



Sebastian Roßner

Master in Computational Science and Engineering



Olha Tupko

Master in Mathematical Finance and Actuarial Science



Master in Management

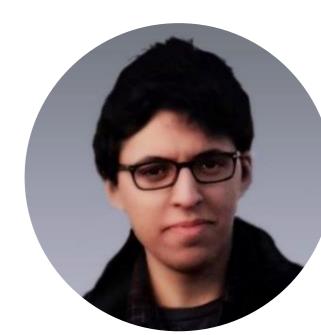
TEAM CELONIS TUM Data Innovation Lab

Keesiu Wong



Margarita Ageeva

Master in Data Engineering and Analytics



Ahmed Ayadi

Master in Electrical and Computer Engineering Master in Management













PHASE II: MACHINE LEARNING







MARKET DEVELOPMENT

Phase I Part 1: Existing Use Cases

MARKET PENETRATION

Existing Products





New Products









PART 2: NEW USE CASES







O PHASE I: DATA ANALYSIS

PHASE II: MACHINE LEARNING



PHASE I: DATA ANALYSIS



PART I: EXISTING USE CASES **PART II: NEW USE CASES**







PHASE I: DATA ANALYSIS

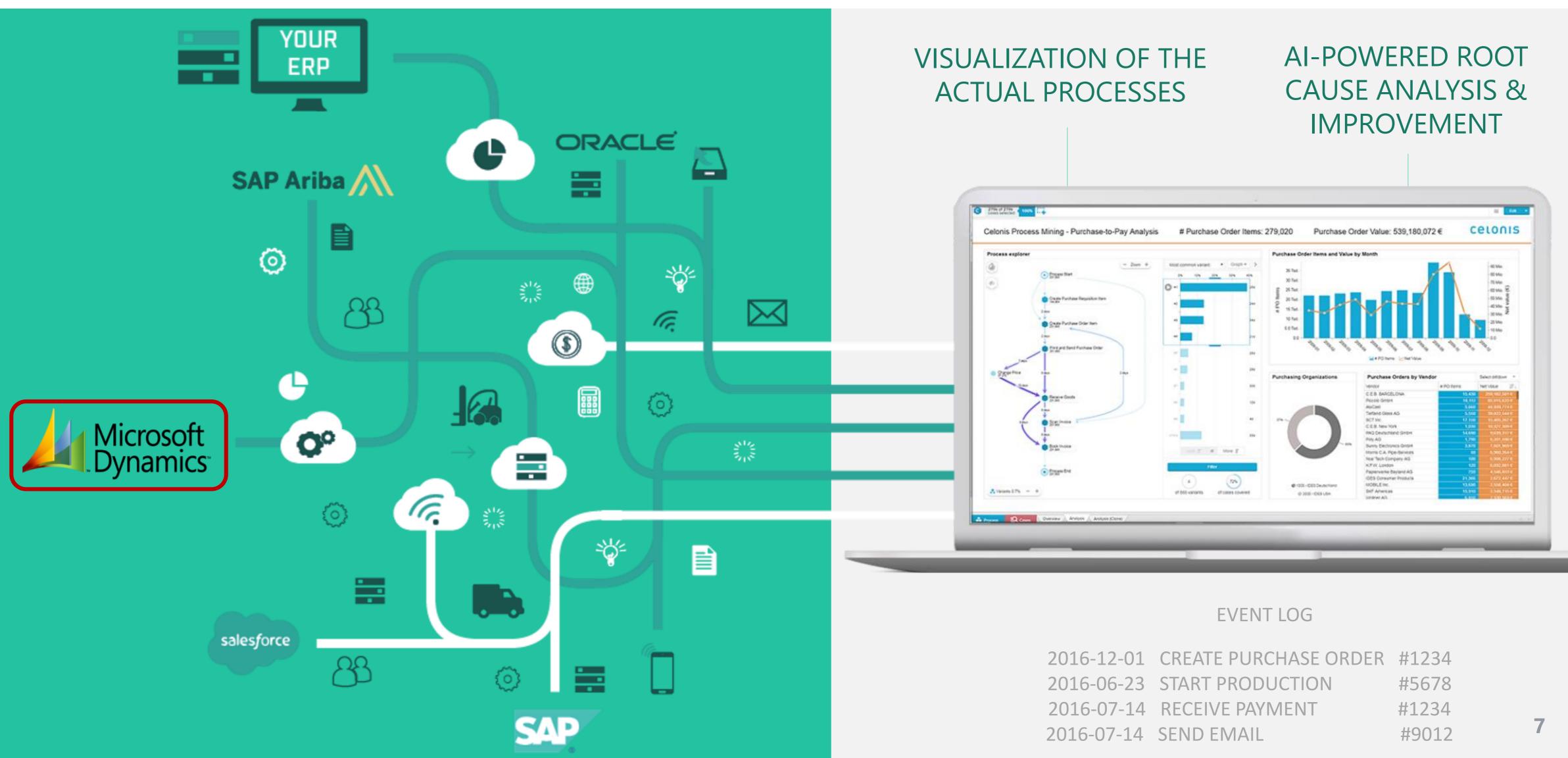
PART I: EXISTING USE CASES











EXISTING USE CASES

11 new Analyses in the Content Store

2016-12-01	CREATE PURCHASE ORDER	#1234
2016-06-23	START PRODUCTION	#5678
2016-07-14	RECEIVE PAYMENT	#1234
2016-07-14	SEND EMAIL	#9012



PHASE I: DATA ANALYSIS



PART I: EXISTING USE CASES **PART II: NEW USE CASES**







PHASE I: DATA ANALYSIS

PART II: NEW USE CASES





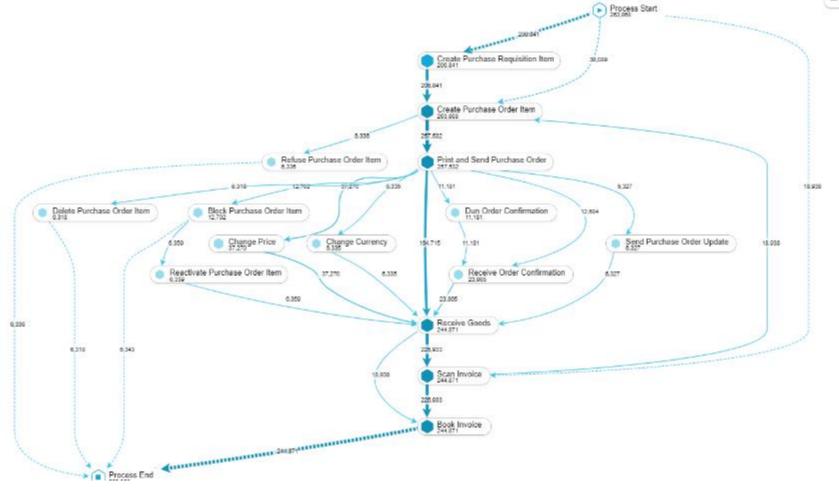


THREE NEW USE CASES for any process, for any ERP platform

COMPLEXITY ANALYSIS

Get a feeling for the complexity for your processes – quantified.





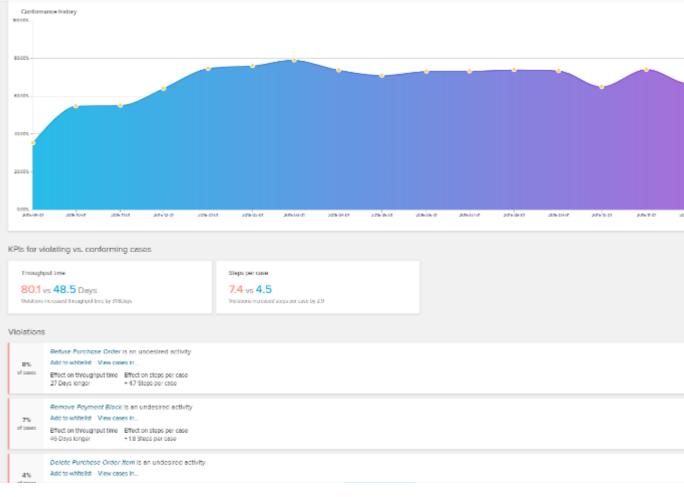
EFFICIENCY ANALYSIS

See exactly which process steps can be improved – *for sure*.



UNDESIRED ACTIVITIE

Analyze where undesired activities happen – and why.





ES	5		
			1
09-12-18	anview • Vinieling covers	eriest Godun	алучна árg сання







COMPLEXITY ANALYSIS Get a feeling for the complexity of your processes – quantified.

Process Start

Create Purchase Requisition Item

Create Purchase Order Item

Record Goods Receipt

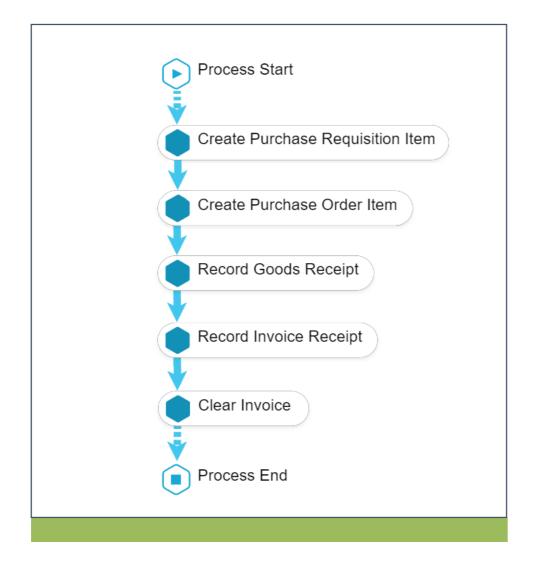
Record Invoice Receipt

Clear Invoice

Process End







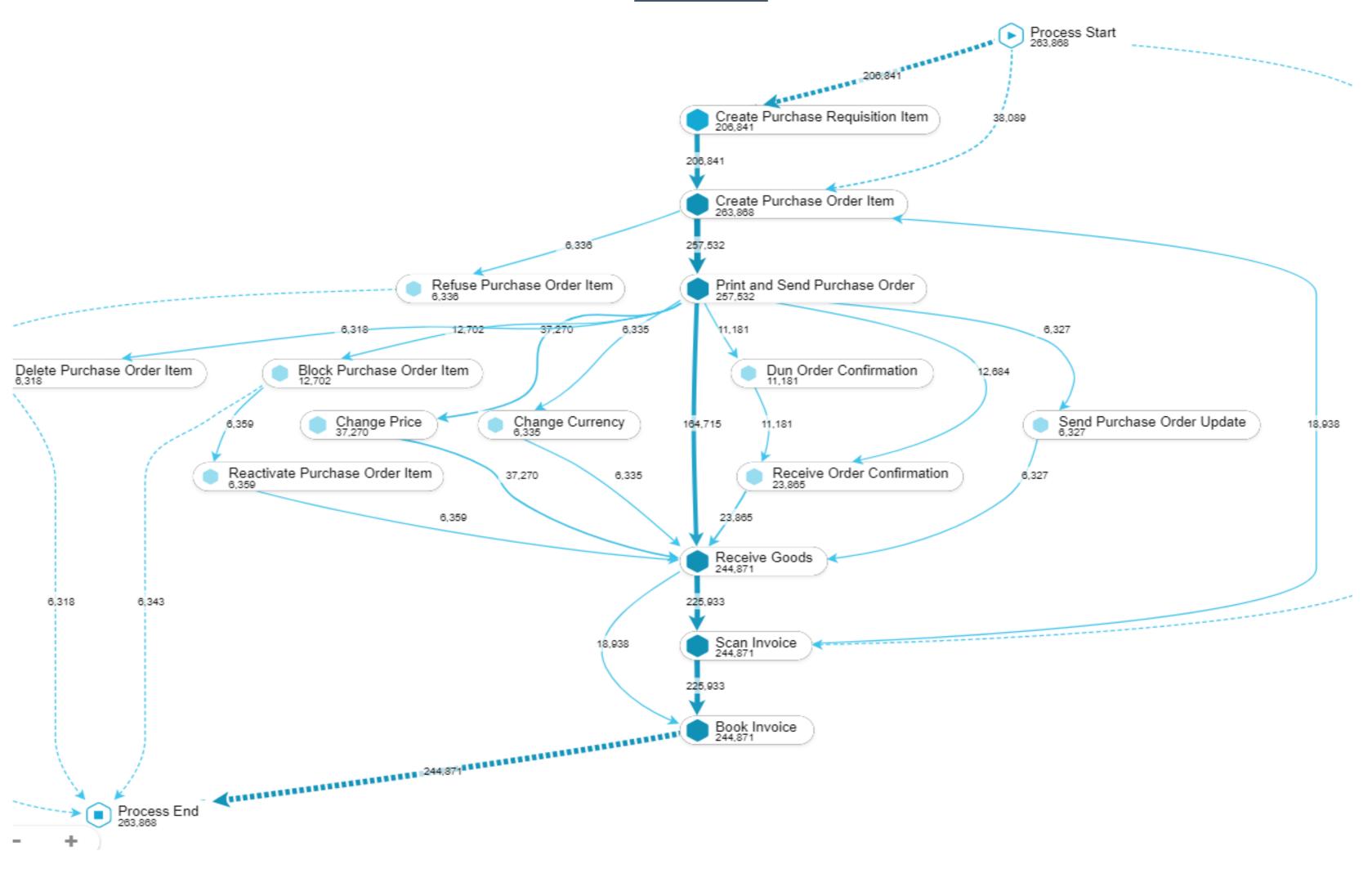
Perfect world



COMPLEXITY ANALYSIS Get a feeling for the complexity of your processes – quantified.







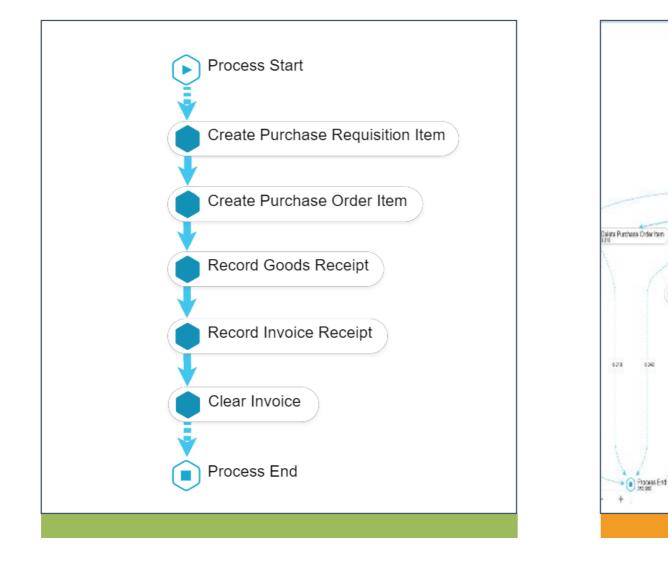


COMPLEXITY ANALYSIS

Get a feeling for the complexity of your processes – quantified.







Perfect world

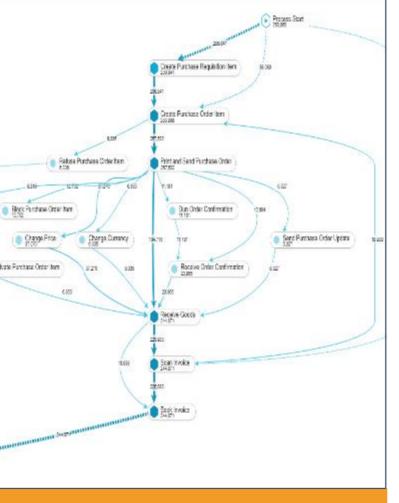


Change Frice

Reactivate Purchase Order Nerr



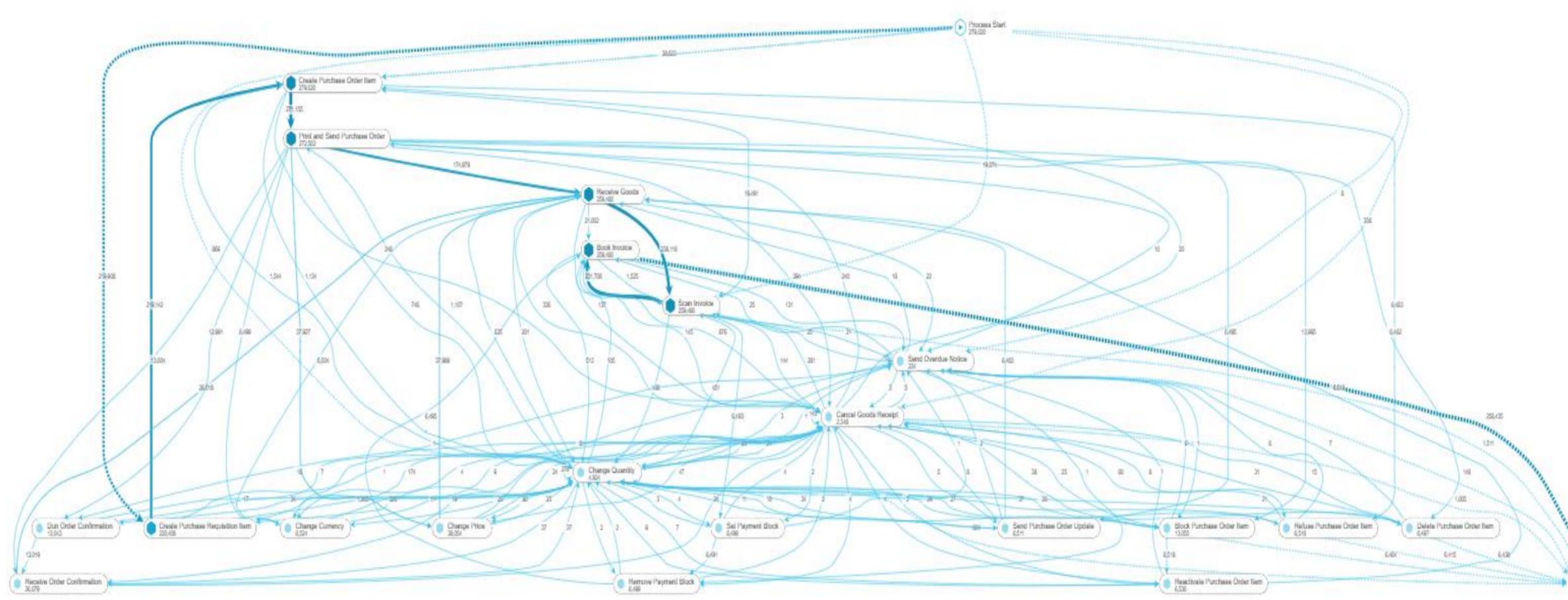
COMPLEXITY ANALYSIS Get a feeling for the complexity of your processes – quantified.



Expectation



COMPLEXITY ANALYSIS Get a feeling for the complexity of your processes – quantified.

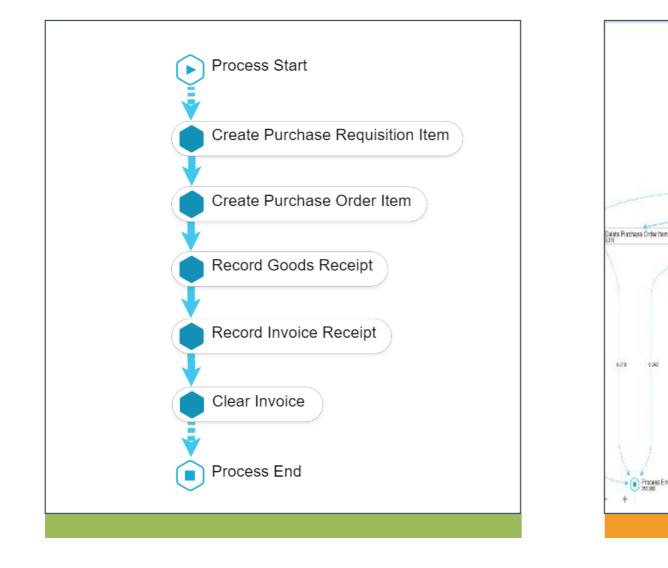










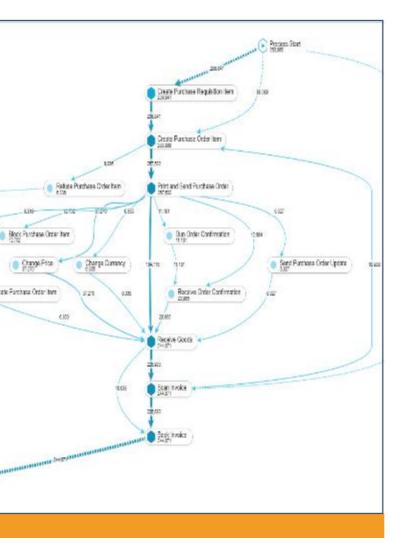


Perfect world





COMPLEXITY ANALYSIS Get a feeling for the complexity of your processes – quantified.





Expectation





COMPLEXITY ANALYSIS Get a feeling for the complexity of your processes – quantified.

celor	115	Average # Users	AVG #A		
Filter Invoice Creation	n Date	2.6	1 5	5.991	
yyyy-MM-dd	Ë				
yyyy-MM-dd	<u>m</u>				
		Variants		Counts	
		Create Purchase Order Ite	m, Print and Send Purch		
Filter by Organization:		Create Purchase Requisiti			
		Create Purchase Requisiti	on Item, Create Purchas m. Print and Send Purch		
a			m. Print and Send Purch		
Company Code	Ŧ		m, Print and Send Purch		
Origin	-		m, Print and Send Purch		
ongin		-	on Item, Create Purchas		
Inventory Model Group	-		m, Print and Send Purch m. Print and Send Purch		
			on Item, Create Purchas		
			on Item, Create Purchas		
Filter by Order Information	tion:		on Item, Create Purchas		
			on Item, Create Purchas on Item, Create Purchas		
		-	m, Send Purchase Order		
Invoice ID					
Sales ID		Average #Activities	over time		
0010010					
Customer	*	900		\wedge	
		800-		/ \	
Customer Group	Ť	700			
Project ID				1	
		600			
Config Group ID	*	500	$\langle \rangle$		
Draduat		400-			
Product		300			
Sales Category	*	200			
		100			
Settings		598-01-02-02-02-02-02-02-02-02-02-02-02-02-02-	5% 1500 151 1500 1500 1500 1500 1500 150	1. Tala antina antina antina	~ (st. 18

s) STD Throughput (Days) STD Throughput (Days) STD Throughput (Days) STD Throughput (Days) Standard Deviation (Days) Average Users O. 665				Edit
0 0.00 2000 22 29.22 2.9056 24 29.95 3.0354 15 28.37 2.8358 11 24.11 2.3206 16 19.63 2.8209 20 35.19 2.9806 34 38.51 2.7855 16 27.19 3.0820 39 46.57 2.7673 32 32.04 2.3361 27 27.81 2.7289 22 26.55 2.845 20 23.83 2.8953 32 34.18 2.6328 0 0.00 2.000				5
22 2922 2905 24 2995 3.0354 15 28.37 2.8358 11 24.11 2.3206 16 19.63 2.8209 20 35.19 2.9806 34 38.51 2.7865 16 27.19 3.0820 39 46.57 2.7673 32 32.04 2.3361 27 27.81 2.7289 22 26.55 2.845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.000	Average Throughput Time (Days)	Standard Deviation (Days)	Average Users	
22 2922 2905 24 2995 3.0354 15 28.37 2.8358 11 24.11 2.3206 16 19.63 2.8209 20 35.19 2.9806 34 38.51 2.7865 16 27.19 3.0820 39 46.57 2.7673 32 32.04 2.3361 27 27.81 2.7289 22 26.55 2.845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.000			-	2,0000
2429.953.03541528.372.83581124.112.32061619.632.82092035.192.98063438.512.78651627.193.08203946.572.76733232.042.3612727.812.72892226.552.88452023.832.89533243.182.632800.002.0000				
15 28.37 2.8358 11 24.11 2.3206 16 19.63 2.8209 20 35.19 2.9806 34 38.51 2.7865 16 27.19 3.0820 39 46.57 2.7673 32 32.04 2.3361 27 27.81 2.7289 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.000				
11 24.11 2.3206 16 19.63 2.8209 20 35.19 2.9806 34 38.51 2.7865 16 27.19 3.0820 39 46.57 2.7673 32 32.04 2.3361 27 27.81 2.7289 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.0000				
1619.632.82092035.192.98063438.512.78651627.193.08203946.572.76733232.042.33612727.812.72892226.552.88452023.832.89533243.182.632800.002.0000				
20 35.19 2.9806 34 38.51 2.7865 16 27.19 3.0820 39 46.57 2.7673 32 32.04 2.3361 27 27.81 2.7289 22 26.55 2.8845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.000				
3438.512.78651627.193.08203946.572.76733232.042.3612727.812.72892226.552.88452023.832.89533243.182.632800.002.0000				
1627.193.08203946.572.76733232.042.3612727.812.72892226.552.88452023.832.89533243.182.632800.002.0000				
39 46.57 2.7673 32 32.04 2.3061 27 27.81 2.7289 22 26.55 2.8845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.0000				
32 32.04 2.361 27 27.81 2.7289 22 26.55 2.8845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.0000				
27 27.81 2.7289 22 26.55 2.845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.0000				
22 26.55 2.8845 20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.0000				
20 23.83 2.8953 32 43.18 2.6328 0 0.00 2.0000				
32 43.18 2.6328 0 0.00 2.0000				
0 0.00 2.0000				2.8953
Average #User *	(0.00	2.0000
		Avera	ige #User	*
		2 ¹ 2 ² 2 ² 2 ² 2 ² 2 ² 2 ² 2 ²	Brand The Brand And The Brand The Br	20.1% 25.00.1% 25.00.2%









PART 1: EXISTING USE CASES

PART 2: NEW USE CASES





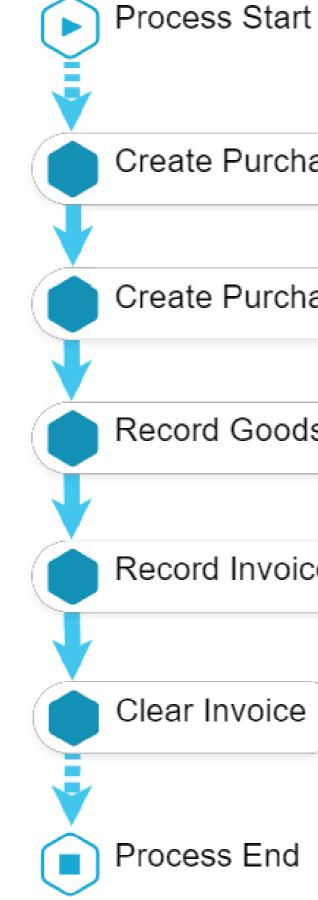


O PHASE I: DATA ANALYSIS

PHASE II: MACHINE LEARNING









ROBLEM where machine learning empowers process mining

Process Start

Create Purchase Requisition Item

Create Purchase Order Item

Record Goods Receipt

Record Invoice Receipt

Clear Invoice







Classical Statistics



Team Celonis TUM Data Innovation Lab 2018





ARD Regression

Bayesian Trade-off

LSTM

Advanced Machine Learning





$(1 - \sum_{i=1}^{p} L^{i})(1 - L^{d})y_{t} = (1 + \sum_{i=1}^{q} \theta_{i}L^{i})\varepsilon_{t}$

Advantages

- Time series specific
- Removes polynomial trend
- Easy to implement
- Computationally fast





- Works with stationary time series only
- Cannot deal with trend other then polynomial



Advantages

- Automatic feature selection
- Confidence intervals
- Easy to implement
- Computationally fast •



ARD Automatic Relevance Determination regression

 $\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\varepsilon}$

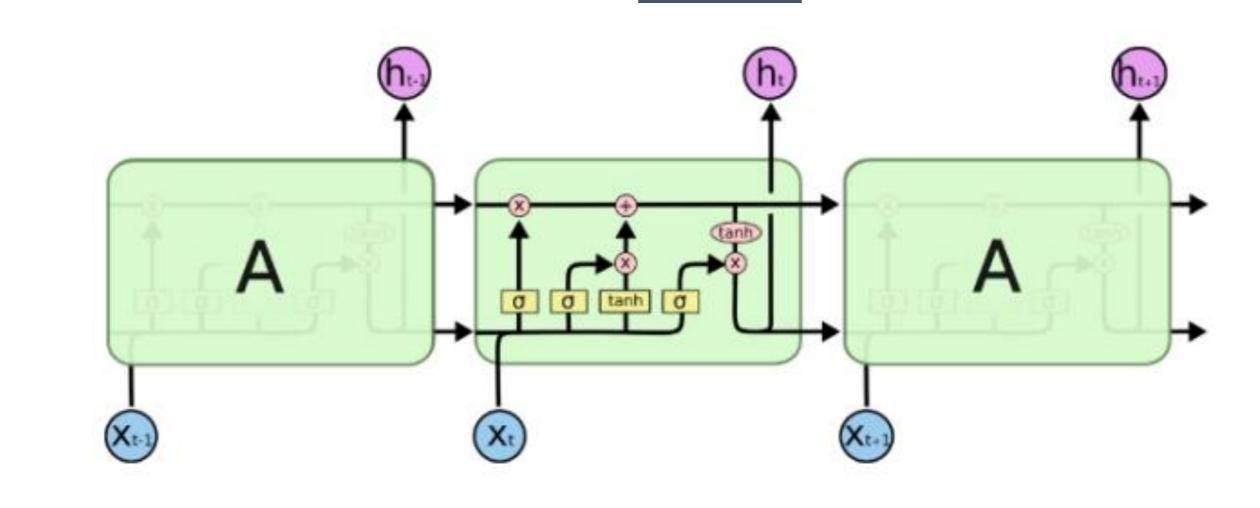
 $p(\mathbf{w}|\boldsymbol{\lambda}) \sim N(\mathbf{0}, \Lambda^{-1}), \Lambda = diag(\lambda_1, ..., \lambda_N)$ $p(\varepsilon_i | \alpha) \sim N(0, \alpha^{-1}), \{\varepsilon_i\}_{i \in \overline{1,N}} - i.i.d.$



- Not times series specific
- Not robust to -distributional changes







Advantages

- Sequential data specific
- Captures long and shortterm dependencies
- Can handle complex data structures





LSTM

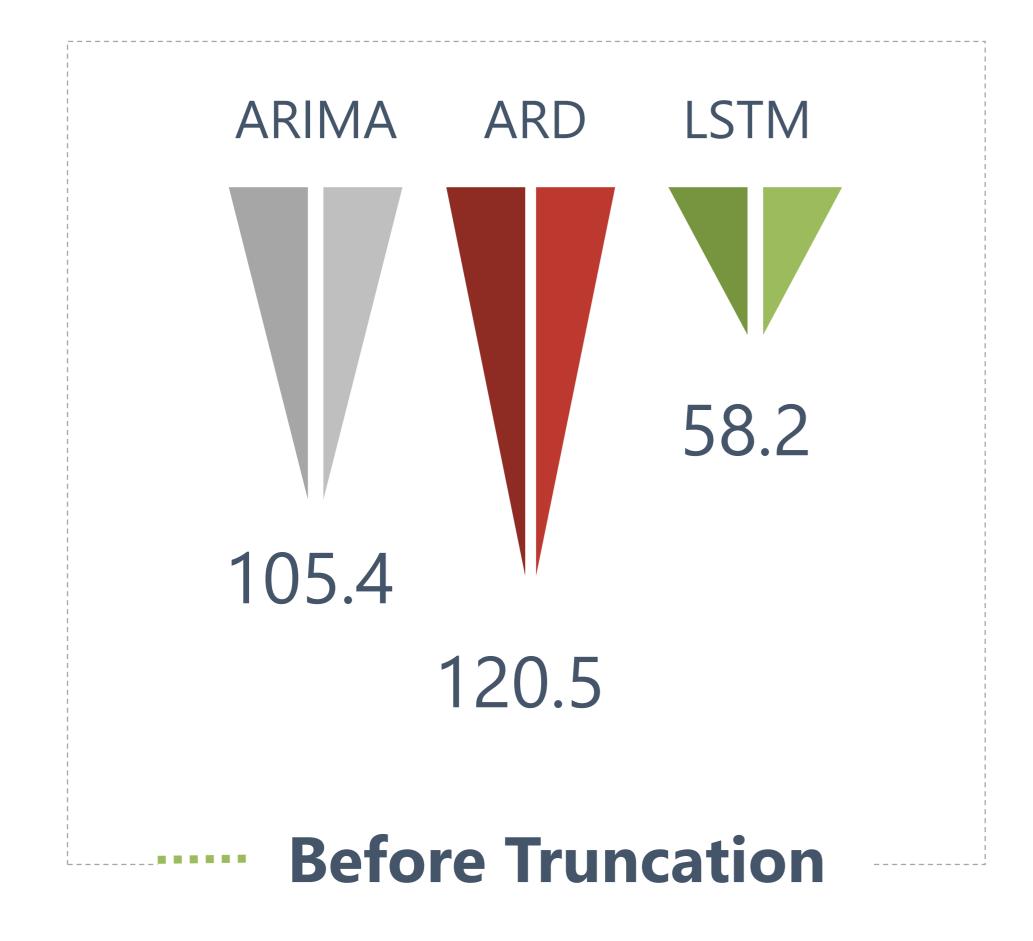
Long Short-Term Memory recurrent neural network



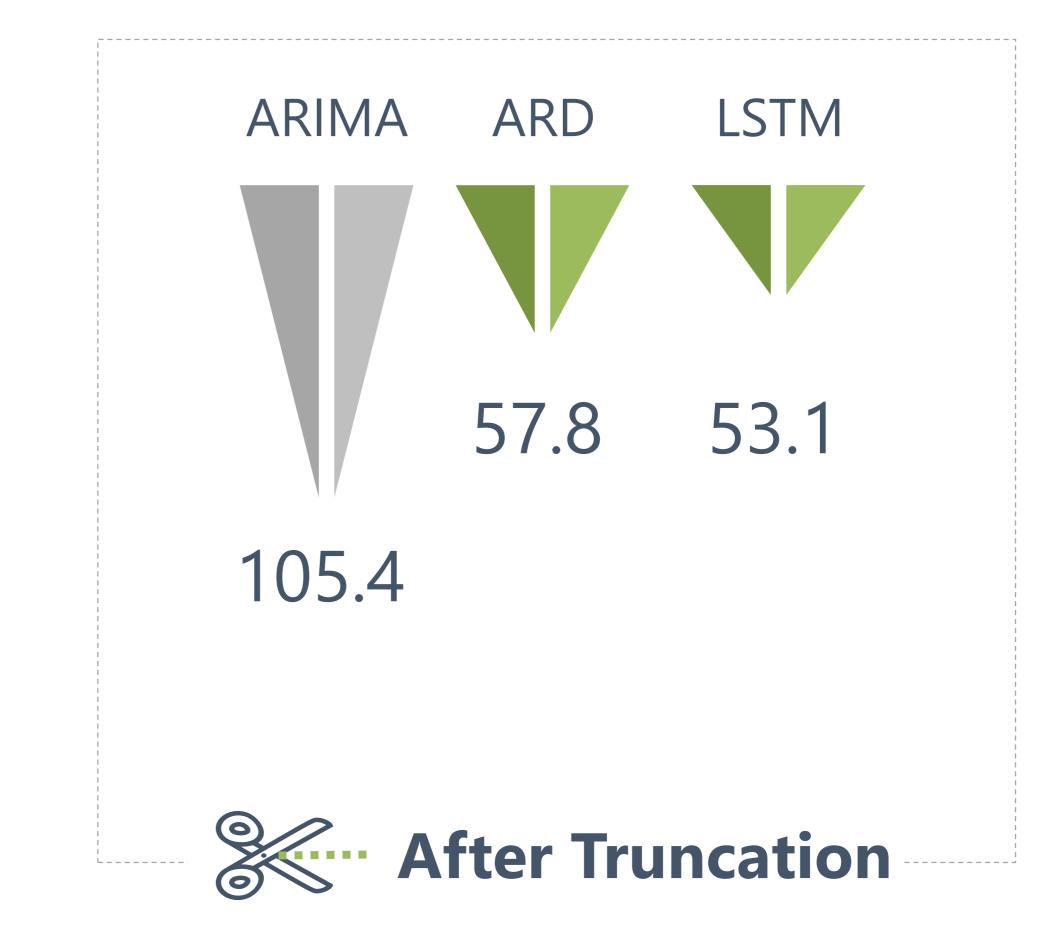
- Gives point estimates
- Requires large datasets
- Computationally expensive















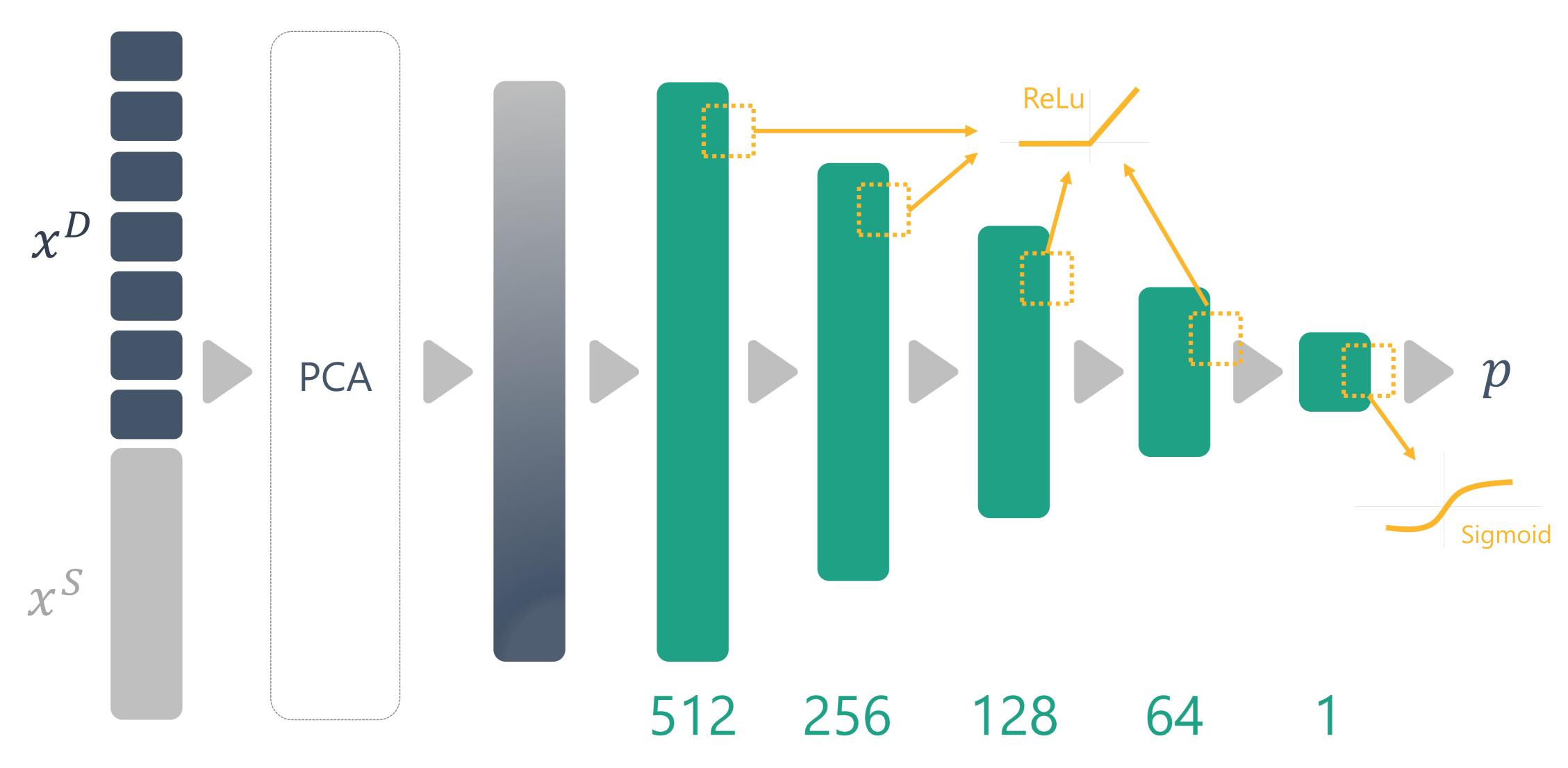


MACHINE LEARNING











	Naïve Bayes	Logistic Regression	KNN	Decision Trees	Random Forests	Neural Networks
1	38	14	59	60	63	85
2	40	37	64	70	74	88
3	42	41	69	75	79	89
4	46	50	75	81	85	85
5	46	63	83	85	89	92
6	47	60	77	79	86	93
7	55	68	81	79	84	90
8	52	72	80	80	86	89
9	50	66	78	79	81	84
10	59	74	78	81	85	83
Overall	43	45	65	74	78	88



	Naïve Bayes	Logistic Regression	KNN	Decision Trees	Random Forests	Neural Networks
1	38	14	59	60	63	85 ★
2	40	37	64	70	74	88 ★
3	42	41	69	75	79	89 ★
4	46	50	75	81	85 ★	85 ★
5	46	63	83	85	89	92 ★
6	47	60	77	79	86	93 ★
7	55	68	81	79	84	90 ★
8	52	72	80	80	86	89 ★
9	50	66	78	79	81	84 ★
10	59	74	78	81	85 ★	83
Overall	43	45	65	74	78	88 ★







MARKET DEVELOPMENT

Phase I Part 1: Existing Use Cases

MARKET PENETRATION

Existing Products





New Products



