



Transforming B2B Sales with Spark Powered Sales Intelligence

Songtao Guo, Wei Di





Business Solutions

We transform how companies



HIRE



MARKET



SELL



WORK



Songtao Guo

Sr. Staff & Tech Lead

Data Analytics Data Mining at

LinkedIn

soguo@linkedin.com



Wei Di

Staff Data Scientist

Data Analytics Data Mining at

LinkedIn

wdi@linkedin.com

Outline

Overview of Sales Intelligence

- ❑ B2B landscape
- ❑ Challenges
- ❑ One-stop solution overview

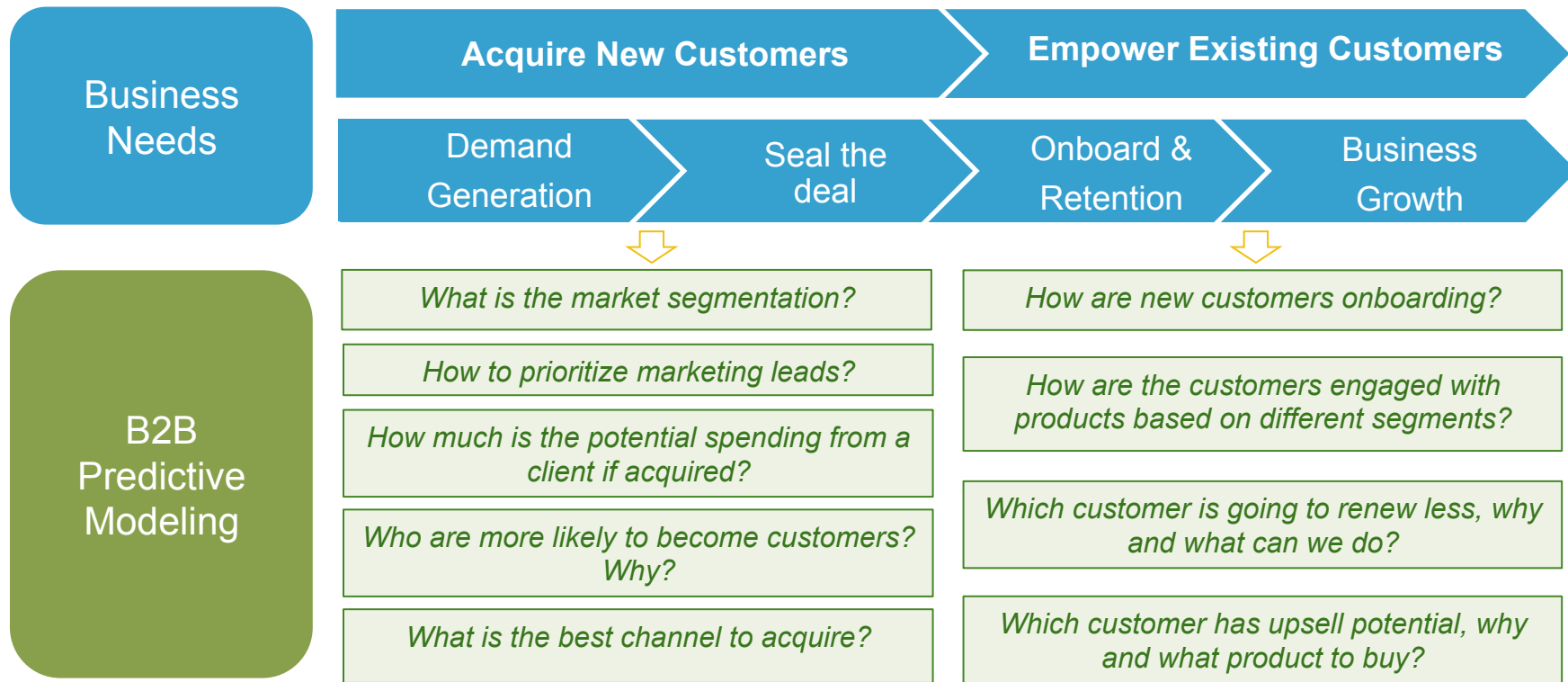
B2B Intelligent Framework Overview & Details

- ❑ Derived insightful features from LinkedIn Graph
- ❑ Centralized Data Mart
- ❑ B2B Intelligent Engine
 - ❑ Model learning
 - ❑ Model reasoning
 - ❑ Model management

Case Study

- ❑ Problem definition
- ❑ Labelling Logic & Interpretation
- ❑ Performance & Lessons learned

B2B Analytics Landscape



Challenges of B2B Data & Application

- Data quality issues
 - Incomplete, sparse, noisy and dynamic over time
 - Missing historical data
- Lack of centralized data covering various needs
- Unclear source of truth
- From score to actionable insights
- Unify existing solutions
 - Multiple owners
 - Multiple predictive scores exist for similar purposes
 - Inconsistent quality
- Scale model building

Change Landscape with Spark

Simplicity is the ultimate sophistication.

Leonardo da Vinci

The Old Landscape

- Chaos of data
- Redundant/repeating work
- No monitoring for feature evolution and model degradation
- No scalability and slow iteration

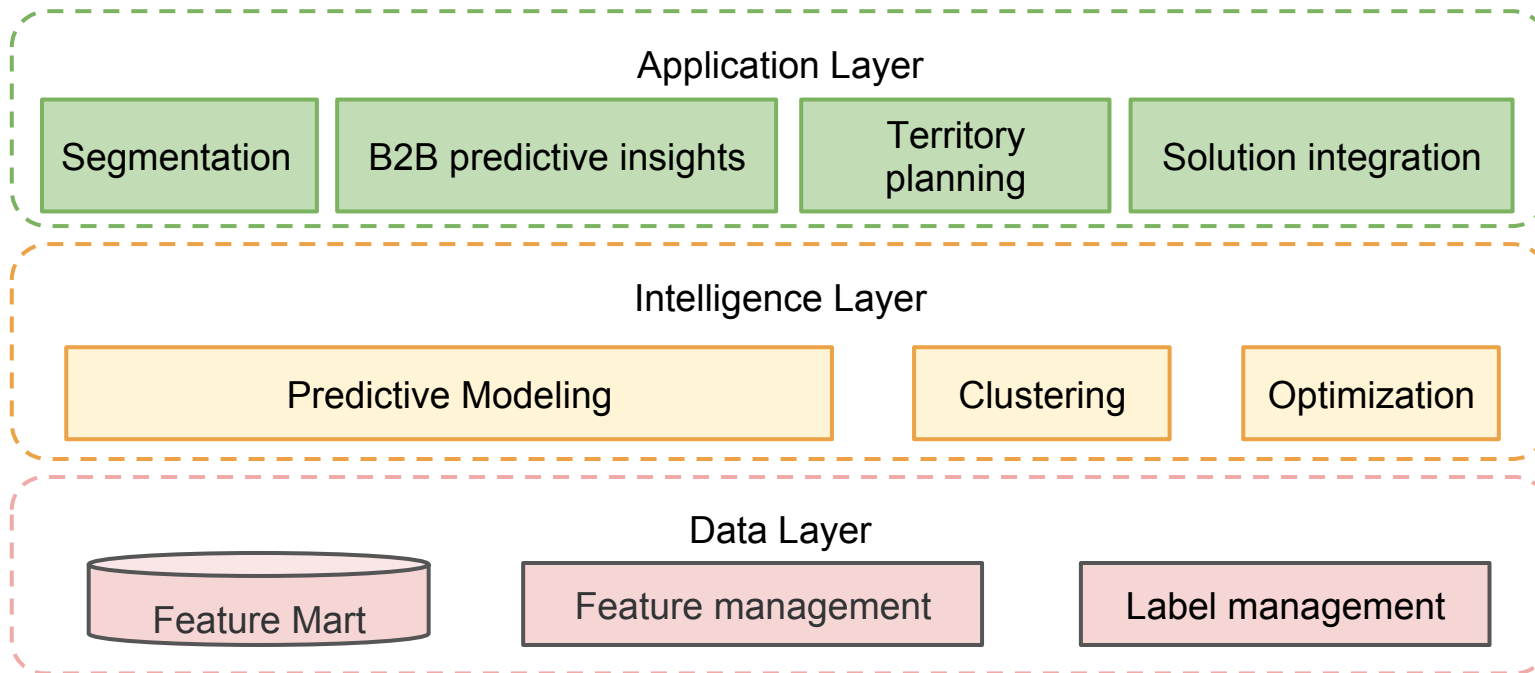


The New Paradigm

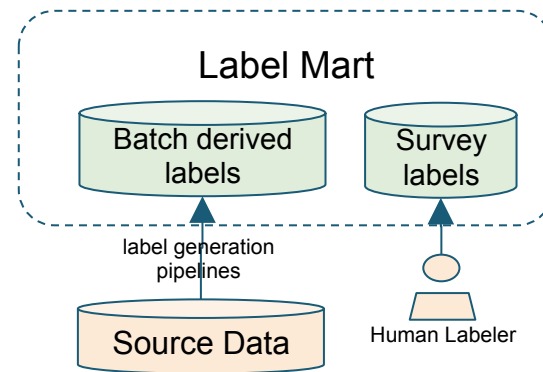
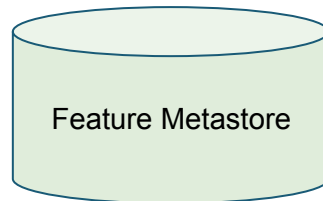
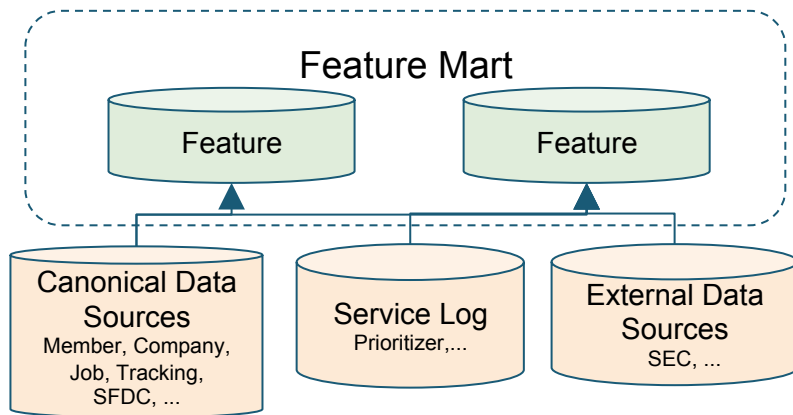
One Stop Solution

- Semi-automatic with flexibility of self-configuration
- Scalable and fast iteration
- Monitoring with closed feedback loop

B2B Intelligent Engine



Data Layer



Feature Mart

- Provide a central feature mart that serves as source of truth for key account level metrics
- Scale feature generation
- Ensure high reliability and consistency (source control, peer-review)
- Cover all the company entities and sufficient long time series

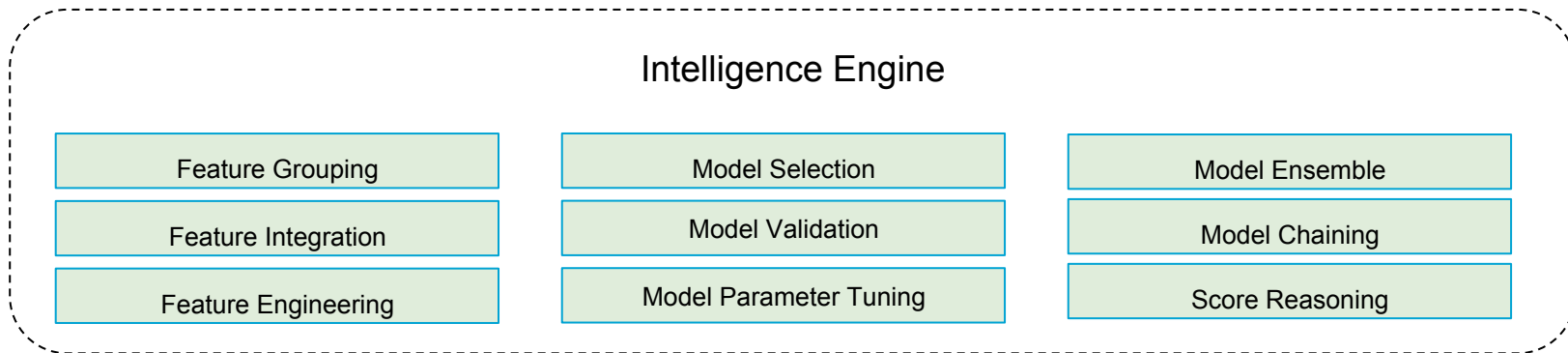
Feature Management

- A standardized feature onboarding process
- Unified metastore to support feature governance and feature application
- Feature search and feature profiling (Spark SQL)

Label Management

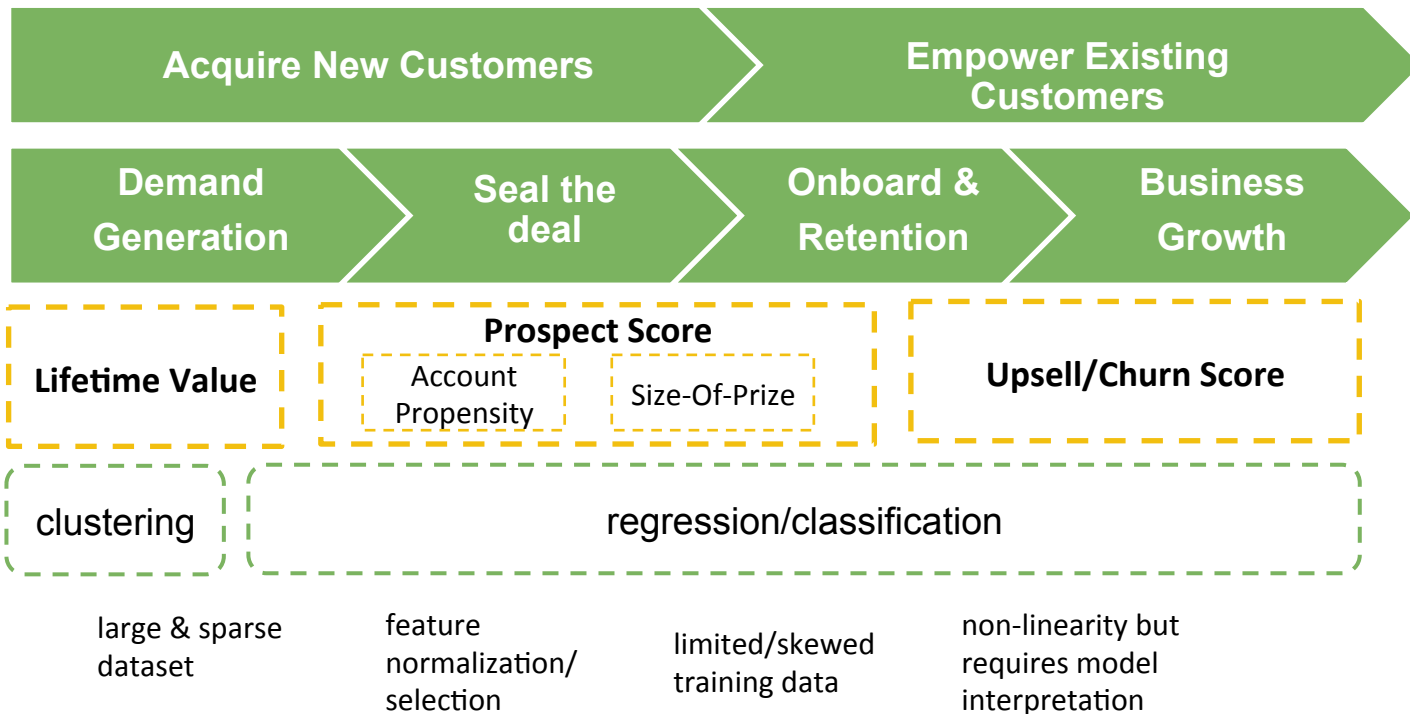
- Standardize problem definition
- Scale label preparation
- Gold dataset management
- Support continuous performance monitoring

Intelligence Layer



Workflows	Feature Engineering	Algorithms
<ul style="list-style-type: none">Highly configurable model building Azkaban workflow<ul style="list-style-type: none">Spark ML PipelinesSpark ParamGridBuilderHighly configurable scoring Azkaban workflow	<ul style="list-style-type: none">Integrate features from various of sources: company level feature mart and user experimental featuresSupport feature representation for different entity typesAdapt different feature granularities: daily, weekly, monthly, etc.	<ul style="list-style-type: none">Regression<ul style="list-style-type: none">RandomForest, GradientBoostedTree, XGBoostTree, LinearRegressionClassification<ul style="list-style-type: none">DecisionTree, RandomForest, GradientBoostedTree, XGBoostTree, LogisticRegression, MultilayerPerceptronClassification

Application Layer



Outline

Overview of Sales Intelligence

- ❑ B2B landscape
- ❑ Challenges
- ❑ One-stop solution overview

B2B Intelligent Framework Overview & Details

- ❑ **Derived insightful features from LinkedIn Graph**
- ❑ **Centralized Data Mart**
- ❑ **B2B Intelligent Engine**
 - ❑ Model learning
 - ❑ Model reasoning
 - ❑ Model management

Case Study

- ❑ Problem definition
- ❑ Labelling Logic & Interpretation
- ❑ Performance & Lessons learned

Derive Features from IN Graph



Company
Growth



Product
Booking



Product
Usage



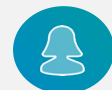
Company
Profile



Product
Performance



Product
Whitespace



Member
Profile



Social



Activity

Centralized Feature/Label Mart

Feature Coverage:

- derived company level knowledge
- generic member/company level
- internal & external data sources
- monthly/daily granularity
- region/functionality segmentation
- enriched meta information
- traceback to years early



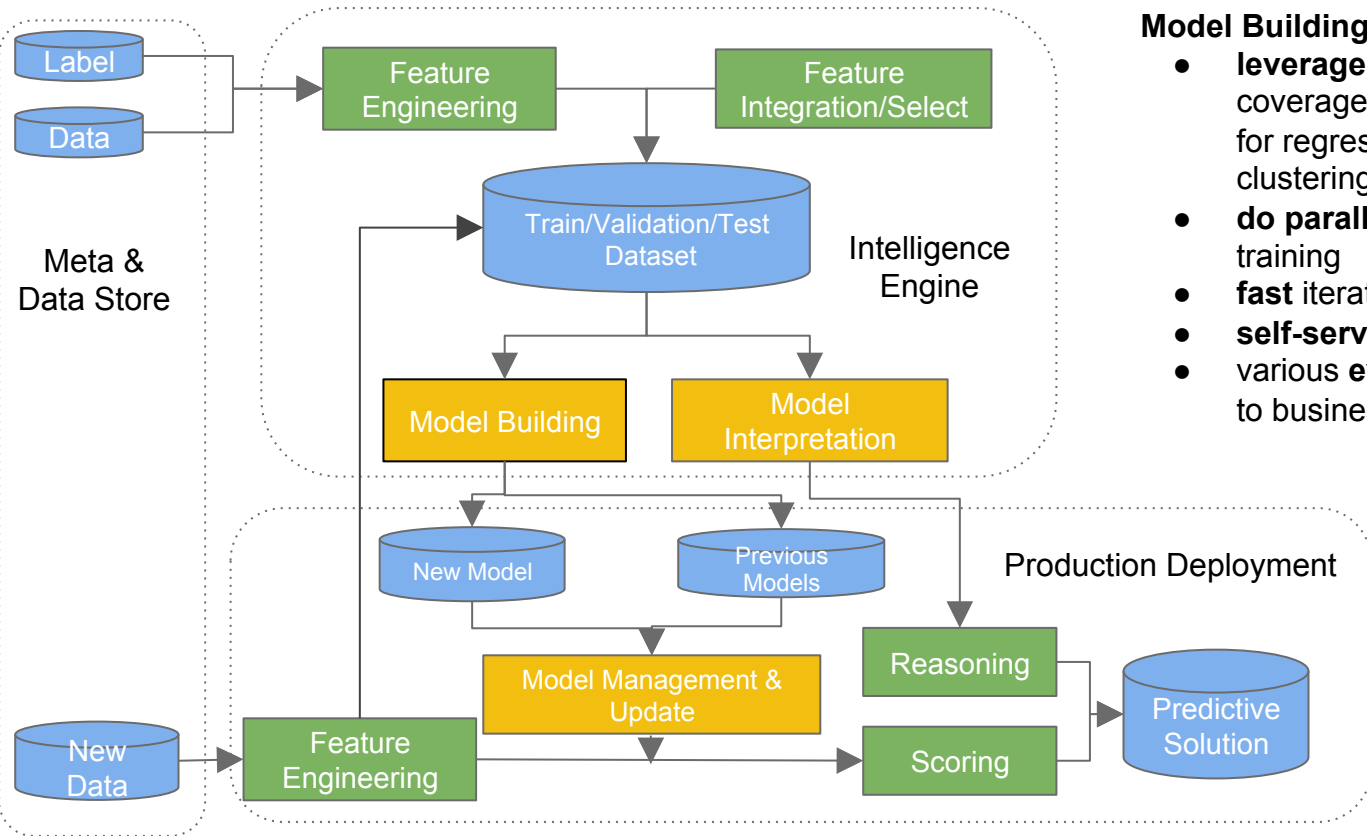
Feature Management:

- feature profiling
- visualized monitoring system
- whitelist/blacklist features as needed
- self-serve onboarding platform

Label Management

- Unified label generation logic for same type of problems across different verticals
- Vertical independent label generation pipeline

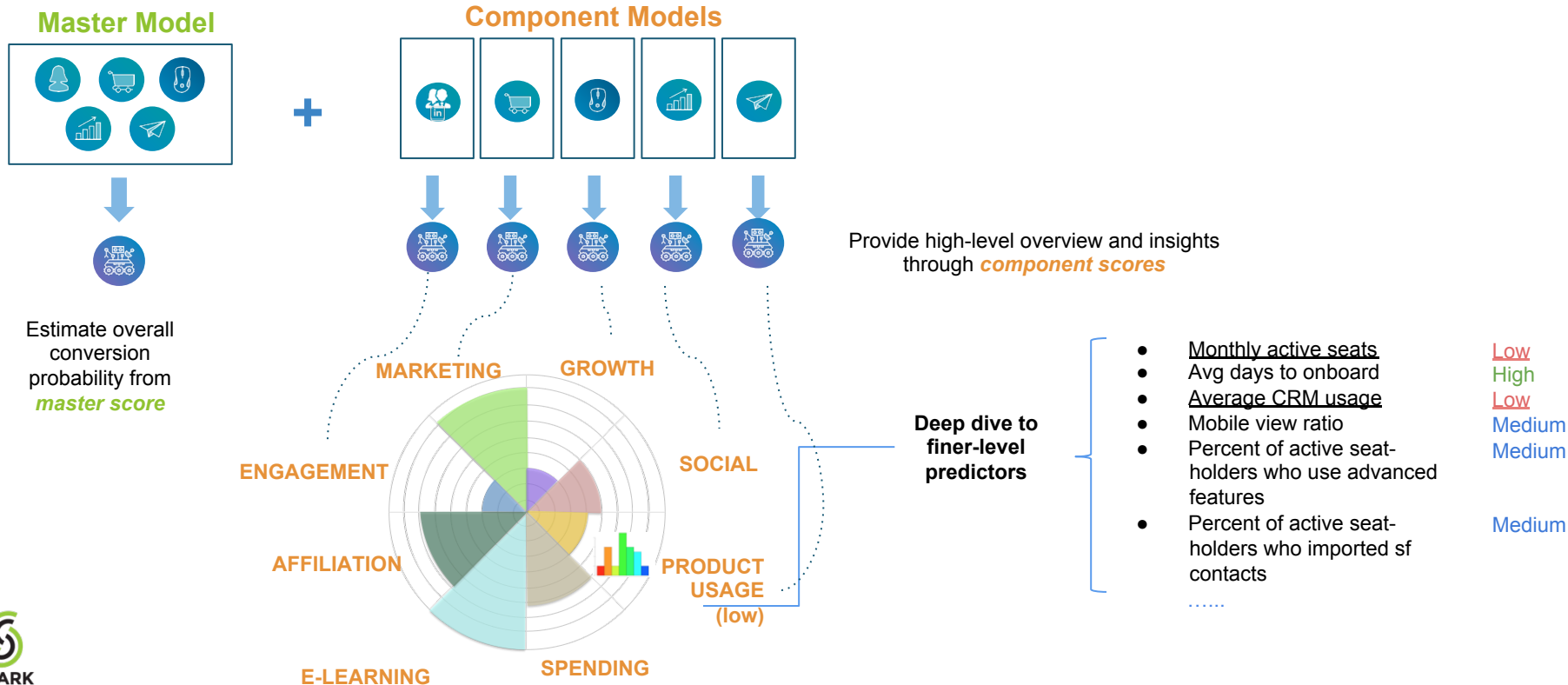
B2B Intelligence Engine



Model Building

- **leverage Spark/ML:** wide coverage of modeling approaches for regression, classification, and clustering
- **do parallel** multi-level model training
- **fast** iteration
- **self-serve** configuration/tuning
- various **evaluation** methods cater to business requirement

Model Interpretation

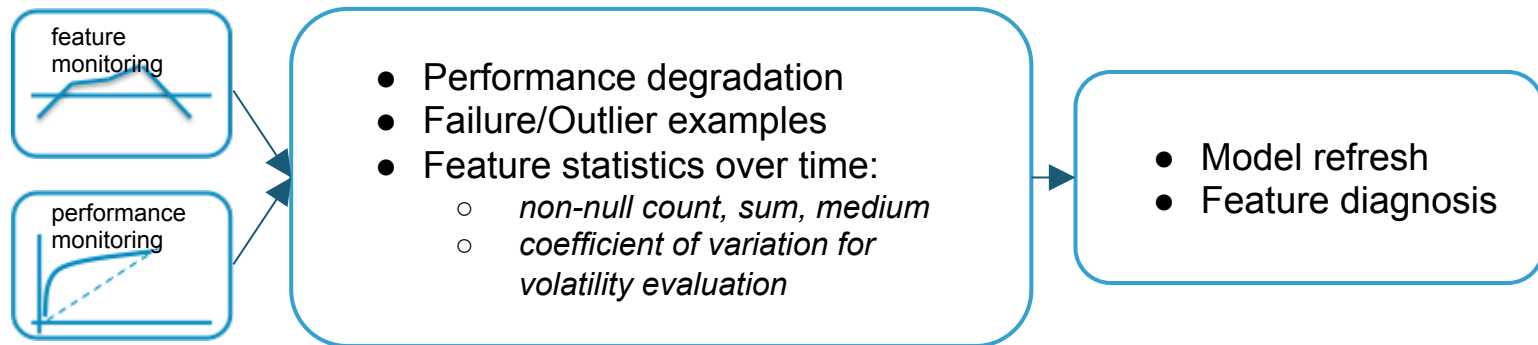


Model Management/Monitoring

- Business customers evolve dynamically
- Products update periodically

- Monitor both **feature/model** performance changes over time
- Centralized model repo with standard format
- Feed in new training data to generate “challenger models” to compete
- Pay attention to feature upstream change

inherent temporal nature



Outline

Overview of Sales Intelligence

- ❑ B2B landscape
- ❑ Challenges
- ❑ One-stop solution overview

B2B Intelligent Framework Overview & Details

- ❑ Derived insightful features from LinkedIn Graph
- ❑ Centralized Data Mart
- ❑ B2B Intelligent Engine
 - ❑ Model learning
 - ❑ Model reasoning
 - ❑ Model management

Case Study

- ❑ **Problem definition**
- ❑ **Labelling Logic & Interpretation**
- ❑ **Performance & Lessons learned**

Case Study

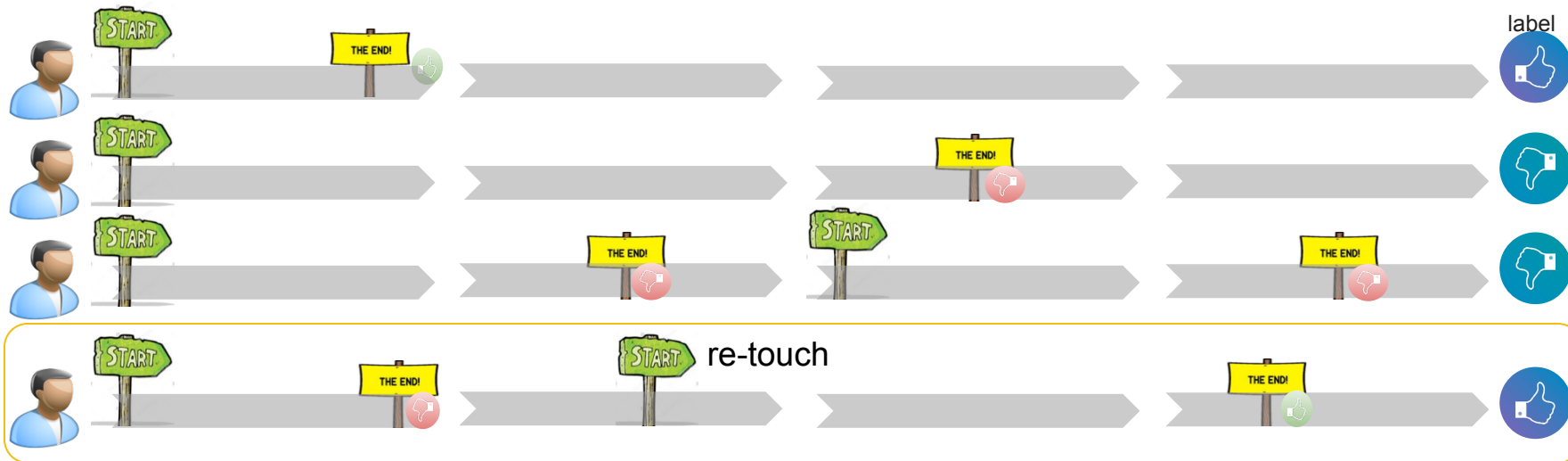
	Account Propensity Score (APS)	Upsell Propensity Score (UPS)
definition	<i>Which enterprise accounts have higher chances of buying the product & why?</i>	<i>Which existing enterprise customers have upsell potential & why ?</i>
common	<ul style="list-style-type: none"> business varies significant across region/product data evolves dynamically, time series events binary propensity model actionable insights on dissect aspects handle 'confuse' samples correctly ($p = 0.5?$) $p(r_{master} f, \dots f) \begin{cases} w_1 * p(r_{compo.} f_1, \dots f_q) \\ w_2 * p(r_{compo.} f_{q+1}, \dots f_{q+n}) \end{cases} \dots$	
differences	<ul style="list-style-type: none"> data sparse, noisy region level, regions that are very small need to borrow information from other regions/ global score accuracy is important for the whole spectrum evaluation on all tiers 	<ul style="list-style-type: none"> data is more reliable given rich product usage features product level score accuracy is important for the top ones evaluation on recommended potentials

Label Generation: APS

Label is defined at (account + region/etc) level

- **Positive** – 1: closed won opportunity.
- **Negative** – 0: closed disengaged opportunity

- lookback from opportunity creation time
- find explicit negatives
- try until won: flip opportunities
- differentiate mid vs. year-end modeling



Opportunity evolves: time series events

Label Generation: UPS

Label is defined at (account + product) level for two scenarios: *Add-On* and *Renewal*

- **Positives – 1:** increased sell & continue contract
- **Negatives – 0:** other cases

- lookback from upsell creation time
- focus on good quality positives
- consider contract renewal/churn for add-on case

Renewal scenario



Add-On scenario



Modeling & Interpretation

Account Propensity Score

	Acct. A	Acct. B	Acct. C
Score	92	65	30
Comp. & Growth	Top Industry	Fast Growth	Decline
Affinity	High	Medium	Low
Booking	High	Medium low	Low
Social Selling	High	Medium High	Medium low

Fast growth reflects potential demands, find out how our product can help the customer further grow and increase social selling & affinity.



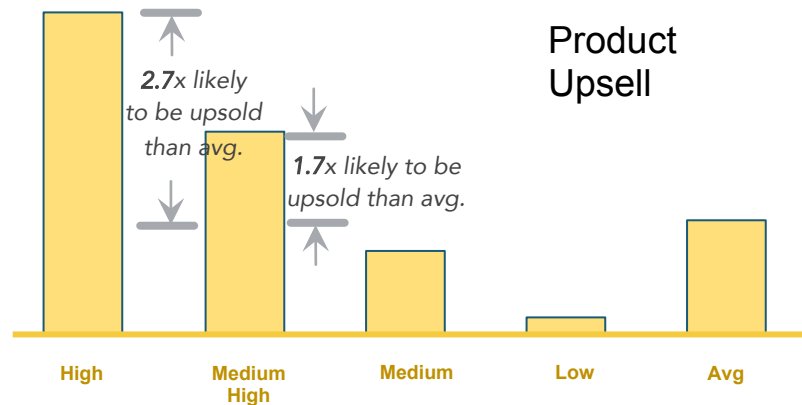
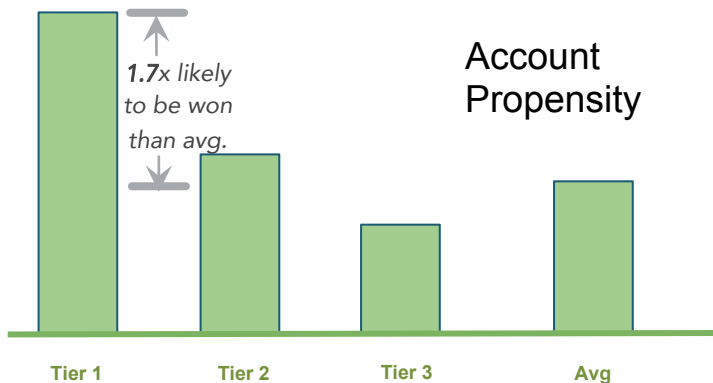
Upsell Propensity Score

	Acct. A	Acct. B	Acct. C
Score	90	75	30
Comp. & Growth	Fast member growth	Hot industry	Edu./Gov
Booking	Product A: \$1000	Product B: \$500	Product B: \$700
Usage	Highly Engaged	Engaged	Less Engaged
Perf.	Good hiring perf.		
White Space	\$2,000	\$500	\$100

Better onboarding & training to boost product usage and performance.



Business Impact



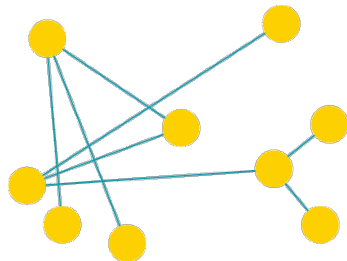
Model numerical validation

- **Account Propensity** model: AUROC of 0.64; High vs. Avg: 1.7x;
- **Product Upsell** model: AUROC of 0.84; High vs. Avg: 2.7x; Medium vs. Avg: 1.7x.

Business impact (Prioritization & Marketing)

- **Account Propensity:** 17Q1: **5.3pp** win-rate boost compared to no prioritization.
- **Product Upsell:**
 - Marketing: 2.0x open rate, and **3.3x CTR** compared to baseline.
 - SMB sales: generated **22% ~ 46%** more opportunities with customers.

Summary



TURNING THIS...



INTO THIS

End-to-end B2B solution

- Centralized large-scale data & advanced intelligence
- Speed up individual model development cycles
- Drive monetization impact and business performance



Thank You.

We are Hiring!

Songtao Guo, Wei Di
soguo,wdi@linkedin.com