IntelliLink©

Seminar Group 6 Team 5

Revitalizing LinkedIn:

A Machine Learning Approach to Enhance User Experience and Engagement

Jiang Lei (U2121557B)
Toh Jing Qiang (U2121442H)
David Tey Bo Yuan (U2121810J)
Nagammai Senthil Kumar (U2120146L)
Fong Ye Xuan (U2120767B)

Business Problem

Current Situation of the Labour Market

Current Situation of the Job Market

- Volatile Job Market characterized by constant changes in job requirements
- **102, 000** employees laid off from U.S. based tech companies in 2023 (Ruby, 2023)

Gap between Available Talents & Demand

Unemployment rate at **3%**
vs
Number of Job vacancies at **104 500**

Job Scams

More than **3500** instances of job scams recorded in 2022, leading to a loss of more than $58 million.

Business Problem

LinkedIn's Position in Current Market

Business Opportunity

- Many layoffs in the job market
- Increase in the number of job seekers & job vacancies
- Job seekers struggle to match their skills with changing job requirements
- Golden opportunity for LinkedIn to address the gap & boost its business

Unemployment Rate (Cue, 2023) **3%**

Annual Job Vacancies (MOM, 2023) **5.1%**

Threat

Post-Covid

Limited in-person interactions has led to shift in remote working and rise of online job searching

Online Job Searching

Provides access to a wide range of job postings online

Vulnerable to Job Scams

70.8 million cases of job scams
Losses of over $1.6 million
(Zamost & Khorram, 2022)

Business Problem

Business Opportunity for LinkedIn

Security
Safeguard users from job scam

Innovation
Stand out from competitors

Effectiveness
Streamline recruitment of passive job seekers

LinkedIn

Adaptability
More adaptable to industry trends amidst the ever changing job market

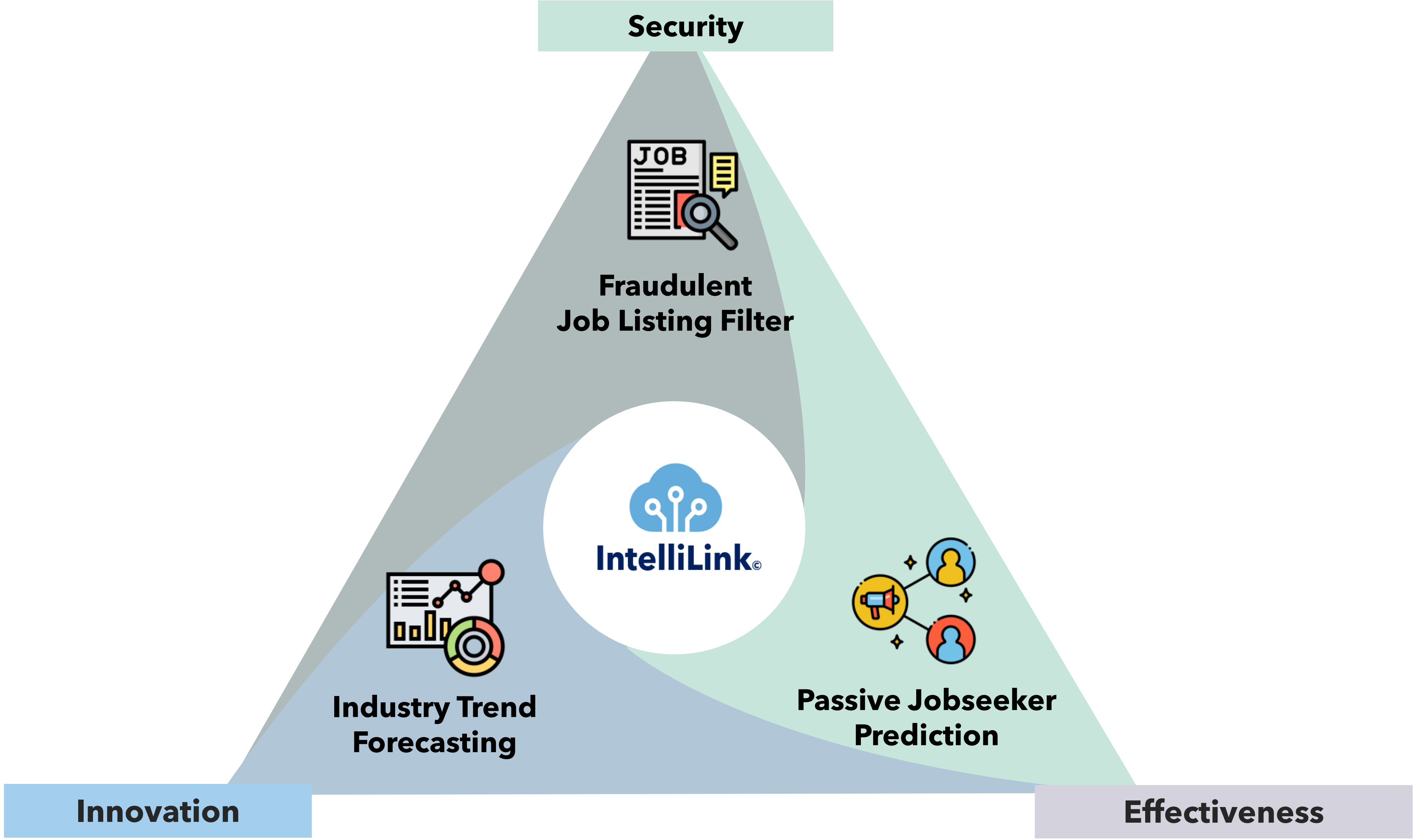
Competitiveness
Gain a competitive edge against its competitors

IntelliLink – Unified Analytical Solution

For Enhancing Security, Innovation & Effectiveness

	Security	Innovation	Effectiveness
Current Solution	Current system mainly flags out **suspicious accounts**	Currently users are encouraged to share **industry news**	Currently only active job seekers are more visible to recruiters through "**open to work**" feature
Gap	System may **leave** sophisticated scams **undetected**	Users must **navigate** changing job market on their **own**	**70%** of workforce are passive job seekers having in-demand skills
Proposed Solution	Automate detection process by **directly** flagging suspicious job postings	Offer **meaningful metrics** on the market trend	**Headhunter** feature to tap the untapped pool of talents – passive jobseekers

Overview of IntelliLink



Jing Qiang

The Technology Behind IntelliLink

Fraudulent Job Listing Filter

Security

Industry Trend Forecasting

Innovation

Passive Jobseeker Prediction

Efficiency

Jing Qiang

Modelling Methodology

Industry Standard & Domain Driven

1

Data Collection
From Reputable Sources
[Kaggle, SingStat, WorldBank]

2

Data Cleaning & Processing
Domain Driven

3

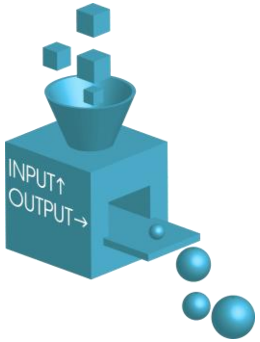
Data Exploration
Understand Data and Obtain Insights

4

Model Training & Optimization
Feature Selection, Hyperparameter Tuning, Feature Importance

5

Model Evaluation
Select best model based on Accuracy & Recall Score



1. Unified Security through Fraudulent Job Listing Classification

Key Technology

Natural Language Processing

Jing Qiang

1. Unified Security through Fraudulent Job Listing Classification

a) Dataset Introduction

Companies → Job Listings → Fraudulent Job Listing Prediction

Company Information	+	Job Information		→	Text Data
Company Profile Has Company Logo Industry		Job Title Job Description Location Department Salary Range Requirements Benefits	Telecommuting Has Questions Employment Type Required Experience Required Education Function		Combination of company information and job information

Jing Qiang

1. Unified Security through Fraudulent Job Listing Classification

b) Word Cloud of Job Listings

Non-fraudulent Job Listing

Common words
"Full time", "bachelor's degree", "Information Technology", and "Senior"

Fraudulent Job Listing

Common words
Entry level", "data entry", "high school", "support", "project", and "design"

1. Unified Security through Fraudulent Job Listing Classification

c) Model Comparison

1 **Oversample Train Dataset (SMOTE)**
- Incentivise model to predict minority, fraudulent cases

2 **Count & TF-IDF Vectorization**
- Make text data structured

	Random Forest (CV)	Support Vector Classifier (TV)	Logistic Regression (TV)	Multinomial Naïve Bayes Classifier (TV)
Overall Accuracy (> 85%)	26.04%	88.50%	97.26%	**89.99%**
False Negative Rate (< 30%)	0.0088%	69.16%	55.07%	**26.43%**
	Very Low Accuracy & Very Low FNR	High Accuracy & High FNR	Very High Accuracy & High FNR	High Accuracy & Fairly Low FNR

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Jing Qiang

1. Unified Security through Fraudulent Job Listing Classification

d) Multinomial Naïve Bayes Classifier with TF-IDF Vectorization – Top Fraudulent Words

Top Fraudulent Words

Ranking	Words
1	Jacksonville job description administrative
2	Restaurant manager awarded
3	Office manager pl
4	Portland sales need
5	Philadelphia administrative 21
6	Welcome require full
7	Referral director product
8	Tampa seek individual
9	Market us md
10	Even faster full

Jing Qiang

1. Unified Security through Fraudulent Job Listing Classification

e) Classification Example 1

Job Listing 1
Company: Food52

Title: Marketing Intern
Location: US, NY, New York
Department: Marketing
Salary: -
Company Profile:
We're Food52, and we've created a groundbreaking and award-winning cooking site.
Description:
Reproducing and/or repackaging existing Food52 content for a number of partner sites, such as Huffington Post, Yahoo, Buzzfeed, and more in their various content management systems. Researching blogs and websites for the Provisions by Food52 Affiliate Program...

Not Fraudulent

1 **Automatically Flags Suspicious Job Listings**

2 **Provides Insights on Fraudulent Listings**
- To make informed decision

3 **Collect feedback from employee**
- To verify classification result
- Enhance model for future prediction

Jing Qiang

1. Unified Security through Fraudulent Job Listing Classification

e) Classification Example 2

Job Listing 2
Company: ???

Title: ADMINISTRATIVE & OFFICE ASSISTANT
Location: US, TX, Houston
Department: Unknown
Salary: -
Company Profile: -
Description:
An exciting growth opportunity for an assistant, who will assist in the daily operations (customer service, office assistant, administrative tasks). Ability to multi-task, prioritize and work on a very dynamic and changing environment. Excellent communication skills, written and oral. Attitude to Solve problems, work INDEPENDENTLY and minimum supervision.

Fraudulent

1 **Automatically Flags Suspicious Job Listings**

2 **Provides Insights on Fraudulent Listings**
- To make informed decision

3 **Collect feedback from employee**
- To verify classification result
- Enhance model for future prediction

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Jing Qiang

1. Unified Security through Fraudulent Job Listing Classification

e) Classification Example 2

Job Listing 2
Company: ???

Title: **ADMINISTRATIVE** & OFFICE ASSISTANT
Location: US, TX, Houston
Department: Unknown
Salary: -
Company Profile: -
Description:
Welcome. This is an **exciting** growth opportunity for an assistant, who will assist in the daily operations (customer service, office assistant, administrative tasks). **Entry level** role perfect for **high school** students. Attitude to Solve problems, work INDEPENDENTLY and **minimum** supervision. Multiple interesting **projects** to choose from.

Fraudulent → Job Listing → Confirm OR Ignore

1 **Automatically Flags Suspicious Job Listings**

2 **Provides Insights on Fraudulent Listings**
- To make informed decision

3 **Collect feedback from employee**
- To verify classification result
- Enhance model for future prediction

2. Innovating with Industry Demand Forecasting

Key Technology

Time Series Forecasting

2. Innovating with Industry Demand Forecasting

a) Dataset Overview

01

Job Vacancy

Singapore

Collected from Government Website

Time Range

2006 April – 2022 July by quarter

Industry

43 industries

Job Vacancy

02

Industry Skill Demand Dataset

Global Context

Collected from WorldBank

Time Range

2015 – 2019 by year

Industry

70 industries

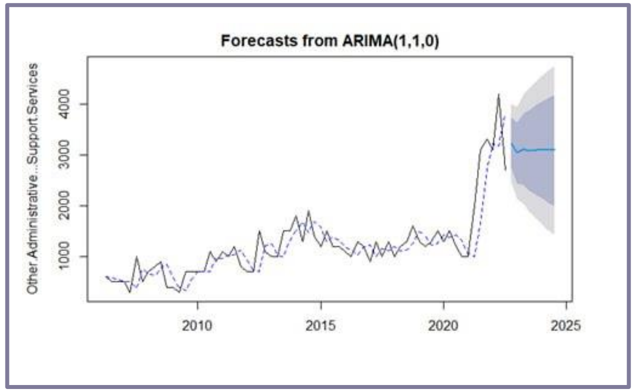
Skill Rank

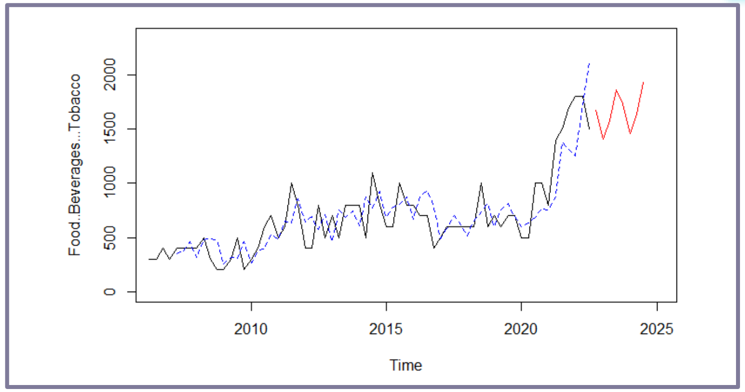
2. Innovating with Industry Demand Forecasting

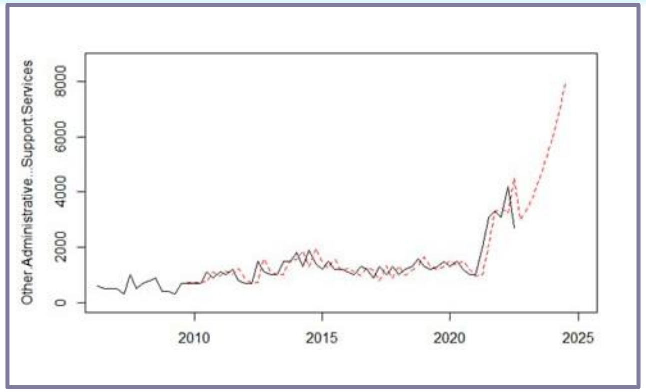
b) Development of Analytics Tool on Job Vacancy Dataset

	ARIMA	Holt-Winters	Taylor Expansion
RMSE	382.84	1485.79	452.39
MAPE	0.2107	2.6021	0.2499
MDA	0.6366	0.4536	0.6341
Decision	NOT chosen due to **invalid assumption**	NOT chosen due to **worst performance**	Chosen due to **interpretability & good performance**

Forecasted Plot







NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

David

2. Innovating with Industry Demand Forecasting

c) Model Evaluation

Assumptions & Interpretability

ARIMA

Assumes **stationary underlying process** e.g., effect of COVID-19 resulting in the coefficients estimated by the ARIMA model **irrelevant** by fitting on **past data**

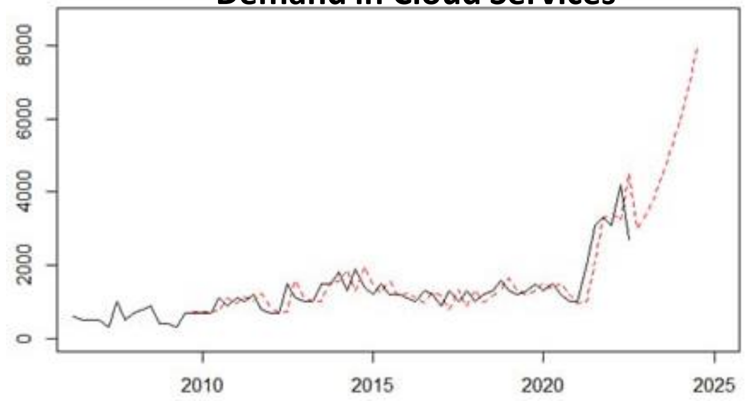
Taylor Expansion

Flexible & interpretable, allow users to incorporate external factors e.g., changes in government policies or natural disaster

Example

1. Context: Cloud services industry **experienced a boom** in the last year

Demand in Cloud Services



2. Model predicts next year will have a similar boom, thus more job vacancies are expected

3. External information: Investment in start-ups been drying up due to a **slowdown in technological breakthroughs**

4. User can **update** his **prior belief** if he thinks that the **underlying model assumption** that the trend will continue is **invalidated** by **external information**

David

2. Innovating with Industry Demand Forecasting

d) Project Job Vacancy Demand to Skill Demand

$$\textit{Industry Forecast Matrix} \times \textit{Skill Weight Matrix}$$

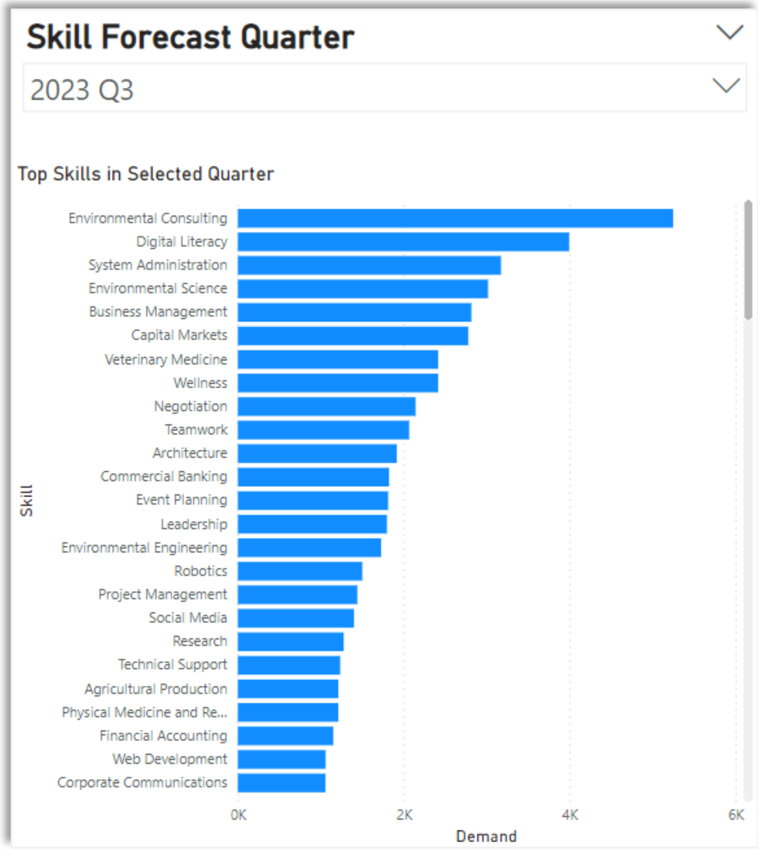
$$= \begin{bmatrix} Q_1, Industry\ 1 & \cdots & Q_1, Industry\ n \\ \vdots & \ddots & \vdots \\ Q_p, Industry\ 1 & \cdots & Q_p, Industry\ n \end{bmatrix} \begin{bmatrix} Industry\ 1, Skill\ 1 & \cdots & Industry\ 1, Skill\ k \\ \vdots & \ddots & \vdots \\ Industry\ n, Skill\ 1 & \cdots & Industry\ n, Skill\ k \end{bmatrix}$$

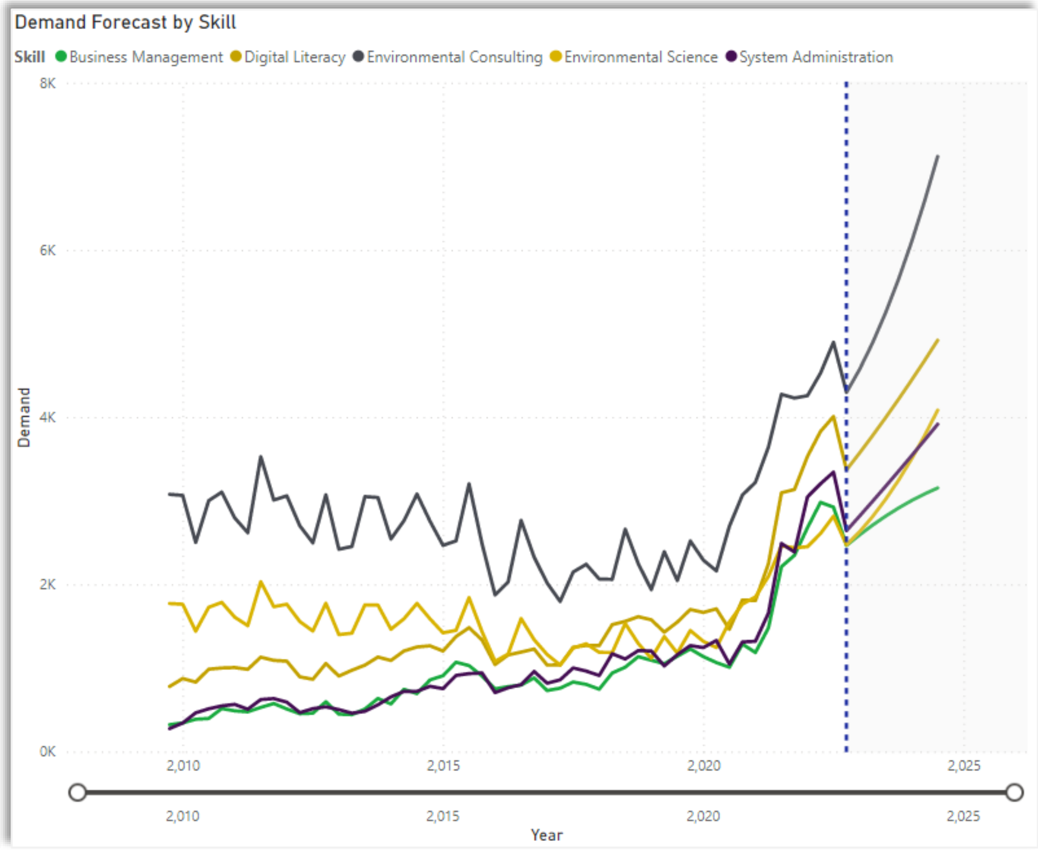
$$= \begin{bmatrix} Q_1, Skill\ 1 & \cdots & Q_1, Skill\ k \\ \vdots & \ddots & \vdots \\ Q_p, Skill\ 1 & \cdots & Q_p, Skill\ k \end{bmatrix} = \textit{Skill Forecast Matrix}$$

1 **Exclusive Insights Sharing**
- Publish reports on the most in-demand skills for different industries

2 **Skill development**
- Develop relevant training courses on Linked Leaning
- Seize business opportunities

3 **Market Analysis**
- Used by companies to inform their hiring and training strategies





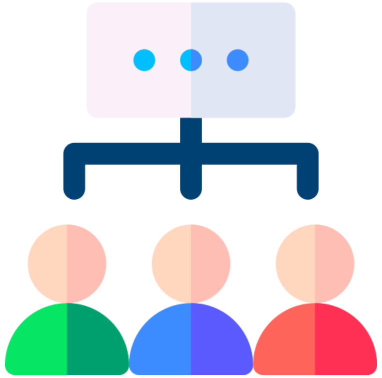
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

Key Technology

Machine Learning Classification



Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

a) Dataset Introduction

U.S. employees → Personal Information → Job Seeker Prediction – willingness of changing jobs

Demographical Variables	Education History	Employment History
Gender **City Code** **City Development Index**	**Enrolled University** *No enrollment/ part-time/ full-time* **Major Discipline** *Arts/ Business/ Humanities/ STEM* **Level of Education** *Highest level of education attainment*	**Total Training Hours** **Years of experience** **Relevant Experience** **Last New Job** *No. of years between previous and current job* **Company Size** **Company Type**

Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

b) Model Comparison

1 **Hyperparameter Tuning**
- Looking for best performer

2 **Reduced Features (8)**
- Adapt to incomplete user profiles in business context.

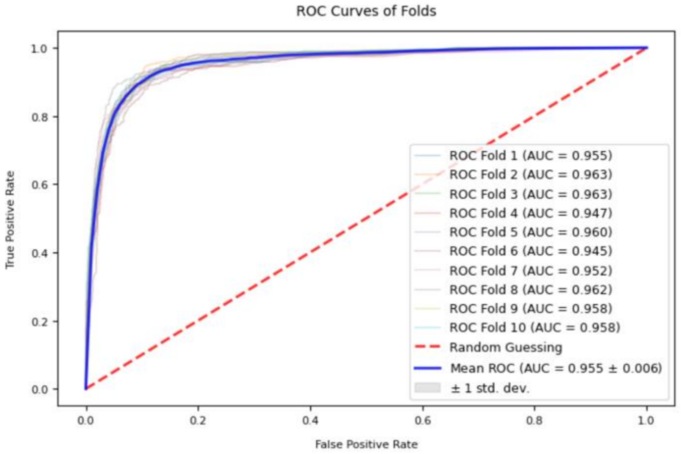
3 **Cross Validation**
- Ensure Consistency & Stability

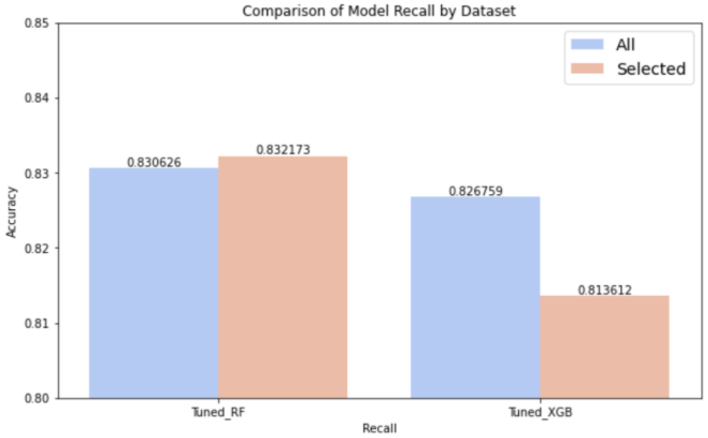
	Support Vector Machine	Logistic Regression	Classification & Regression Trees	Extreme Gradient Boosting	Random Forest
Overall Accuracy (> 90%)	79.95%	86.56%	90.08%	**91.44%**	91.14%
False Negative Rate	21.19%	12.37%	18.17%	17.32%	**16.47%**
Recall	78.81%	87.63%	81.83%	82.68%	**83.53%**
	Worst FNR Lowest Accuracy & Lowest Recall	Best FNR Low Accuracy & Highest Recall	Poor FNR High Accuracy & Fair Recall	Poor FNR Highest Accuracy & Fair Recall	Fair FNR High Accuracy & Fair Recall

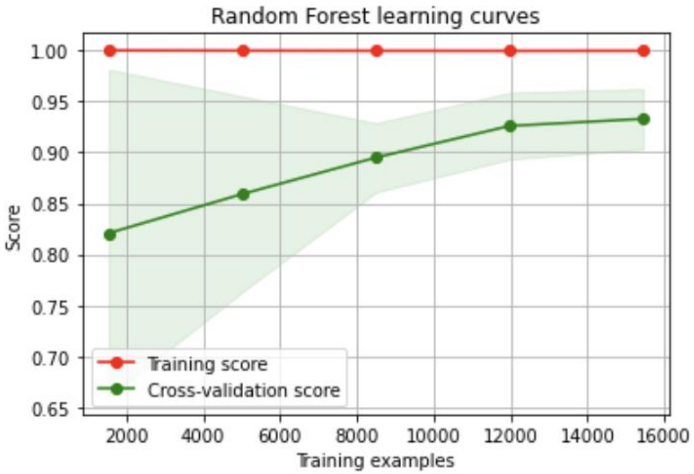
Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

c) Selected Model Evaluation – random forest on 8-feature dataset





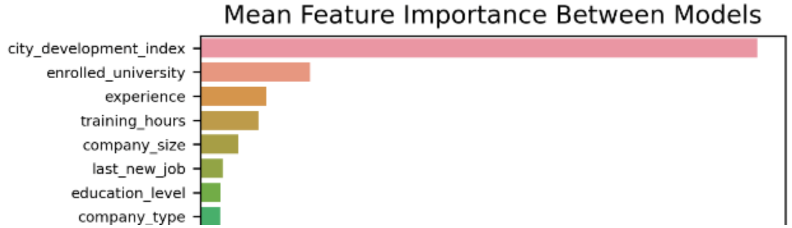


Analytic Insights	**Stable performance** with low standard deviation across different folds	Best Performer on both **initial** and **selected** dataset	Expected to perform well on **unseen** data when **trainset size increases**
Business Implication	**Recruiter** can confidentially **count on the model's prediction** of job seekers who are most likely to accept their job offers	Despite **incomplete profile information**, the model can robustly predict whether he/she is a potential jobseeker	To gather more data, recruiters be prompted to select whether the predicted user is a **genuine jobseeker**, **to feedback model**

Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

d) What user profile information should be prioritized for collection?



1 Geographical Data

- City development index
  (Most significant predictor)

2 Education Experience

- Enrolled university type
- Education level

3 Employment History

- Previous work experience
- Current company size
- Type of company
- Hours of training
- Time elapsed since last job change

Decision:

- Incorporate a **dynamically updated dataset** for mapping location to development index
- Remind users of **updating** their locations

Decision:

- Ensure the **accuracy** of users' educational profiles
  - **Detect** overlapping timeline & mismatched information.
  - Provide supporting educational documentation if necessary

Decision:

- **Verify** job experiences through company email
- **Monitor** for suspicious activity or inconsistencies in employment history

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

e) Prediction Example

Profile 1
Gender: Male
Location: New York City

Profile 2
Gender: Female
Location: Portland

	Background Information	**Background Information**
Enrolled university type	Full-time	Part time
Education level	Graduate	Graduate
Previous work experience	14 years	19 years
Current company size	500-999	5000-9999
Type of company	Private Limited	Private Limited
Hours of training	148	141
Time elapsed since last job change	1 year	> 4 years
Jobseeker Score	**98%** (Recommended to Recruiter)	**54%** (In the rec list, but not in the front page)

Jiang Lei

3. Enhancing Effectiveness through Passive Jobseeker Identification

e) Prediction Example

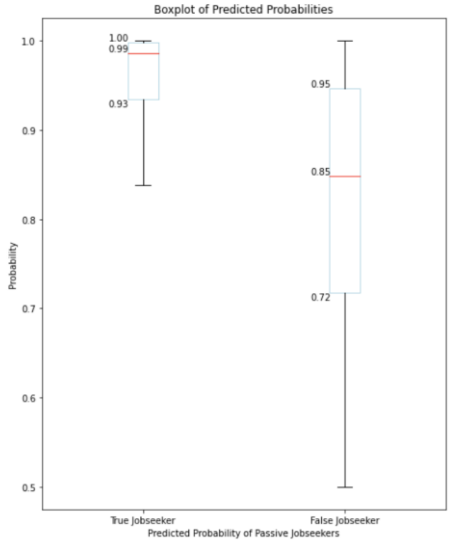
Gender: Male
Location: New York City

	Background Information
Enrolled university type	Full-time
Education level	Graduate
Previous work experience	14 years
Current company size	500-999
Type of company	Private Limited
Hours of training	148
Time elapsed since last job change	1 year
Jobseeker Score	**98%** (Recommended to Recruiter)

1

Recommendation Priority

- Recommend profiles for recruiters with prediction probability **>0.95**
- Sort recommendation by probability



2

Reveal Details for Recruiter

- To make informed decision combined with recruiter's experience
- Tailored offer: an experienced individual who has recently started a new job and may be seeking improved compensation packages

3

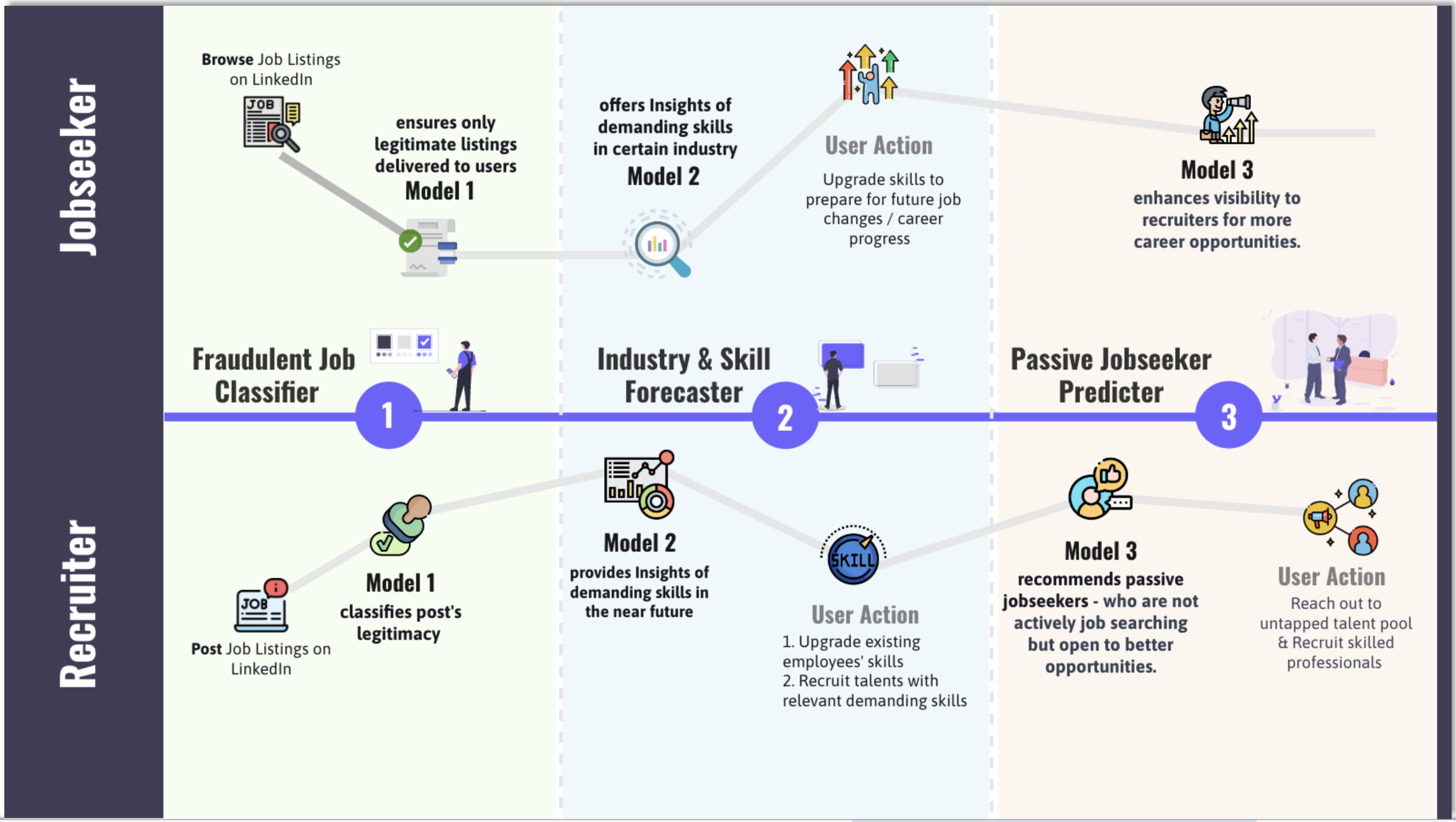
Collect feedback from recruiter to verify prediction result

- "May I have more details about this position" -> genuine jobseeker
- No / Negative response -> wrongly prediction
- Feedback to model for enhancements

Jiang Lei

Business Implementation

How will IntelliLink achieve the goals outlined in LinkedIn's Opportunity Statement?



How can our IntelliLink models optimize LinkedIn Services?

① Talent Solutions division

Safeguard Users from Job Scam

② LinkedIn Learning

Provide Exclusive Industry Insights & Tailored Skill Training

③ Premium Subscription

Premium Head-Hunter Service

Business Implementation

1: Safeguard Users from Job Scam

01 Auto-detected fraudulent listings were sent to the admin for **prompt review**

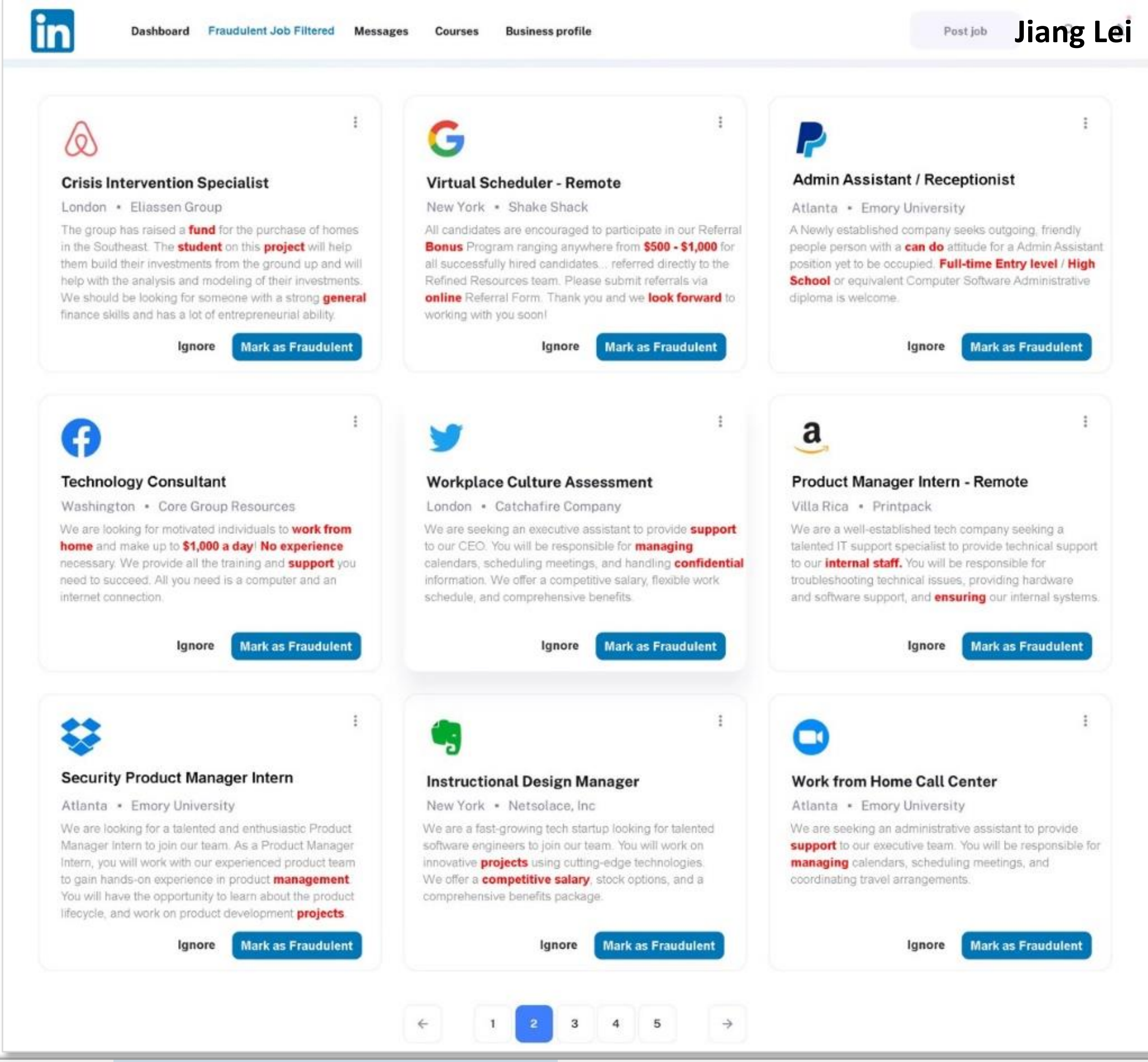
02 Suspicious keywords **highlighted** for easy identification

03 If an account has three flagged posts, it will be **suspended**.

04 **Retrain** the model on misclassified listings

Rigorously improve model's capability to effectively combat the ceaseless influx of job scams

Jiang Lei



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Business Implementation

2: Provide Exclusive Industry Insights & Tailored Skill Training

Jiang Lei

Provide insight into future skill demand

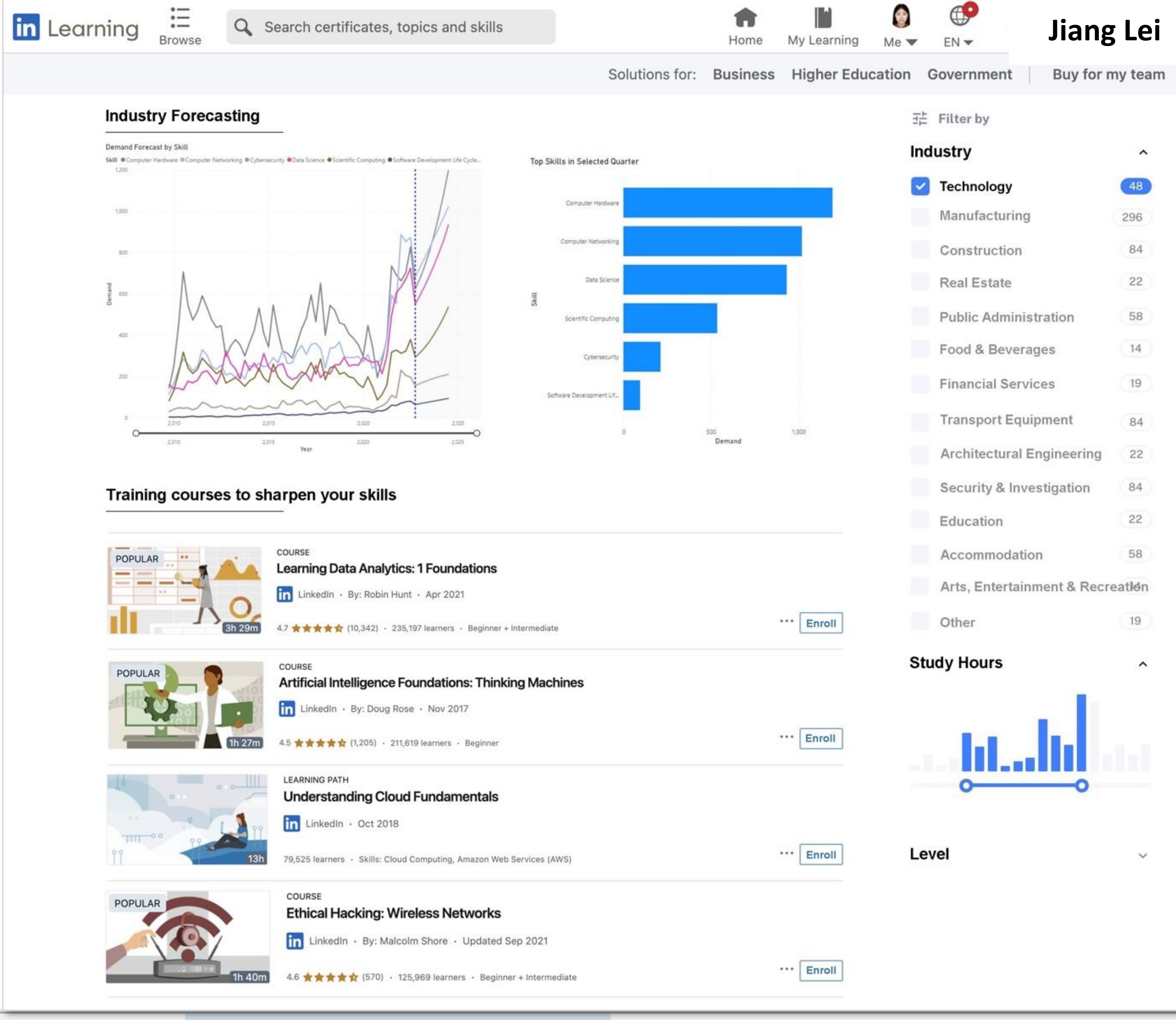
- Market understanding for jobseekers
- Adaptive recruitment strategies for recruiters

Offer relevant training courses

- Remain resilient in an ever-evolving industry by constant upskilling
- Enhance course attractiveness & introduce new courses

Optimize Forecasting Model

- Collect employer feedback to supplement forecasting model (e.g., skill demand accuracy, pre-planning effectiveness).



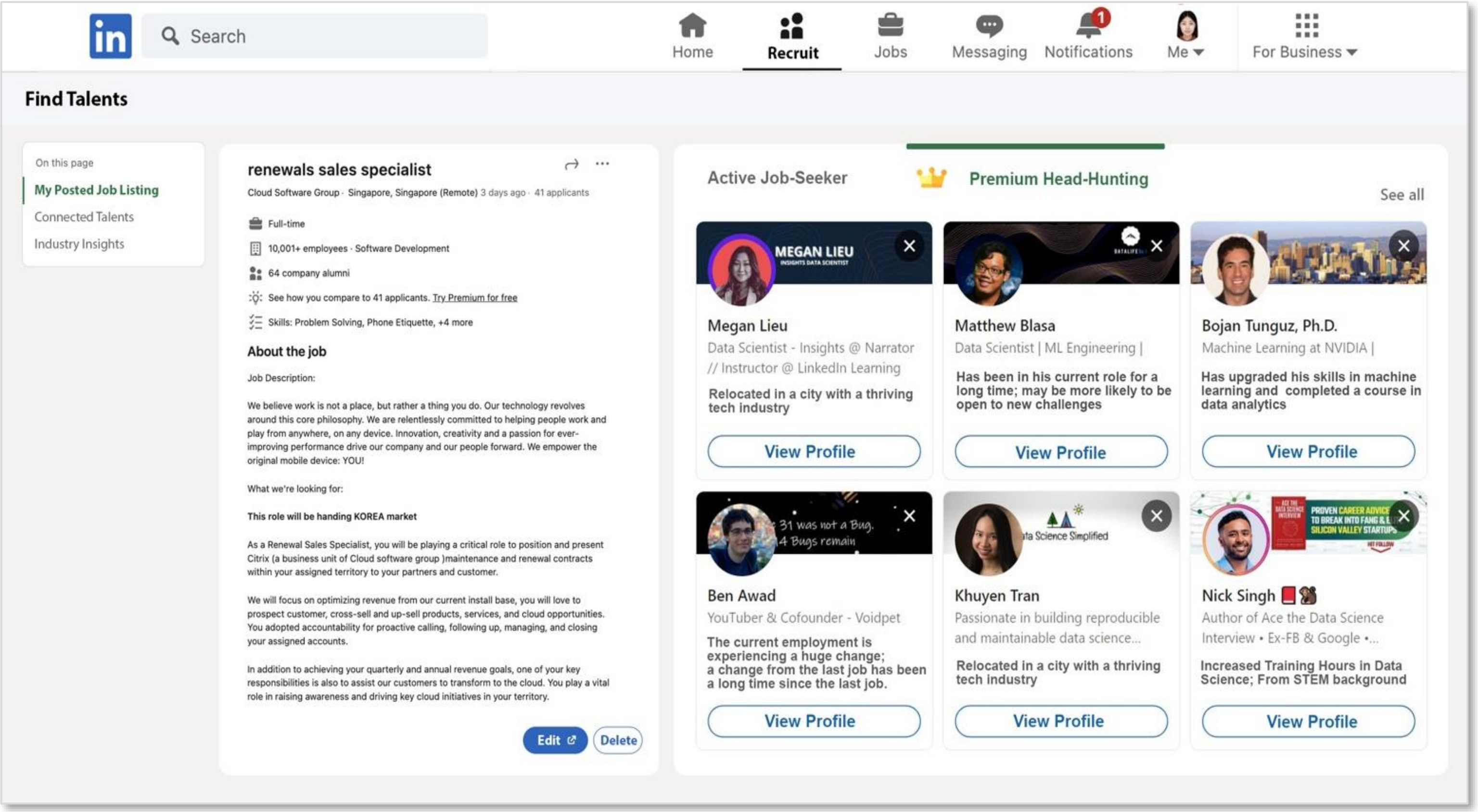
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Jiang Lei

Business Implementation

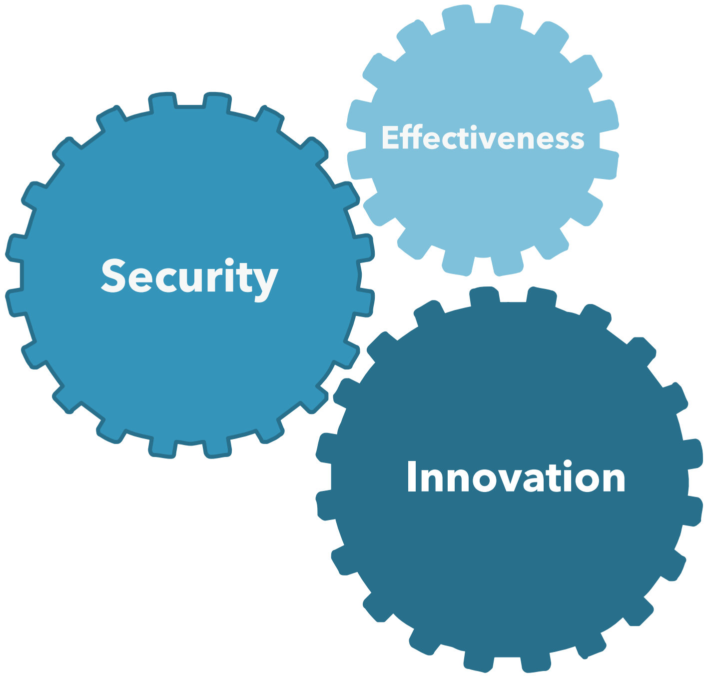
3: Premium Head-Hunting Service

- Premium Head-Hunting feature for **Recruiter**
- Recommend potential job seekers identified by model, **sorted** by predicted probability.
- Provide **insights** on reasons for a candidate's job change willingness.
- Understand the candidate's **motivations** and provide a more **tailored** offer
- Negative & Proactive candidates will be **feedbacked** to model for enhancement
  - *After reach out, the recruiter checks if candidate seeks job change and feedback to system*



Jiang Lei

Expected Business Outcomes



01 Address common recruiting pain points

- Secure 58 million companies & 57% of job seekers from scams
- Bridge talent gap by offering in-demand upskilling courses.

02 Enhance LinkedIn's Work Efficiency

- Automate detection of suspicious job postings
- Ensure efficiency despite a surge in listings, seekers, and recruiters.

03 Boost Revenue Stream

Headhunting premium feature (new)

- Improve the success rate of hiring highly qualified candidates
- Increase Recruiter Lite subscription rates

LinkedIn Learning

- Increase course enrolment by tapping into a large untapped user base
- Expand business by providing more relevant courses based on forecasted in-demand skills
- Partner with employers.

Ken

Future Improvements

Limitations and Concerns

Limitations	Example	Mitigating Strategies
1 **Potential Bias**	• Individuals from certain background **more** likely to be **visible** to recruiters. • Only **certain** types of scams identified	• **Constantly** collect & update data • Ensuring dataset is **diverse** & **representative** • Monitor & gather user feedback **regularly**

Future Improvements

Limitations and Concerns

Limitations	Example	Mitigating Strategies
1 Potential Bias	• Individuals from certain background **more** likely to be **visible** to recruiters. • Only **certain** types of scams identified	• **Constantly** collect & update data • Ensuring dataset is **diverse** & **representative** • Monitor & gather user feedback **regularly**
2 Privacy Concerns	• Vast amount of user data collected to predict if the user is a job seeker • Risk of **data leakage**	• Only collect **necessary data** • Ensure **transparency** in data collection

Future Improvements

Limitations and Concerns

Limitations	Example	Mitigating Strategies
1 Potential Bias	• Individuals from certain background **more** likely to be **visible** to recruiters. • Only **certain** types of scams identified	• **Constantly** collect & update data • Ensuring dataset is **diverse** & **representative** • Monitor & gather user feedback **regularly**
2 Privacy Concerns	• Vast amount of user data collected to predict if the user is a job seeker • Risk of **data leakage**	• Only collect **necessary data** • Ensure **transparency** in data collection
3 Inaccurate User Information	• Users entering **incorrect** information • Users **failing** to **update** their profiles	• Provide more **detailed instructions** for user input • **Outlier detection** to filter potentially inaccurate information

Future Improvements

How can our models be further improved?

Enhancing Model Accuracy

Prediction Model	Enhancements
Fraudulent Job Listing Filter	1. Utilise additional predictor variables 2. Hyperparameter tuning to maximise models' capabilities 3. Alternative word embedding technique – Word2Vec
Industry Trend Forecasting	1. Increase number of data sources 2. Utilise more sophisticated time series models
Passive Jobseeker Prediction	1. Take advantage of LinkedIn's extensive user database 2. Incorporate more pertinent and useful data to provide greater transparency in prediction

Ken

Future Improvements

Further Considerations

Localising Data to LinkedIn

01 **Retrain** the machine learning models using **LinkedIn-specific data**

02 Enhances the models' **prediction accuracy**

03 Results are more **reliable** and **applicable** to LinkedIn's context

Improving Users' Trust

01 Machine Learning models are **"black boxes"**

02 Users may find it difficult to **trust the predictions**

03 Implement a **model interpreter**, such as **SHAP** (SHapley Additive explanation)

Detecting and Removing Fake Profiles and Information

01 Presence of **fake profiles** and **inaccurate information**

02 **Detection** can mitigate impact of **false** or **misleading data**

03 Ensure that IntelliLink's models **perform accurately and reliably**

04 Maintain LinkedIn's **credibility**

Future Improvements

Ending Remarks

IntelliLink©

Security, Innovation, Effectiveness

+ Advantages
+ Enhancements

LinkedIn

Machine Learning Models

- Fraudulent Job Listing Filter
- Industry Trend Forecasting
- Passive Jobseeker Prediction

Enhanced Recruiting Experience

Increased Business Profits

Improved Efficiency

Strengthened Recruitment Security