

## Perhatian - Perhatian..

Kami, kelompok 5, sangat membutuhkan dukungan kalian semua dengan cara berpartisipasi dalam presentasi kami.

Dalam presentasi kami kali ini, **para juri dan audience** berperan sebagai sebuah **anak perusahaan penyalur kerja** yang baru saja buka, dan **bergerak dalam perekrutan khusus bidang Data Scientist**. Anda menyiapkan Data Scientist baru untuk perusahaan client-client Anda, karena sekarang ini, permintaan akan Data Scientist (dari client Anda) sangatlah tinggi.

Case :

Anda baru saja melakukan **pelatihan dan penyaluran kandidat ke perusahaan client Anda** pada batch pertama, namun hasilnya kurang memuaskan karena mengalami kerugian yang disebabkan oleh **rendahnya conversion rate**.

**Kelompok kami** berperan sebagai sebuah **perusahaan konsultan** dalam bidang recruitment. **Anda sedang menyewa jasa konsultasi dari kami** untuk menangani masalah kerugian perusahaan Anda pada batch pertama.

Kami akan bekerja segiat mungkin untuk memecahkan masalah Anda dan mencoba untuk meningkatkan conversion rate serta keuntungan dari perusahaan Anda. :)



# HR Analytics : Job Change of Data Scientist

Final Project by J.A.R.V.I.S

# Our Team



**Wilmay Armianty  
Golden Utomo**



**Mim Hanifah  
Permana**



**Mutia Dewi  
Kurniasih**



**Rizca  
Zahra**



**Nadia Putri  
Prabandini**



**Fikih  
Imam R**



Just A Really Very Intelligent System

# J.A.R.V.I.S

## Talent Recruitment Consultant

(Spesialist Recruitment)

J.A.R.V.I.S adalah sebuah perusahaan konsultan yang membantu klien dalam permasalahan perekrutan kandidat yang tepat.





# Outlines:

01

## Problem Statement

Permasalahan apa yang sedang kita hadapi dan tujuan yang akan dilakukan berdasarkan business metrics

02

## Data Analysis

Mengekplorasi dataset untuk menemukan business insight.

03

## Modelling

Membuat model machine learnig dengan pendekatan supervised learning untuk memprediksi target

04

## Simulation & Recommendation

Membahas gambaran hasil model melalui simulasi dan memberikan rekomendasi kepada stakeholder

# Problem Statement

Business Understanding, Business Purpose,  
Business Metrics



# **“ Data Scientist: The Sexiest Job of the 21st Century ”**

Harvard Business Review

Berdasarkan artikel Harvard Business Review, Data Scientist merupakan pekerjaan yang paling diminati di abad 2021 (Davenport, 2021).

**Median Base Salary: \$130,000**

**Job Openings (YoY Growth): 4,000+ (56%)**

**Career Advancement Score (out of 10): 9**

( Based on LinkedIn article by Pattabiraman, 2019 )



## Mengapa Data Scientist dibutuhkan oleh Perusahaan sekarang ini?

“Data Science Matters  
and How It Powers Business in  
2022”

- Empowering to Make Better Decisions - Data driven Evidence
- Help to Define Goals
- Identifying Opportunities
- Recruiting the Right Talent for the Company

# Latar Belakang Masalah



Banyak perusahaan saat ini yang kesulitan mencari talent data scientist yang tepat (skillful). Oleh karena itu, Kandidat yang ingin melamar ke perusahaan membutuhkan pelatihan dengan harapan agar direkrut.

Dari banyaknya orang-orang yang mendaftar training, perusahaan tersebut ingin mengetahui kandidat mana yang serius mencari pekerjaan baru karena hal ini akan membantu mengurangi **biaya dan waktu perekrutan**.



Just A Really Very Intelligent System

**Bagaimana  
Kami dalam  
membantu  
Problem  
Perusahaan Ini**



# Objective & Tujuan

## Analysis

Mengevaluasi training batch-batch sebelumnya dan memberikan saran untuk batch selanjutnya.



## Machine Learning

Memprediksi kandidat yang benar-benar ingin beralih menjadi Data Scientist



Memahami karakterisasi kandidat training yang serius mencari pekerjaan (*Job Seeker*)



## Our Business Metrics

# Conversion Rate (%)

Presentase kandidat yang akhirnya bergabung dengan perusahaan



$$\frac{\text{Successful hire made}}{\text{Total Candidate}} \times 100\%$$



“

If you have a low conversion rate, this means that you're not getting a return on investment.

(Betterteam, 2021)

Simple Portfolio

Presentation



# Why?

1.

## Faktor Internal

Kita belum memahami profil para calon kandidat yang mengikuti training

Kita belum mengetahui faktor lain yang memotivasi kandidat untuk beralih pekerjaan

## Faktor Eksternal

2.



# Data Analysis

Exploratory Data  
Data Pre Processing  
Business Insight



# Dataset Overview

- Dataset terdiri dari 19158 baris dan 14 kolom

- Sebagian besar data bersifat kategorik  
10 kolom kategorikal  
4 kolom numerikal

## Variabel Target:



**Non job-seeker**

**Target 0 – Not searching for new job**  
Peserta training yang tidak benar-benar ingin bergabung dengan perusahaan



**job-seeker**

**Target 1 – Searching for new job**  
Peserta yang ingin bergabung dengan perusahaan

## Berapa banyak yang mencari pekerjaan?

We have imbalance data  
Kandidat didominasi oleh Non Job-Seeker

75%

Non Job-Seeker

25%

Job-Seeker

# Data Processing

## Handling Missing Value

Missing data menggunakan machine learning imputer

## Feature Encoding

Label Encoder

## Handling Duplicates

Tidak ada data yang terduplikat

Outlier masih dalam batas wajar

## Handling Outliers

Standarization

## Feature Transformation

## Modelling Machine Learning

Data Train : 21.571 baris  
Data Test: 7.191 baris

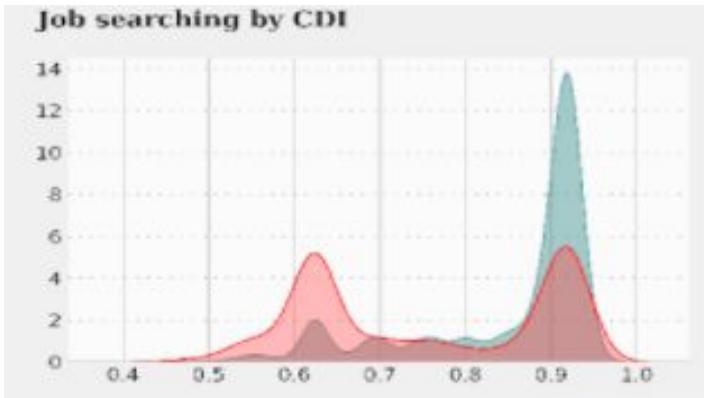
SVM Smote

## Split Data Train & Test 75:25

## Handling Class Imbalance

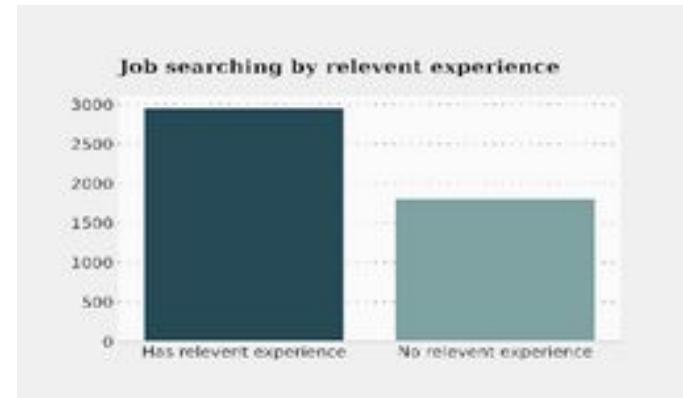
# Siapa sajakah yang sedang mencari pekerjaan?

## City Development Index



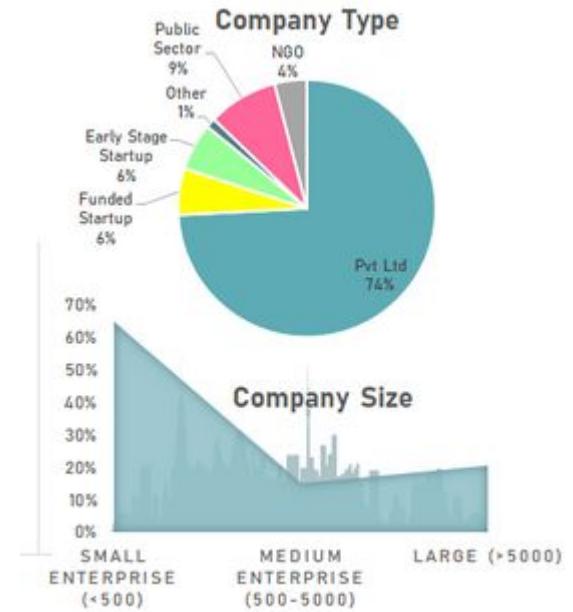
Data Scientist Seekers cukup banyak berasal dari kota dengan skor CDI 0.6 dan 0.9

## Relevant Experience



Kandidat dengan pengalaman yang relevan, lebih serius mengikuti pelatihan

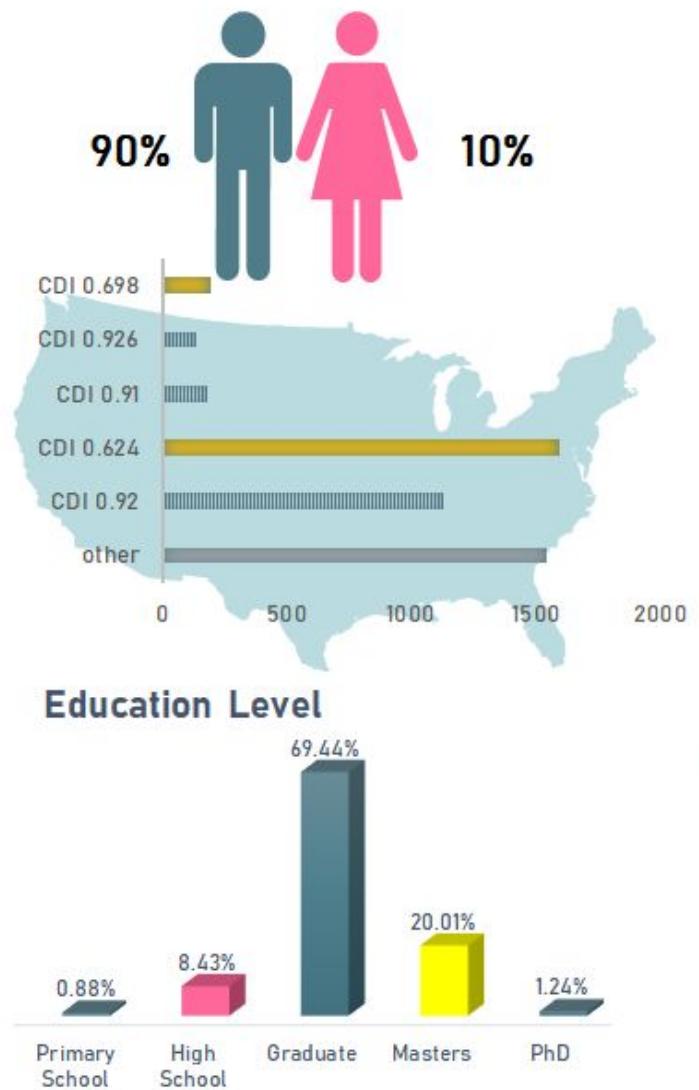
## Company Type & Company Size



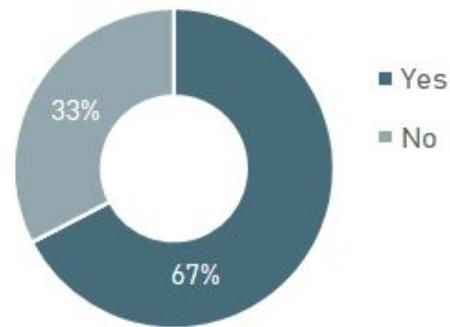
Kandidat yang bekerja di perusahaan kecil (small enterprise) dan di sektor Pvt Ltd lebih serius mengikuti pelatihan

# Resume Profil

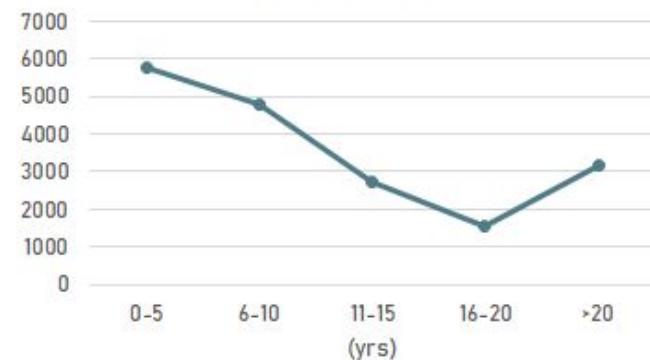
## Kandidat Job-Seeker



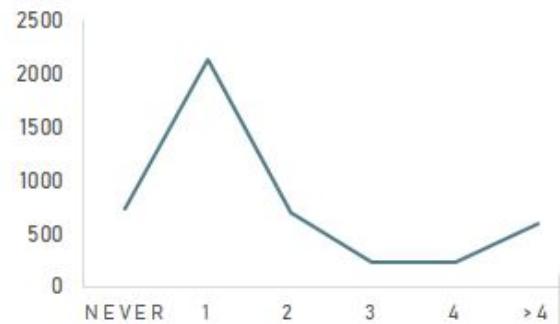
## Relevent Experience



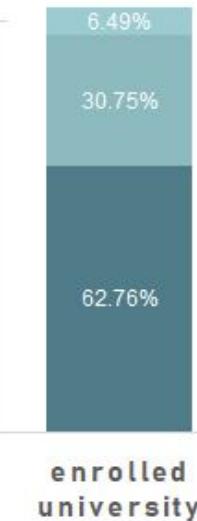
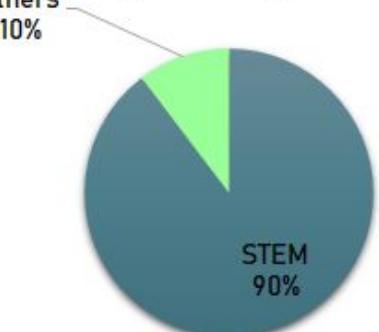
## Experience



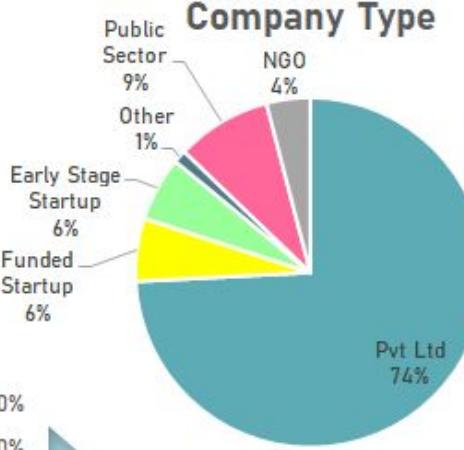
## Last New Job



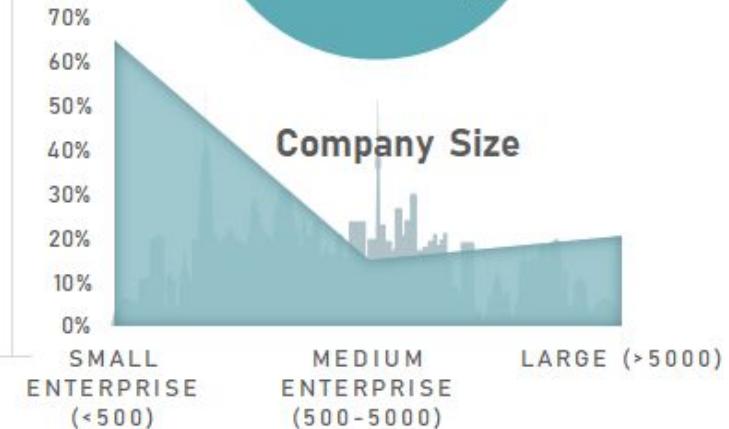
## Major Disciplines



## Company Type



## Company Size



# Modelling



# Model Evaluation

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	TP	FP
	Negative	FN	TN

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

## F1 Score

Distribusi data target tidak seimbang (25:75). *F1 Score* memperhitungkan nilai **Precision** dan **Recall**, sehingga *F1-score* memperhitungkan pentingnya nilai **False Positive** dan **False Negative**.

# Hasil Modelling

Model	Accuracy	F1 Score	AUC Score
Logistic Regression	0.71	0.70	0.7101
Linear SVC	0.71	0.69	0.7086
rbfSVC	0.78	0.79	0.7902
K-Neighbors Classifier	0.78	0.79	0.8543
Random Forrest Classifier	0.88	0.88	0.9831
Decision Tree Classifier	0.84	0.84	0.9832
Gradient Boosting Classifier	0.87	0.87	0.8698
Gausian NB	0.70	0.69	0.6998
Easy Ensemble Classifier	0.90	0.90	0.9074
<b>Easy Ensemble Classifier (after Tuned)</b>	<b>0.90</b>	<b>0.90</b>	<b>0.9075</b>

Algorima terbaik:

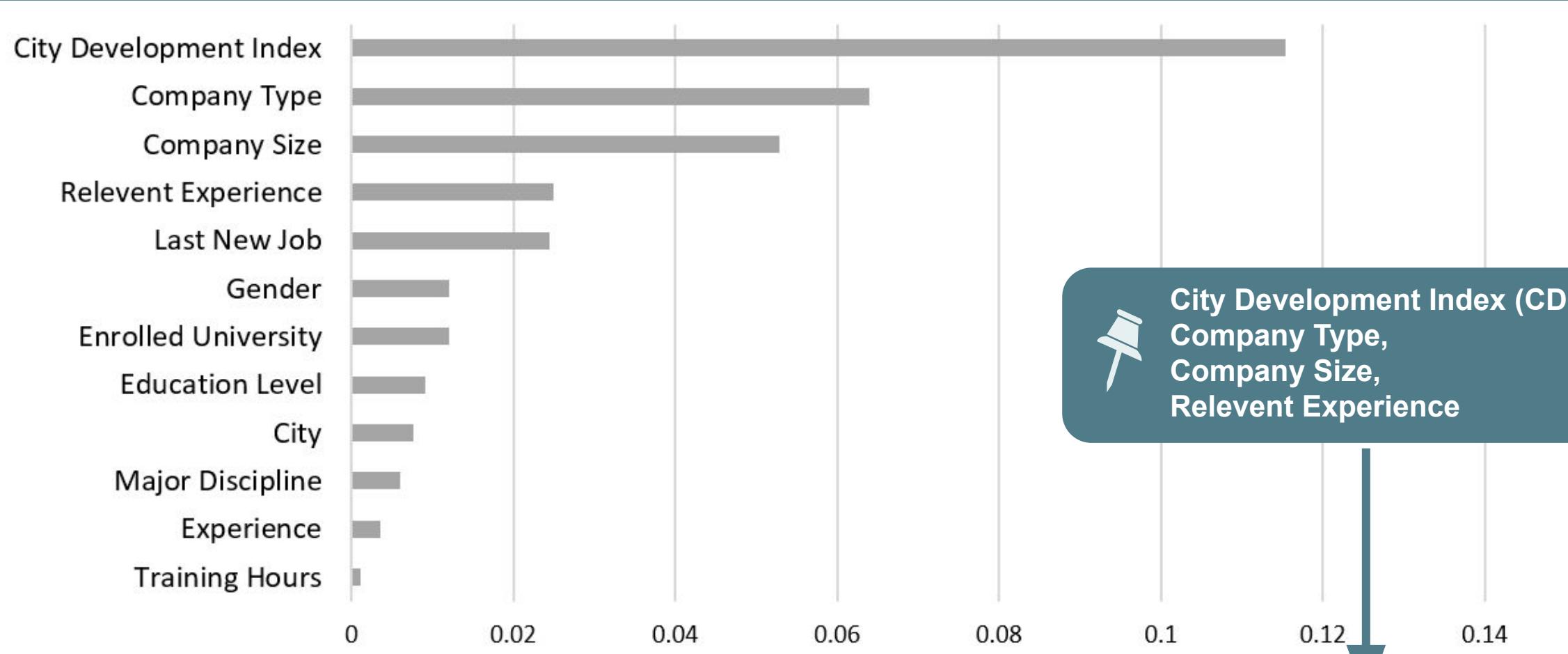
**Easy Ensemble Classifier  
(with Hyperparameter Tuning)**



Meningkatkan score  
AUC sebesar **0.01%**

Prediksi	Aktual	
	Will Join DS	Not Join DS
Will Join DS	3327	361
Not Join DS	368	3235

# Feature Importance



**City Development Index (CDI)  
Company Type,  
Company Size,  
Relevent Experience**



**Business  
Recommendation**

# Simulation & Recommendation



# Referral Fees

(sampel sebesar 100 orang)

## Harga Karyawan yang Lolos Diterima Perusahaan Client

(Estimate: 20 juta/orang)

### A Case : Batch Sebelumnya (tanpa machine learning)

#### Conversion Rate



#### Revenue



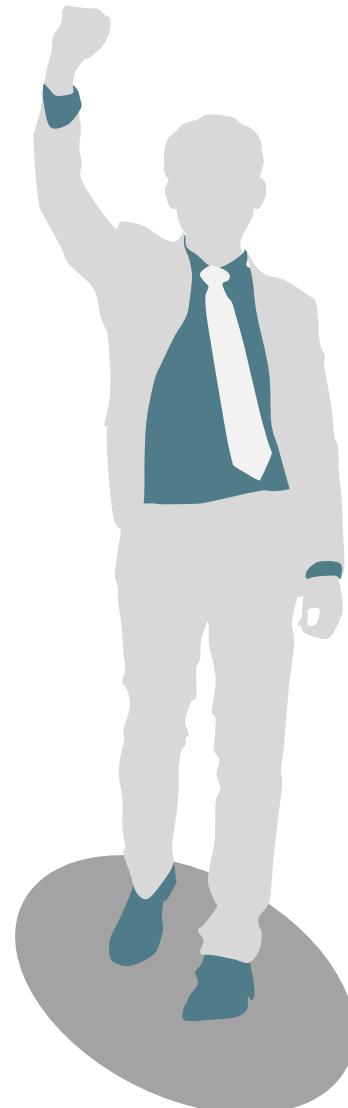
#### Calcution

Total biaya pelatihan/Jumlah orang yang terekrut.

Biaya training: 936 Juta

$$\text{Cost/Training} = 936\text{jt}/25 \text{ orang} \\ = 37.44 \text{ juta/orang}$$

(Mahal)



### B Case: Dengan Implementasi Machine Learning

#### Conversion Rate



#### Revenue



#### Calcution

Total biaya pelatihan/Jumlah orang yang terekrut.

Biaya training: 936 Juta

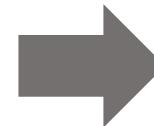
$$\text{Cost/Training} = 936\text{jt}/90 \text{ orang} \\ = 10,4 \text{ juta/orang}$$

(Murah)

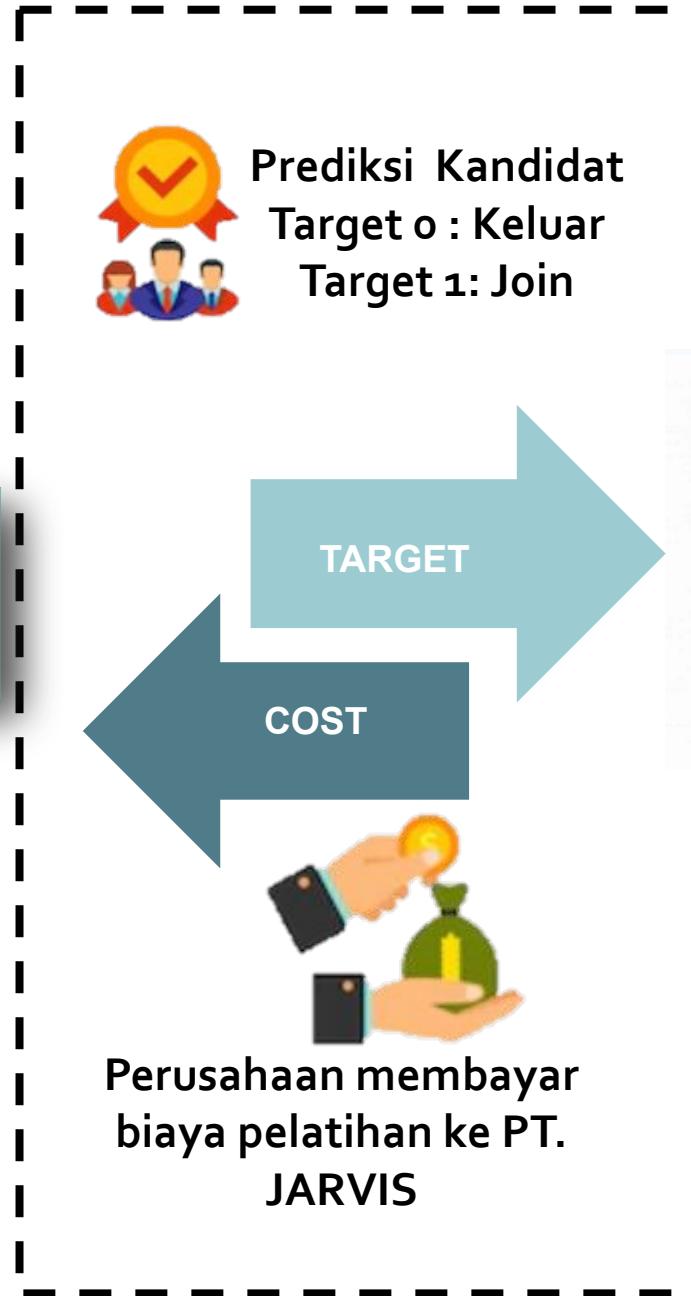
# Project Milestone



Peserta kandidat  
Data Scientist



Penyalur Kerja



Perusahaan  
Client

Simulation



Prediksi Kandidat  
Target 0 : Keluar  
Target 1: Join



Perusahaan membayar  
biaya pelatihan ke PT.  
JARVIS

## Business Problem

### Target 0 (Rate Tinggi)

- Cost untuk trainer lebih banyak
- Inefisiensi waktu

### Target 1 (Rate Rendah)

- Revenue dari payment perusahaan menjadi rendah karena yang join perusahaan menjadi sedikit
- Cost per Hire menjadi mahal

# Project Milestone

<p><b>Key Partners</b></p> <ul style="list-style-type: none"> <li>- Platform Developer</li> <li>- Professional Firms</li> <li>- PR / Marketing Firm</li> <li>- Key Advisors</li> <li>- Strategic Partners</li> <li>- Investors</li> <li>- Equity/Debt Financing</li> </ul>	<p><b>Key Activities</b></p> <ul style="list-style-type: none"> <li>- Career Management</li> <li>- Management Recruiting</li> <li>- Executive Search</li> <li>- Franchise Opportunities</li> </ul>	<p><b>Value Propositions</b></p> <ul style="list-style-type: none"> <li>- Lifetime Career Management &amp; Coaching</li> <li>- Superior and loyal clientcentric talent acquisition services</li> <li>- Relentless pursuit of talent</li> <li>- Entrepreneurial opportunities</li> </ul>	<p><b>Customer Relationship</b></p> <ul style="list-style-type: none"> <li>- Lifetime relationship from college to retirement</li> <li>- Personalized coaching</li> <li>- Long term and loyal client and candidate relationships</li> <li>- Contact management portal</li> <li>- Franchisor training and support to franchisees.</li> </ul>	<p><b>Customer Segment</b></p> <ul style="list-style-type: none"> <li>- Individuals from college to retirement</li> <li>- Companies (Clients) seeking talent</li> <li>- People (Candidates) seeking Data Scientist opportunities</li> <li>- Entrepreneurs seeking a business opportunity</li> </ul>
	<p><b>Key Resources</b></p> <ul style="list-style-type: none"> <li>- Experienced management team</li> <li>- Technology portal</li> <li>- Existing marketing channels</li> <li>- Proven concepts generating revenue</li> </ul>		<p><b>Channels</b></p> <ul style="list-style-type: none"> <li>- Data Scientist focused organizations</li> <li>- Professional organizations</li> <li>- Industry specific companies</li> <li>- Start Up companies</li> <li>- External recruiting firms</li> <li>- Direct sales and recruiting franchise marketing organizations</li> <li>- Social Media</li> </ul>	
<p><b>Cost Structure</b></p> <ul style="list-style-type: none"> <li>- Completion of portal</li> <li>- Trainers for coaching sessions</li> <li>- Acquiring participants through social media</li> <li>- Marketing</li> <li>- Direct sales and promotion</li> <li>- Company needs</li> </ul>	<p><b>Revenue Streams</b></p> <ul style="list-style-type: none"> <li>- Events</li> <li>- Referral fees</li> <li>- Franchise fees</li> <li>- Royalties</li> <li>- Partnership member fees</li> <li>- Management fees</li> </ul>			

Simulation

## Review Berdasarkan Data Batch Sebelumnya

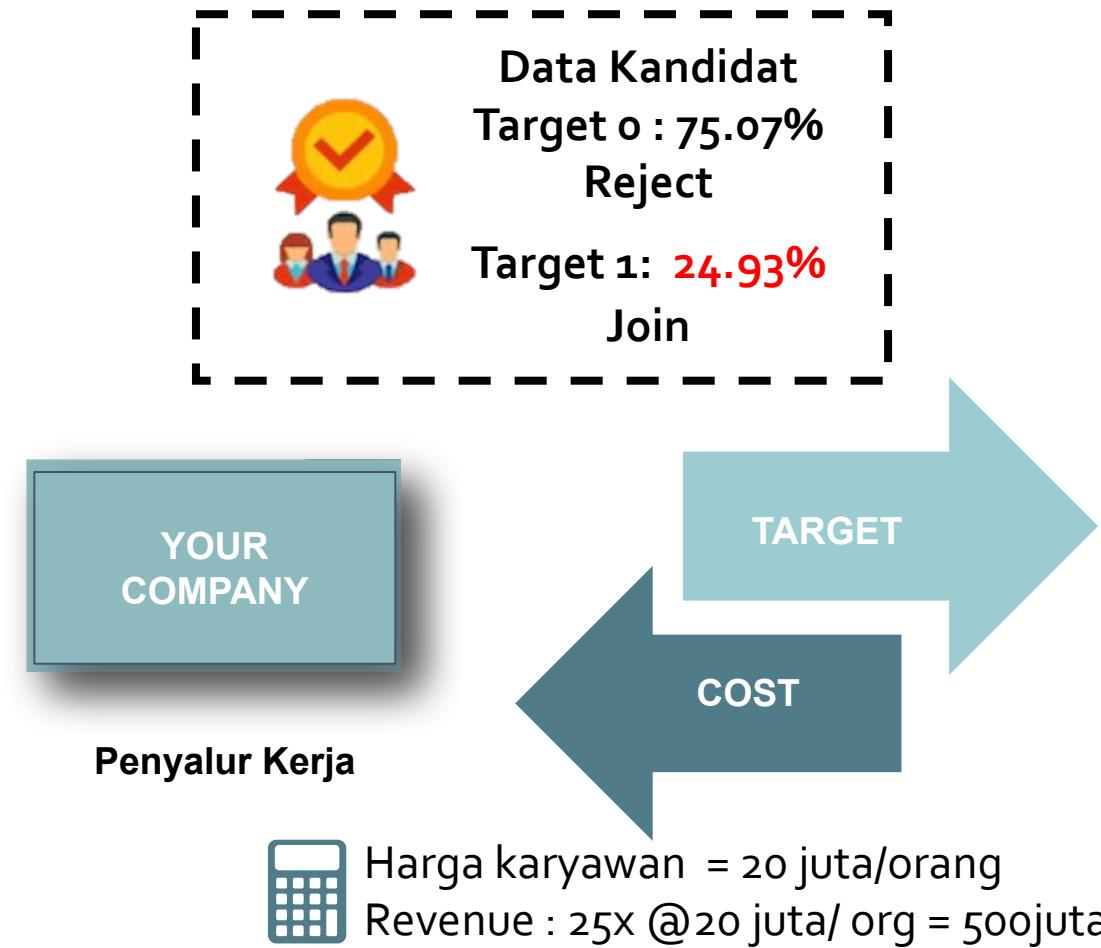


100 orang  
Peserta



YAH RUGI..

Simulation



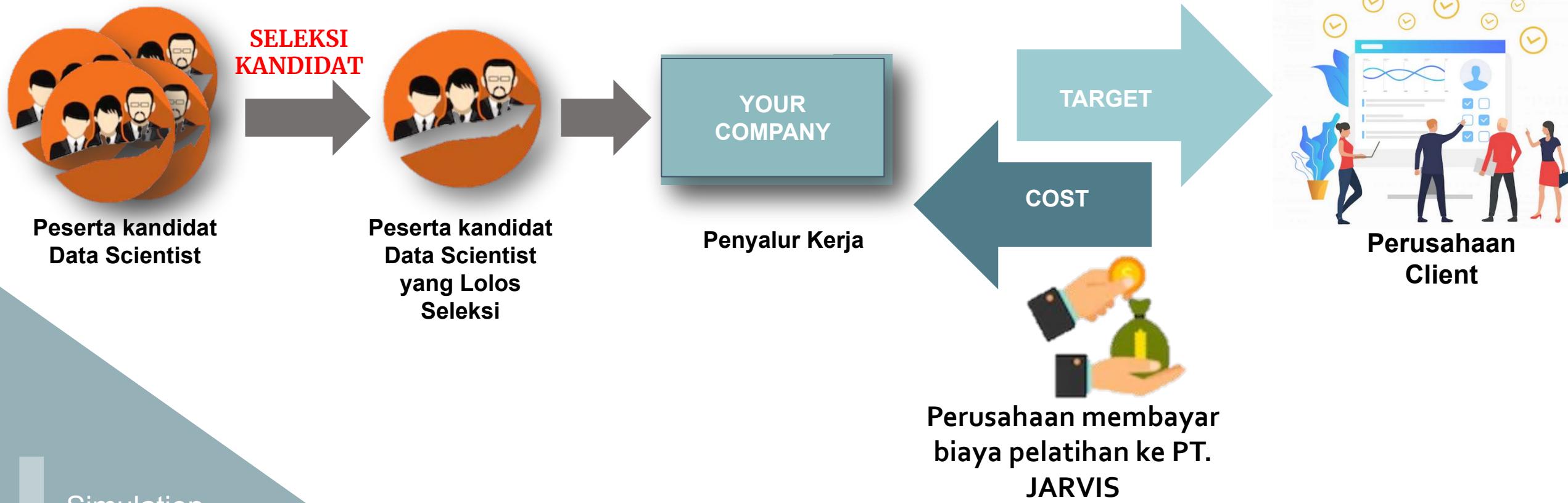
25 join dan  
lolos  
rekrutmen

### Perhitungan Biaya Pelatihan 3 bulan (90 sesi):

1. Trainer 4 kelas :  $4 \times 2 \text{ juta} \times 90 \text{ sesi} = 720 \text{ juta}$
2. Gaji Project Manager :  $8 \text{ juta} \times 3 \text{ bulan} = 24 \text{ juta}$
3. Gaji 2 Supervisor :  $2 \times 4 \text{ juta} \times 3 \text{ bulan} = 24 \text{ juta}$
4. Biaya pengembangan perusahaan dll :  $56 \text{ juta} \times 3 \text{ bulan} = 168 \text{ juta}$

Total pengeluaran : **936 juta** > Revenue : **500 juta**

# Implementasi Machine Learning



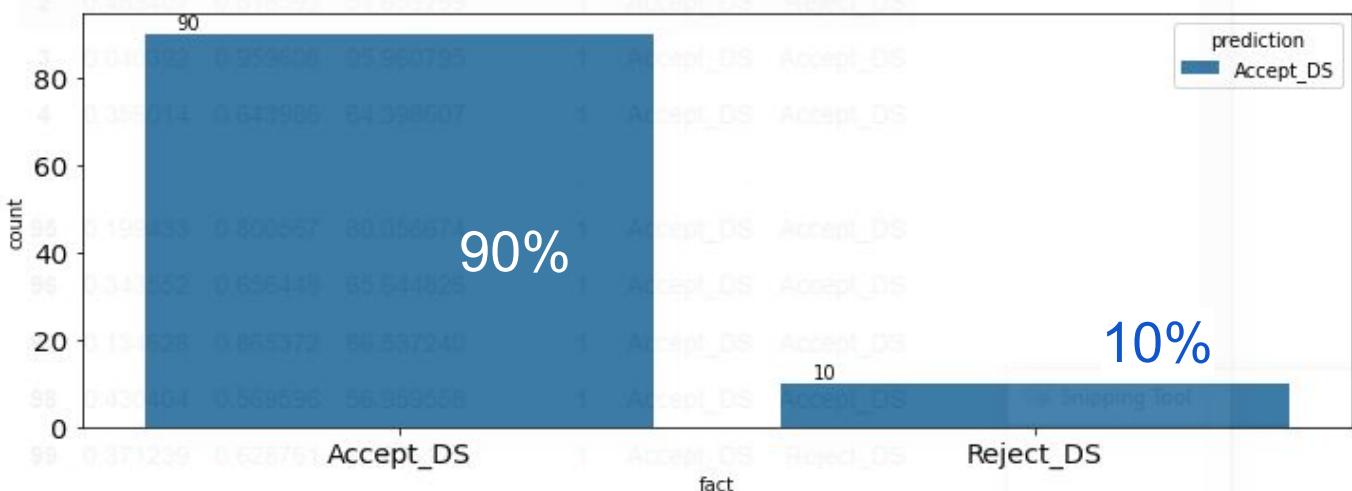
# Simulasi Tahap Seleksi Kandidat

## Proses Seleksi Kandidat

enrollee_id	city	city_development_index	gender	relevent_experience	enrolled_university	education_level	major_discipline	experience	company_size	company_type	last_new_job	fact
5764	city_21	0.624	Nan	Has relevent experience	no_enrollment	Graduate	STEM	2	5000-9999	Pvt Ltd	2	Accept_DS
9006	city_21	0.624	Male	Has relevent experience	no_enrollment	Graduate	STEM	5	100-500	Pvt Ltd	never	Accept_DS
33057	city_103	0.92	Male	Has relevent experience	no_enrollment	Graduate	STEM	>20	Nan	Nan	>4	Reject_DS
13256	city_21	0.624	Male	Has relevent experience	no_enrollment	Graduate	STEM	4	Nan	Nan	2	Accept_DS
2478	city_21	0.624	Male	Has relevent experience	no_enrollment	Graduate	STEM	8	100-500	Pvt Ltd	2	Accept_DS
31767	city_21	0.624	Nan	No relevent experience	no_enrollment	Graduate	STEM	<1	10000+	Pvt Ltd	1	Accept_DS
9921	city_103	0.92	Female	Has relevent experience	no_enrollment	Graduate	STEM	9	Nan	Public Sector	4	Accept_DS
32708	city_103	0.92	Male	Has relevent experience	no_enrollment	Graduate	Humanities	>20	Nan	Nan	1	Accept_DS
33150	city_103	0.92	Male	Has relevent experience	no_enrollment	Graduate	STEM	2	Nan	Nan	never	Accept_DS
28019	city_103	0.92	Female	Has relevent experience	no_enrollment	Graduate	No Major	2	Nan	Nan	2	Accept_DS

Hasil prediksi job-seeker menggunakan Machine Learning

	0	1	prob_1_%	predict	prediction	fact
0	0.054684	0.945316	94.531643	1	Accept_DS	Accept_DS
1	0.071064	0.928936	92.893647	1	Accept_DS	Accept_DS
2	0.483407	0.516593	51.659299	1	Accept_DS	Reject_DS
3	0.040392	0.959608	95.960795	1	Accept_DS	Accept_DS
4	0.356014	0.643986	64.398607	1	Accept_DS	Accept_DS
...	...	...	...	...	...	...
95	0.199433	0.800567	80.056674	1	Accept_DS	Accept_DS
96	0.343552	0.656448	65.644826	1	Accept_DS	Accept_DS
97	0.134628	0.865372	86.537240	1	Accept_DS	Accept_DS
98	0.430404	0.569596	56.959558	1	Accept_DS	Accept_DS
99	0.371239	0.628761	62.876126	1	Accept_DS	Reject_DS



Simulation

# Dengan Implementasi Machine Learning



100 orang  
Peserta lolos  
seleksi

++Denda:  $10 \times 20 \text{ juta} = 200 \text{ Juta}$



Penyalur Kerja

TARGET

COST

Harga karyawan = 20 juta/orang  
Revenue :  $90 \times @20 \text{ juta/ org} = 1.8 \text{ M}$

Perhitungan Biaya Pelatihan 3 bulan (90 sesi):

1. Trainer 4 kelas :  $4 \times 2 \text{ juta} \times 90 \text{ sesi} = 720 \text{ juta}$
2. Gaji Project Manager :  $8 \text{ juta} \times 3 \text{ bulan} = 24 \text{ juta}$
3. Gaji 2 Supervisor :  $2 \times 4 \text{ juta} \times 3 \text{ bulan} = 24 \text{ juta}$
4. Biaya pengembangan perusahaan dll :  $56 \text{ juta} \times 3 \text{ bulan} = 168 \text{ juta}$



90 join dan  
lolos  
rekrutmen

Simulation

Total pengeluaran : 936 juta < Revenue : 1.8 Milliar + denda

YES UNTUNG..



# Business Recommendation

- Untuk batch selanjutnya, penggunaan **Model Easy Ensemble Classifier** dapat membantu dalam memprediksi kandidat job-seeker dan meningkatkan conversion rate sebesar 90%

## Conversion Rate

**25% >> 90%**

Reduce Cost	Increase Revenue
Pengurangan jumlah kelas menjadi 2-3 kelas (30-50 peserta per kelas)	memberlakukan penalty fee (kontrak denda) untuk yang tidak menyelesaikan pelatihan sebesar 20 juta memberlakukan seleksi kandidat sebelum training



## Rekomendasi dalam seleksi kandidat Job-Seeker

No	Feature	Karakterisasi	Action Item
1.	<b>CDI</b>	CDI 0.6 dan 0.9	Memprioritaskan kandidat berasal dari kota dengan CDI 0.6 dan 0.9
2.	<b>Relevant Experience</b>	Has Relevant Experience	Menawarkan program internship untuk kandidat yang belum mempunyai pengalaman yang sebidang
3.	<b>Major Discipline</b>	STEM	Memberi jaminan lolos rekrut kepada kandidat lulusan pendidikan STEM
4.	<b>Education Level</b>	Graduate dan Masters	Kriteria pendaftaran dengan pendidikan minimal High School atau memiliki skill excel dan python
5.	<b>Company Size &amp; Company Type</b>	Small Enterprise/ Pvt Ltd	Memprioritaskan kandidat yang pernah bekerja di Small Enterprise dan sektor Pvt Ltd
6.	<b>Training Hours</b>	66 jam	Training dilakukan selama 3 bulan (66 jam) dengan durasi belajar 5,5 jam/perminggu (include PR, Exam dan Class)
7.	<b>Experience</b>	0-5 Tahun	Memprioritaskan kandidat yang memiliki kurang dari 5 tahun pengalaman bekerja



Thank You

# Appendix

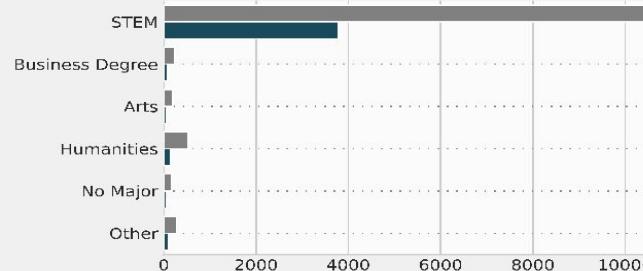
## Siapa saja yang sedang mencari pekerjaan?

Overall

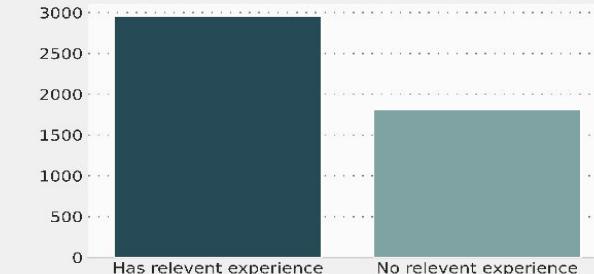
### Job searching by gender



### Job searching by major discipline



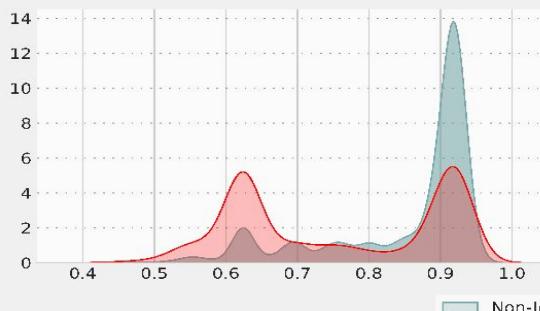
### Job searching by relevant experience



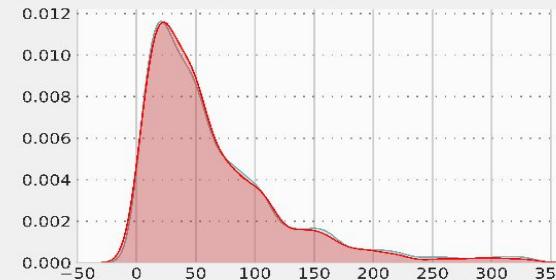
## Bagaimana pengaruh City Development Index dan Training Hours?

Menariknya, Job Seekers cukup banyak berasal dari kota dengan skor CDI rendah

### Job searching by CDI



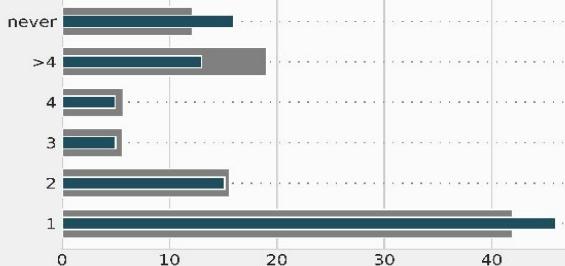
### Job searching by Training Hours



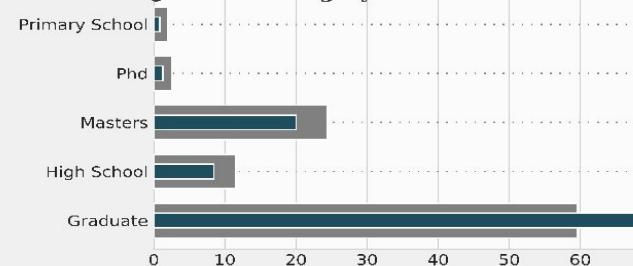
## Bagaimana dengan pekerjaan sebelumnya dan level pendidikannya?

Kandidat dengan pengalaman kurang dari 1 tahun dan lulusan sarjana adalah job seeker terbanyak

### Job searching by Last job change (yrs)



### Job searching by Education level



## Insight

Data dinominasi (proporsi lebih dari 50% data) oleh kaum laki-laki, berpengalaman dan STEM. 88.90% job-seeker adalah laki-laki dan hanya 9.62% merupakan job-seeker perempuan. 89.98% berlatar pendidikan STEM (Matematika dan 61.98% memiliki pengalaman yang relevan

Grafik City Development Index (CDI).

Terdapat dua puncak pada grafik job-seeker. Puncak dengan CDI rendah(0.6) dan CDI tinggi(0.9).

Dapat kita sumskan bahwa;  
Job seeker dengan CDI tertinggi terdorong untuk mencari peran baru menjadi data scientist.  
Sedangkan, job seeker dengan CDI rendah berharap untuk meningkatkan kualitas hidup mereka melalui pekerjaan baru menjadi DS.

Semua ini hanya duguan tetapi menarik untuk diperhatikan

Berbeda dengan grafik training hours;  
Ternyata tidak terdapat perbedaan dalam jumlah training hours antara job-seeker dan non job-seeker yaitu 46-48 jam.

Hal yang menarik dari grafik last new job;  
Kandidat yang memiliki perbedaan 1 tahun antara pekerjaan lama dan pekerjaannya sekarang cenderung akan mencari pekerjaan baru (45.97%) dan kandidat yang belum pernah mencari pekerjaan (15.99%) tampaknya siap untuk tantangan baru.

Grafik education level, bila dibandingkan dengan education level lainnya terlihat bahwa job-seeker dengan lulusan sarjana (69.44%) lebih banyak daripada non job-seeker (59.56%).