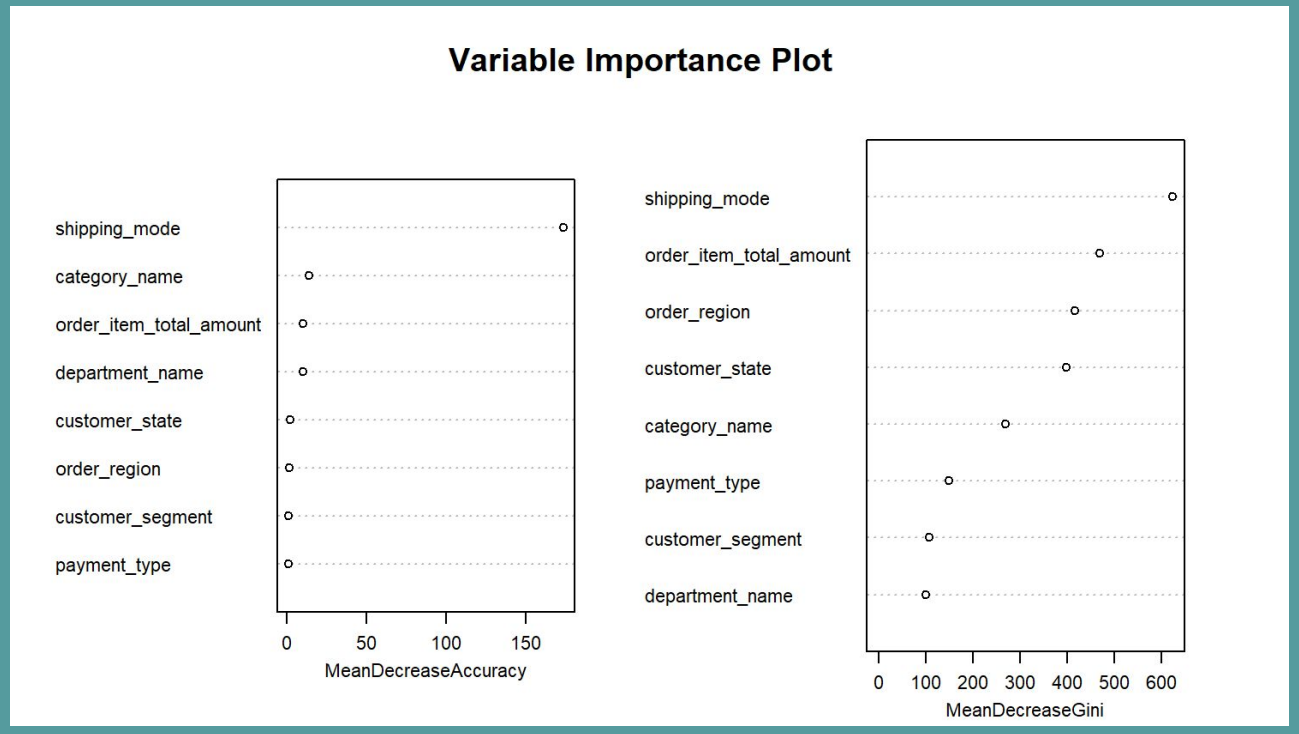


Time Series - ARIMA

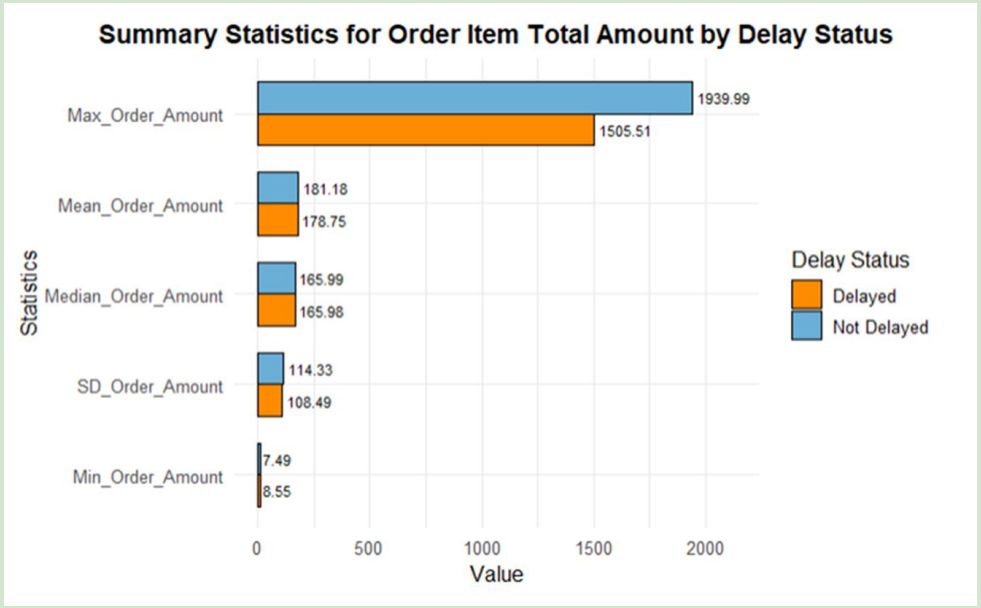
```
Coefficients:
          ma1     sar1    drift
      0.3923  -0.5433  -8.8379
s.e.  0.1623   0.2074   2.3312

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 1.968449 101.5968 69.70962 -4.29687 12.39157 0.448951 0.06160744
```

Practical Implication

Q1: Analyse the factors that will result in an order delay

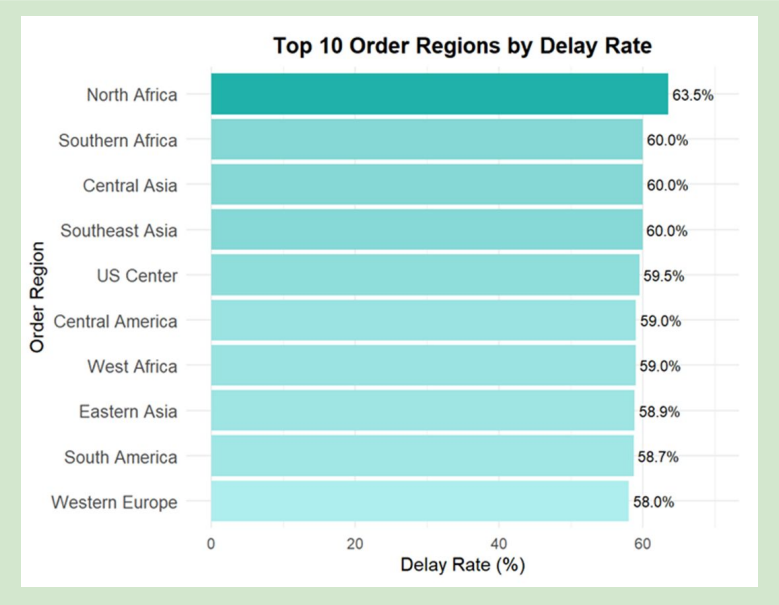


- Shipping mode, Order item total amount, Order region, Customer state and Category, these five variables demonstrate notable importance in the context of the model.
- In terms of shipping mode, first Class has the highest delay rate, while Standard Class has the lowest delay rate.
- For order item total amount, it may not be as influential for delay prediction; its prominence could be due to a few high-value cases rather than a consistent trend.



Practical Implication

Q1: Analyse the factors that will result in an order delay

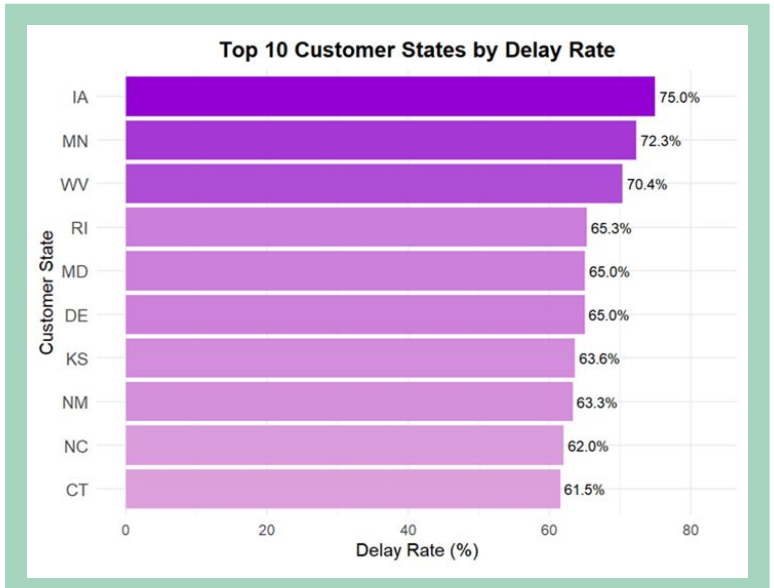


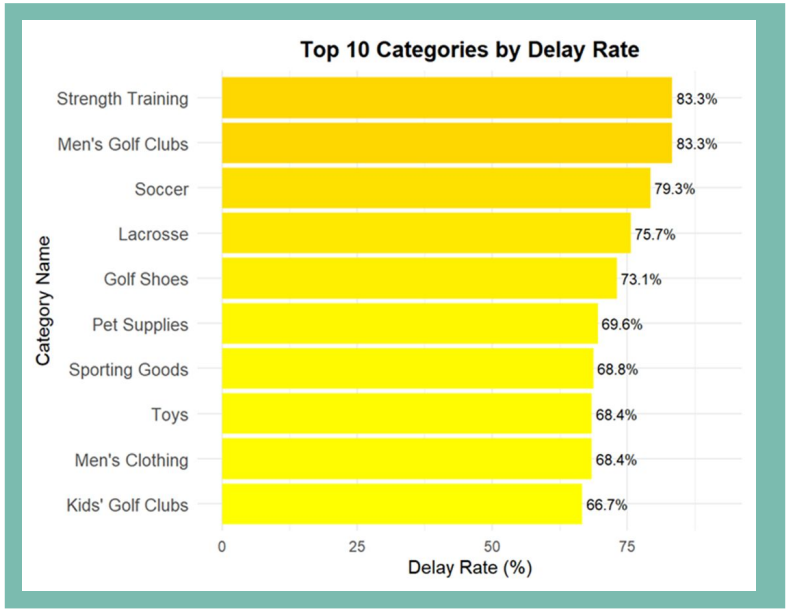
Order Region

North Africa facing significantly higher delays than other regions

Customer State

Iowa(IA), Minnesota(MN), and West Virginia(WV) have the highest delay rates, highlighting potential regional issues affecting order timeliness.



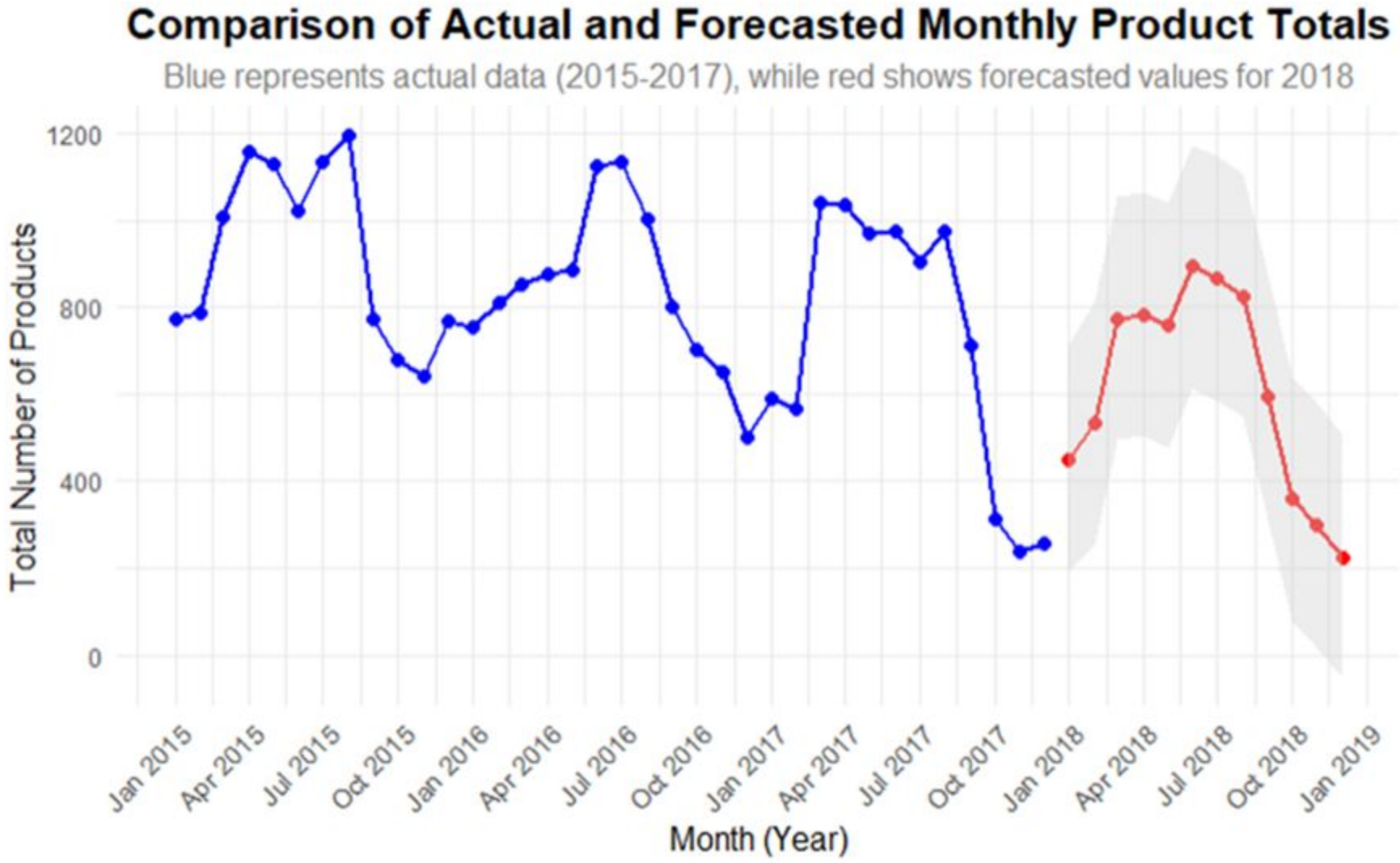


Category

Sporting goods have the highest delay rates. Sellers should focus on these categories, as high delay rates may impact customer satisfaction and efficiency.

Practical Implication

Q2: Predict the future order demand



The line chart sees the forecasted product volume for 2018.
The projection indicates a **peak** in products **from June to August**, suggesting that sellers should increase inventory and ensure adequate logistics resources during this period.
The forecasted product volume **from October to December** shows a **downturn**, indicating that sellers should consider reducing inventory levels during these months to optimise storage and reduce potential overstock costs.

Recommendations

Customer Recommendations

With urgent needs, choose Same Day service while less time-sensitive needs choose Standard Class.

Delay Management by Region

Partner with local providers, use multimodal transport, set up warehouses, and offer real-time tracking.

Seasonal Inventory Adjustment

Boost stock in peak months, reduce in low-demand months to control costs.

High-Delay Product Monitoring

Collaborating with suppliers to maintain stock and optimize processing.

THANK YOU