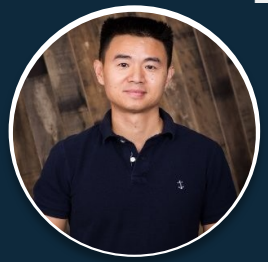
Democratizing AI

Back-fitting end-to-end machine learning at LinkedIn *at scale!*



Bo Long

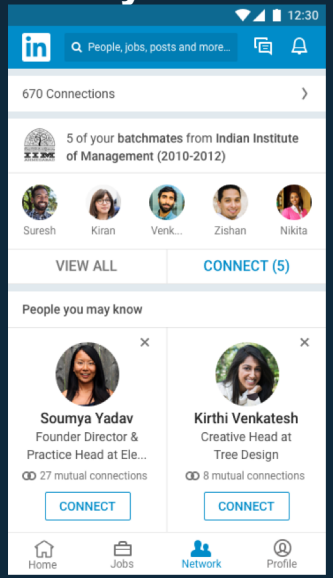
Joel Young

Today's agenda

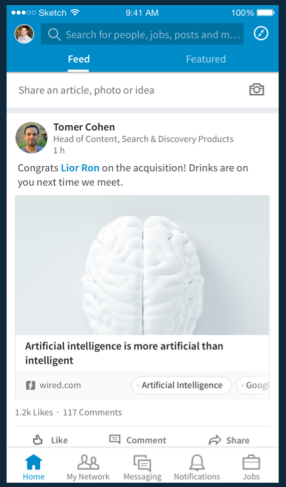
1:30	Introduction: LI Products & AI Overview
1:40	The AI Problem Space at LinkedIn
1:45	Deep Dive: Scaling Our Technology: Pro-ML
1:55	Scaling Our People: AI Academy
2:00	Conclusion/Q&A

AI for LinkedIn products

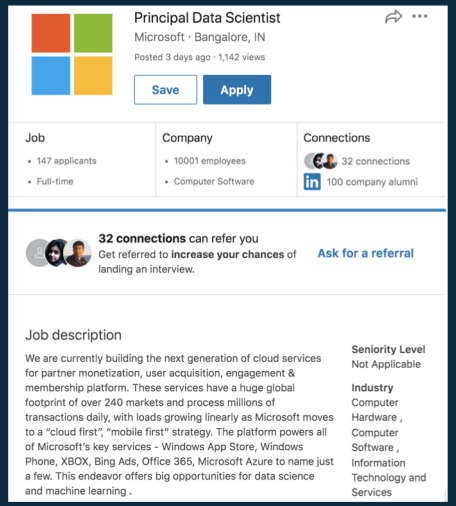
People You May Know



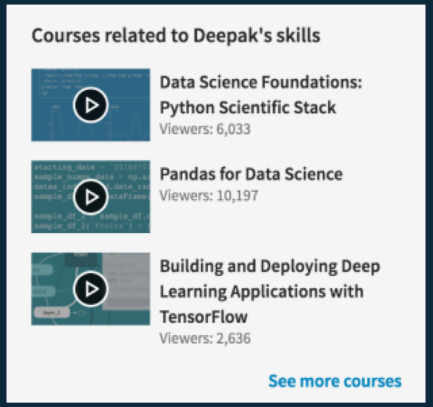
Feed



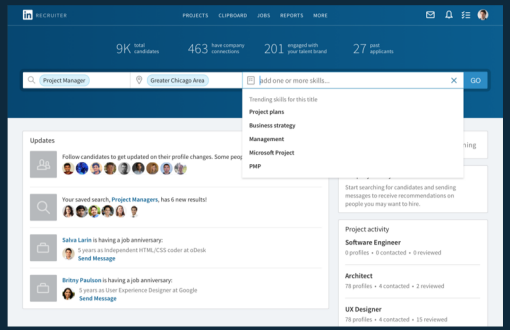
Jobs



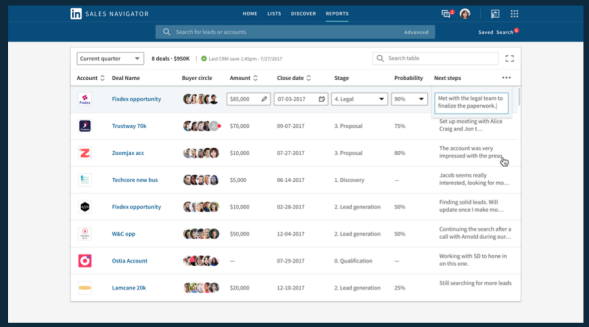
Learning



Recruiter Search



Sales



Product & Business understanding, Analytics, Data science

Machine Learning platform, automation, monitoring,...

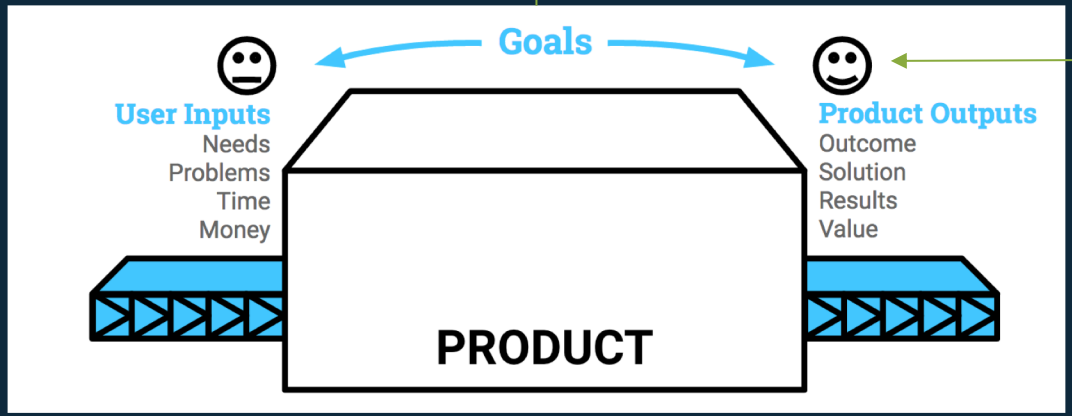
Product Metrics

Relevance Metrics

ML Optimization Framework

e.g., high-quality job recommendations

e.g., job applies, hires, interactions with Jobs



A/B Testing Framework

Scientific rigor, statistical methodology

Opportunities for AI at LinkedIn

Machine Learning Contributes Everywhere!

Opportunities for AI at LinkedIn

Many Technologies!

Opportunities for AI at LinkedIn

All Mission Critical!

Opportunities for AI at LinkedIn

Machine Learning
Is Becoming
Democratized!

Opportunities for AI at LinkedIn

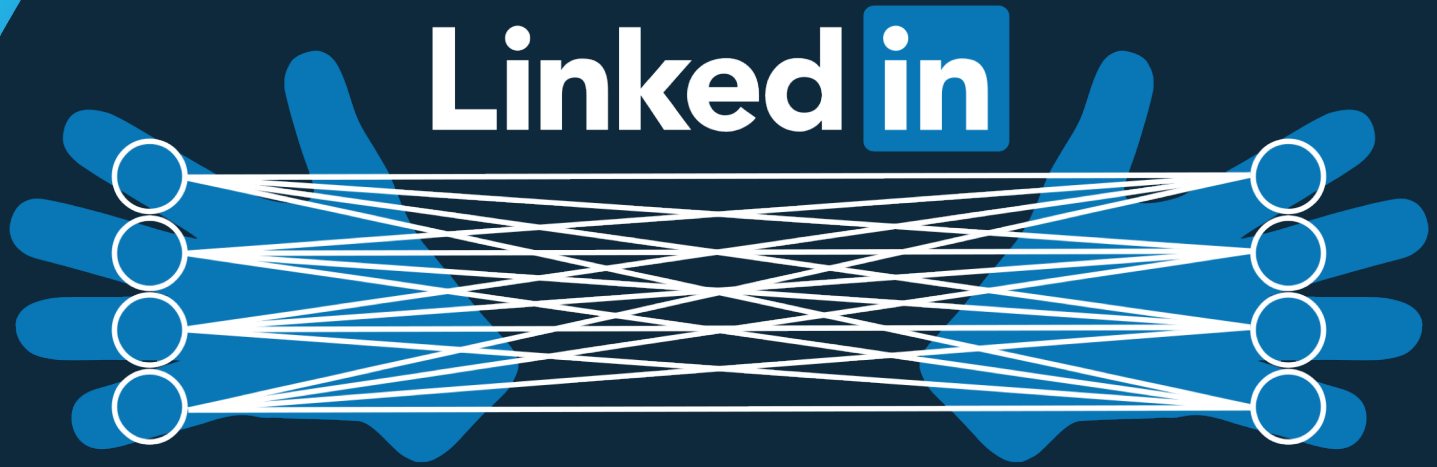
Too Much Friction!

Opportunities for AI at LinkedIn

State of the Art Advancing Quickly!

Opportunities for AI at LinkedIn

Which Stack Will Win?

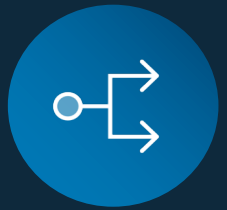


Productive Machine Learning

Pro-ML

Goal of Pro-ML

Make the end-to-end ML experience
easy, fast, robust, and automatic



Model Creation

Explore a large space of different variations of a model

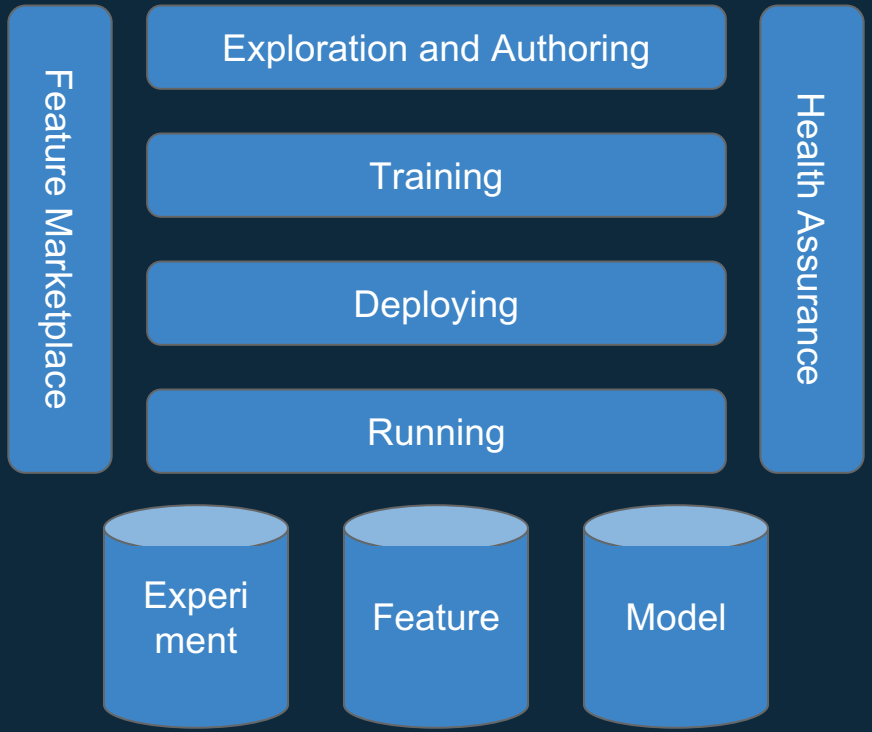
Deployment

Deploy different components of the model and features in different places in production

Maintenance

Continuously monitor model health, data quality and detect anomalies

Pro-ML Layers

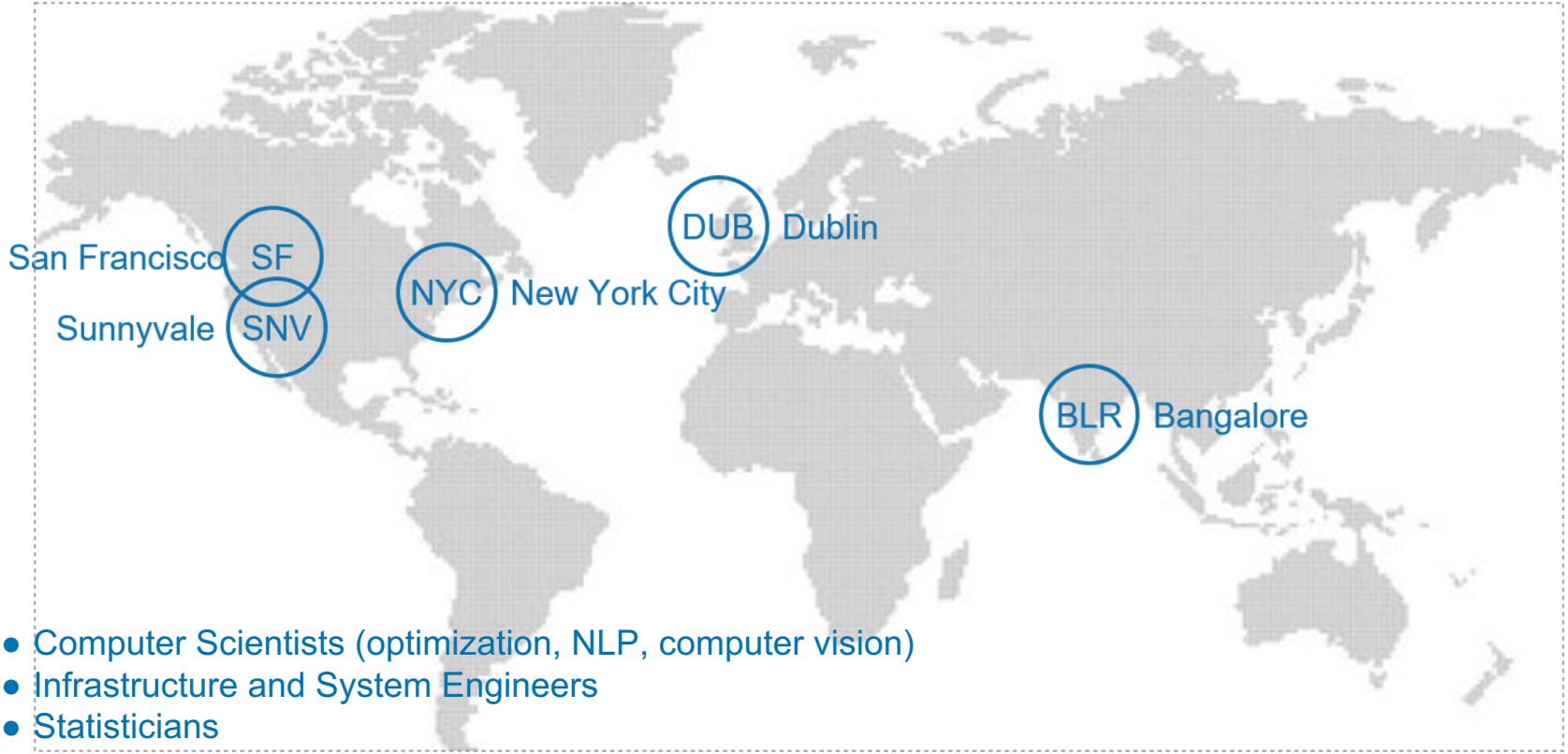


Organizationally

Team of Teams

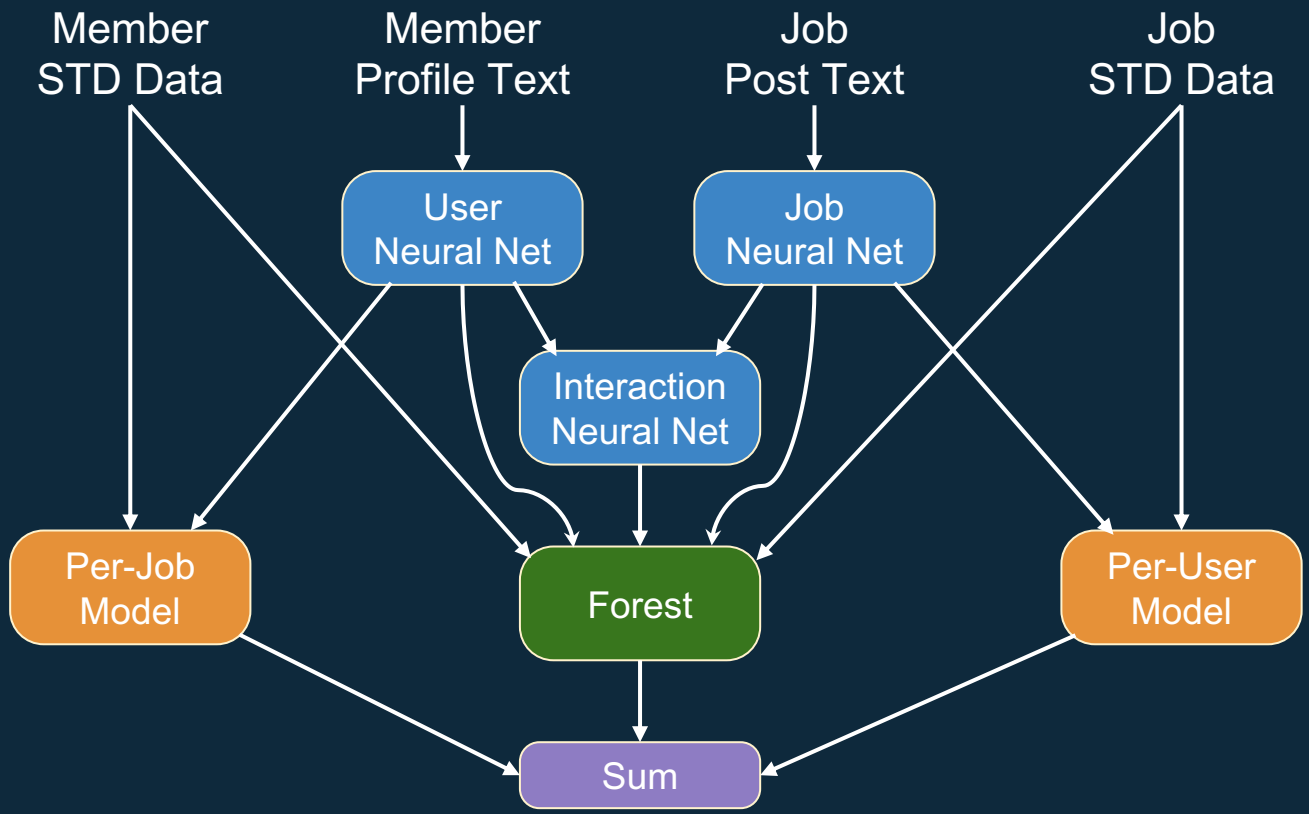
Finding a more
global optima!

Team of Teams



- Computer Scientists (optimization, NLP, computer vision)
- Infrastructure and System Engineers
- Statisticians

What is a Model?


Member
STD Data
Member
Profile Text
Job
Post Text
Job
STD Data
User
Neural Net
Job
Neural Net
Interaction
Neural Net
Per-Job
Model
Forest
Per-User
Model
Sum

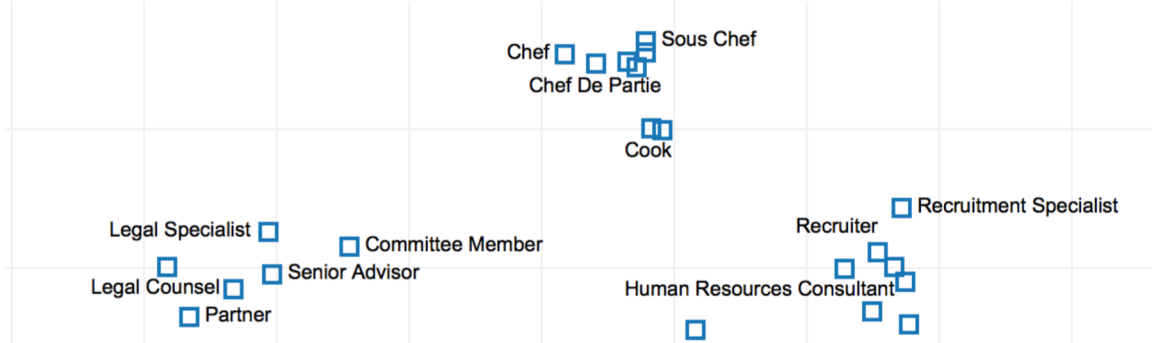
Exploring and Authoring - Jupyter

File Edit View Insert Cell Kernel Widgets Help Trusted ProML

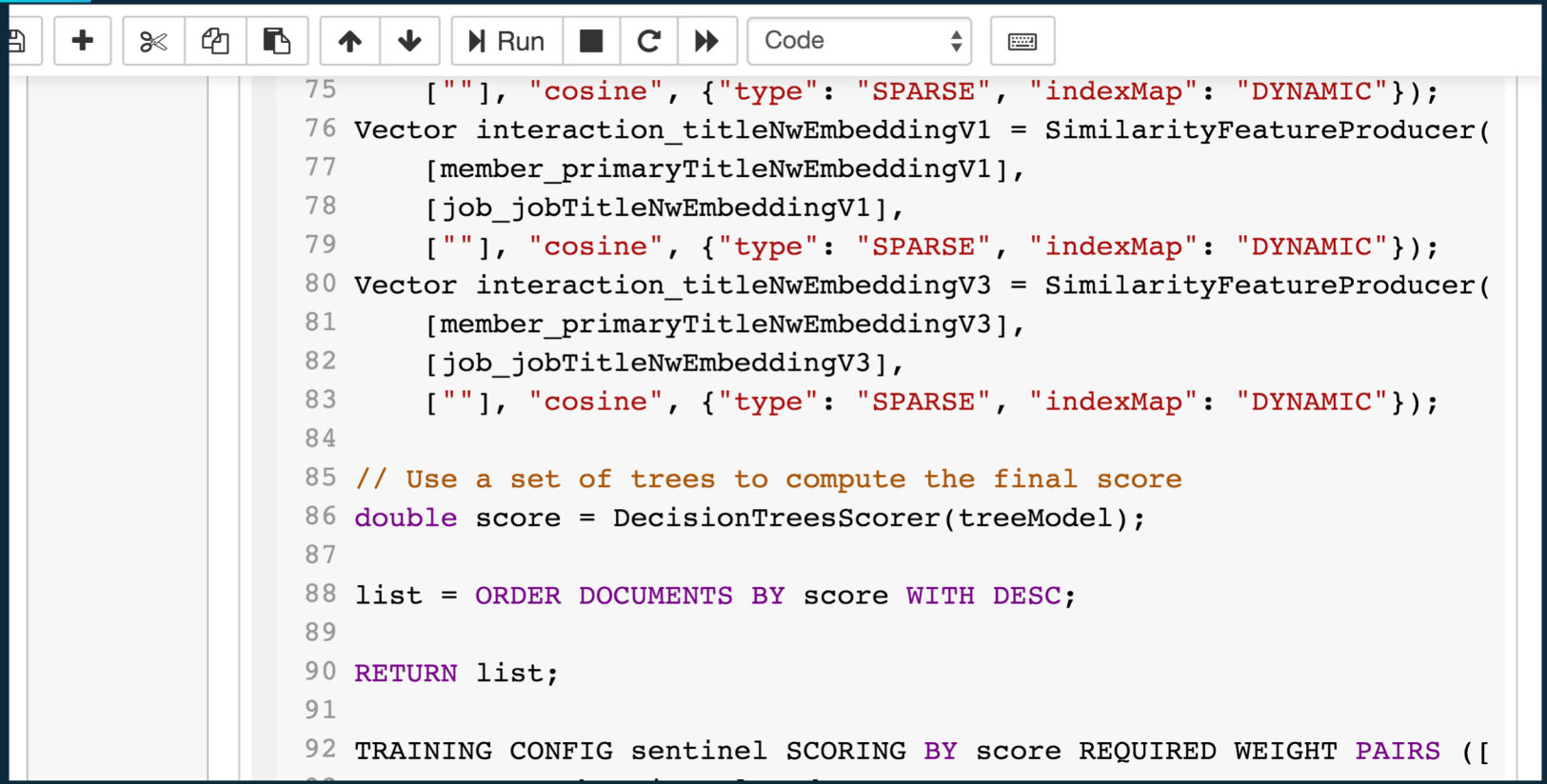
Run Code

Title Embedding Features

The features that we will look at are title embedding features. Using an advanced neural network, we learn a vector representation (called an **embedding vector**) for each title from a large amount of data, so that similar titles are close to one another in the vector space. The following figure shows a few examples. For instance, "Recruiter" and "Human Resources Consultant" are very close to each other, although they have not words in common. Text-based features would not be able to capture this relationship.



Exploring and Authoring - Quasar



Exploring and Authoring - Plotting

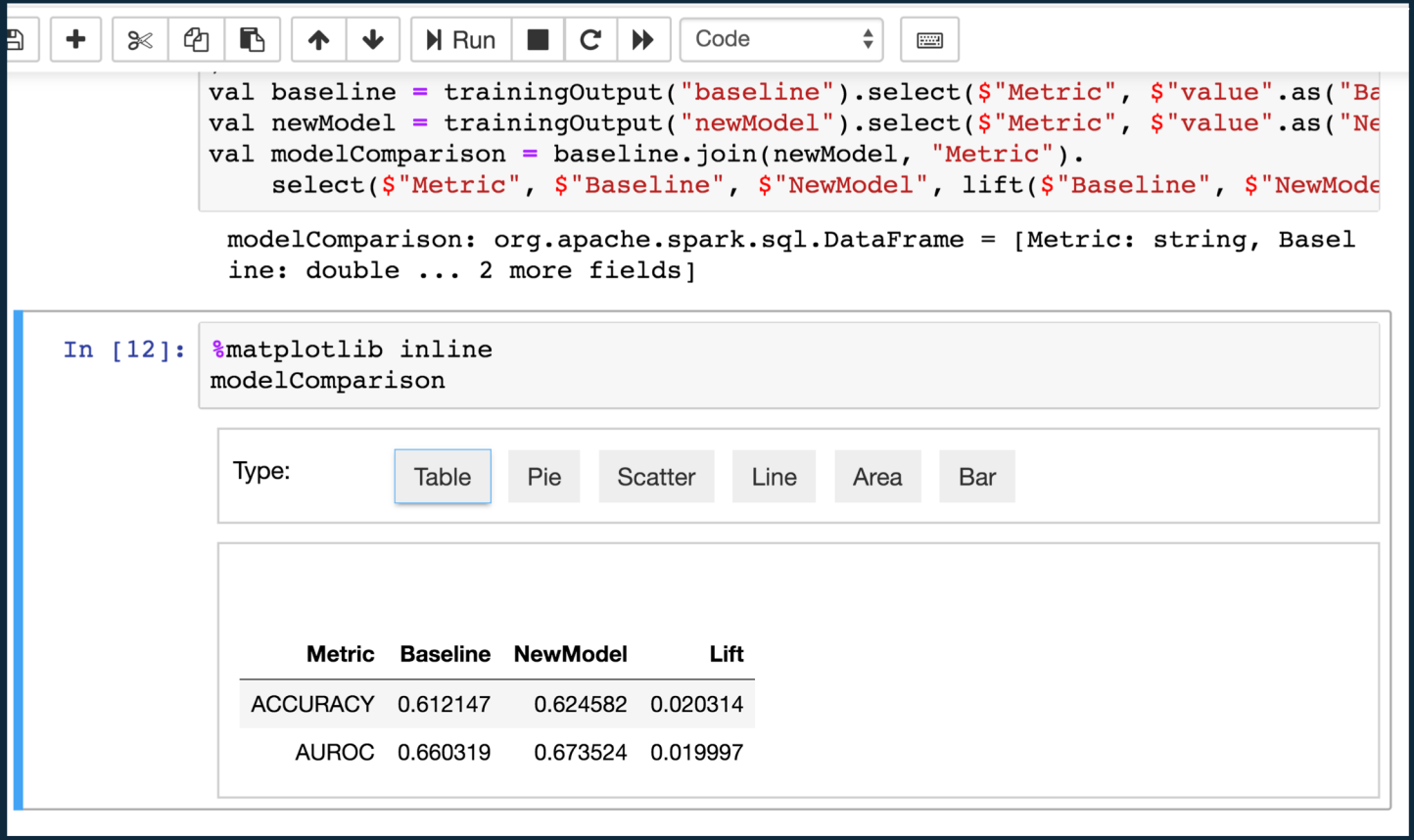
Run Code

In [*]:

```
%matplotlib inline

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

fig, ax = plt.subplots(figsize=(8,7))

for key, grp in featureStats.groupby(['term']):
    ax = grp.plot(ax=ax, kind='line', x='bucket', y='dismissRate', label=key

ax.set_xlabel('Feature value (title similarity score)')
ax.set_ylabel('Dismiss rate')
plt.legend(loc='best')
plt.show()
```

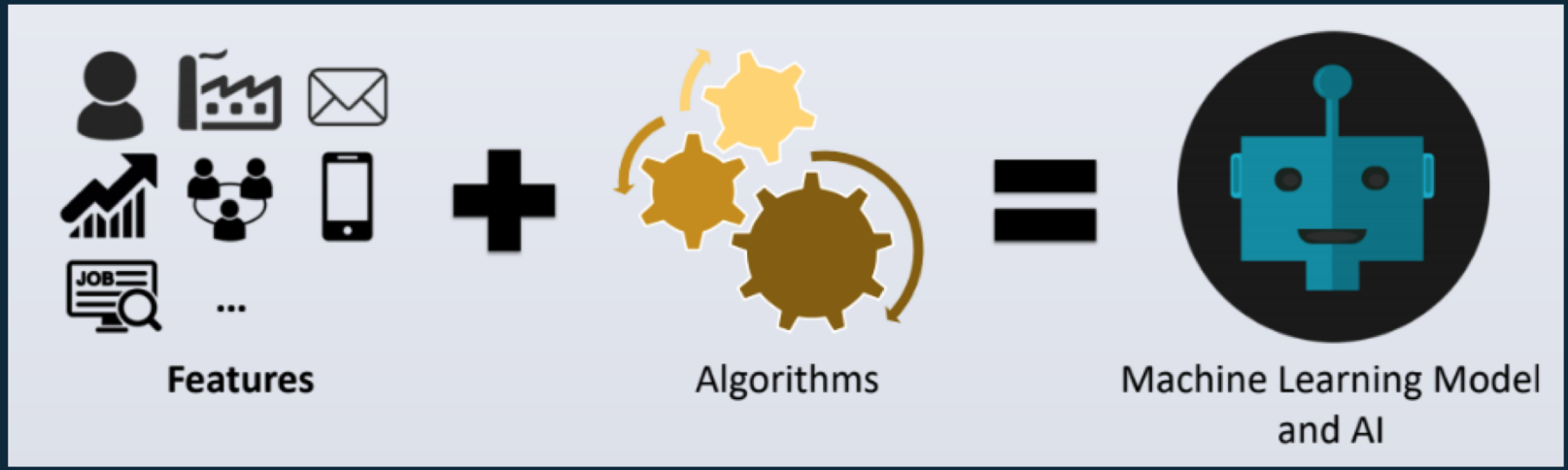
In the above figure, each curve correspond to one feature (each of which corresponds to a way of

Exploring and Authoring - Spark



How to Train the Model

Features and Algorithms



Feature Marketplace

MODOOP FEATURES

Search features...

Categories

company	7304
member	5841
company_region	248
contract	20

data_type

double	8377
long	3522
int	1405
string	100
float	5
bigint	4

feature_type

numerical	13106
binary	211
categorical	96

is_whitelisted

1	13413

aggregation_length

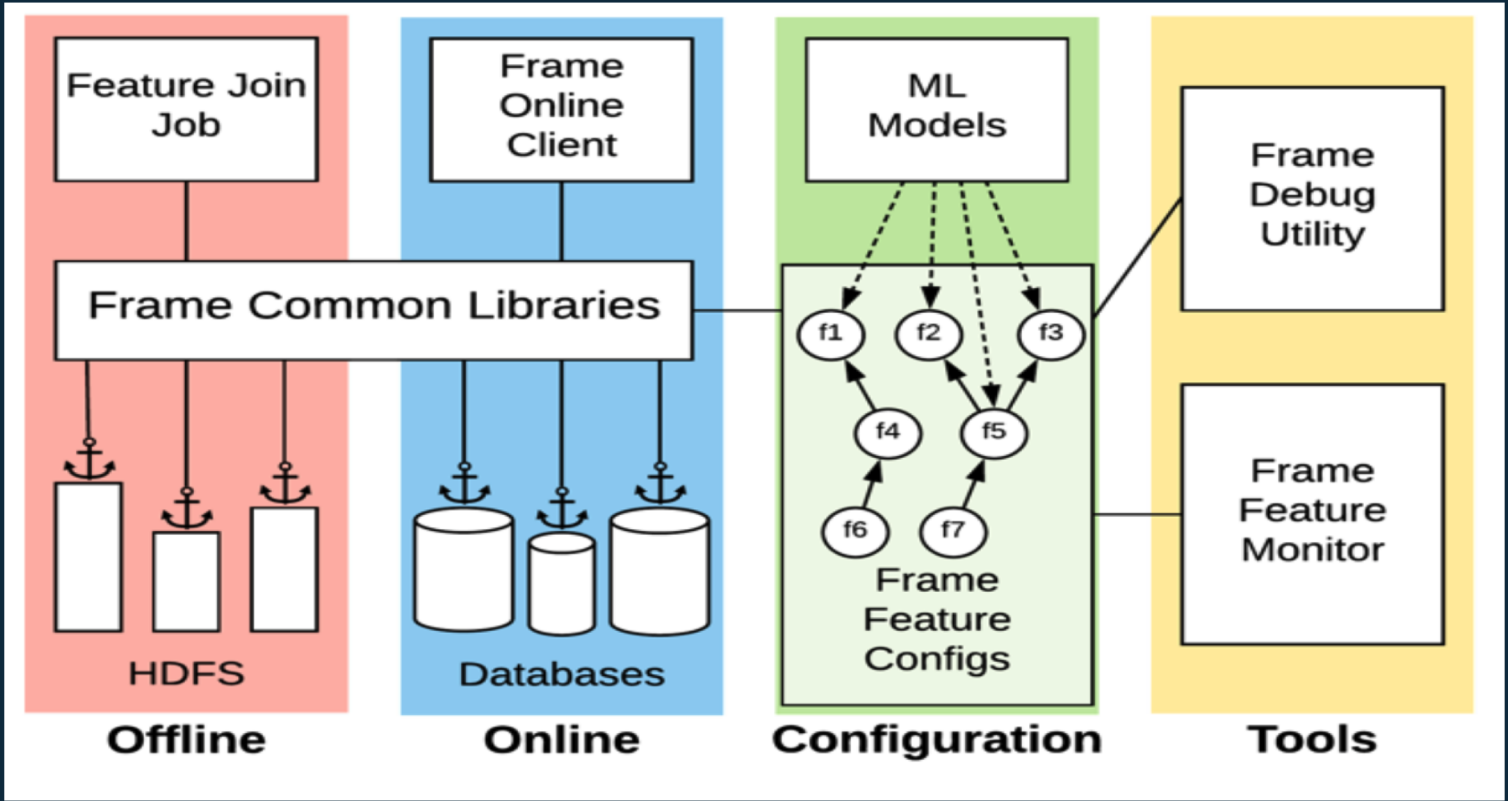
all,last 7 days	5530
All	4203
	3040
all	198
yearly	56
monthly	15

13413 results found in 25ms

Clear all filters

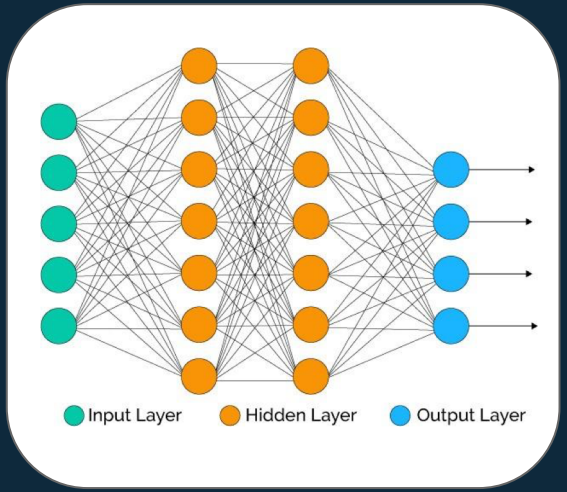
Feature ID	Feature Name	Data Type	Feature Type	Description	Aggregati Length
1	imported_contacts	long	numerical	total imported contacts	
2	imported_contacts_l07d	long	numerical	total imported contacts within last 7 days	
3	imported_contacts_l30d	long	numerical	total imported contacts within last 30 days	
4	is_uploaded_abook_l07d	long	binary	has uploaded abook within last 7 days	
5	is_uploaded_abook_l30d	long	binary	has uploaded abook within last 30 days	
6	is_uploaded_abook_l90d	long	binary	has uploaded abook within last 90 days	
7	conn_uploaded_abook	long	numerical	total abook uploads from connections	all,last 7 da
8	email_appeared	long	numerical	times of profile appeared in others abook email	all,last 7 da

Feature Access Consistency

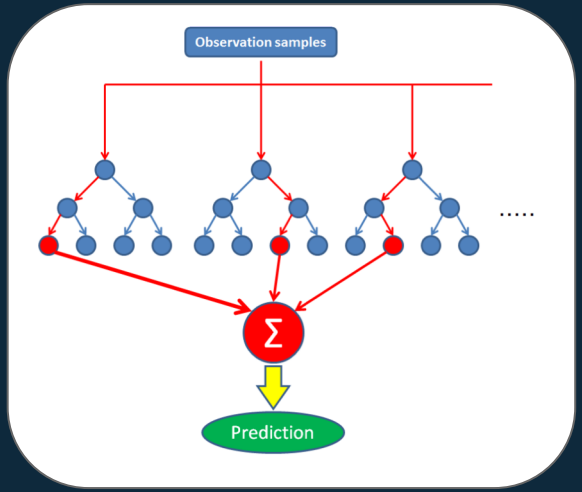


Algorithms

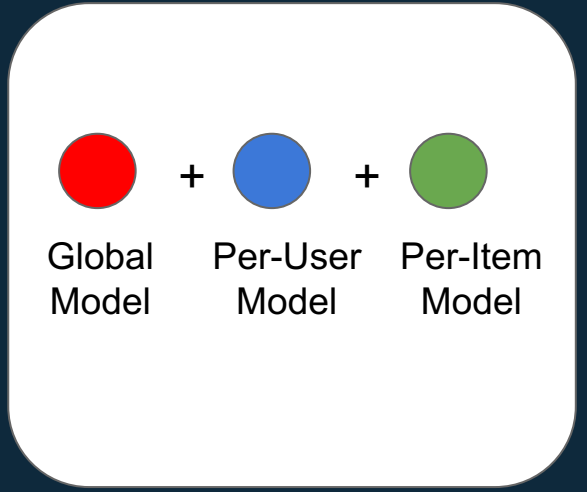
Deep Learning



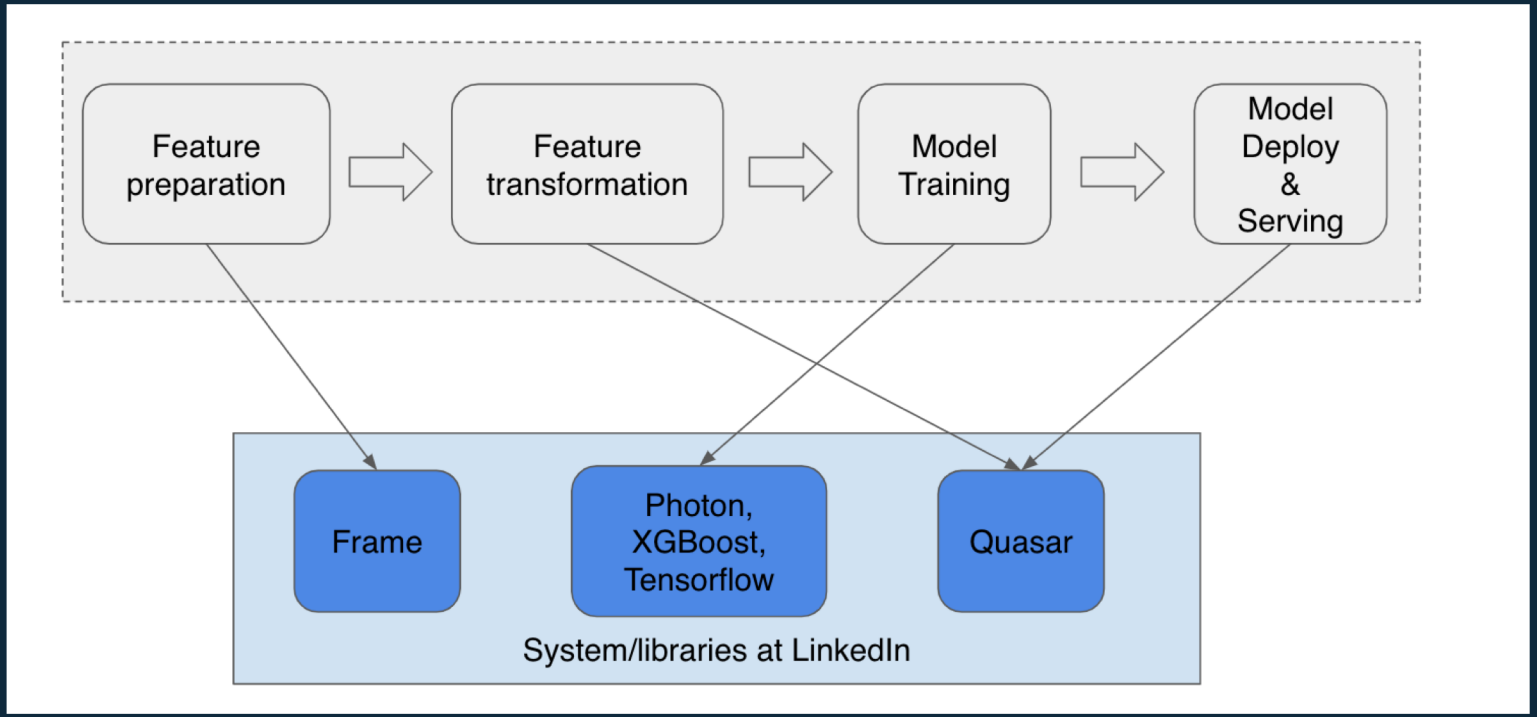
Trees

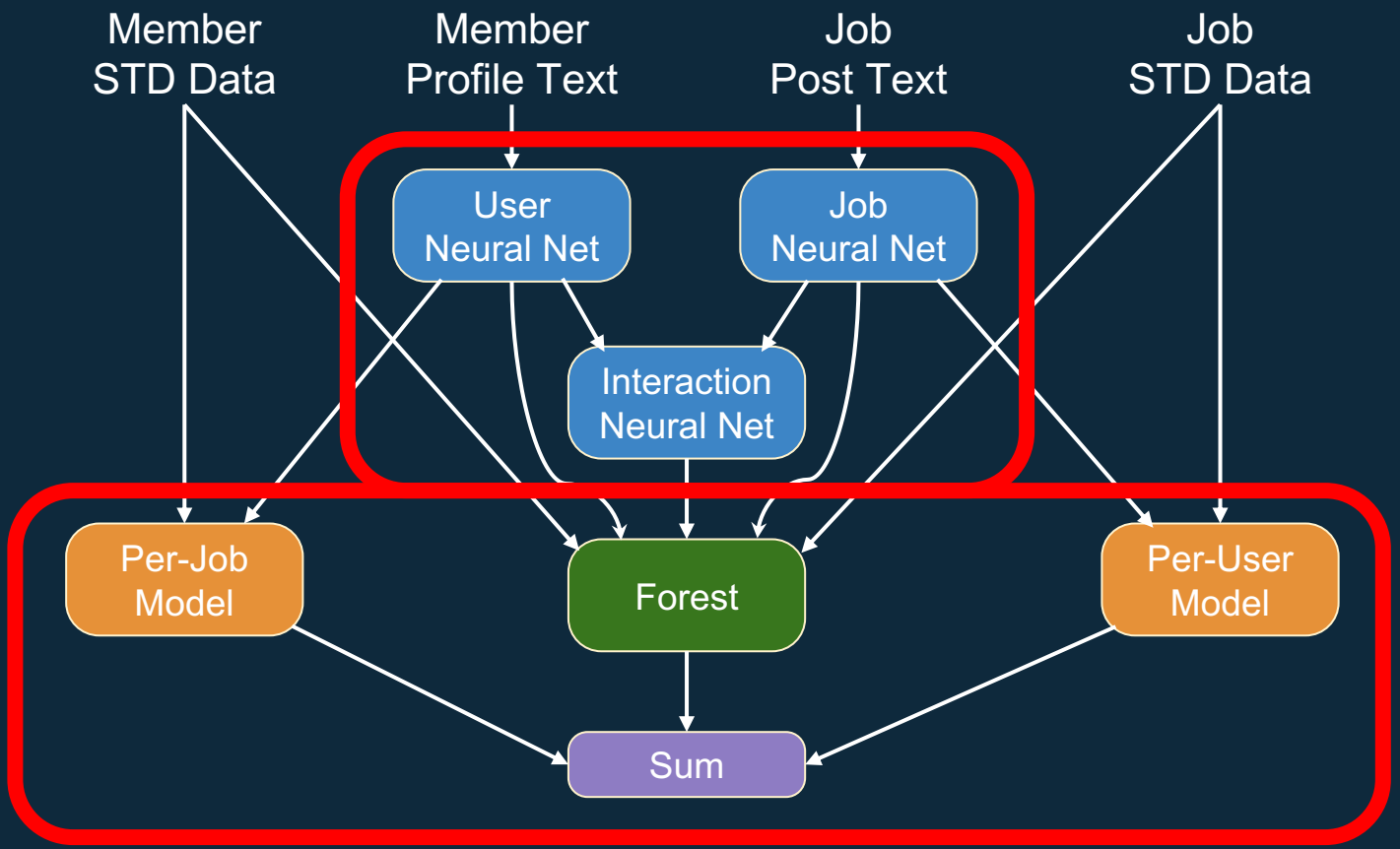


General Linear Mixed Models

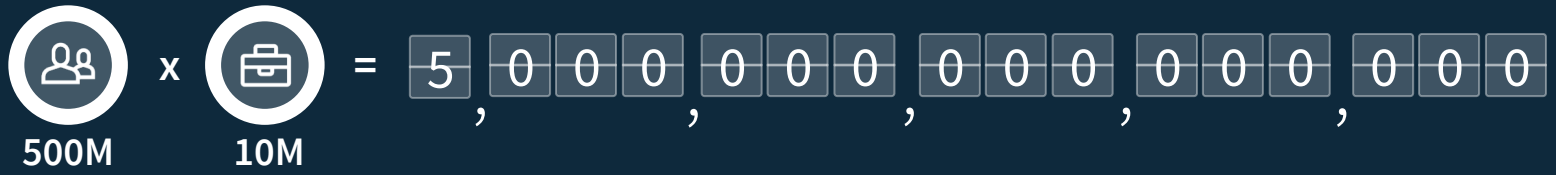


Training Engine: Photon-Connect




Member
STD Data
Member
Profile Text
Job
Post Text
Job
STD Data
User
Neural Net
Job
Neural Net
Interaction
Neural Net
Per-Job
Model
Forest
Per-User
Model
Sum

The Problem is Huge



How We Learn From Data

Global Component

Macro patterns

Member Component

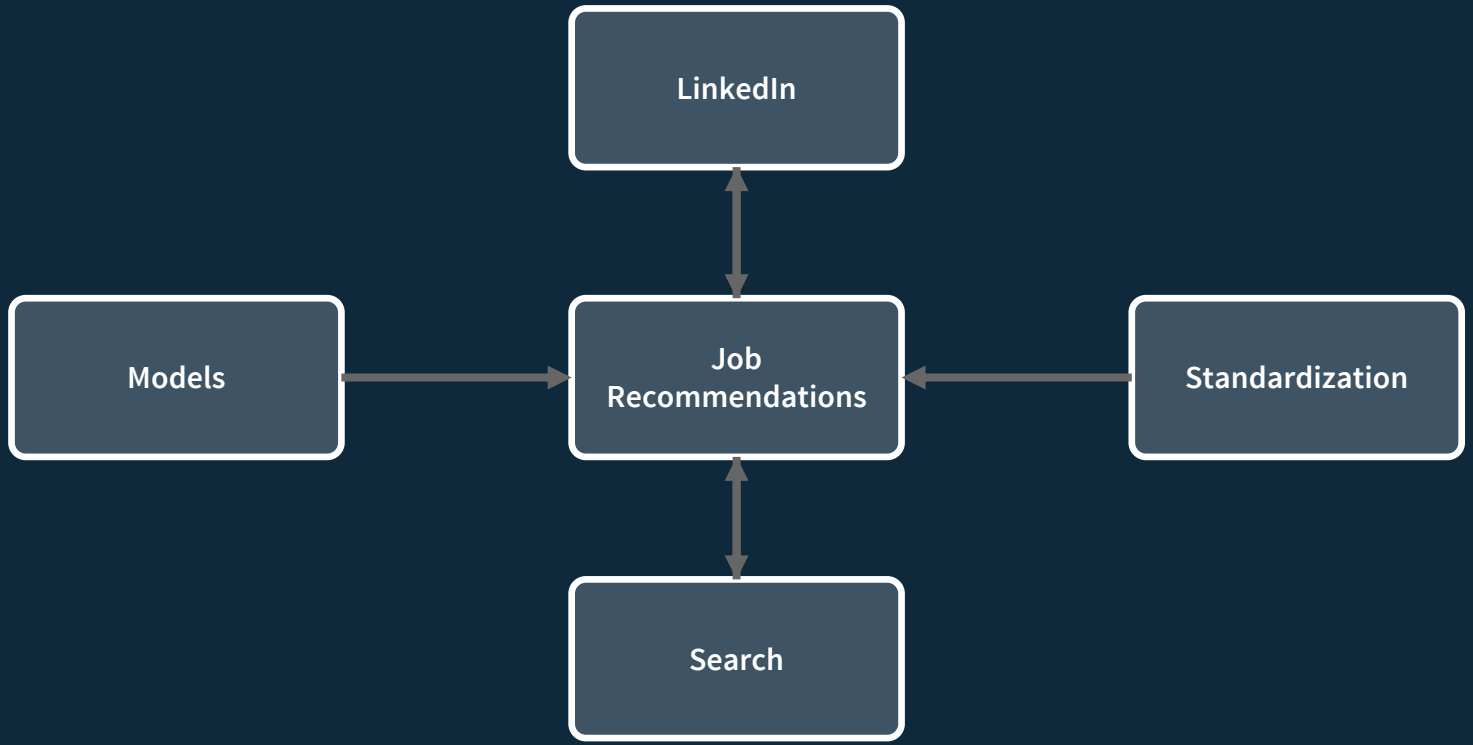
Member specific micro patterns

Job Component

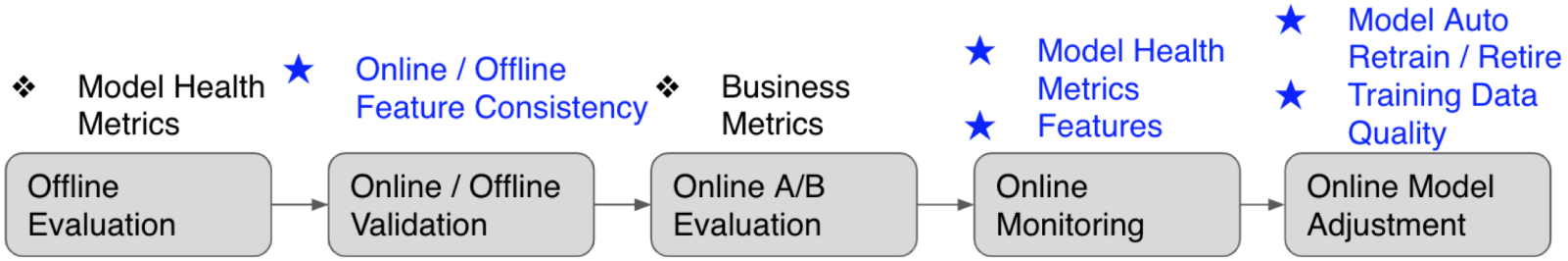
Job specific micro patterns

It’s Looking Better!

How We Make it Happen

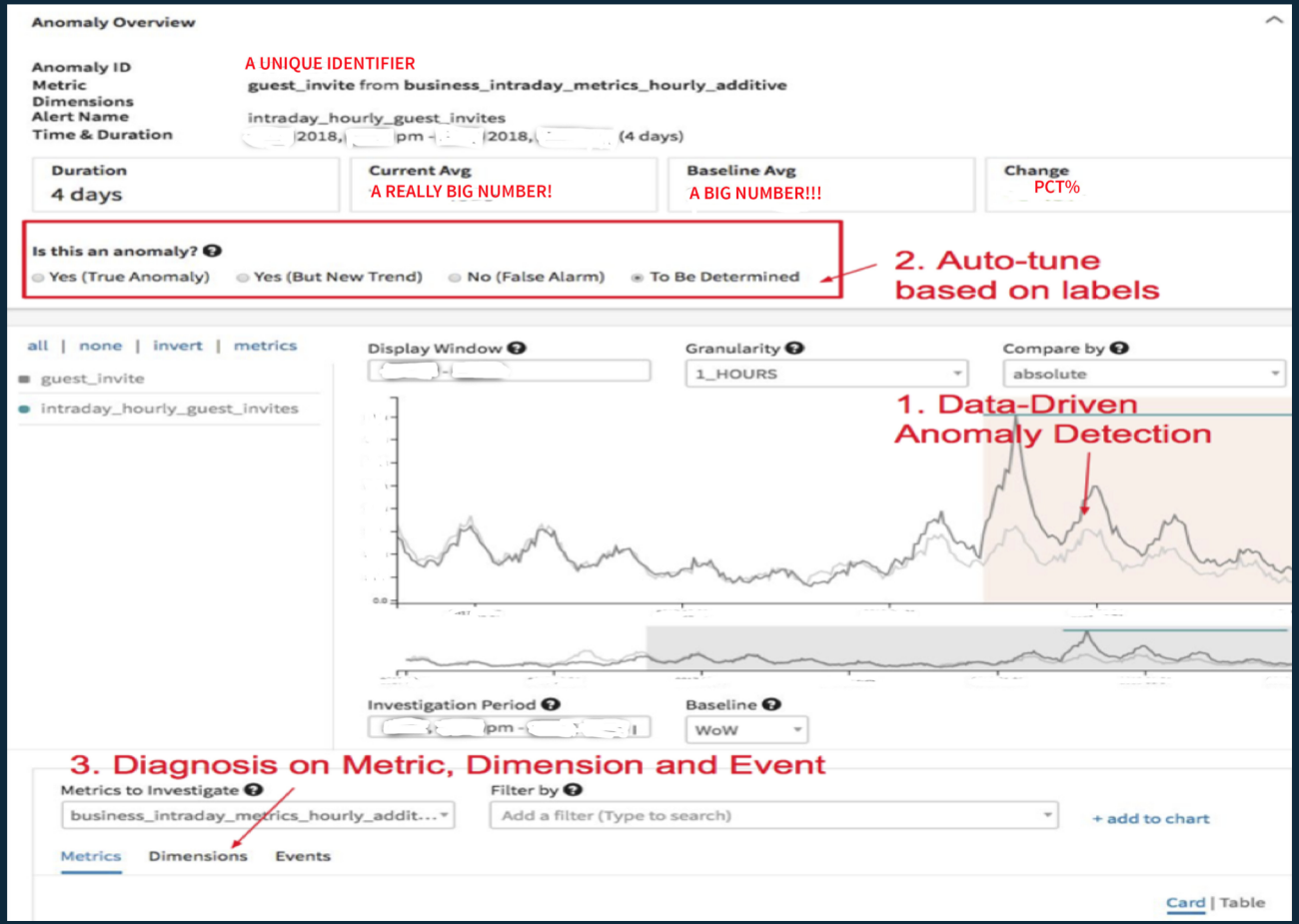


Health Assurance
Model Development Cycle



- Configuration not consistent between offline and online, e.g., model, transformation
- Data not consistent between offline and online, e.g., features
- Ranking and serving system failure
- Feature distribution change over time
- Data quality issues
- Model obsolete
- ...

Anomaly Detection



AI Academy
Accelerated Democratization

Using LinkedIn tools and infra, train all engineers to use basic supervised ML in their day-to-day job

Partial Curriculum:
- AI 100 – Managing AI-driven Products
- AI 200 – AI For Software Engineers
- AI 300 – AI Engineering in Depth

Continuing education via LinkedIn Learning

LinkedIn
AI Academy

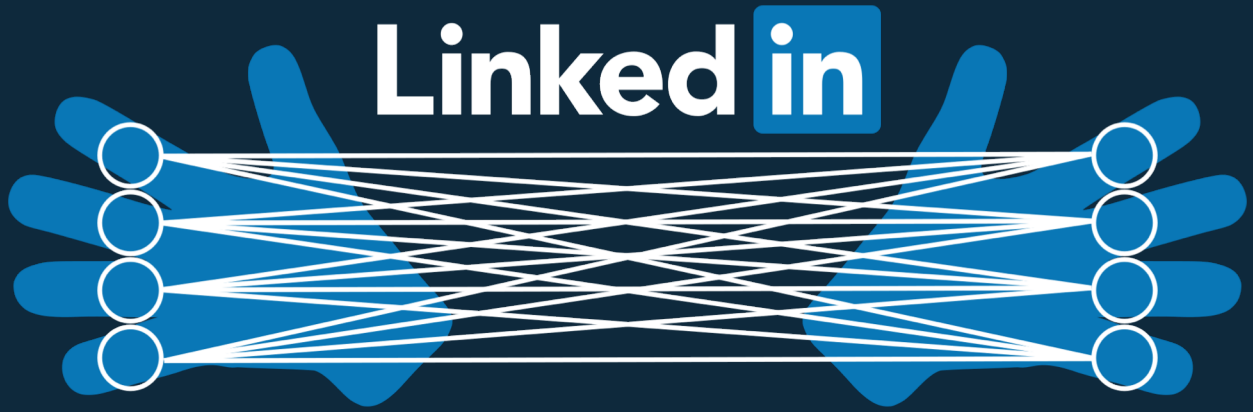
Adoption

Success Traps!

Support Trap

Democratizing AI

Back-fitting end-to-end machine learning at LinkedIn *at scale!*



https://engineering.linkedin.com/blog