Table 1 Pseudo Algorithm for Solving Unexecuted Truckloads

Step 1: Begin
Step 2: Update the starting and ending dates of training data using a sliding time window to avoid stale model performance
Step 3: Pre-processing test data

1. Start with the assumption of non-missing features and non-missing transit times in test data. Both conditions set to False. This step yields how many lanes have missing features and/or missing transit times.
2. Query for missing features from the planning system using truckload ID, change the condition for missing features as True, rerun 1.
3. Change condition for missing transit times as True, read in a static transit-time table, rerun 1.
4. Check how many loads are still missing features and transit times, exclude from modeling process.

Step 4: Split data into train, validation, and test datasets for modeling
Step 5: Select between different sampling algorithms
    Run SMOTENN
    Run Near Miss
    Run SMOTETomek
    Run SMOTE
Step 6: Parameter tuning for different models to find the best parameters for each model
    Tune XGB parameters
    Tune random forecast parameters
    Run Logistics regression
Step 7: Define and collect model measures for the outputs from Step 6
Step 8: Model selection based on model measures from Step 7
Step 9: Model prediction – predict probability on the test data
Step 10: End

Table 2 Data Summary

	Train (80% of data from 2020-06-15 to 2020-11-30)	Validate (20% of data from 2020-06-15 to 2020-11-30)	Test (2020-12-01 and 2020-12-02)	Total
No. of Observations	153,814	38,454	2,139	194,407
0 in y	54,483	24,955	1,144	80,582
1 in y	99,331	13,499	995	113,825

Table 3 Under Sampling and Over Sampling for Train Data

	Train Data Before Sampling	SMOTE	SMOTENN	SMOTETomek	Near Miss
No. of Observations	153,814	198,662	110,001	182,774	108,240
0 in y	54,483	99,331	58,906	91,387	54,483
1 in y	99,331	99,331	51,095	91,387	53,757

Table 4 Logistic Regression Output

	coefficient	s.e.	z	P>|z|
Departure: Hour of Day-Sin	-0.073	0.019	-3.85	0.000
Departure: Hour of Day-Cos	0.220	0.021	10.45	0.000
Departure: Hour of Day	0.026	0.001	34.41	0.000
Departure: Day of Week	0.121	0.003	45.69	0.000
Departure: Week of Year	-0.017	0.000	-33.78	0.000
Hook Trailer Time	0.001	0.000	1.74	0.081
Drop Trailer Time	0.010	0.000	27.36	0.000
Transit Time	0.002	0.000	25.18	0.000
Miles	-0.001	0.000	-14.97	0.000
Check-in Time Window	0.000	0.000	6.04	0.000
Total Block Time	0.000	0.000	17.50	0.000
Number of Available Blocks	-0.015	0.000	-48.86	0.000
Origin Zip	0.001	0.000	37.76	0.000
Destination Zip	0.000	0.000	-15.41	0.000
Time to departure	-0.006	0.000	-88.63	0.000

Table 5 Four Model Performance Measures

Model Performance Measures Used	Definition	Rationale of using this performance measure
Accuracy	Measure of how well a model performs on a dataset, defines as the proportion of correct predictions made by the model out of all the predictions made.	Accuracy is appropriate for balanced data. We have a balanced data by using the SMOTE, accuracy is a proper performance measure given our data structure.
Brier score	Measures the mean squared error between the predicted probabilities and the true outcomes. The lower the Brier score, the better the predictions.	The Brier score has several appealing properties. First, it is a proper scoring rule, meaning that it is always optimal to make predictions that maximize the Brier score. Second, the Brier score is closely related to the log-loss, another popular metric for evaluating probabilistic predictions. In fact, the Brier score is simply the mean squared error of the log-loss. Third, the Brier score can be decomposed into a component that measures the calibration of the predictions and a component that measures the sharpness of the predictions. This decomposition is useful for understanding the sources of error in probabilistic predictions.
Log-loss	Log-loss is a measure of the accuracy of classifier predictions. The log-loss measure penalizes false negatives more heavily than false positives, thus, is commonly used in binary classification problems where the goal is to identify the positive class. A lower log-loss value means better predictions.	The Log-loss has several advantages. First, it is differentiable, which makes it easier to optimize. Second, the log-loss function is convex, which means that there is only one global minimum. The log-loss function is also scale-invariant, meaning that it is not affected by changes in the scale of the data, especially for logistic regression where data needs to be scaled sometimes. Lastly, the log-loss function is robust to outliers, meaning that it is not affected by a few extreme values. Given these advantages of log-loss measure, we include it as our third measure.
Area under the curve (AUC)	AUC represents the probability that a model will correctly classify a positive instance as positive and a negative instance as negative. The AUC ranges from 0 to 1, with a higher AUC indicating a better model.	The AUC has several desirable properties. First, the AUC is scale-invariant, meaning that it is not affected by changes in the distribution of the data. Second, the AUC is insensitive to the class imbalance, meaning that it is not affected by changes in the proportion of positive and negative instances. Third, the AUC is invariant to the specific choice of threshold, meaning that it can be used to compare models with different threshold values. Therefore, we include AUC as our fourth measure.

Table 6 Model Performance Comparison

	Random Forest	XGB Classifier	Logistic Regression
Training Data			
Accuracy Score	0.892	0.930	0.645
Brier Score	0.089	0.054	0.215
Log-loss Score	0.301	0.186	0.616
AUC	0.962	0.983	0.711
Validation Data			
Accuracy Score	0.827	0.869	0.643
Brier Score	0.125	0.095	0.217
Log-loss Score	0.394	0.311	0.622
AUC	0.901	0.936	0.711
Test Data			
Accuracy Score	0.799	0.837	0.605
Brier Score	0.143	0.116	0.227
Log-loss Score	0.439	0.375	0.638
AUC	0.883	0.915	0.679

Table 7 Precision, Recall, and F-1 Score for the Test Data

	Precision			Recall			F1-Score			Total Observation
	Random Forest	XGB	Logistic Regression	Random Forest	XGB	Logistic Regression	Random Forest	XGB	Logistic Regression	
1	0.82	0.86	0.66	0.81	0.84	0.55	0.81	0.85	0.60	1144
0	0.78	0.82	0.56	0.79	0.84	0.67	0.79	0.83	0.61	995
Weighted Average	0.80	0.84	0.61	0.80	0.84	0.60	0.80	0.84	0.60	2139

Table 8 Calibrated Model Performance on Test Data

	Random Forest	XGB Classifier	Logistic Regression
Test Data			
Accuracy Score	0.799	0.837	0.605
Brier Score	0.143	0.116	0.227
Log-loss Score	0.439	0.375	0.638
AUC	0.883	0.915	0.679
Model Calibration on Test Data			
Sigmoid			
Accuracy Score	0.793	0.835	0.613
Brier Score	0.142	0.124	0.225
Log-loss Score	0.447	0.418	0.634
AUC	0.883	0.915	0.679
Isotonic			
Accuracy Score	0.800	0.837	0.598
Brier Score	0.141	0.120	0.224
Log-loss Score	0.570	0.469	0.631
AUC	0.883	0.914	0.679

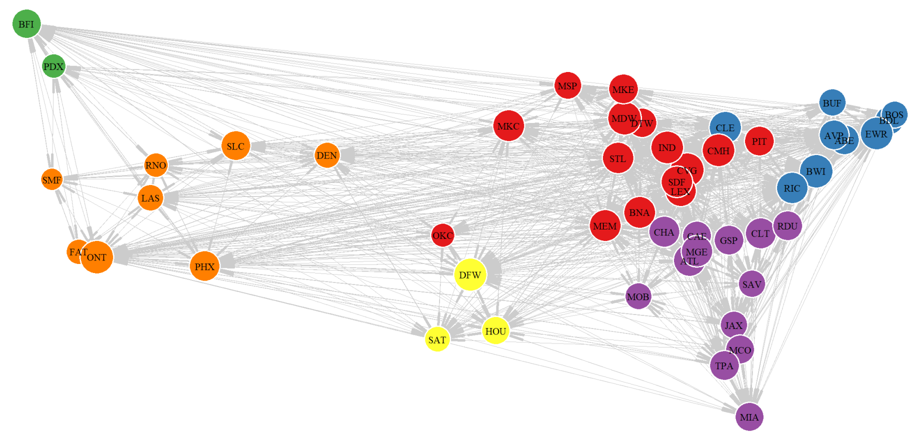
Table 9 Post Hoc Analysis of Cut-off Probability

	Week 50 (50% cutoff)	Week 51(60% cutoff)	Week 52(70% cutoff)
Truckloads Withheld for Self-planning System	246	334	421
Actually being Executed by Self-own Fleet	162	139	207
Execution Rate	65.85%	58.38%	50.83%

Table 10 Cost Savings from Pilot Run Weeks in 2021

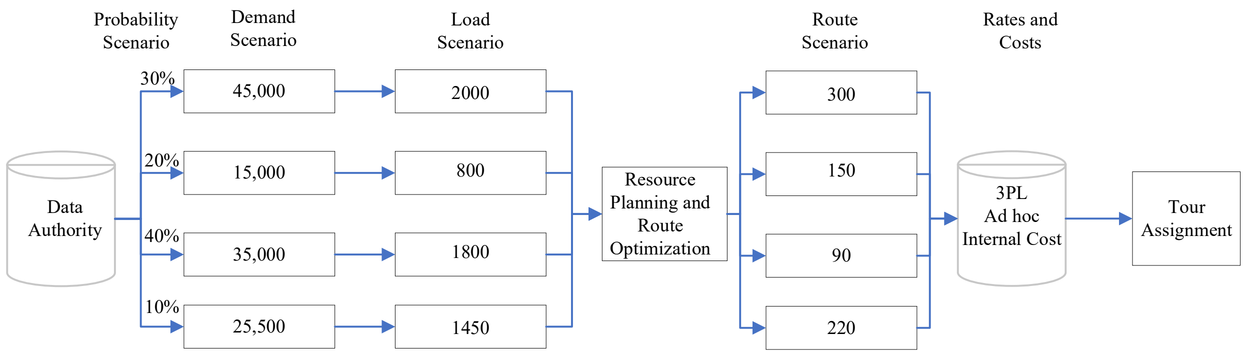
	W1	W2	W3	W4	W5	W6
# Truckloads Piloted	1,158	2,944	2,838	1,842	3,750	1,993
Cut-off Threshold	50%	50%	50%	30%	40%	40%
# Truckloads Retained	407	724	942	699	1321	551
# Truckloads Executed	195	267	655	293	688	202
# Truckloads Unexecuted	212	457	287	406	633	349
Mean Distance of Truckloads Executed	762	831	818	801	849	860
Mean Distance of Truckloads Unexecuted	790	883	792	729	778	842
Premium paid to 3PL for Unexecuted	$ 28,185	$ 169,614	$ 60,159	$ 94,623	$ 140,113	$ 53,839
Savings from Executed Truckloads	$ 247,807	$ 369,428	$ 978,890	$ 389,677	$ 986,973	$ 304,258
Net Gain	$ 219,622	$ 199,814	$ 918,731	$ 295,055	$ 846,861	$ 250,419

Figure 1 Shipment Flow of Order fulfillment and Replenishment



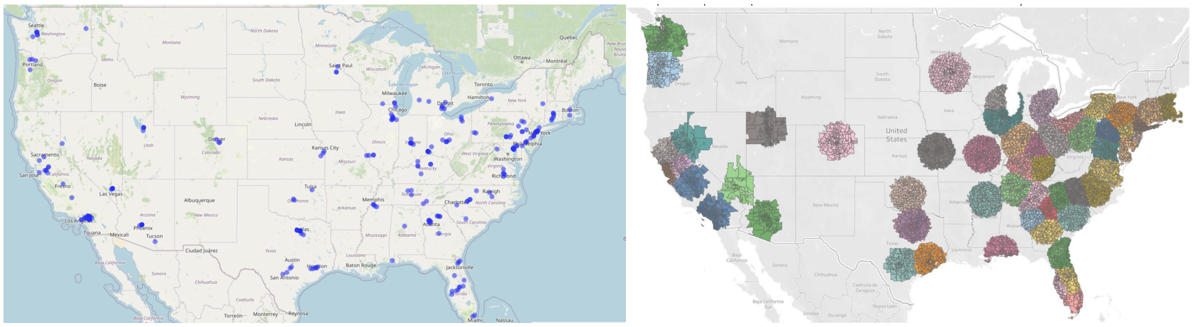
Note: Figure 1 presents an average weekly truckload flow among different warehouses of the focal company. Each warehouse is represented by a single node. The size of each node corresponds to the volume of shipments (inbound + outbound) in each warehouse. Different color represents different regions. Nodes with the same color belong to the same region. There are altogether six regions for the focal company.

Figure 2 Demand Driven Planning Process (For Illustration Purpose)



Note: Figure 2 demonstrates how demand flows through the resource planning process. First, different demand scenarios (such as outbound, inbound, warehouse transfer, etc.) are generated from database. Then, each demand scenario is converted into possible truckloads. Third, these truckloads will be optimized into routing using demand clusters in Figure 2. The last step is the tendering process where set of routes are tendered to 3PLs for execution.

Figure 3 Resource Planning Based on Demand Clusters



Note: Figure 3 demonstrates the concept of “Demand Clusters”. The graph on the left shows all Tier I warehouses. Each blue dot is a Tier I warehouse. The graph on the right shows how demands from both Tier II and Tier I warehouses were clustered together based on their geographical proximity.

Figure 4 Class Distribution Before Sampling for Train Data

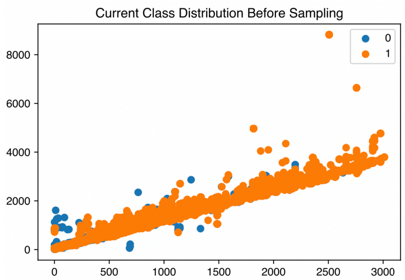


Figure 5 Class Distribution After Sampling for Train Data

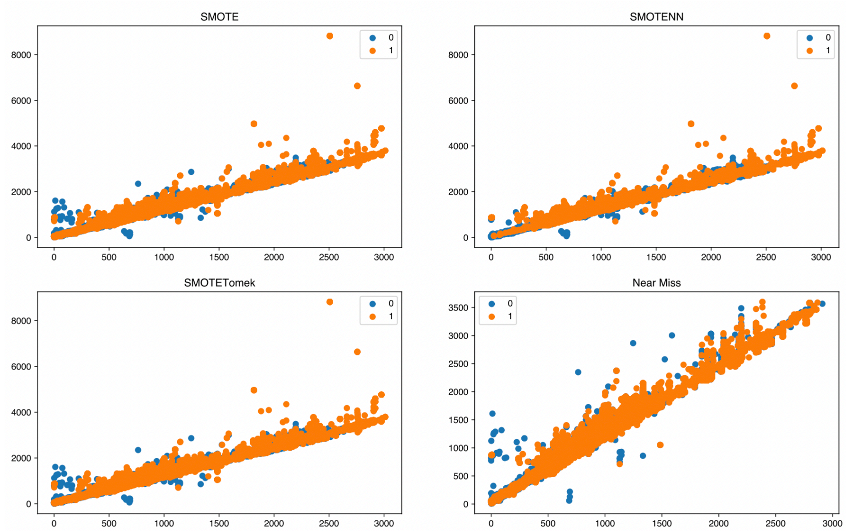


Figure 6 ROC Curve of the Three Models

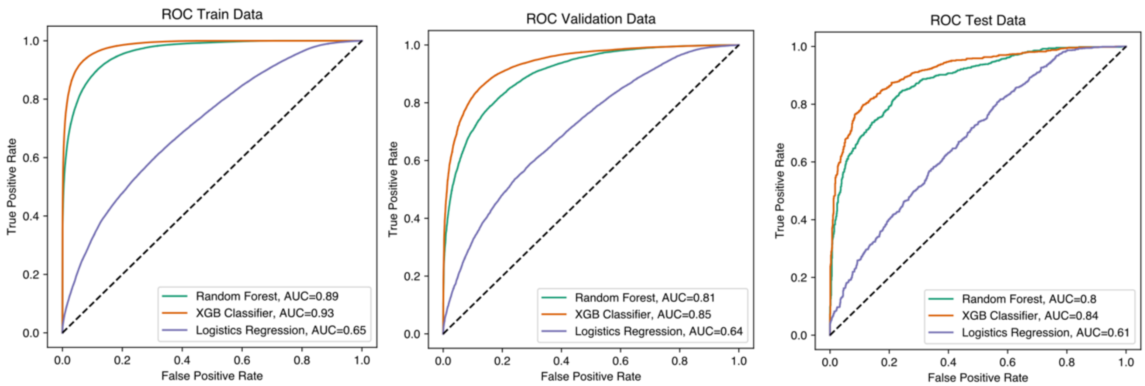


Figure 7 Confusion Matrix of the Three Models

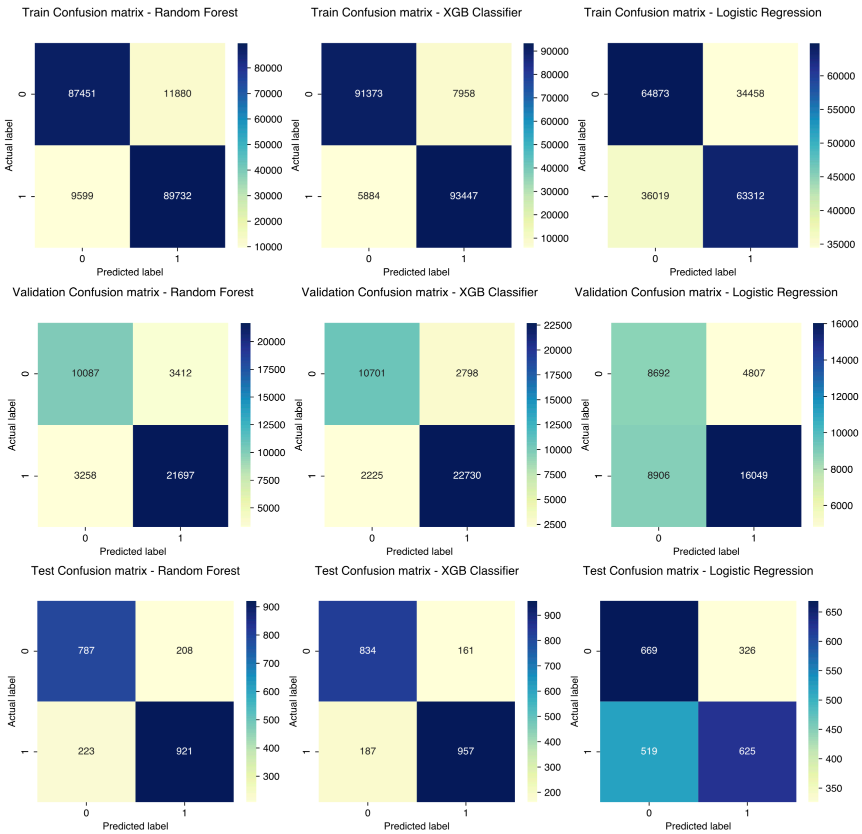


Figure 8 Random Forest Predicted Probability for Actual Executed VS Actual Unexecuted

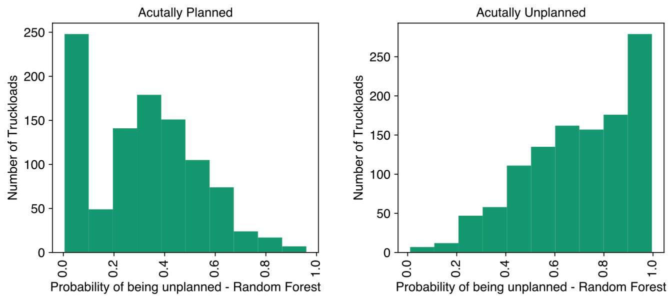


Figure 9 XGBoost Predicted Probability for Actual Executed VS Actual Unexecuted

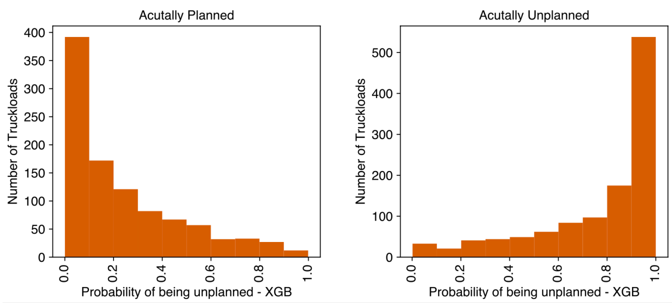


Figure 10 Logistic Regression Predicted Probability for Actual Executed VS Actual Unexecuted

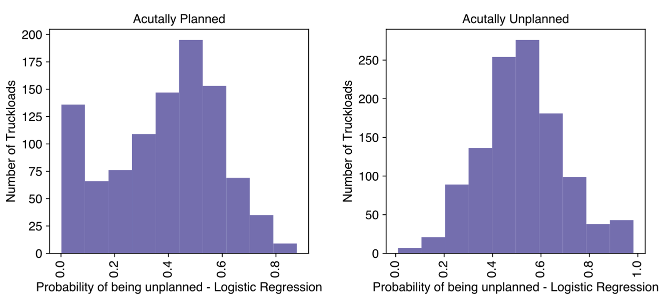


Figure 11 Partial Dependence Plot for Test Data

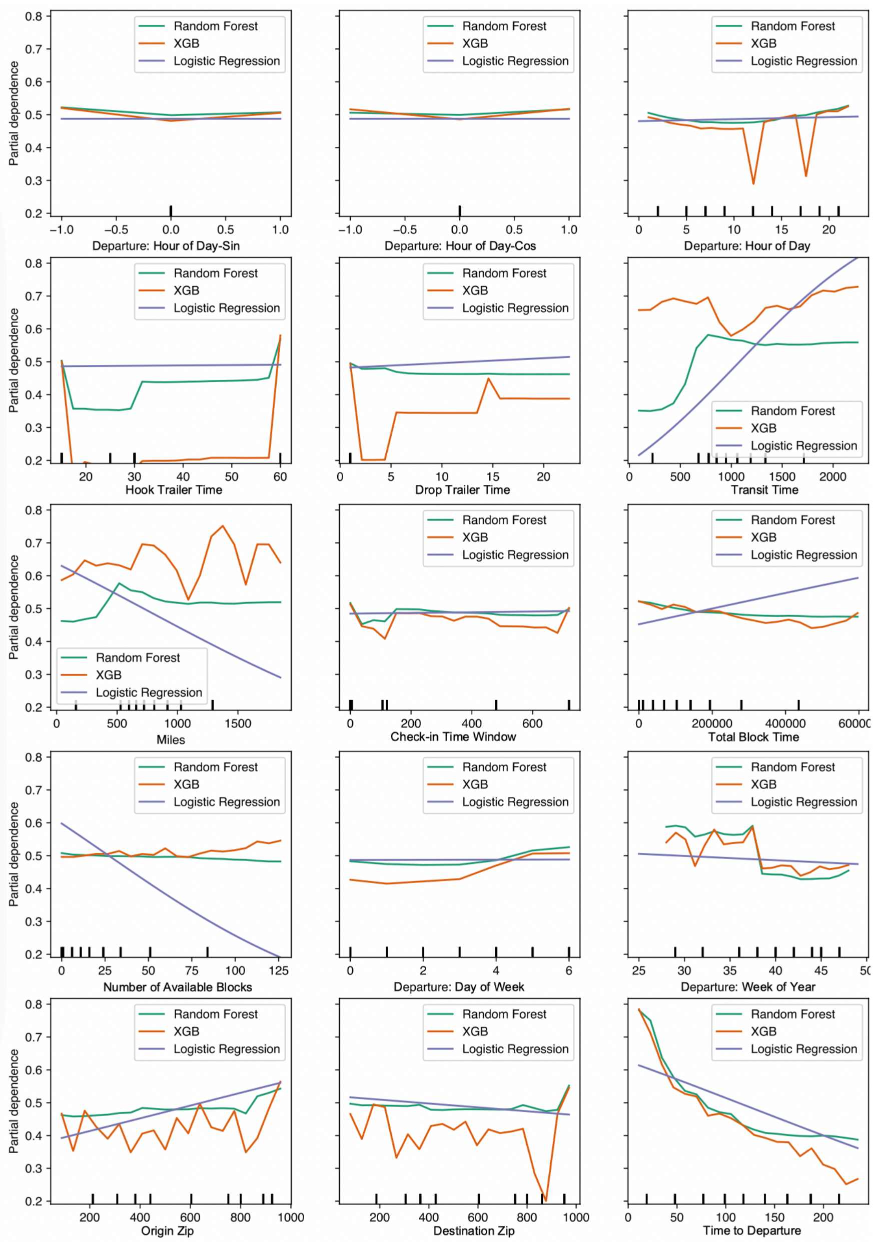


Figure 12 Permutation Feature Importance – Test Data – Random Forest

Weight	Feature
0.1126 ± 0.0138	Time to Departure
0.0827 ± 0.0083	Check-in Time Window
0.0488 ± 0.0073	Departure: Hour of Day-Sin
0.0410 ± 0.0106	Total Block Time
0.0379 ± 0.0034	Origin Zip
0.0251 ± 0.0070	Destination Zip
0.0225 ± 0.0038	Departure: Hour of Day-Cos
0.0149 ± 0.0032	Departure: Hour of Day
0.0115 ± 0.0093	Number of Available Blocks
0.0066 ± 0.0058	Transit Time
0.0044 ± 0.0063	Departure: Day of Week
0.0020 ± 0.0015	Miles
0.0014 ± 0.0054	Departure: Week of Year
0 ± 0.0000	Drop Trailer Time
0 ± 0.0000	Hook Trailer Time

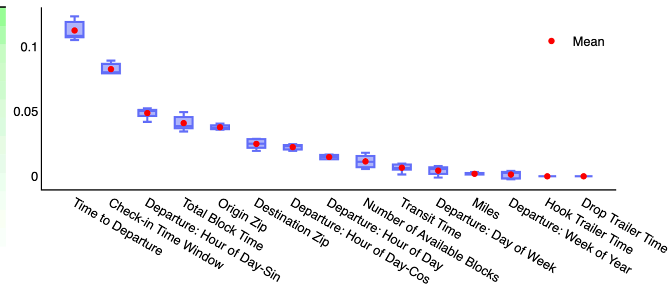


Figure 13 Permutation Feature Importance – Test Data – XGB

Weight	Feature
0.1796 ± 0.0138	Total Block Time
0.1193 ± 0.0030	Check-in Time Window
0.1075 ± 0.0054	Time to Departure
0.0555 ± 0.0123	Origin Zip
0.0410 ± 0.0101	Destination Zip
0.0354 ± 0.0083	Departure: Hour of Day-Sin
0.0341 ± 0.0082	Departure: Hour of Day-Cos
0.0263 ± 0.0107	Departure: Hour of Day
0.0197 ± 0.0022	Number of Available Blocks
0.0132 ± 0.0053	Departure: Day of Week
0.0112 ± 0.0046	Departure: Week of Year
0.0041 ± 0.0052	Transit Time
0.0009 ± 0.0017	Miles
0 ± 0.0000	Drop Trailer Time
0 ± 0.0000	Hook Trailer Time

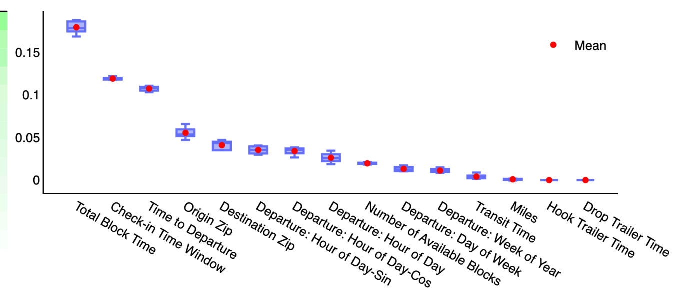


Figure 14 Permutation Feature Importance – Test Data – Logistic Regression

Weight	Feature
0.0554 ± 0.0132	Departure: Week of Year
0.0456 ± 0.0122	Time to Departure
0.0362 ± 0.0147	Check-in Time Window
0.0220 ± 0.0111	Total Block Time
0.0148 ± 0.0149	Origin Zip
0.0112 ± 0.0120	Departure: Day of Week
0.0003 ± 0.0023	Departure: Hour of Day
0.0002 ± 0.0005	Departure: Hour of Day-Cos
0 ± 0.0000	Drop Trailer Time
0 ± 0.0000	Hook Trailer Time
0 ± 0.0000	Departure: Hour of Day-Sin
-0.0004 ± 0.0016	Miles
-0.0007 ± 0.0015	Transit Time
-0.0020 ± 0.0023	Number of Available Blocks
-0.0038 ± 0.0075	Destination Zip

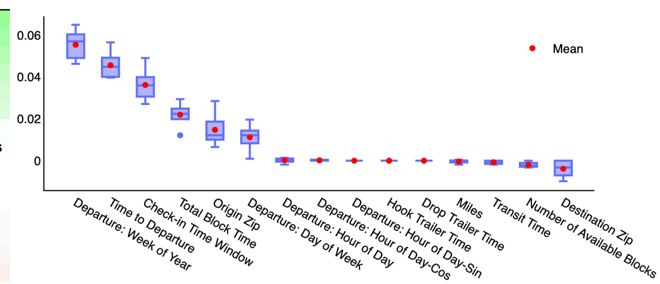


Figure 15 Probability Calibration Curve on Train Data – Random Forest

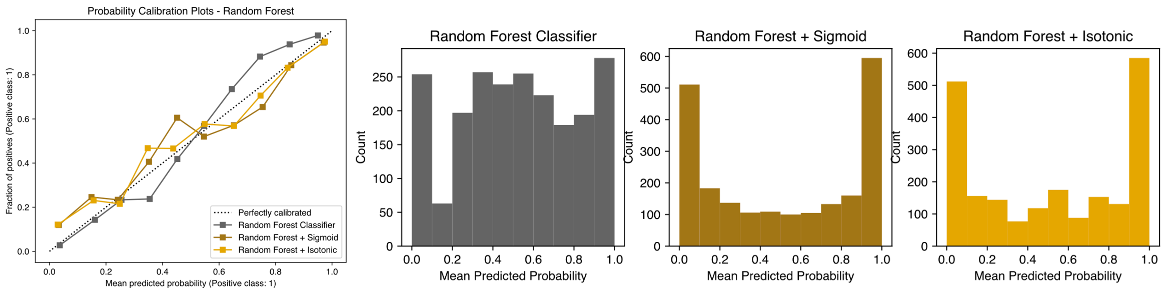


Figure 16 Probability Calibration Curve on Train Data – XGB

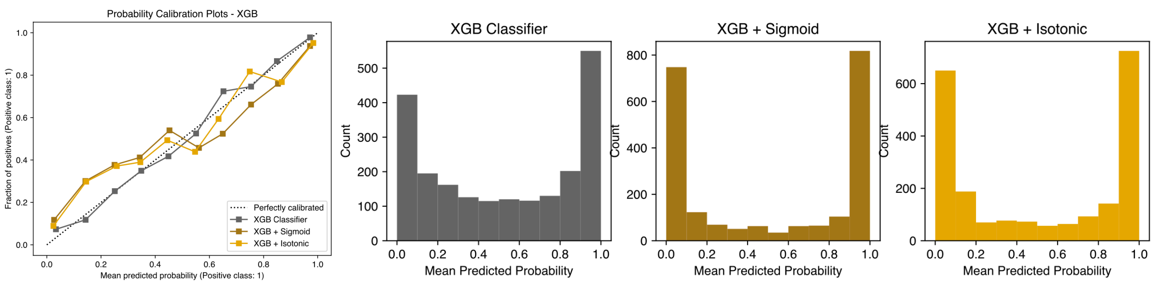
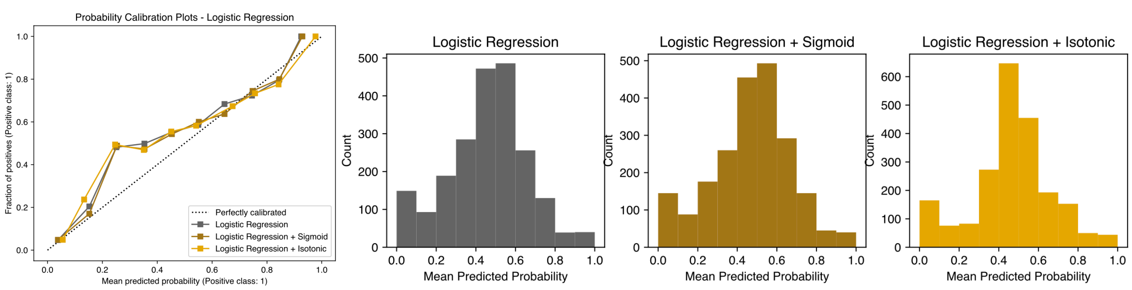


Figure 17 Probability Calibration Curve on Train Data – Logistic Regression



Appendix A

Author	Journal	Type of article	Topic	Data sources	Method	Real cases involved (if any)	Context
Abrahams et al., (2014)	POMS	Empirical analysis	Conduct text mining on user generated content and proposed an integrated text analytic framework for product defect discovery	Extracted data from user generated content from various platforms (Honda-Tech.com; ToyotaNation.com and ChevroletForum.com	Two cases studies by applying the proposed content for product defect	Extracted data from user generated content from actual discussion threads	Retail and consumer engagement
Araz et al., (2020)	DSJ	Literature review and topic introduction	Big data analysis in operational risk management	NA	NA	NA	NA
Bansal et al., (2020)	JSCM	Literature review and topic introduction	Propose a framework of using qualitative big data to conduct topic modeling	NA	Case study with three previous articles	NA	Retail and consumer engagement
Bertsimas et al., (2016)	POMS	Empirical analysis	Combining actual sales data from a global media vendor with search/rating from multiple platforms to inform the vendor inventory related decisions	Vendor's sale and inventory records across the network of retailers, information about each of the locations, and information about each of the items; Search record and ratings from multiple entertainment platforms.	Local weighting approach based on random forest weights	Sales and inventory data come from one of the three largest global media vendors in the world	Inventory management
Boone et al., (2018)	POMS	Empirical analysis	Examine how to general sale forecast by combining data from retailer and Google trend	Data from a specialty retailer of food and cookware and consumer search queries from Google trend	ARIMA model as the base line and incorporate the "trend" component from Google trend	Part of the data from a specialty retailer of food and cookware	Forecasting
Chan et al., (2016)	POMS	Empirical analysis	Propose a framework for conducting text mining with social media data with a case study	Extracted data from user generated content from various platforms	Cluster analysis	Extracted data from the official Samsung Mobile Facebook page	Retail and consumer engagement
Chang et al., (2020)	POMS	Empirical analysis	Using machine learning to predict and identify true offenders in environmental sustainability	Continuous Emission Monitoring System (CEMS); Violation and Punishment data (VPD)	PU Learning	The CEMS data provides hourly pollution emission data for all industries in 30 province-level regions of mainland China from January 1, 2016 to June 30, 2017. VPD is publicly available data	Sustainability
Choi et al., (2018)	POMS	Literature review and topic introduction	Big data analysis to operations management	NA	Case study	Several big companies with real-world applications of big data analytics in operations	NA

Chuang et al., (2021)	JOM	Empirical analysis	Develops and further implement a framework to help improve prediction accuracy under high demand volatility for procurement managers.	Data provided by leading electronics distributors in the world	Several econometrics modelling and machine learning models (ensemble learning algorithm utilizing gradient boosting machines) have been used	Yes. Data was provided by one of the leading electronics distributors in the world	Forecasting
Corbett (2018)	POMS	Topic introduction and application opportunities	Big data analysis in sustainability	NA	NA	NA	NA
Feng and Shanthikumar (2017)	POMS	Topic introduction and application opportunities	Big data analysis in demand management and manufacturing	NA	Analytical	NA	NA
Fisher and Raman (2018)	POMS	Topic introduction and application opportunities	Big data analysis in retailing	NA	NA	NA	NA
Foster et al., (2018)	POMS	Empirical analysis	Using data from emergency physician management network for physician evaluations	Large scale data from emergency physician management networks in US	Cluster analysis with k-means; 2SLS	Large scale data from large private emergency physician management networks EPMN, which includes numerous physicians across 84 facilities in 14 US states from 2010-2014	Health care
Geva et al., (2020)	POMS	Empirical analysis	Propose a data-driven approach for producing effective rankings based on the decision quality of expert workers	Data from simulations relying on three datasets corresponding to real-world decisions	Propose a framework regarding how to apply machine learning technique to make REQ (ranking expert quality) decisions	NA	Decision quality evaluation
Gopalakrishnan et al., (2022)	DSJ	Empirical analysis	Propose a framework and pricing strategy for a monopoly firm selling multiple substitute products to customers characterized by their different social network degrees	A sample of 1,000,000 users who made purchases on a video-sharing and live-streaming platform with for two year span.	Multinomial logit model	Yes	Pricing strategy
Guha, and Kumar (2018)	POMS	Topic introduction and application opportunities	Big data analysis in information systems, operations management, and healthcare	NA	NA	NA	NA
Hopp et al., (2018)	POMS	Topic introduction and application opportunities	Big data analysis in healthcare	NA	NA	NA	NA

Ko et al., (2019)	JOM	Empirical analysis	Explore the relationship between operational efficiency and patient satisfaction by examining one million physician reviews across 17 medical specialties	Web scraping data from Vitals (http://www.vitals.com) for 1,560,639 physician reviews	Censored (Tobit) regression model and topic model (latent Dirichlet allocation LDA)	Yes. Data was craped from online review platforms, covering all 50 states of the United States and spanning 60 medical specialties	Health care
Lam et al., (2016)	JOM	Empirical analysis	Examines the impact of social media initiatives on firms' operational efficiency and innovativeness.	Performance measures collected from Compustat and Fortune databases; Social media announcements collected from Factiva	Dynamic panel data (DPD) models with GMM estimation		Retail and consumer engagement
Lau et al., (2018)	POMS	Empirical analysis	Conduct sentiment analysis with consumer comments for sale forecasts	Product comments from multiple social media platforms (US and China)	Parallel aspect-oriented sentiment analysis		Retail and consumer engagement
Lim et al., (2020)	DSJ	Empirical analysis	Examine the mechanism of sustained patient engagement in the context of online health infomediaries	Data collected from an online infomediary and covers for one year span	Multi-State Markov Chain Model	Yes	Health care
Merrick et al., (2021)	POMS	Proposed framework with application example	Develop a prototype risk prediction system to provide early alerts of elevated risk levels to vessel traffic managers and operators by incorporating incident and accident data	Case study performed for the United States Coast Guard (USCG)	Compare seven generic classification algorithms from the machine learning (LR, DT, RF, kNN, NN, GBT, SGBT)	Yes	Risk management
Nenova and Shang (2022a)	POMS	Empirical analysis	Conduct machine learning technique to optimize patient monitoring frequencies and improve treatment	Data from 11 U.S. Department of Veterans Affairs hospitals, containing 68,513 CKD patients from Jan 2009 to February 2016	Finite-horizon Markov Decision Process (MDP)	Data from Veterans Affairs (VA) hospitals with patients with chronic kidney disease	Health care
Nenova and Shang (2022b)	POMS	Empirical analysis	Propose an iCBR model to better predict kidney disease progression	Data from three patients across seven year's time frame (Jan 2009 to Feb 2016)	Build on the eGFR Smoothing Model to develop the intelligent CBR Model	Data from electronic health records from the Veterans Affairs (VA) hospitals	Health care
Sanders and Ganeshan (2018)	POMS	Editorial and topic introduction	Big data analysis in supply chain management	NA	NA	NA	NA
Shen and Sun (2021)	JOM	Empirical analysis	Analyzes the impact of the COVID-19 pandemic on supply chain resilience, summarizes the challenges in retail industry in China.	Quantitative operational data obtained from JD.com	Identify indicators for resilience based on machine learning technique-identified performance indicators	Yes. Access to all transaction data in the daily operations of JD.com.	Risk management

Smyth et al., (2018)	JBL	Tutorial with application example	Propose a tutorial how academic can work with practitioners to derive value from big data (from actual firm)	Sales forecasts and plan replenishment data came from a restaurant firm with 15,000 retail outlets within the US.	Cluster analysis; Principal component analysis; Factor analysis; Linear regression	Yes	Forecasting
Sun et al., (2019)	DSJ	Tutorial with application example	Propose the Partially Profiled LASSO model to understand the effects of price promotions	Sale, promotion data from a large retailer in North China that sells a wide variety of consumer goods	Partially Profiled LASSO model	Yes	Pricing strategy
Sun et al., (2020)	DSJ	Tutorial with application example	Propose a data driven bicriteria mathematical model	2 years of flight data was provided by a major flight data service provider	Analytical modeling combines with a column generation-based algorithm	Yes	System optimization
Swaminathan (2018)	POMS	Topic introduction and application opportunities	Big data analysis in humanitarian operations	NA	NA	NA	NA
Wang and Wu (2020)	DSJ	Empirical analysis	Develop and analyze a driver dispatching system for a control center that minimize passengers' waiting time in a ride sharing context	Ride share requests data collected from the Taxi and Limousine Commission (TLC) in one month period in 2018	Conduct a simulation study based on actual past taxi order data; Rolling Time Horizon approach that incorporates hybrid forecasting model and a heuristic algorithm	Yes	System optimization
Zhao (2021)	DSJ	Tutorial with application example/Empirical analysis	Propose a content analysis approach to analyze reviews, and predict the overall attitudes of consumers	Data was scraped from the orbitz.com	Support vector machine (SVM) approach	Yes	Retail and consumer engagement
Zhu et al., (2021)	POMS	Proposed framework with application example	Propose a novel demand forecasting framework by combining historical data with machine learning technique	Test the framework with two large datasets from major pharma manufacturers	One of the MA, LR, ETS as the baseline model, recurrent neural network (RNN)	Yes	Forecasting