Machine Learning in Finance Workshop 2021

Improving Bond Trading Workflows by Learning to Rank RFQs

Andy Almonte

Bloomberg

Improving Bond Trading Workflows by Learning to Rank RFQs

Bloomberg

Engineering

Machine Learning in Finance 2021
September 17, 2021

Andy Almonte
Team Lead, AI Signal Extraction

TechAtBloomberg.com

Overview: Stock Market vs. Fixed Income Market

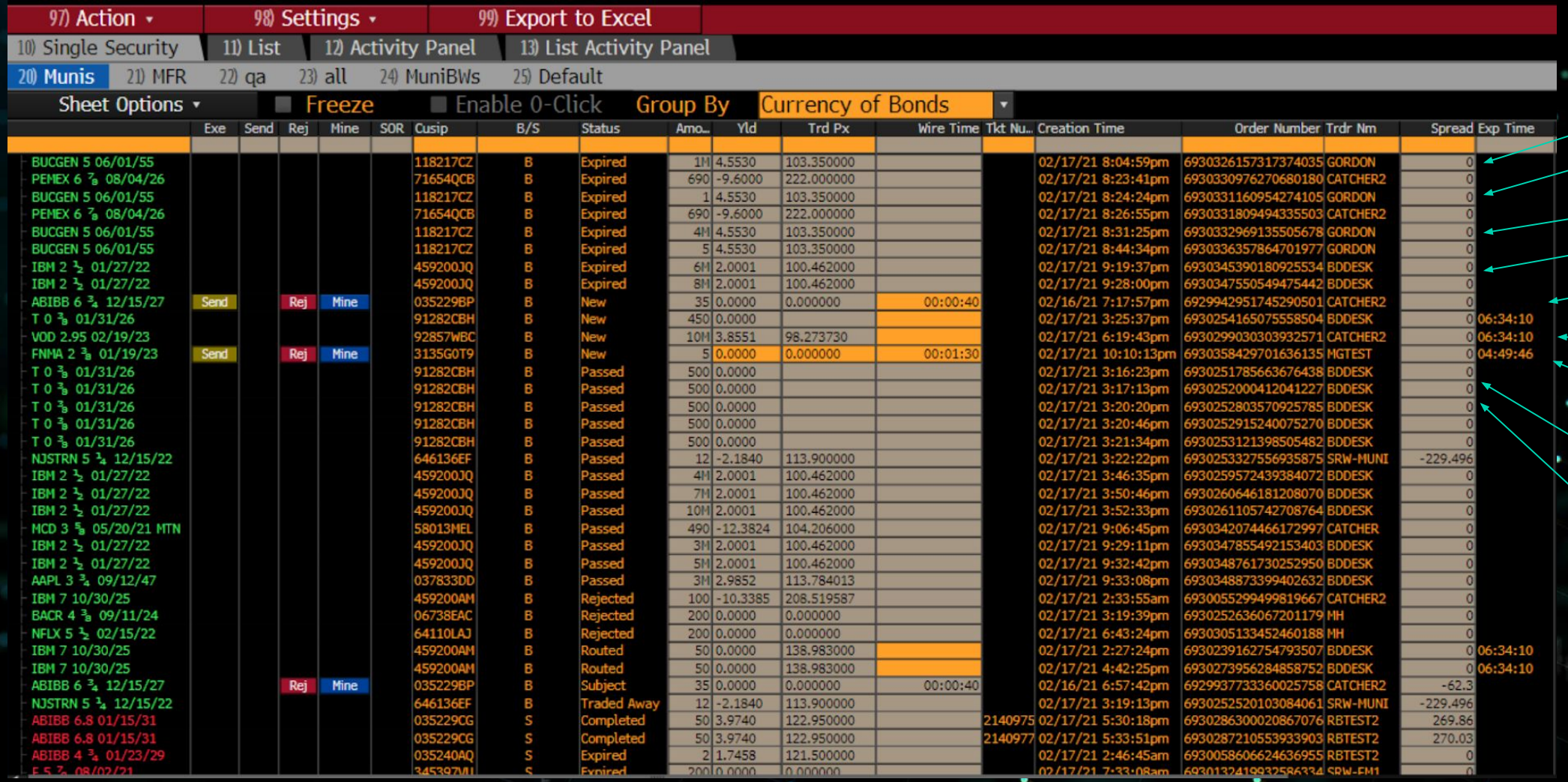
Stock Market

- Trades on centralized exchanges
  - Ex. NYSE, NASDAQ, London Stock Exchange
- Strict trade reporting requirements
- Higher volatility
- Largely automated

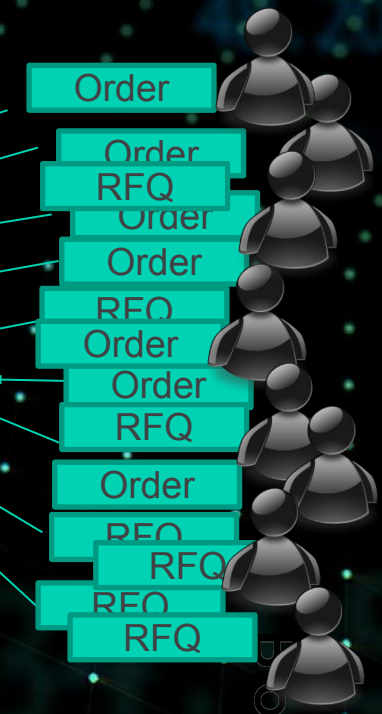
Fixed Income Market

- Trades over-the-counter, largely through major banks (termed the “Sell Side”)
  - Ex. JP Morgan, Wells Fargo
- Looser trade reporting requirements
- Lower volatility
- Human effort required to facilitate trades

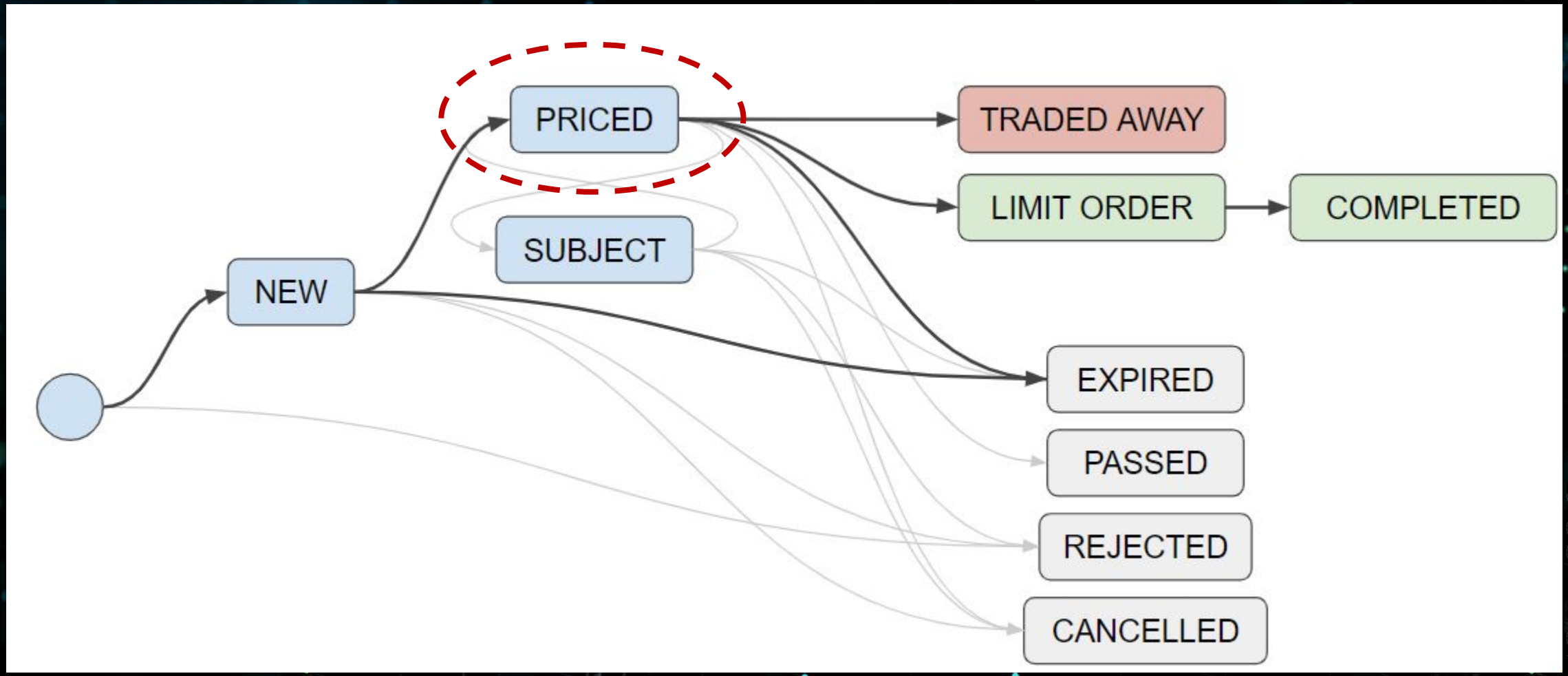
The Life of a Bond Trader: SSOX <GO>



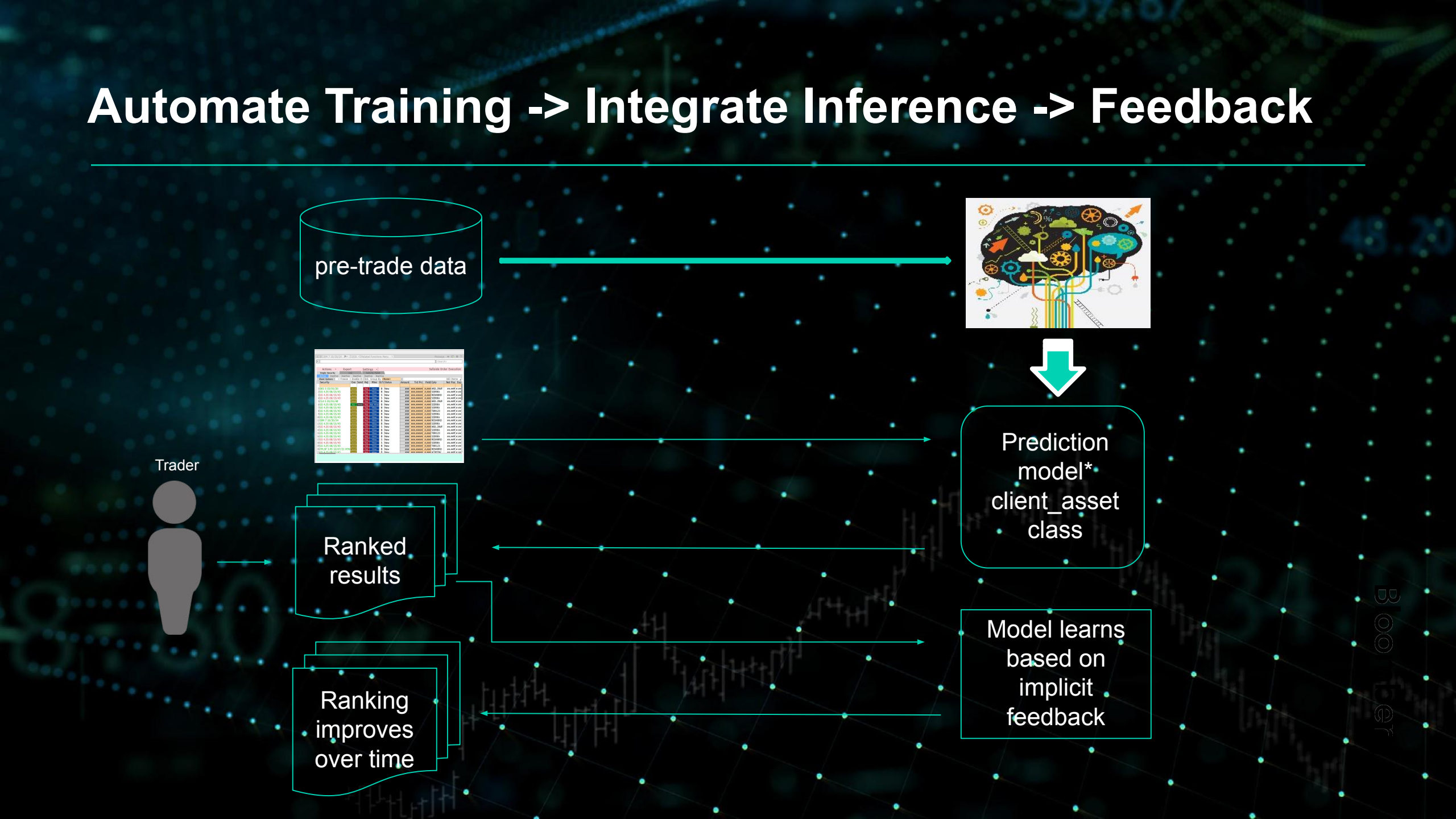
	Exe	Send	Rej	Mine	SOR	Cusip	B/S	Status	Amo…	Yld	Trd Px	Wire Time	Tkt Nu…	Creation Time	Order Number	Trdr Nm	Spread	Exp Time
BUCGEN 5 06/01/55						118217CZ	B	Expired	1M	4.5530	103.350000			02/17/21 8:04:59pm	6930326157317374035	GORDON	0	
PEMEX 6 7/8 08/04/26						71654QCB	B	Expired	690	-9.6000	222.000000			02/17/21 8:23:41pm	6930330976270680180	CATCHER2	0	
BUCGEN 5 06/01/55						118217CZ	B	Expired	1	4.5530	103.350000			02/17/21 8:24:24pm	6930331160954274105	GORDON	0	
PEMEX 6 7/8 08/04/26						71654QCB	B	Expired	690	-9.6000	222.000000			02/17/21 8:26:55pm	6930331809494335503	CATCHER2	0	
BUCGEN 5 06/01/55						118217CZ	B	Expired	4M	4.5530	103.350000			02/17/21 8:31:25pm	6930332969135505678	GORDON	0	
BUCGEN 5 06/01/55						118217CZ	B	Expired	5	4.5530	103.350000			02/17/21 8:44:34pm	6930336357864701977	GORDON	0	
IBM 2 1/2 01/27/22						459200JQ	B	Expired	6M	2.0001	100.462000			02/17/21 9:19:37pm	6930345390180925534	BDDESK	0	
IBM 2 1/2 01/27/22						459200JQ	B	Expired	8M	2.0001	100.462000			02/17/21 9:28:00pm	6930347550549475442	BDDESK	0	
ABIBB 6 3/4 12/15/27	Send		Rej	Mine		035229BP	B	New	35	0.0000	0.000000	00:00:40		02/16/21 7:17:57pm	6929942951745290501	CATCHER2	0	
T 0 3/8 01/31/26						91282CBH	B	New	450	0.0000				02/17/21 3:25:37pm	6930254165075558504	BDDESK	0	06:34:10
VOD 2.95 02/19/23						92857WBC	B	New	10M	3.8551	98.273730			02/17/21 6:19:43pm	6930299030303932571	CATCHER2	0	06:34:10
FNMA 2 3/8 01/19/23	Send		Rej	Mine		3135G0T9	B	New	5	0.0000	0.000000	00:01:30		02/17/21 10:10:13pm	6930358429701636135	MGTEST	0	04:49:46
T 0 3/8 01/31/26						91282CBH	B	Passed	500	0.0000				02/17/21 3:16:23pm	6930251785663676438	BDDESK	0	
T 0 3/8 01/31/26						91282CBH	B	Passed	500	0.0000				02/17/21 3:17:13pm	6930252000412041227	BDDESK	0	
T 0 3/8 01/31/26						91282CBH	B	Passed	500	0.0000				02/17/21 3:20:20pm	6930252803570925785	BDDESK	0	
T 0 3/8 01/31/26						91282CBH	B	Passed	500	0.0000				02/17/21 3:20:46pm	6930252915240075270	BDDESK	0	
T 0 3/8 01/31/26						91282CBH	B	Passed	500	0.0000				02/17/21 3:21:34pm	6930253121398505482	BDDESK	0	
NJSTRN 5 1/4 12/15/22						646136EF	B	Passed	12	-2.1840	113.900000			02/17/21 3:22:22pm	6930253327556935875	SRW-MUNI	-229.496	
IBM 2 1/2 01/27/22						459200JQ	B	Passed	4M	2.0001	100.462000			02/17/21 3:46:35pm	6930259572439384072	BDDESK	0	
IBM 2 1/2 01/27/22						459200JQ	B	Passed	7M	2.0001	100.462000			02/17/21 3:50:46pm	6930260646181208070	BDDESK	0	
IBM 2 1/2 01/27/22						459200JQ	B	Passed	10M	2.0001	100.462000			02/17/21 3:52:33pm	6930261105742708764	BDDESK	0	
MCD 3 5/8 05/20/21 MTN						58013MEL	B	Passed	490	-12.3824	104.206000			02/17/21 9:06:45pm	6930342074466172997	CATCHER	0	
IBM 2 1/2 01/27/22						459200JQ	B	Passed	3M	2.0001	100.462000			02/17/21 9:29:11pm	6930347855492153403	BDDESK	0	
IBM 2 1/2 01/27/22						459200JQ	B	Passed	5M	2.0001	100.462000			02/17/21 9:32:42pm	6930348761730252950	BDDESK	0	
AAPL 3 3/4 09/12/47						037833DD	B	Passed	3M	2.9852	113.784013			02/17/21 9:33:08pm	6930348873399402632	BDDESK	0	
IBM 7 10/30/25						459200AM	B	Rejected	100	-10.3385	208.519587			02/17/21 2:33:55am	6930055299499819667	CATCHER2	0	
BACR 4 3/8 09/11/24						06738EAC	B	Rejected	200	0.0000	0.000000			02/17/21 3:19:39pm	6930252636067201179	MH	0	
NFLX 5 1/2 02/15/22						64110LAJ	B	Rejected	200	0.0000	0.000000			02/17/21 6:43:24pm	6930305133452460188	MH	0	
IBM 7 10/30/25						459200AM	B	Routed	50	0.0000	138.983000			02/17/21 2:27:24pm	6930239162754793507	BDDESK	0	06:34:10
IBM 7 10/30/25						459200AM	B	Routed	50	0.0000	138.983000			02/17/21 4:42:25pm	6930273956284858752	BDDESK	0	06:34:10
ABIBB 6 3/4 12/15/27			Rej	Mine		035229BP	B	Subject	35	0.0000	0.000000	00:00:40		02/16/21 6:57:42pm	6929937733360025758	CATCHER2	-62.3	
NJSTRN 5 1/4 12/15/22						646136EF	B	Traded Away	12	-2.1840	113.900000			02/17/21 3:19:13pm	6930252520103084061	SRW-MUNI	-229.496	
ABIBB 6.8 01/15/31						035229CG	S	Completed	50	3.9740	122.950000		2140975	02/17/21 5:30:18pm	6930286300020867076	RBTEST2	269.86	
ABIBB 6.8 01/15/31						035229CG	S	Completed	50	3.9740	122.950000		2140977	02/17/21 5:33:51pm	6930287210553933903	RBTEST2	270.03	
ABIBB 4 3/4 01/23/29						035240AQ	S	Expired	2	1.7458	121.500000			02/17/21 2:46:45am	6930058606624636955	RBTEST2	0	
F 5 7/8 08/02/21						345397WU	S	Expired	200	0.0000	0.000000			02/17/21 7:33:08am	6930132419932586334	SRW-FM1	0	



RFQ Workflow



Automate Training -> Integrate Inference -> Feedback



Machine Learning Problem Statement

Task

Rank RFQs in a useful way

Approach

Rank by the probability that an RFQ will be priced. This transforms our ranking problem into **binary classification**.

Implementation

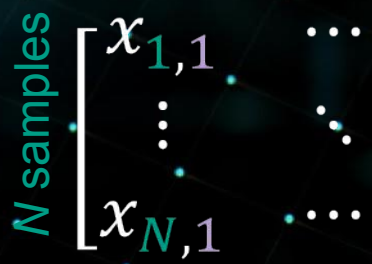
We use a **random forest** model to predict the **probability** that an RFQ will be priced based on its characteristics (amount, price, side, etc.) trained on RFQs from the previous months. We retrain periodically.

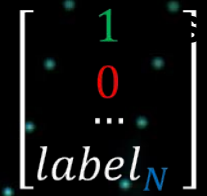
Training / Inference Pipeline

Input Data

- Trade attributes
- Dealer attributes

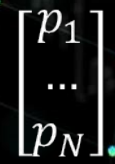
Featurization



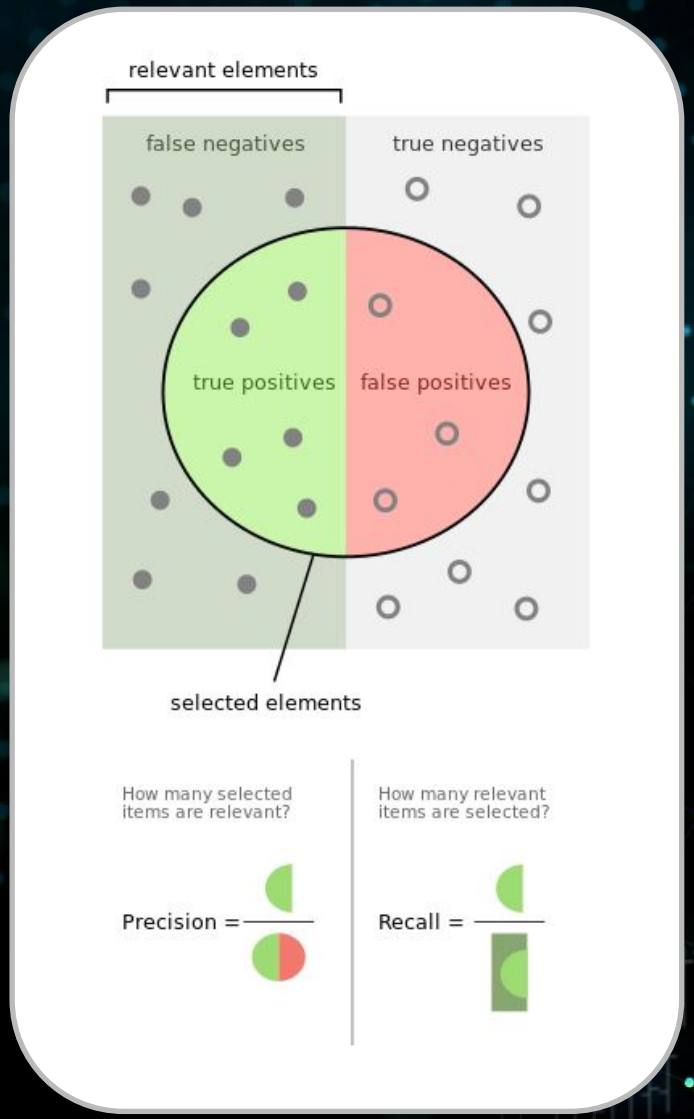


ML Model
(Random Forest)

ML Score



Binary Classification Metrics: Precision and Recall



Threshold = 0.5
Precision = 2/2 = 1.0
Recall = 2/7 = 0.29

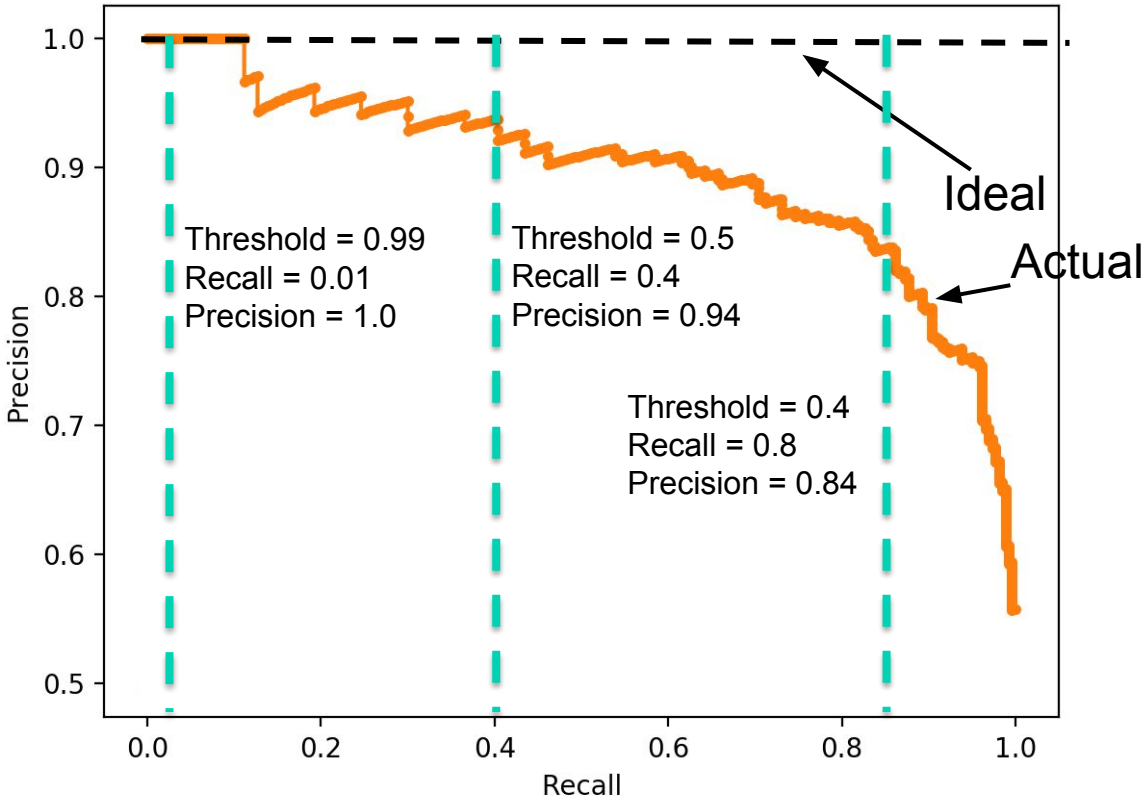
Threshold = 0.1
Precision = 5/6 = 0.83
Recall = 5/7 = 0.71

Threshold = 0.0
Precision = 7/13 = 0.54
Recall = 7/7 = 1.0

	Outcome
0.67	PRICED
0.52	PRICED
0.49	PRICED
0.43	PRICED
0.31	NOT PRICED
0.19	PRICED
0.09	NOT PRICED
0.05	NOT PRICED
0.03	PRICED
0.03	NOT PRICED
0.02	NOT PRICED
0.01	PRICED
0.00	NOT PRICED

Binary Classification Metrics: AUC-PR

Area Under the Precision Recall Curve (AUC-PR) measures how well the model performs across all possible confidence thresholds



- Each confidence threshold leads to a specific precision and specific recall
- AUC-PR measures the area under the curve (optimal = 1.0)

Ranking Metrics – Precision @ N

$$Precision@N = \frac{P_N}{N}$$

$P_N = True\ positives\ in\ N\ highest\ ranked\ samples$

Example 1
Precision@10 = 0.6

ML Score	Outcome
0.67	PRICED
0.52	PRICED
0.49	PRICED
0.43	PRICED
0.31	NOT PRICED
0.19	PRICED
0.09	NOT PRICED
0.05	NOT PRICED
0.03	PRICED
0.03	NOT PRICED
0.02	NOT PRICED
0.01	PRICED
0.00	NOT PRICED

Example 2
Precision@10 = 0.3

ML Score	Outcome
0.98	PRICED
0.92	PRICED
0.71	PRICED
0.10	NOT PRICED
0.09	NOT PRICED
0.09	NOT PRICED
0.09	NOT PRICED
0.05	NOT PRICED
0.04	NOT PRICED
0.02	NOT PRICED
0.02	NOT PRICED
0.00	NOT PRICED
0.00	NOT PRICED

Discounted Cumulative Gain (DCG) @ N

$$DCG_{@N} = \sum_{i=1}^{N} \frac{(Scored\ Correctly)}{\log_2(i+1)}$$

Higher ranked orders matter more

Example 1

ML Score	Outcome
0.67	PRICED
0.52	PRICED
0.49	PRICED
0.43	PRICED
0.31	NOT PRICED
0.19	PRICED

Precision@5 = 0.8

DCG@5 = 2.56

Example 2

ML Score	Outcome
0.78	NOT PRICED
0.61	PRICED
0.42	PRICED
0.39	PRICED
0.35	PRICED
0.21	PRICED

Precision@5 = 0.8

DCG@5 = 1.95

Ranking Metrics – NDCG @ N

NDCG = Normalized DCG
(Discounted Cumulative Gain)

$$NDCG@N = \frac{DCG_{@N}}{IDCG_{@N}}$$

$DCG_{@N} = DCG\ for\ N\ highest\ ranked\ samples$

$IDCG_{@N} = Best\ possible\ DCG\ for\ N\ samples$

Example 1
NDCG@10 = 0.97

ML Score	Outcome
0.67	PRICED
0.52	PRICED
0.49	PRICED
0.43	PRICED
0.31	NOT PRICED
0.19	PRICED
0.09	NOT PRICED
0.05	NOT PRICED
0.03	PRICED
0.03	NOT PRICED
0.02	NOT PRICED
0.01	PRICED
0.00	NOT PRICED

Example 2
NDCG@10 = 1.0

ML Score	Outcome
0.98	PRICED
0.92	PRICED
0.71	PRICED
0.10	NOT PRICED
0.09	NOT PRICED
0.09	NOT PRICED
0.09	NOT PRICED
0.05	NOT PRICED
0.04	NOT PRICED
0.02	NOT PRICED
0.02	NOT PRICED
0.00	NOT PRICED
0.00	NOT PRICED

Calibration

- If a group of orders have probability of approximately 20%, then approximately 20% of them should be priced

- A separate model monotonically transforms the output of the Random Forest “probability” so that the above statement becomes more correct, while preserving order

ML Score	Outcome
0.67	PRICED
0.52	PRICED
0.49	PRICED
0.43	PRICED
0.31	NOT PRICED
0.19	PRICED
0.09	NOT PRICED
0.05	NOT PRICED
0.03	PRICED
0.03	NOT PRICED
0.02	NOT PRICED
0.01	PRICED
0.00	NOT PRICED

ML Score	Outcome
0.82	PRICED
0.81	PRICED
0.80	PRICED
0.79	PRICED
0.78	NOT PRICED
0.51	PRICED
0.50	NOT PRICED
0.50	NOT PRICED
0.49	PRICED
0.34	NOT PRICED
0.33	NOT PRICED
0.32	PRICED
0.00	NOT PRICED

Recap: Model Performance Metrics

AUC-PR: The **most direct measure** of how the model performs its **stated task**

Precision@10: An **interpretable ranking metric** that measures how well our top ranked RFQs perform

NDCG@10: Incorporates weights for order and is relative to an idealized ranking, making it our **best metric** for describing **ranking performance**

Results Conclusion

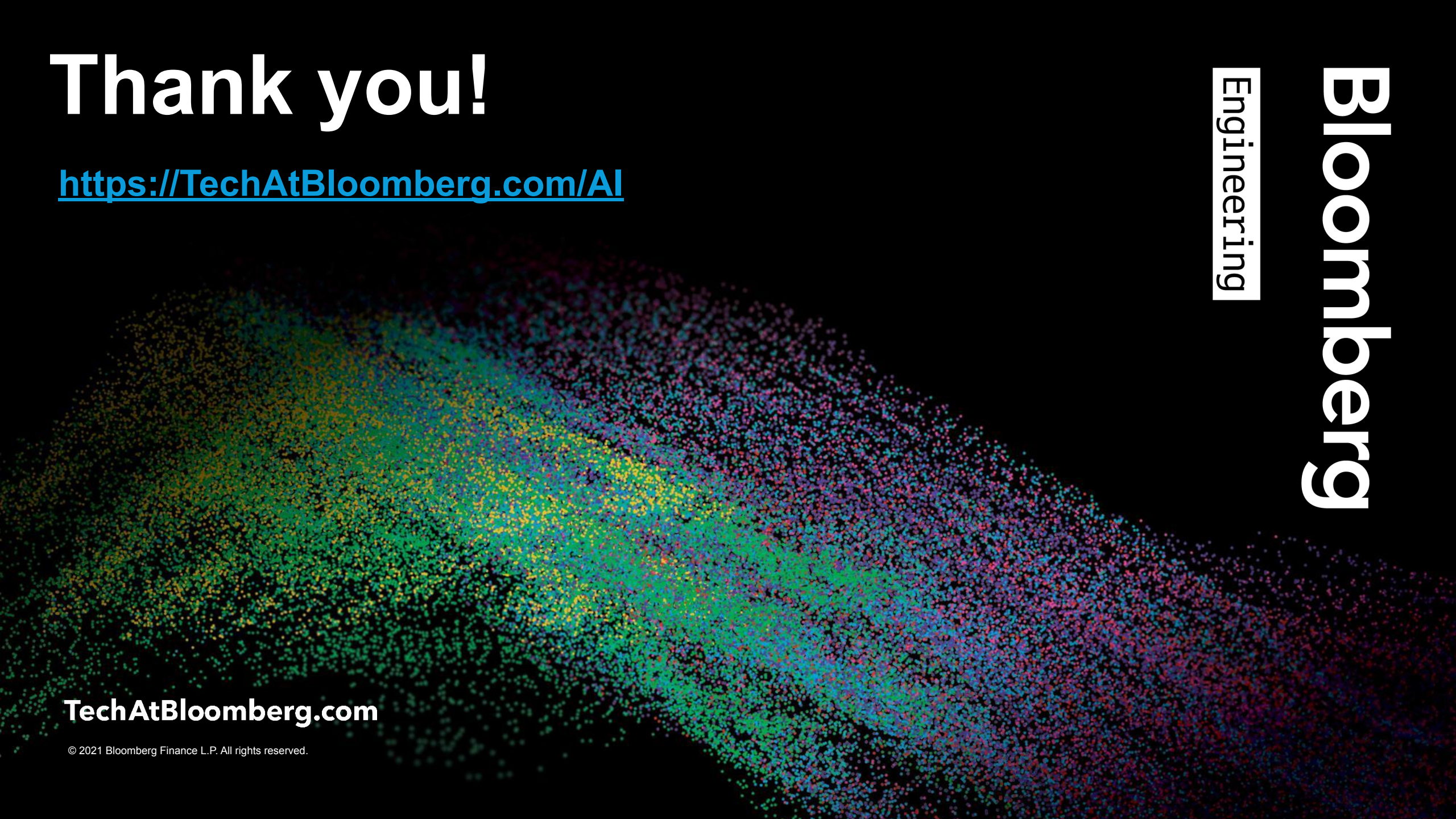
All models perform significantly better than the baseline* across all metrics

Client	AUC-PR (baseline)	Precision@10 (baseline)	NDCG@10 (baseline)
A	**0.087** (0.014)	**0.100** (0.022)	**0.115** (0.022)
B	**0.637** (0.174)	**0.804** (0.211)	**0.853** (0.226)
C	**0.278** (0.017)	**0.331** (0.023)	**0.389** (0.033)
D	**0.521** (0.150)	**0.592** (0.156)	**0.609** (0.161)
E	**0.175** (0.044)	**0.148** (0.044)	**0.365** (0.117)
F	**0.781** (0.062)	**0.712** (0.084)	**0.869** (0.101)
G	**0.884** (0.624)	**0.751** (0.508)	**0.913** (0.605)

*Baselines are by amount and by timestamp: best performing baseline is chosen for comparison
Best performing baseline for E and G is **amount**, for all others it is **timestamp**

Thank you!

https://TechAtBloomberg.com/AI



Bloomberg

Engineering

TechAtBloomberg.com