NCD Breach Prediction

TOGETHER,
WE DARE
TO MAXIMISE.

Reservation Node Model - Revisit for BBD

Training Model - Base Wk 37 with Historic Wk 36 (no rolling 7 days)

Prediction/Test Data - 22nd Sep BBD EA orders and 26th Sep BBD orders

BBD - 22nd Sep	Prediction	
cpd	0	1
20220923	142	34
20220924	219,066	3,039
20220925	775,550	5,825
20220926	542,355	2,771
20220927	435,644	1,431
20220928	358,293	836
20220929	242,328	491
20220930	147,871	341
20221001	84,088	190
20221002	70,148	127
20221003	47,205	90
20221004	11,631	20
20221005	1,719	4
20221006	14,219	15
20221007	2,330	3
20221008	1,305	NaN
20221009	680	2
20221010	428	NaN
20221011	240	1
20221012	157	5
20221013	57	0
Total	2,955,456	15,225
		0.52%

BBD - 26th Sep	Prediction	
CPD	0	1
20220927	48	15
20220928	33535	510
20220929	311376	3331
20220930	470980	3889
20221001	336566	2564
20221002	372835	2216
20221003	409381	2099
20221004	313095	1338
20221005	127072	386
20221006	178518	492
20221007	163208	360
20221008	107385	219
20221009	49510	78
20221010	41237	46
20221011	20850	13
20221012	7462	8
20221013	2587	2
20221014	1032	3
20221015	384	4
20221016	159	25
20221017	254	0
20221018	103	7
Total	2,947,577	17,605
		0.60%

Key Highlights:

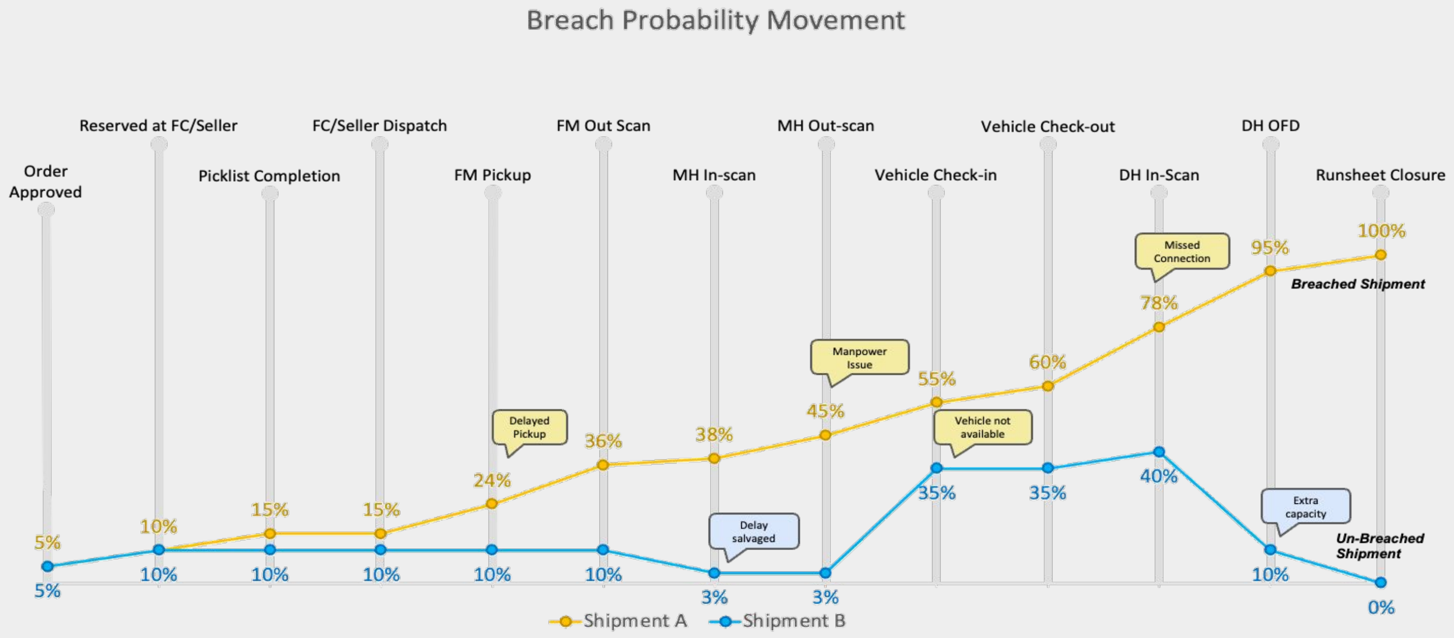
- Model is predicting very few NCD breach probability.
- Even if we say we have high precision, recall is very less(~25%) as actual EKL breach for these dates were 2.1% and 2.97% respectively

Scope of improvement of the existing model:

- We were not considering 7 day rolling average in the historic information, while training the model
- **Buffer** information was not added in the model
- There were very **few** variables that were being considered
- Also, considered overlapping **variables** eg , sla days and expected hrs at each leg might have led to higher **precision**
- Prediction was Based on orderid , rather than unit id
- **RTO** by promise shipments were included in the model, which in turns reduce the actual breach rate

Business Context - Breach

- Overall **breach** trends ~7-8 %, where non-customer dependent breach is ~3-4%
- Today, mostly actions are taken, post the shipments get breached(either functionally or cpd level), due to **data visibility issues**
- Need of the hour to take proactive measures to control breaches, which will help in improving the customer experience
- This can be accomplished using **Breach Prediction Model** which will help in identifying potential breach shipment at different **nodes** of the supply chain proactively



Model Potential Output-
Shipment Journey

Shipment Journey Considered - Dispatch Node Model

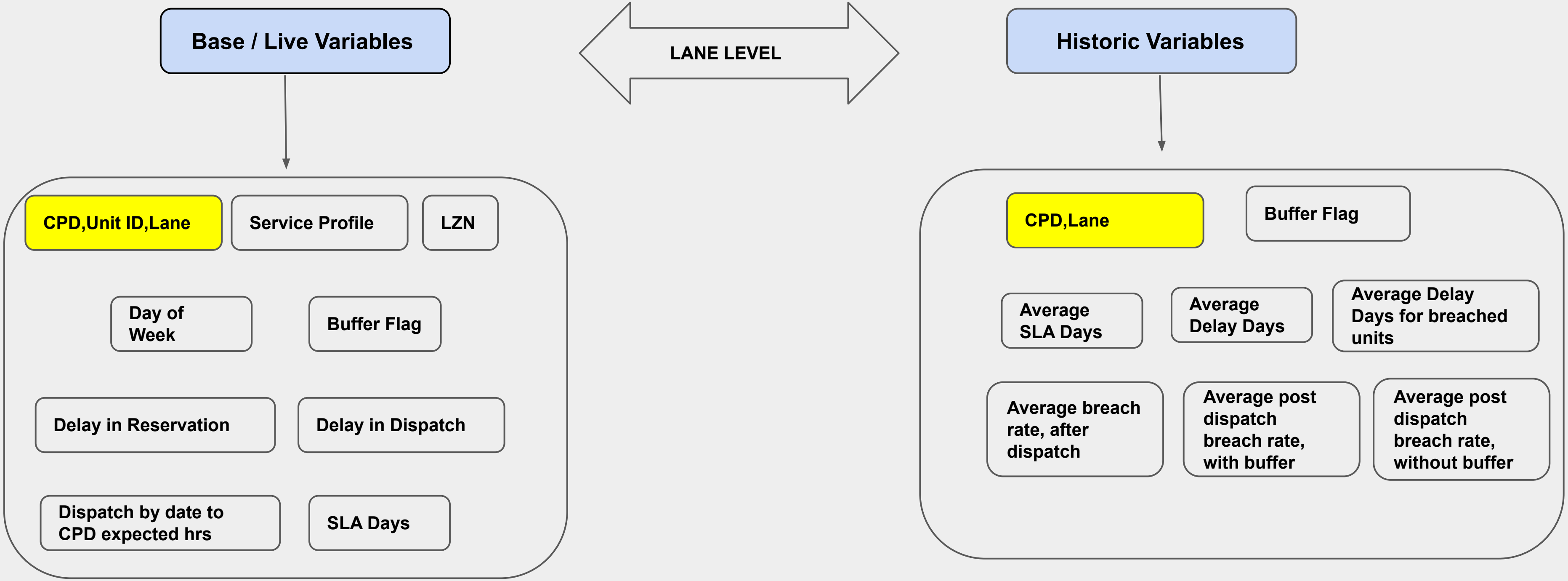
Order Placed → Order Reserved → Order Dispatched → Order Shipped → Order OFD

Data Assumptions

- **Breach Flag :**
  If first out for delivery/RTO is higher than Initial Promise Date then it is an **NCD** breach
- **Shipments** :
  Non Large, Regular, Forward, Flipkart, FSD Shipment
- **Address/Slot Change CPD :**
  Ignored such cases ~8% of shipments as they involve different solve
- **RTO by promise :**
  RTO made before CPD with no OFD, has been excluded from the analysis
- **Boundary Conditions(Lane):**
  Considering **source** MH and **Destination** DH to define the lane
- **Training Data:**
  Considered 1 week **base data** - 6 Nov’22 to 12 Nov’22 **CPD** shipments and made prediction at **Unit-id** level , with **historic** data from 31 Oct’22 to 11 Nov’22

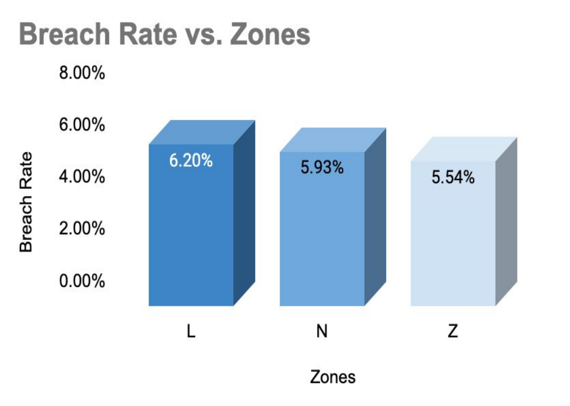
***Base Fact :** bigfoot_external_neo.scp_fulfillment__fulfillment_unit_hive_fact*

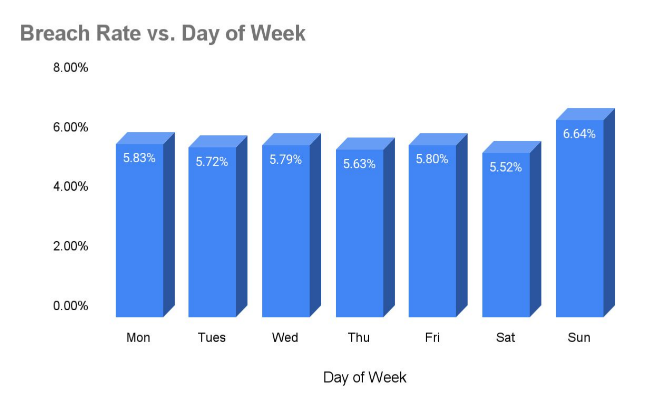
Data Variables for Model

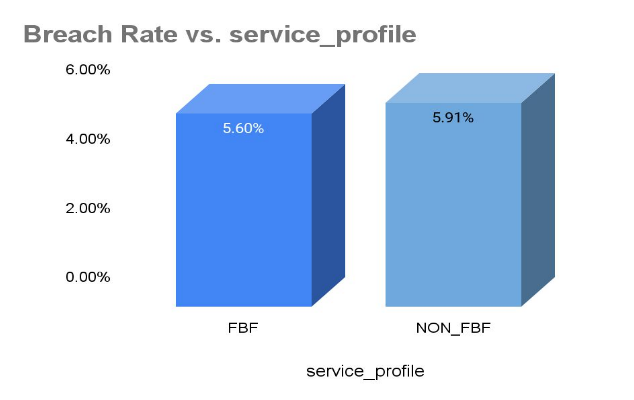


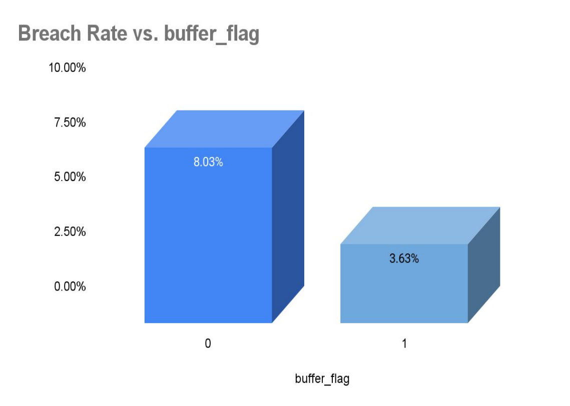
- **Historic Lane Information:** Considered 1 week **rolling average** for **on-time dispatch** shipments(considering **post dispatch** breaches)
- Joined Base and historic information based on **Lane level**
- **Also, considered difference between sla days on base and historic, to check if any buffer/capacity actions have been taken and how it impact breaches**

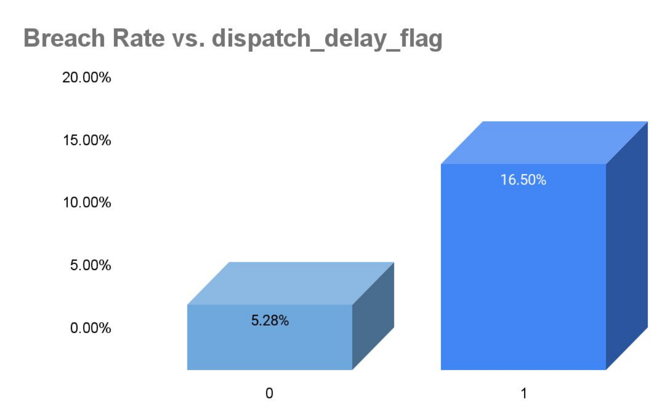
Exploratory Data Analysis

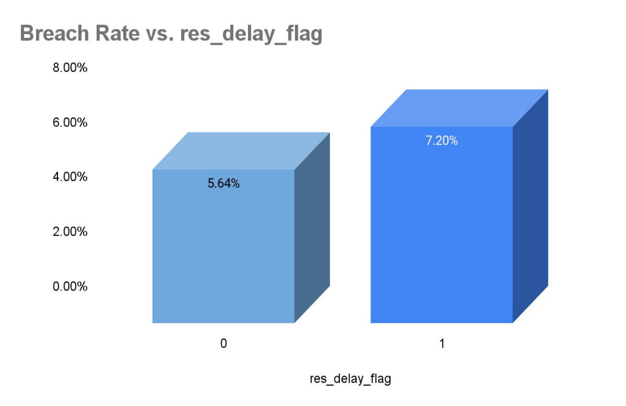












Data Pre-Processing and Modelling Approach

Missing Values

- Initially,we had **20.2M** units in the training dataset, with ~ 5% breach data
- **Dropped** ~3% units where **LZN** information was null and ~.19 % units where there was no historic information available
- Finally, had **19.5M** units used in training the model

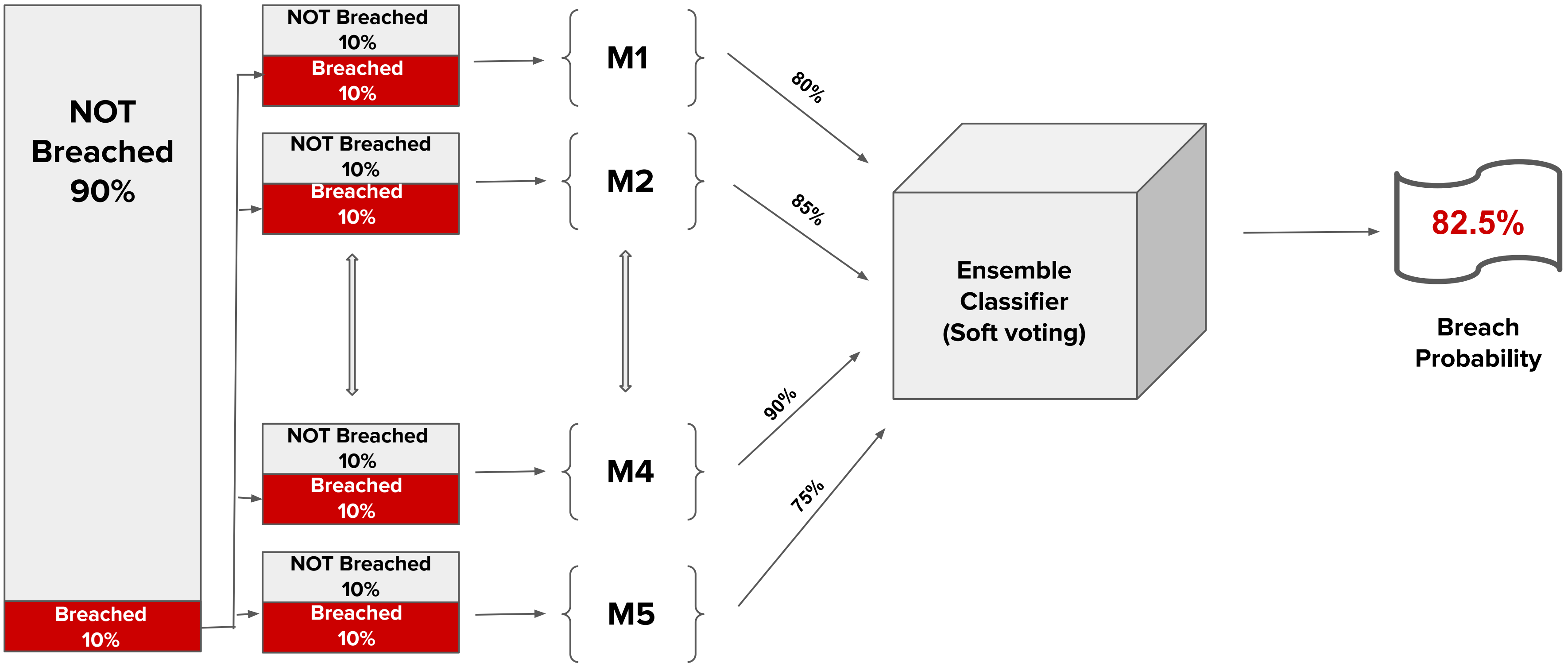
Outlier Treatment

- Used **Five Number summary method** to detect the outliers
- Replaced the outlier values by lower and upper bound
- For some of the variables, also looked at the 99th and 1st percentile and replaced the outliers by these values

Model

- We have a class **Imbalance** issue as Dependent variables has only ~**5% positive values**
- Modelling Technique - **Balanced Balancing Classifier - Ensemble Technique**

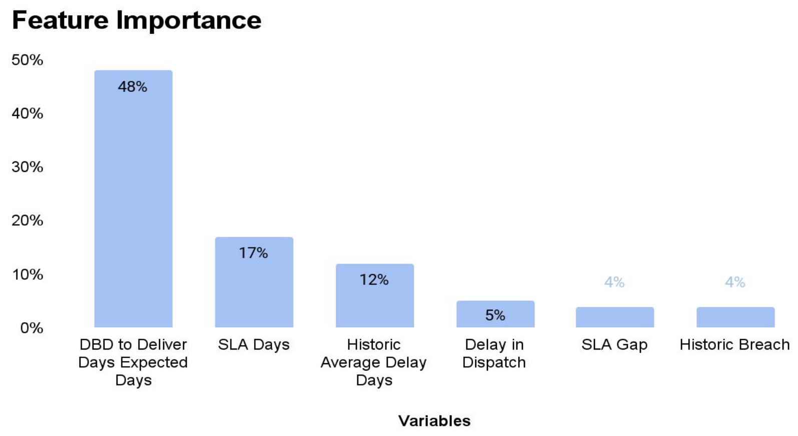
Ensemble model architecture - Bagging technique



Model Results

Training Data Accuracy

	precision	recall	f1-score	support
0.0	1.00	0.98	0.99	18417730
1.0	0.77	0.99	0.87	1143051
accuracy			0.98	19560781
macro avg	0.89	0.99	0.93	19560781
weighted avg	0.99	0.98	0.98	19560781



Out of Sample Accuracy

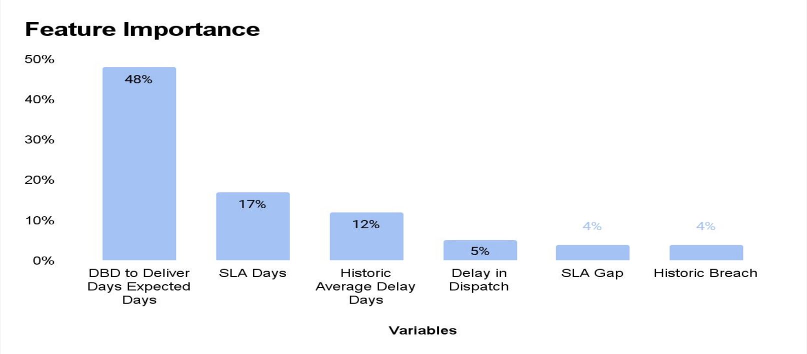
We considered data for **13M** shipments that have been **dispatched** from 13th Nov’22 to 19th Nov’22(with **CPD** within 1 week) and took one week rolling historical information(6th Nov’22- 18th Nov’22)

	precision	recall	f1-score	support
0.0	1.00	0.99	0.99	12328964
1.0	0.79	0.92	0.85	710830
accuracy			0.98	13039794
macro avg	0.89	0.95	0.92	13039794
weighted avg	0.98	0.98	0.98	13039794

Model Results & Accuracy (*Out of Sample*)

Model was **trained** from 6 Nov’22 to 12 Nov’22 **CPD** shipments and tested the model on **13M** shipment, **dispatched** from 13th Nov’22 to 19th Nov’22 (with **CPD** within 1 week) and took one week rolling historical information(6th Nov’22- 18th Nov’22)

	precision	recall	f1-score	support
0.0	1.00	0.99	0.99	12328964
1.0	0.79	0.92	0.85	710830
accuracy			0.98	13039794
macro avg	0.89	0.95	0.92	13039794
weighted avg	0.98	0.98	0.98	13039794

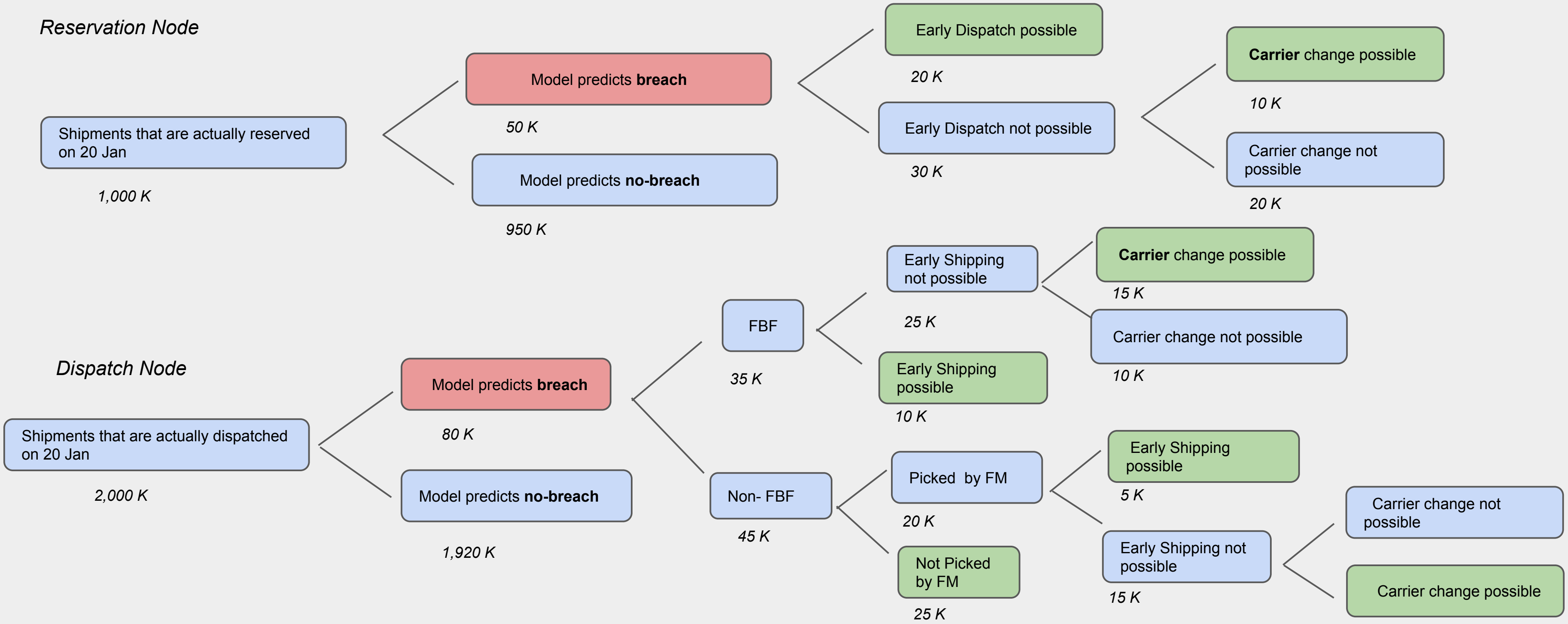


Recall	CPD - Dispatch Date Difference							
Dispatch Date	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
13/11/22	0.90	0.90	0.94	0.94	0.95	0.94	0.88	0.93
14/11/22	0.91	0.76	0.92	0.92	0.94	0.88	0.88	0.89
15/11/22	0.91	0.88	0.94	0.92	0.90	0.91	0.92	0.91
16/11/22	0.82	0.89	0.90	0.90	0.92	0.95	0.92	0.91
17/11/22	0.88	0.90	0.91	0.91	0.96	0.95	0.90	0.93
18/11/22	0.86	0.90	0.92	0.96	0.96	0.94	0.92	0.92
19/11/22	0.87	0.92	0.96	0.96	0.96	0.95	0.91	0.94
Overall	0.88	0.87	0.93	0.93	0.94	0.93	0.90	

Precision	CPD - Dispatch Date Difference							
Dispatch Date	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
13/11/22	0.98	0.96	0.93	0.94	0.90	0.79	0.68	0.87
14/11/22	0.98	0.94	0.95	0.93	0.90	0.88	0.23	0.64
15/11/22	0.98	0.95	0.96	0.94	0.91	0.63	0.56	0.80
16/11/22	0.99	0.96	0.95	0.92	0.80	0.77	0.63	0.82
17/11/22	0.99	0.95	0.95	0.93	0.88	0.75	0.66	0.82
18/11/22	0.98	0.95	0.95	0.93	0.85	0.73	0.51	0.85
19/11/22	0.99	0.95	0.95	0.89	0.86	0.72	0.56	0.80
Overall	0.98	0.95	0.95	0.93	0.87	0.75	0.47	

Closer the gap between CPD and dispatch date, better the Precision

Action Plan



- **Action Playbook -** Need to identify actions to be taken for reservation and dispatch node, for high potential breach shipments (*CT - Ops Dependency*)
- **Model Development** for rest of the nodes, where considerable actions can be taken
- Enable **prediction model** into live systems for real time visibility(*Product Dependency*)

Results to be shared at D-1 level

Overall level structure

Customer Promise Date

Shipments with High Breach Probability (as on 20th Jan)

21 JAN	22 JAN	23 JAN	24 JAN	25 JAN	26 JAN
10K	20k	30k	50k	45k	32k

Reservation Node (20 Jan)	5K
Dispatch Node(20 Jan)	5K
Shipment Node(20 Jan)	10K
DH Node(20 Jan)	10K

Further breakdown at **Lane/Facility** Level

Accuracy Level across different cuts

Training data - 6 Nov’22 to 12 Nov’22 **CPD** shipments

Recall	CPD - Dispatch Date Difference							
Dispatch Date	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
13/11/22	0.90	0.90	0.94	0.94	0.95	0.94	0.88	0.93
14/11/22	0.91	0.76	0.92	0.92	0.94	0.88	0.88	0.89
15/11/22	0.91	0.88	0.94	0.92	0.90	0.91	0.92	0.91
16/11/22	0.82	0.89	0.90	0.90	0.92	0.95	0.92	0.91
17/11/22	0.88	0.90	0.91	0.91	0.96	0.95	0.90	0.93
18/11/22	0.86	0.90	0.92	0.96	0.96	0.94	0.92	0.92
19/11/22	0.87	0.92	0.96	0.96	0.96	0.95	0.91	0.94
Overall	0.88	0.87	0.93	0.93	0.94	0.93	0.90	

Precision	CPD - Dispatch Date Difference							
Dispatch Date	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
13/11/22	0.98	0.96	0.93	0.94	0.90	0.79	0.68	0.87
14/11/22	0.98	0.94	0.95	0.93	0.90	0.88	0.23	0.64
15/11/22	0.98	0.95	0.96	0.94	0.91	0.63	0.56	0.80
16/11/22	0.99	0.96	0.95	0.92	0.80	0.77	0.63	0.82
17/11/22	0.99	0.95	0.95	0.93	0.88	0.75	0.66	0.82
18/11/22	0.98	0.95	0.95	0.93	0.85	0.73	0.51	0.85
19/11/22	0.99	0.95	0.95	0.89	0.86	0.72	0.56	0.80
Overall	0.98	0.95	0.95	0.93	0.87	0.75	0.47	

Closer the gap between CPD and dispatch date, better the Precision

What are the Next Steps ?

Customer Promise Date

	21 JAN	22 JAN	23 JAN	24 JAN	25 JAN	26 JAN
# Shipments with High Breach Probability (as on 20th Jan)	10K	20k	30k	50k	45k	32k

Reservation Node (20 Jan)	5K
Dispatch Node(20 Jan)	5K
Shipment Node(20 Jan)	10K
DH Node(20 Jan)	10K

Further breakdown at **Lane/Facility** Level

Model Results

Without SLA

Training Data Accuracy

	precision	recall	f1-score	support
0.0	1.00	0.97	0.98	18417730
1.0	0.65	0.99	0.78	1143051
accuracy			0.97	19560781
macro avg	0.82	0.98	0.88	19560781
weighted avg	0.98	0.97	0.97	19560781

	var	prop
0	0.004014	day_of_week_Thu
1	0.004131	day_of_week_Wed
2	0.004630	day_of_week_Sat
3	0.004757	day_of_week_Tues
4	0.004905	day_of_week_Mon
5	0.007457	day_of_week_Sun
6	0.015001	delay_in_reservation
7	0.016896	lzn_Z
8	0.024640	service_profile_NON_FBF
9	0.037980	lzn_N
10	0.047147	buffer_flag
11	0.049728	sla_gap
12	0.053976	delay_in_dispatch
13	0.059605	historic_breach
14	0.059798	hist_avg_delay_days_breached_units
15	0.178103	hist_avg_delay_days
16	0.427233	dbd_to_Deliver_days

Out of Sample Accuracy

We considered data for **12.5M** shipments that have been **dispatched** from 13th Nov'22 to 19th Nov'22(with **CPD** within 1 week) and took one week rolling historical information(6th Nov'22- 18th Nov'22)

	precision	recall	f1-score	support
0.0	0.98	0.96	0.97	12328964
1.0	0.53	0.72	0.61	710830
accuracy			0.95	13039794
macro avg	0.75	0.84	0.79	13039794
weighted avg	0.96	0.95	0.95	13039794