Predicting Late Parcel Delivery

Authors

- Kevin Alkindy
- Annisa Dwinda F. H.
- Arief Maulana
- Aris Rigel P.
- Kindy Nurhakim
- M. Rezadiandra

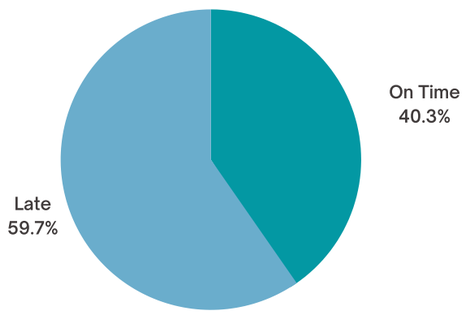
Affiliations

Rakamin Academy
School for Career Acceleration

1 Business Understanding

PROBLEM STATEMENT

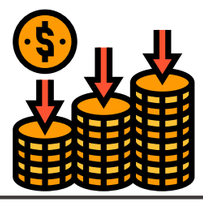
The electronic e-commerce has experienced delivery problem. 59.67% of the parcels were delivered late. Thus, affecting the customer's satisfaction which will potentially cause loss of revenue.



OBJECTIVES

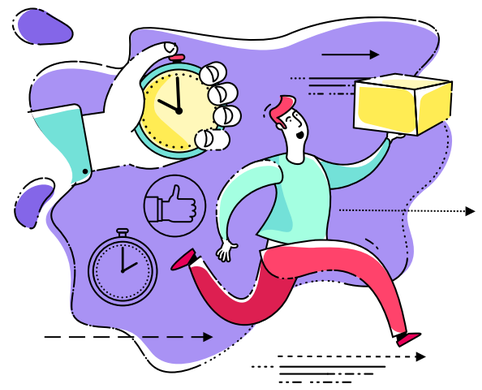
- To increase delivery performance.

- To minimize the potential loss of e-commerce revenue.

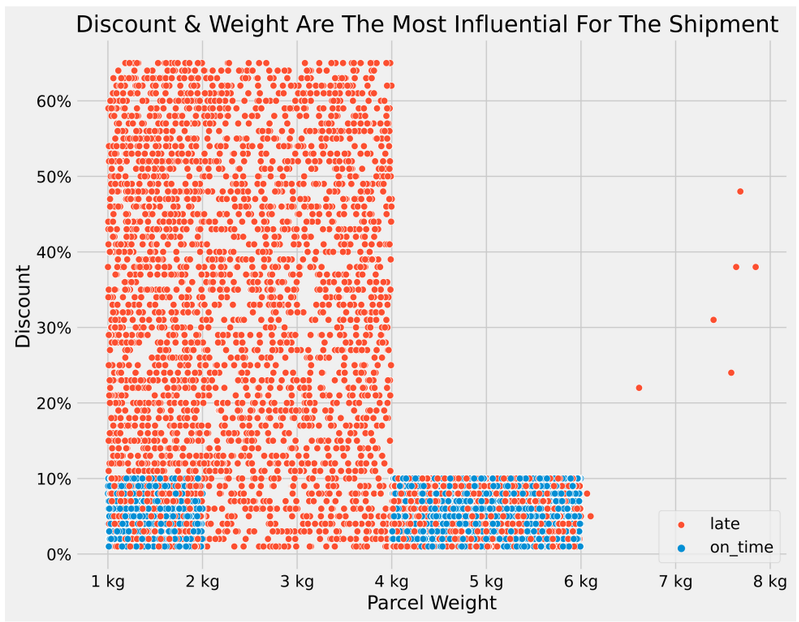


BUSINESS METRIC

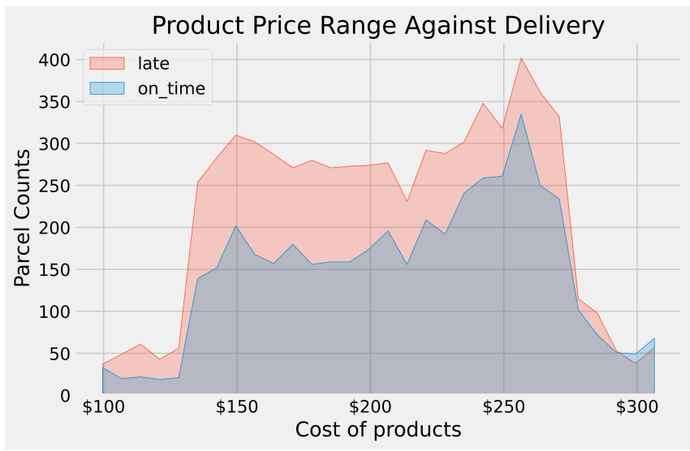
Late ratio.



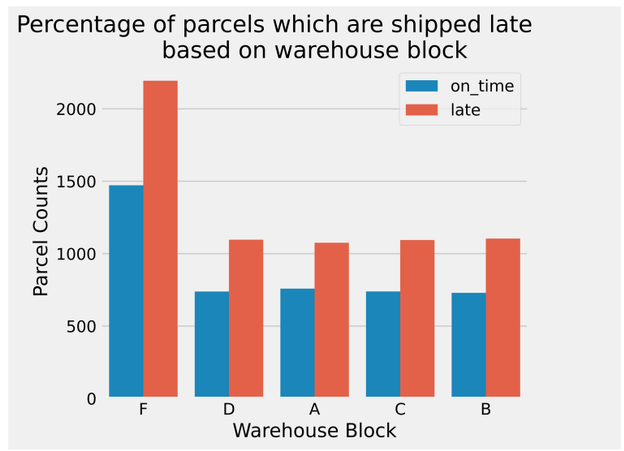
2 Exploratory Data Analysis



The shipment tends to deliver on time when the discount under 10% and the weight is 1 - 2kg and 4 - 6kg.



Most of products that cost around $300 delivered on time, others late.



Most of parcels are stored in warehouse block F.

The dataset consists of 10999 rows & 12 columns.
Feature discount, weight, cost of product, and shipment priority are influential to the shipment.

3 Machine Learning Modelling

The dataset is split **70%** for training & **30%** for testting.

DATA CLEANSING

MISSING VALUES	DUPLICATED VALUES	OUTLIERS
No missing values in the datasaet	No duplicated values in the datasaet	29% outlier values handled using capping outlier

FEATURE ENGINEERING

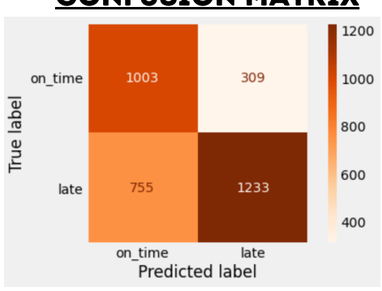
LOG TRANSFORMATION	FEATURE SCALING : STANDARDIZATION	FEATURE SELECTION CHI SQUARED	FEATURE ENCODING ONE HOT ENCODING
Apply to all numeric values	Apply to all numeric values	Feature **product_importance** with high category **are correlated** to the target.	We break feature **product_importance** using one hot encoding to get **product_importance_high** and drop others.

MODELLING & EVALUATION

	Accuracy (Train)	Accuracy (Test)	Precision (Test)	Recall (Train)	Recall (Test)	F1 Score (Test)	AUC (Test)
Logistic Regression	0.63	0.63	0.67	0.75	0.75	0.75	0.5
Decision Tree	0.67	0.68	0.8	0.62	0.62	0.7	0.69
Random Forest	1	0.67	0.76	1	0.66	0.71	0.67
KNN	0.76	0.64	0.72	0.77	0.67	0.69	0.64
SVC	0.68	0.66	0.88	0.52	0.51	0.65	0.70
XG Boost	0.93	0.66	0.75	0.92	0.67	0.71	0.66

Among all models, **Decision tree with hyperparameter tuned** has a good balance between its score, also the model is neither Underfitting nor Overfitting. We focuss on **recall,** to prevent incorrectly predicting parcel that get delayed but marked as delivered on time.

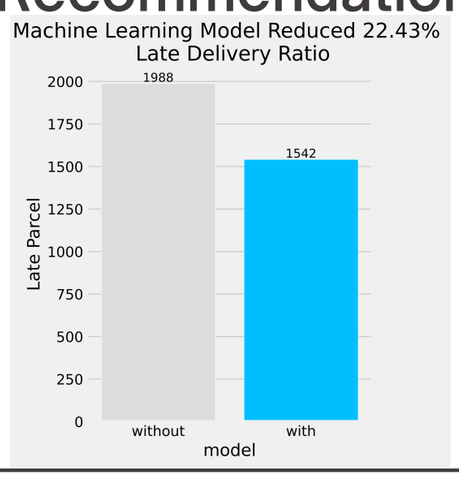
CONFUSION MATRIX



3 Business Impact & Recommendation

MODEL IMPACT ON BUSINESS METRIC

Our model is proved to reduce 22.43% late ratio. Based on research,[2] the customers will stop shopping when they get their parcels late, hence it can minimize the potential loss of revenue by 62% or equal to $255.587.



RECOMMENDATION

- The operation team should add more manpower when there is a sale program especially for the discount more than 10% and the parcel weight is 1 - 4kg.
- The parcel should not be centralized in the warehouse block F, so that the handling is not too crowded which can cause the late shipment.

Adding more features can improve model performance, such as delivery time estimation, delivery date, customer address, courier.

Appendix :
1.https://www.kaggle.com/prachi13/customer-analytics
2.https://www.hollingsworthllc.com/how-late-deliveries-impact-customer-retention/