

Job Recommendation System, Youths4Jobs Platform

Thesis Stage 2 Presentation

December 19, 2022

Presenter : Kamal Shrestha, cs21mtech16001

Thesis Supervisor : Dr. Maunendra Sankar Desarkar

**Department of Computer Science and Engineering (CSE)
IIT Hyderabad**

Stage 1: Conditional Natural Language Generation for Dialogue Systems

Dialogue
History

[PERSON 1:] Hi
[PERSON 2:] Hello ! How are you today ?
[PERSON 1:] I am good thank you , how are you.
[PERSON 2:] Great, thanks ! My children and I were just about to watch GOT
[PERSON 1:] Nice ! How old are your children?
[PERSON 2:] I have four that range in age from 10 to 21. You?
[PERSON 1:] I do not have children at the moment.
[PERSON 2:] That just means you get to keep all the popcorn for yourself.
[PERSON 1:] And Cheetos at the moment!
[PERSON 2:] Good choice. Do you watch Game of Thrones?

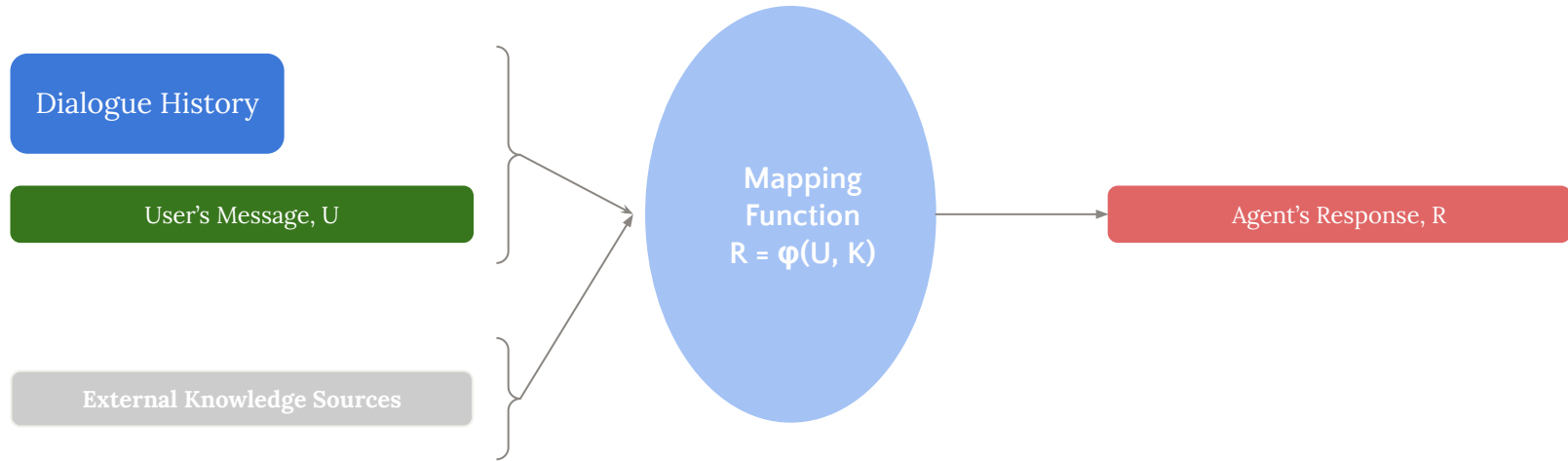
User's Message

[PERSON 1:] No, I do not have much time for TV.

Agent's Response

[PERSON 2:] I usually spend my time painting: but, I love the show.

Knowledge Grounded Systems use an external knowledge such as common-sense knowledge as a significant source of information when organizing an utterance.

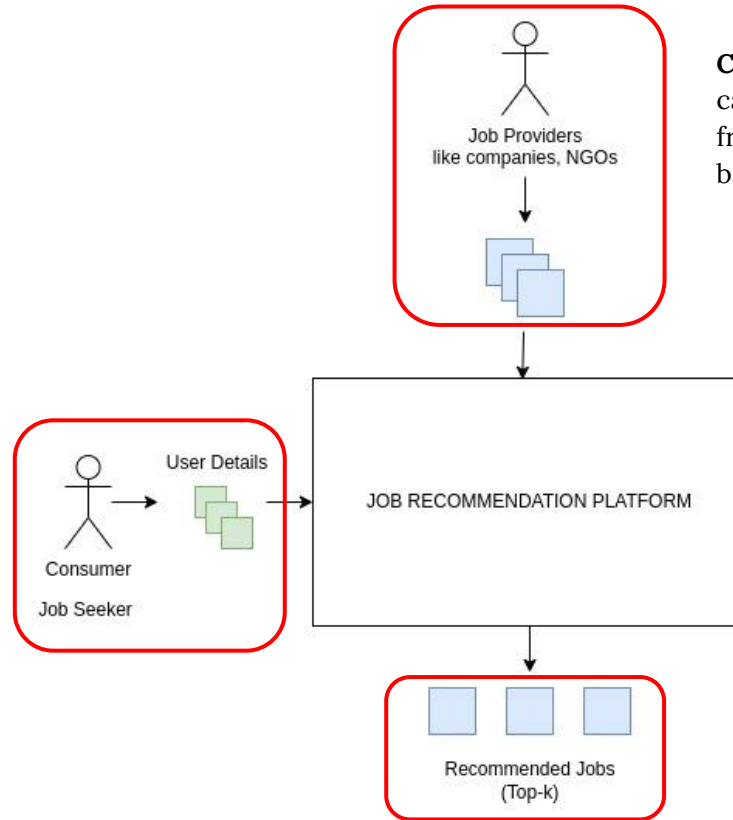


Dialogue Systems with external knowledge (k)

$$U = \{u^{(1)}, u^{(2)}, \dots, u^{(i)}\}, K \longrightarrow R = \varphi(U, K) \longrightarrow R = \{r^{(1)}, r^{(2)}, \dots, r^{(j)}\}$$

Parallely: A heuristic based **jobs recommender systems**

Consumer or job seekers can find appropriate jobs based on multiple features like skills sets, locations, experiences and training **to get hired**



Companies or other job providers can register for odd jobs like working on front desks, assistants, guards **to hire** people based on their requirements and user's skills

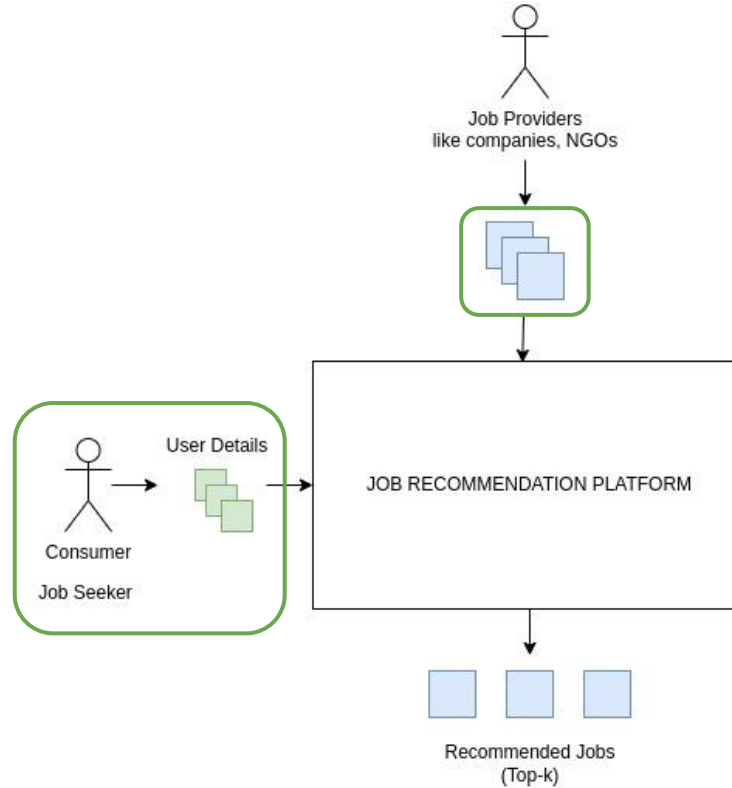
The differentiation with naukari.com or linkedin lies in the **target audience** (less privileges, differently abled audience)

People with disabilities

Less privileged with education

Good Work Ethics and commitment

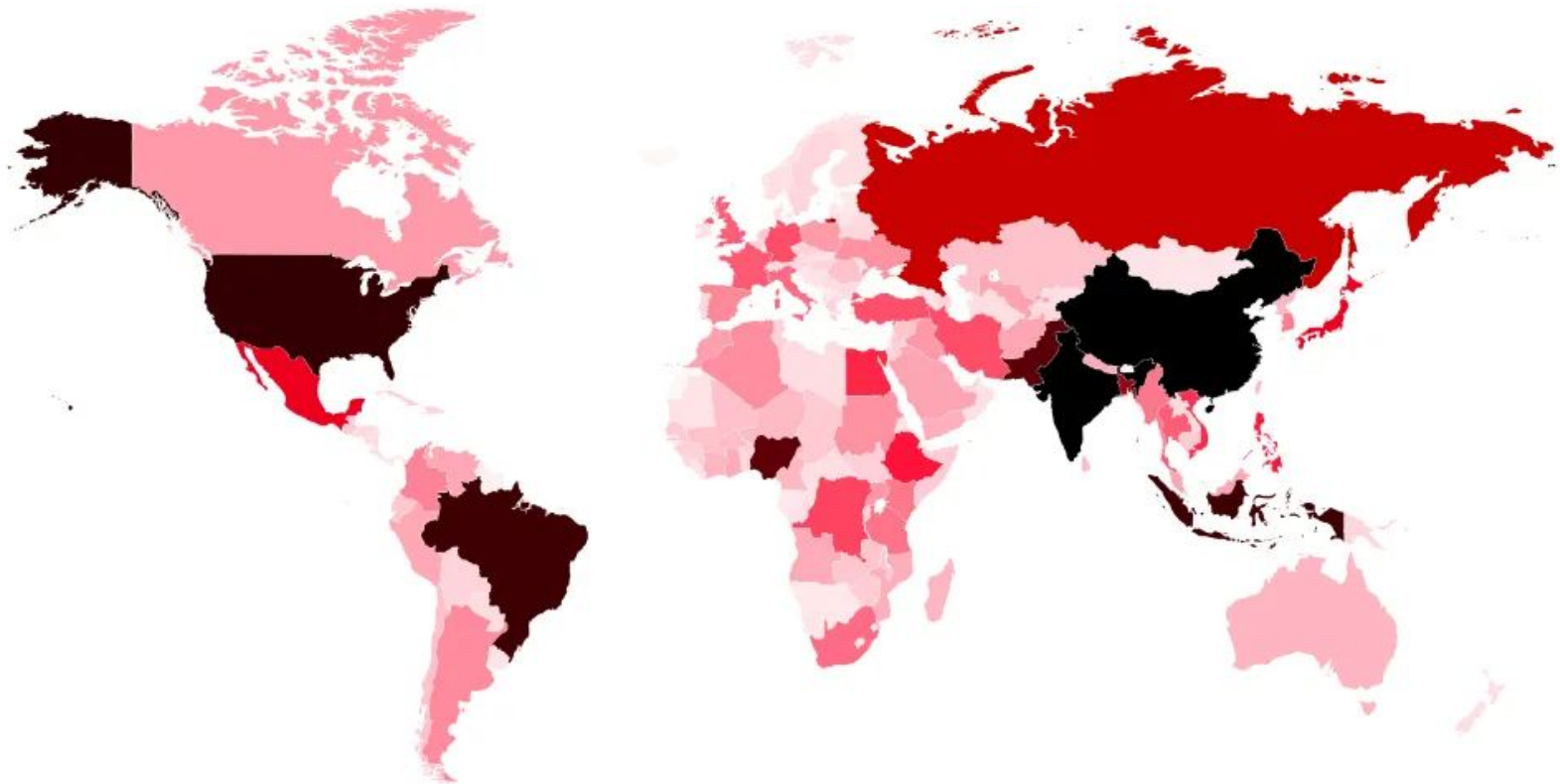
Willing to learn a skill and work



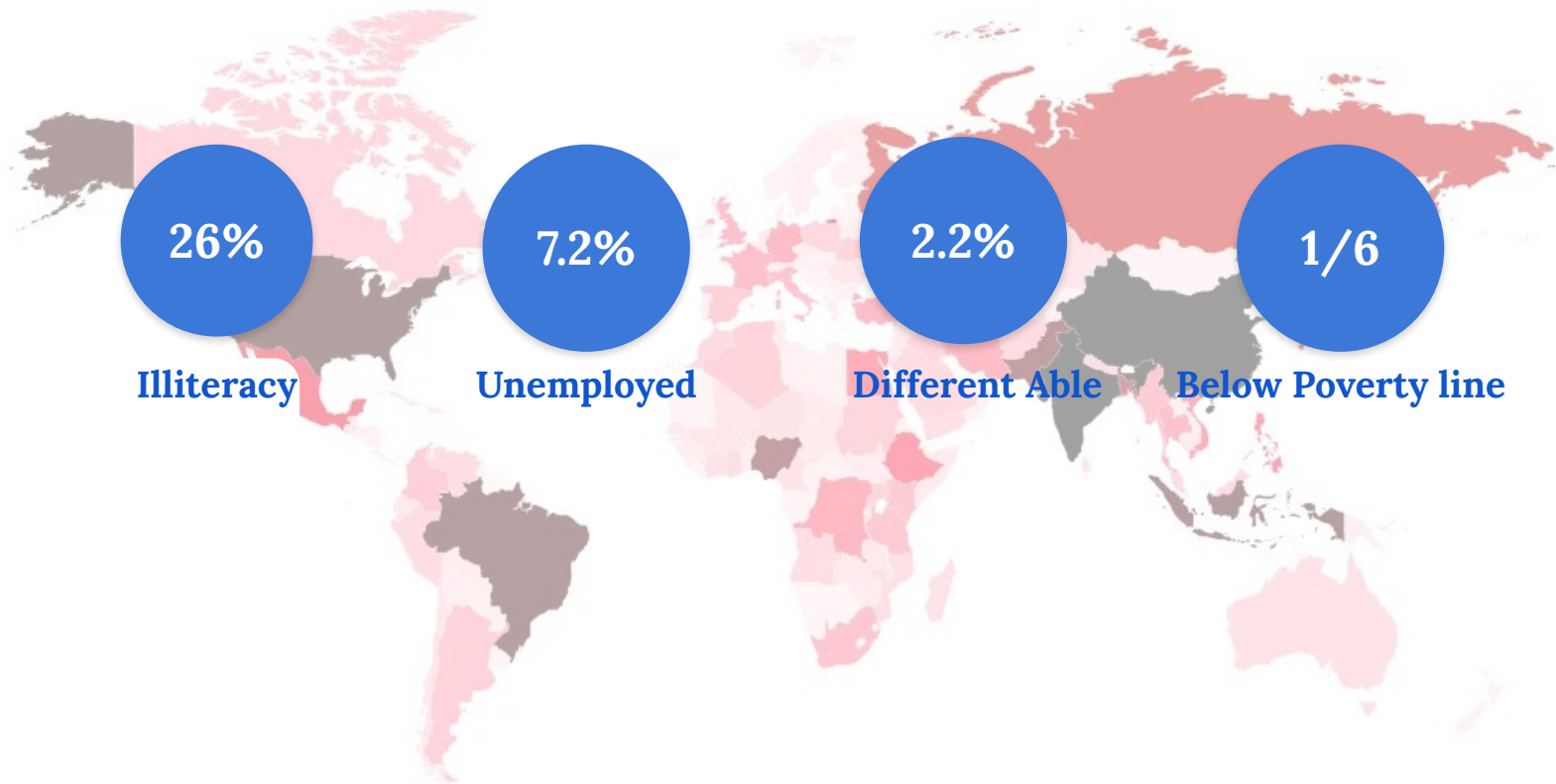
naukri.com



India is the second **most populous country** in the world, after China with 17.71% of the world and a total of 1.39 billion people



In India, **16.4% of people are considered to be poor**, and 4.2% of people are considered to be extremely poor. 18.7% of people are considered to be at risk of poverty.





HOME FOR EMPLOYERS FOR INSTITUTIONS ADMIN

Our Partners



uatstagingy4j.vqindia.com



INDIA'S  FIRST
JOB PORTAL WITH **AI**
FOR PERSONS WITH DISABILITIES

BEST IN CLASS
ACCESSIBILITY FEATURES
TO SUPPORT SOCIAL INCLUSION

Disability Type



Location



Sector



Find Jobs

REGISTER

LOGIN



A CSR Initiative by



Are you a Person with Disability (PwD)? Register yourself here.

REGISTER



Recent Jobs

Recent

Full Time

Part Time

Intern

Apprenticeship

Front Office KFC

₹ Not Specified
📍 Agra, Ahmednagar, Bangalore

Apply

Full Time

Grassroot Fellow Youth4Jobs

₹ Not Specified
📍 Belgaum, Mumbai

Apply

Full Time

Grassroot Coordinator Youth4Jobs

₹ Not Specified
📍 Belgaum, Bangalore, Mumbai

Apply

Full Time

SLI Trainer Youth4Jobs

₹ Not Specified
📍 Bangalore, Mumbai

Apply

Full Time

Regular Trainer Youth4Jobs

₹ Not Specified
📍 Mumbai, Bangalore

Apply

Full Time

Multi-trade technician Youth4Jobs

₹ Not Specified
📍 Mumbai, Ahmedabad, Bangalore

Apply

Full Time

REGISTER

LOGIN



Project Requirements and Expectations

Quick Response Time
(Low Latency)

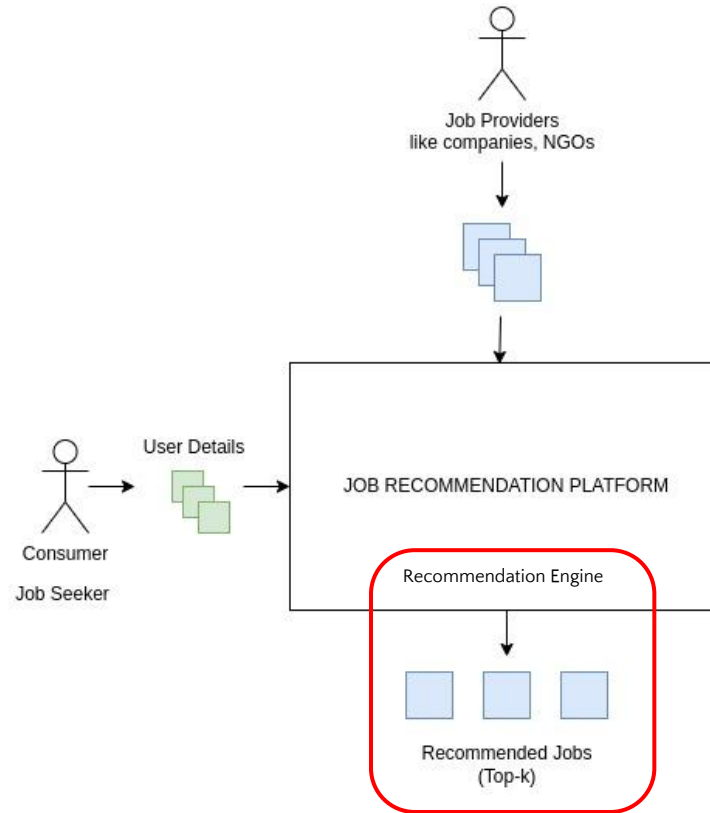
Reliability

Dynamic

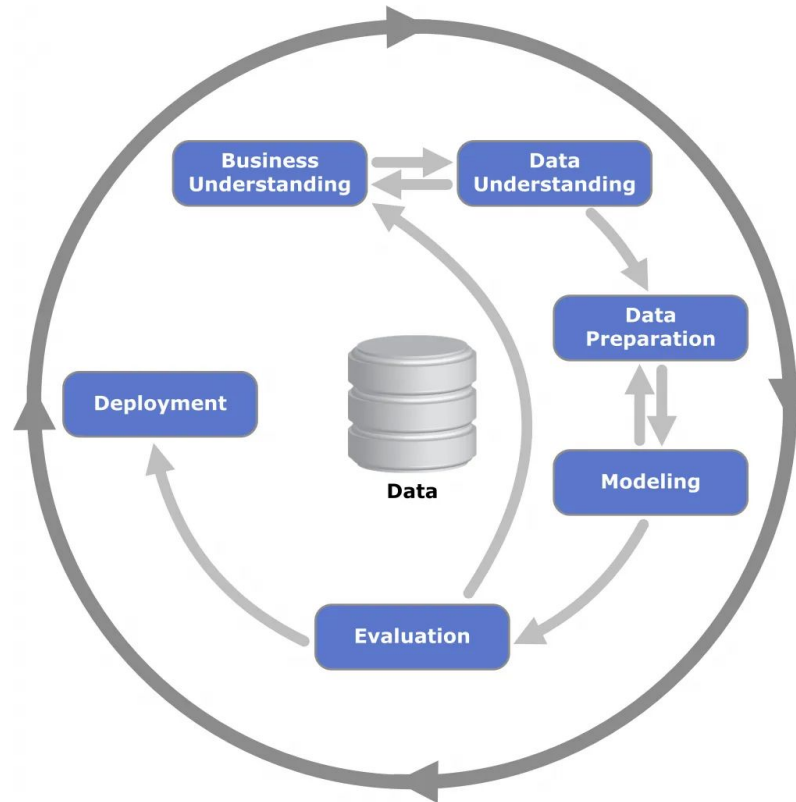
Incremental

Minimal Space
Requirements

Automated



Project Lifecycle: **CRISP-DM** (Cross-Industry Standard Process for data mining) with **Agile Methodology**



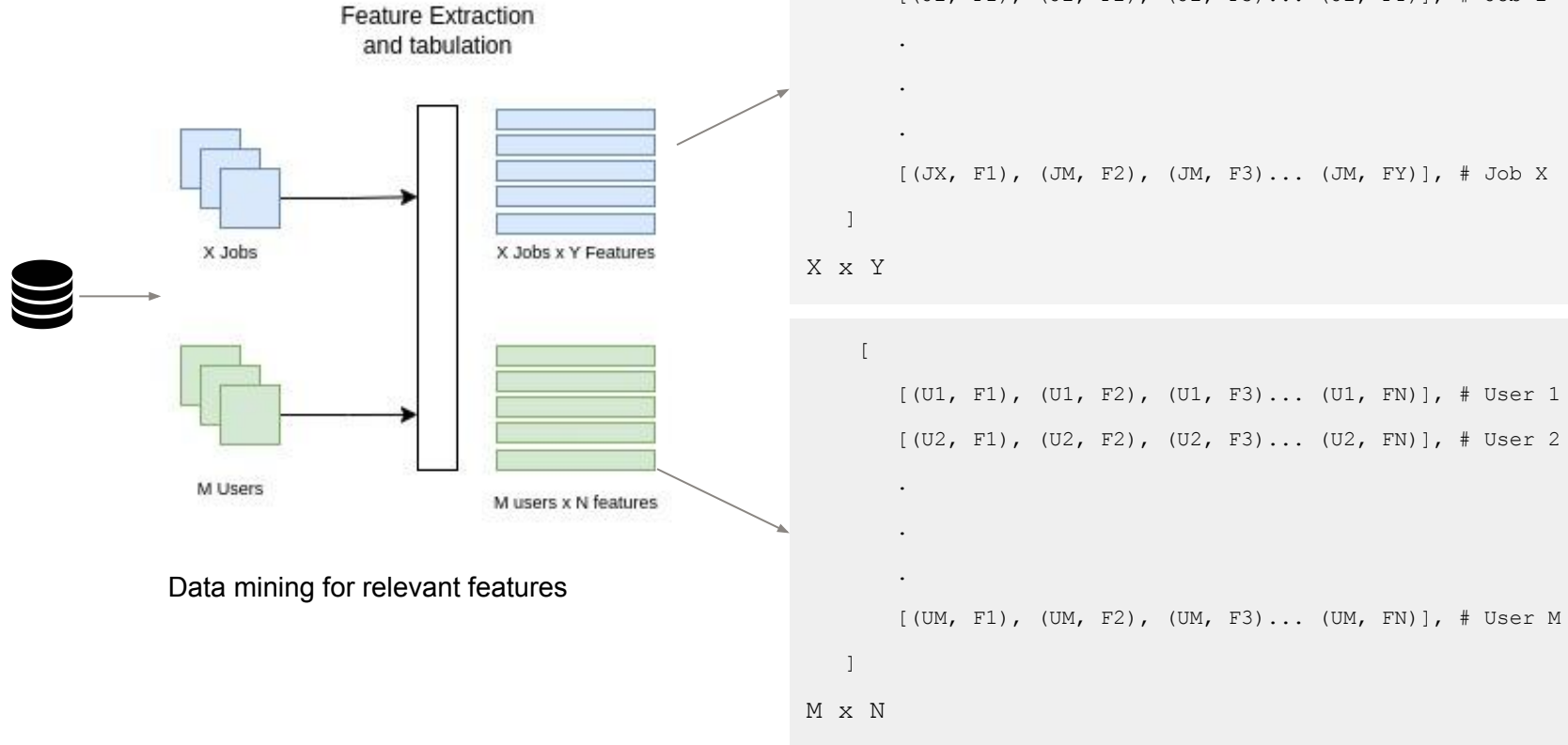
A distilled version of table schema for storing each candidate and job openings

CandidateDetails	
FIELD NAME	DATA TYPE
CandidateId (Primary Key)	int identity
Name	nvarchar
Aadhaar	nvarchar
PanNumber	nvarchar
DateOfBirth	datetime
Gender	tinyint
SocialCatId (Foreign Key)	int
Address	nvarchar
PreferredJobLocation	int
Religion	nvarchar
MaritalStatus	tinyint
DisabilityTypeId	int
DisabilityPercent	tinyint
Neurodiversity	bit
Phone	nvarchar
MotherTounge	tinyint
Skills	tinyint
DomainSkillId	int
Y4JRecommends	nvarchar
ExpectedSal	nvarchar
Trainings	nvarchar
SkillsCovered	nvarchar

JobOpenings	
FIELD NAME	DATA TYPE
JobId (Primary Key)	int identity
JobTitle	nvarchar
JobDesc	ntext
CompanyId (Foreign Key)	int
JobType	tinyint
Experience	tinyint
JobLocation	int
NoOfVacancies	smallint
DisabilityTypeId	tinyint
startdate	datetime
enddate	datetime
LocationIds	nvarchar
Age	int
Gender	tinyint
Address	nvarchar
SkillSet	nvarchar
Responsibilities	nvarchar
Languages	nvarchar
AnnualSalary	tinyint



Algorithm, Step 1: Database SQL Query and Feature Extraction/Profiling



Algorithm, Step 2: Weighted Similarity Score Calculation for Users and Jobs

Given Two users, U_m and U_n

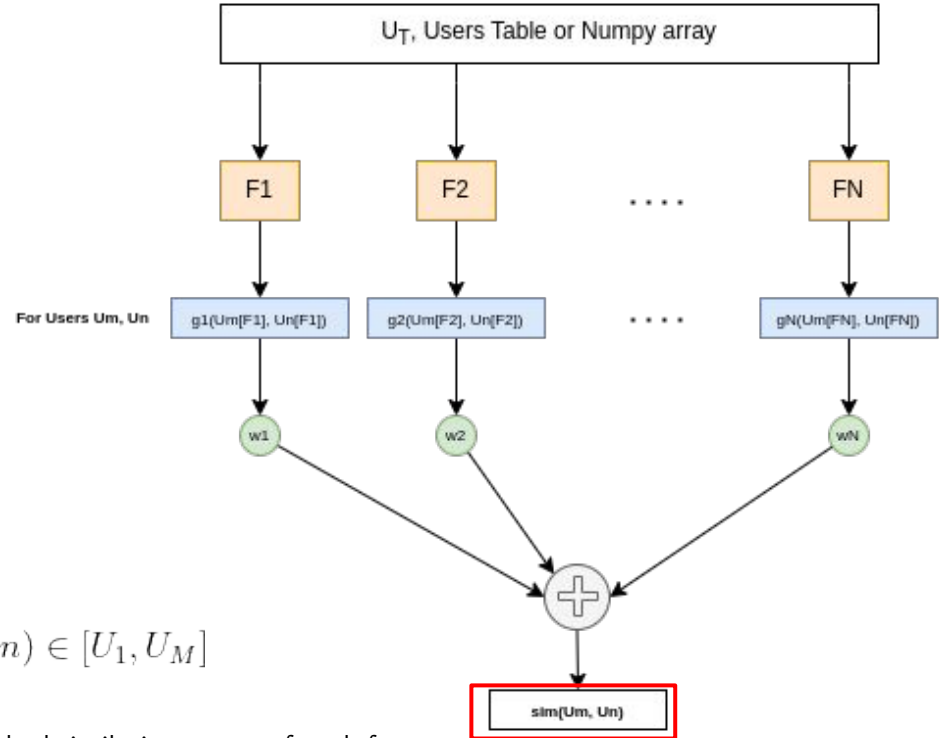
```
[  
  [(U1, F1), (U1, F2), (U1, F3)... (U1, FN)], # User 1  
  [(U2, F1), (U2, F2), (U2, F3)... (U2, FN)], # User 2  
  .  
  .  
  .  
  [(Um, F1), (Um, F2), (Um, F3)... (Um, FN)], # User M  
]  
M x N
```

$W = \{w_1, w_2, \dots, w_N\}$ defined heuristically.

Similarity Functions = $G = \{g_1, g_2, \dots, g_N\}$, feature understanding

$$\text{sim}(U_m, U_n) = \sum_{f \in F} w_f * g(U_T[U_m][f], U_T[U_n][f]) \forall (m, n) \in [U_1, U_M]$$

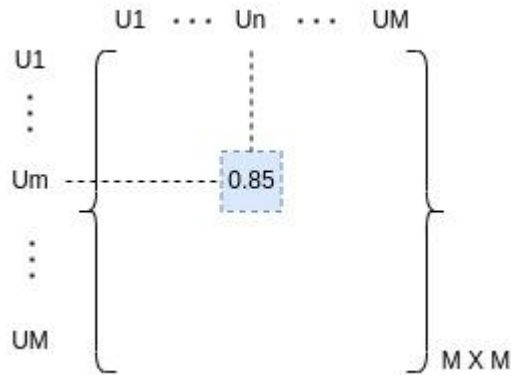
Combined similarity value = weighted sum of individual similarity scores of each features



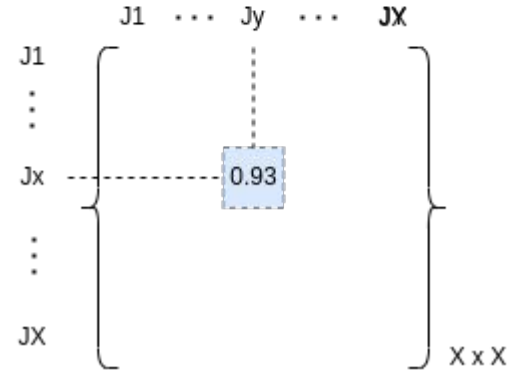
Algorithm, Step 2: Weighted Similarity Score Calculation

$$\text{sim}(U_m, U_n) = \sum_{f \in F} w_f * g(U_T[U_m][f], U_T[U_n][f]) \forall (m, n) \in [U_1, U_M]$$

$$\text{sim}(J_x, J_y) = \sum_{f \in F} w_f * g(J_T[J_x][f], J_T[J_y][f]) \forall (x, y) \in [J_1, J_X]$$



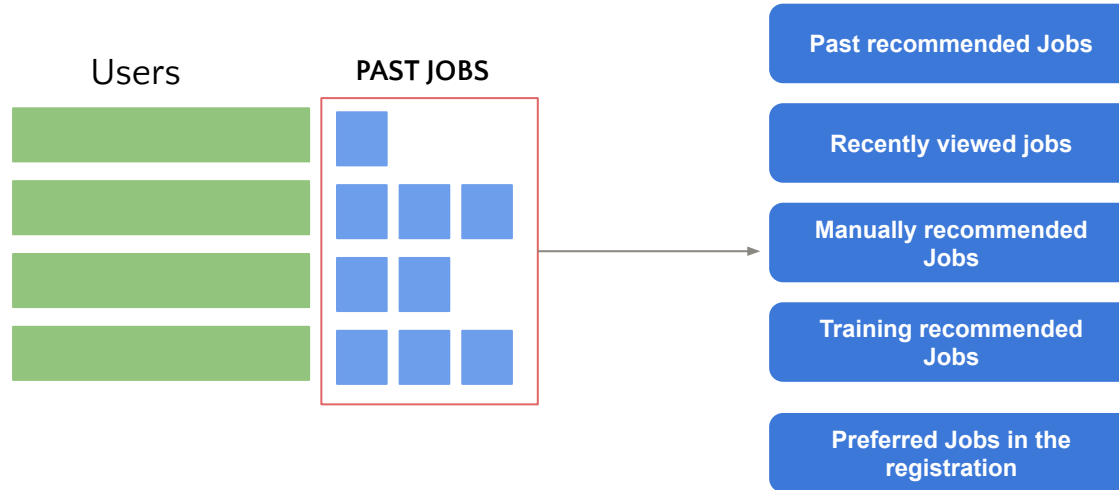
User - User Similarity Matrix



Job - Job Similarity Matrix

Cold Starting the recommendation engine

Before any recommendation occurs,



Each user is associated with a number of past jobs

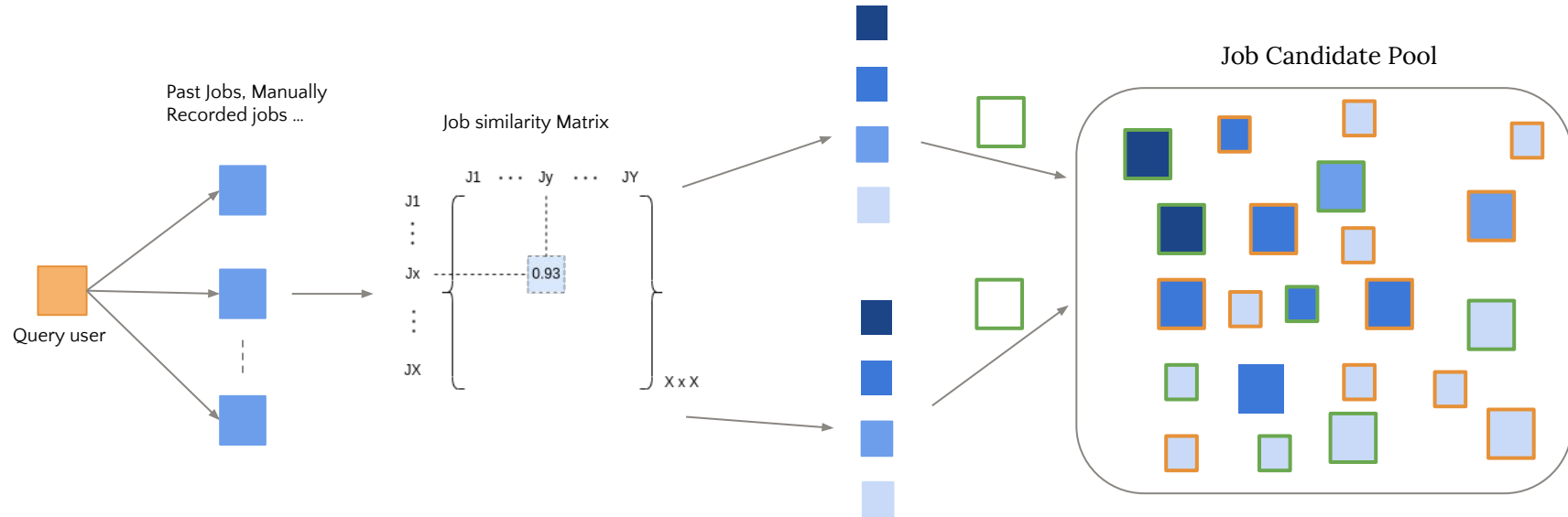
Algorithm, Step 3: Job Candidates Pooling using User Similarity matrix



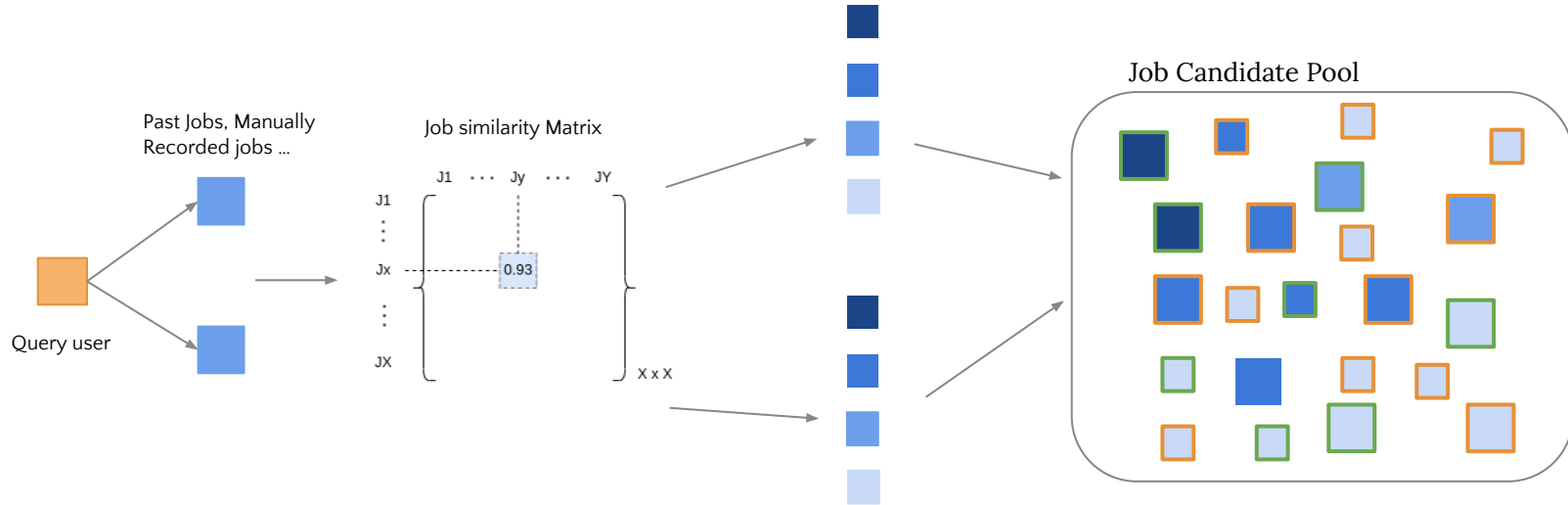
Each of the jobs are associated with their corresponding user similarity score.

```
{  
    s1*[J11, J12, J13 ... J1K],  
    s2*[J21, J22, J23 ... J2K],  
    .  
    .  
    .  
    sp*[JP1, JP2, JP3 ... JPK],  
}
```

Algorithm, Step 4: Job Candidates Pooling using Job Similarity matrix



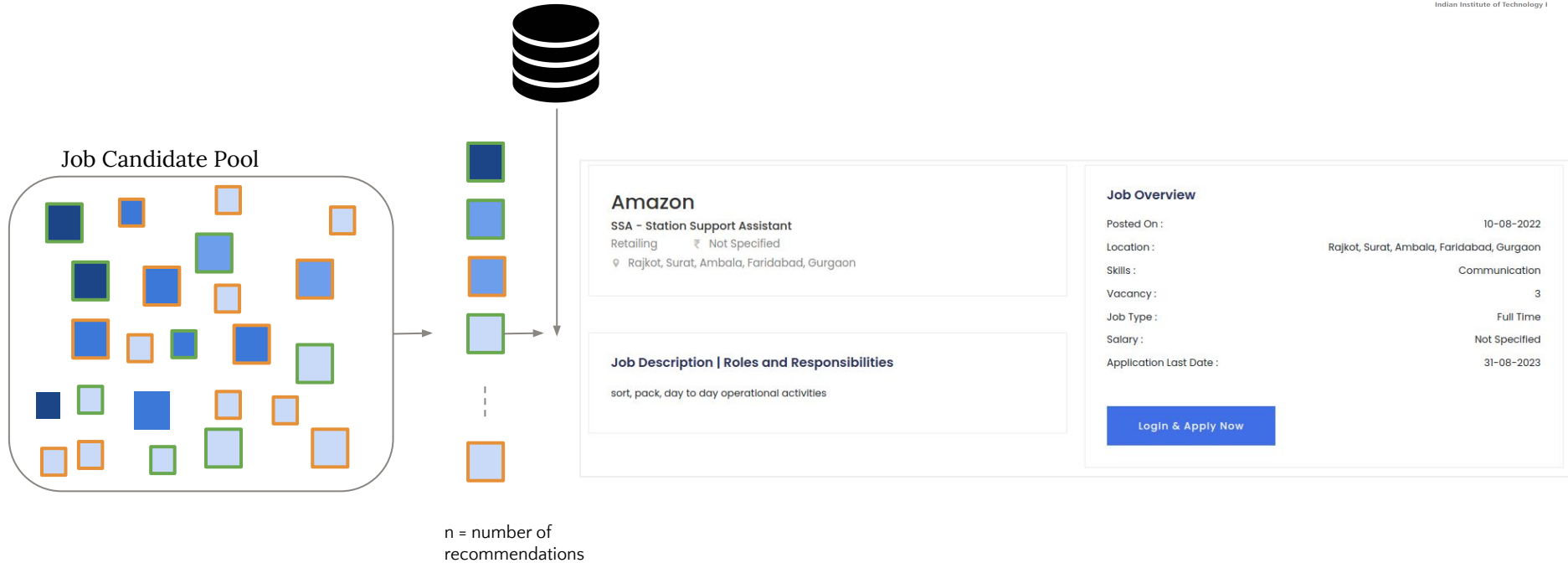
Algorithm, Step 4: Job Candidates Pooling using Job Similarity matrix



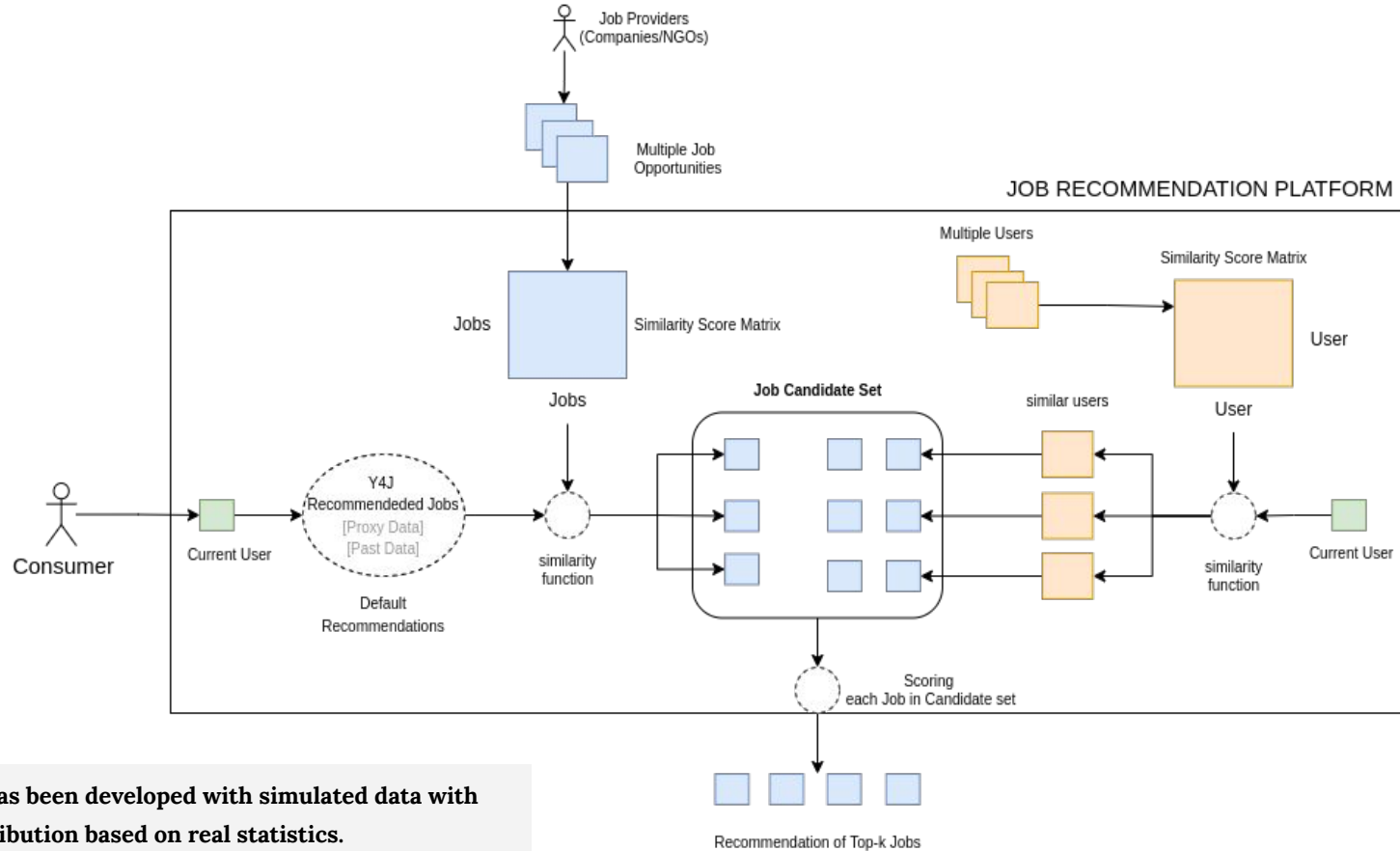
Each jobs are associated with their own corresponding job similarity score with past job that it is associated with

```
{  
  [s11*J11, s12*J12, s13*J13 ... s1P*J1P ]  
  [s21*J21, s22*J22, s23*J23 ... s2P*J2P ],  
  .  
  .  
  .  
  [sk1*JK1, sk2*JP2, sk3*JP3 ... sKP*JKP ],  
}
```

Algorithm, Step 5: Top-N recommendation

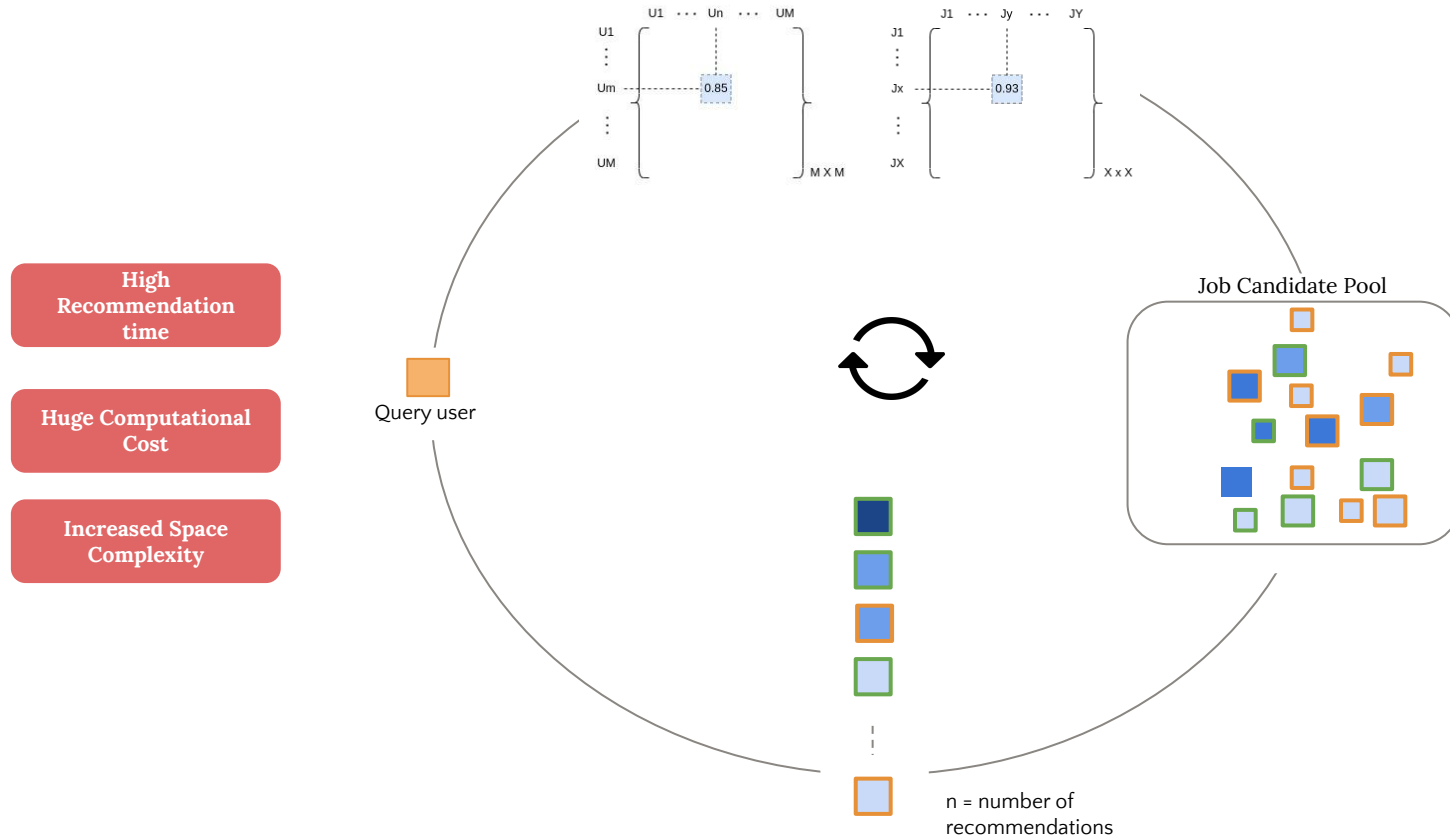


System diagram for Job Recommendation System

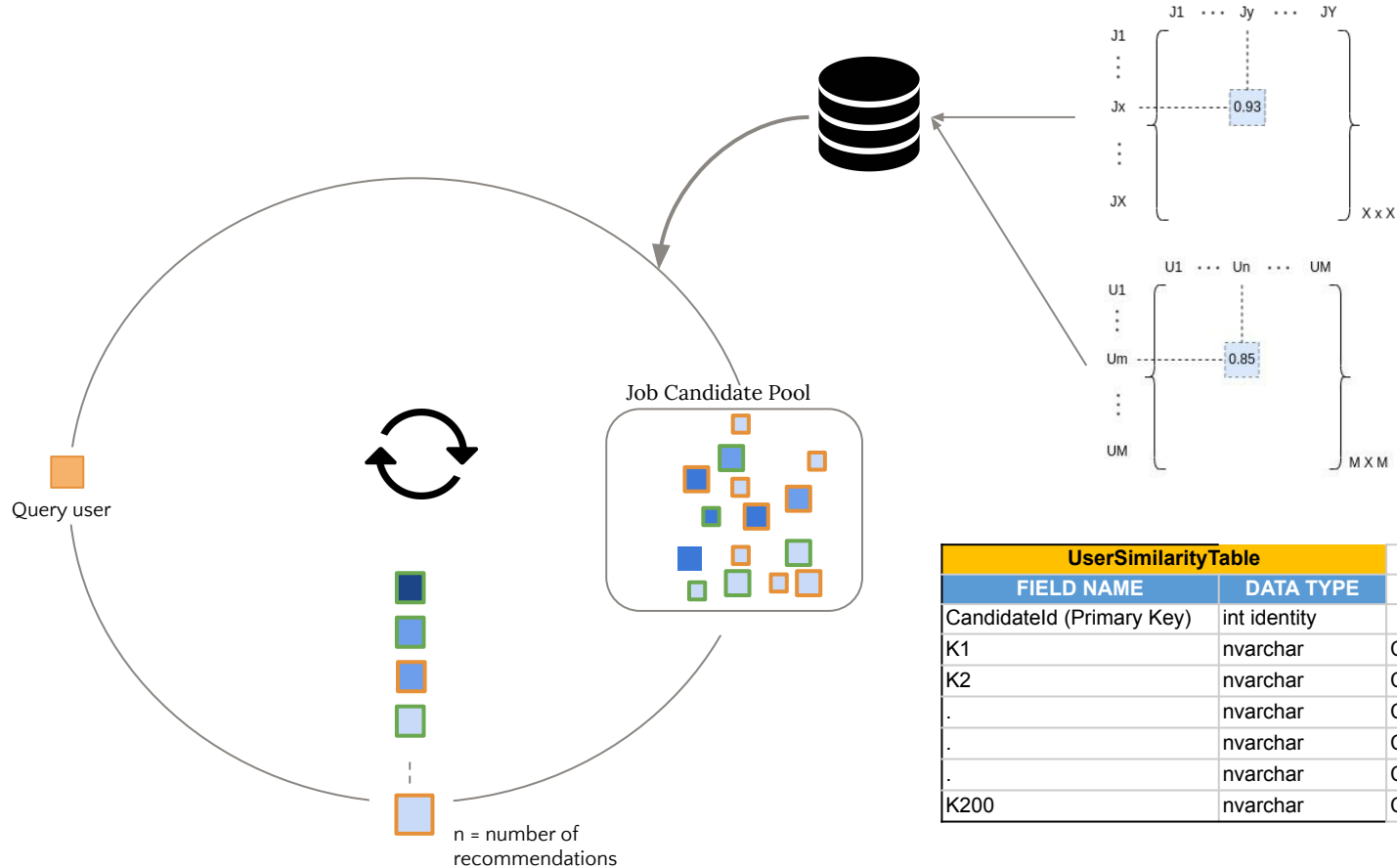


Current system has been developed with simulated data with Multinomial distribution based on real statistics.

System Constraints: Computation of Similarity Matrices



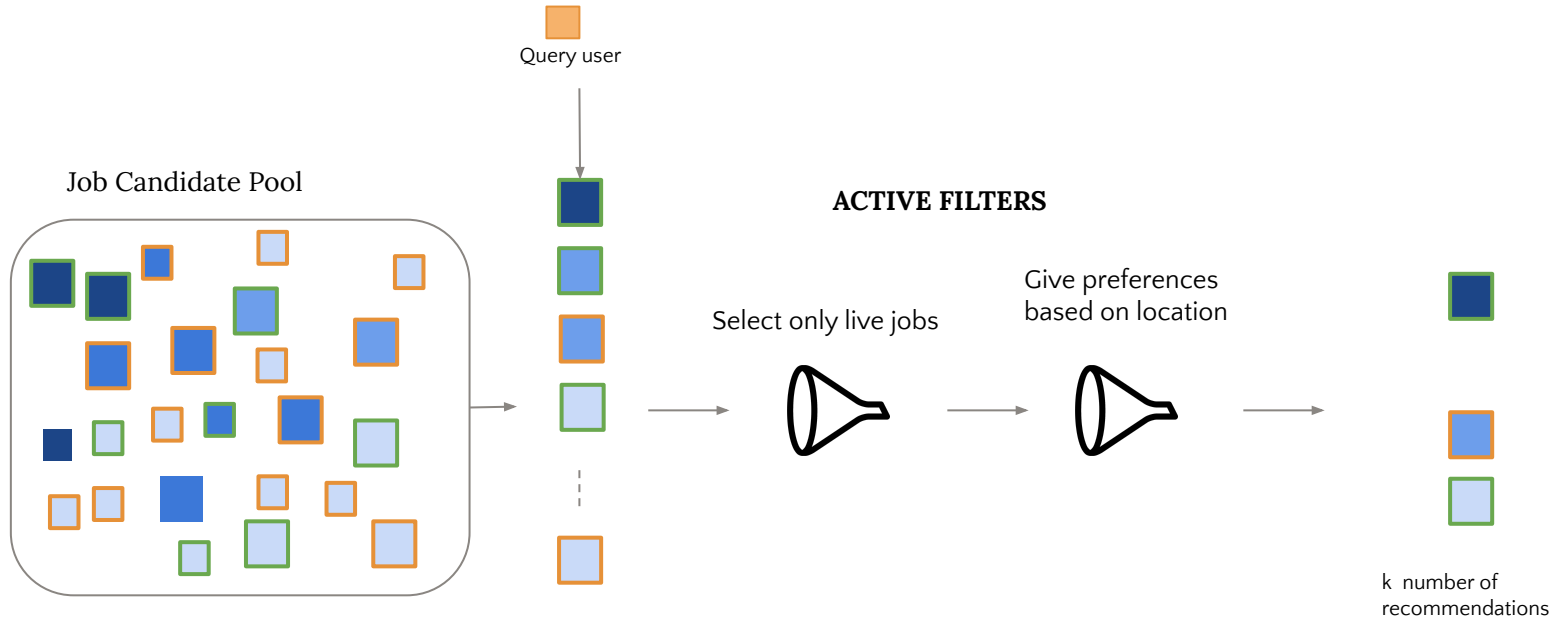
System Constraints: Computation of Similarity Matrices



UserSimilarityTable		
FIELD NAME	DATA TYPE	
CandidateId (Primary Key)	int	identity
K1	nvarchar	CandidateId, score
K2	nvarchar	CandidateId, score
.	nvarchar	CandidateId, score
.	nvarchar	CandidateId, score
.	nvarchar	CandidateId, score
K200	nvarchar	CandidateId, score

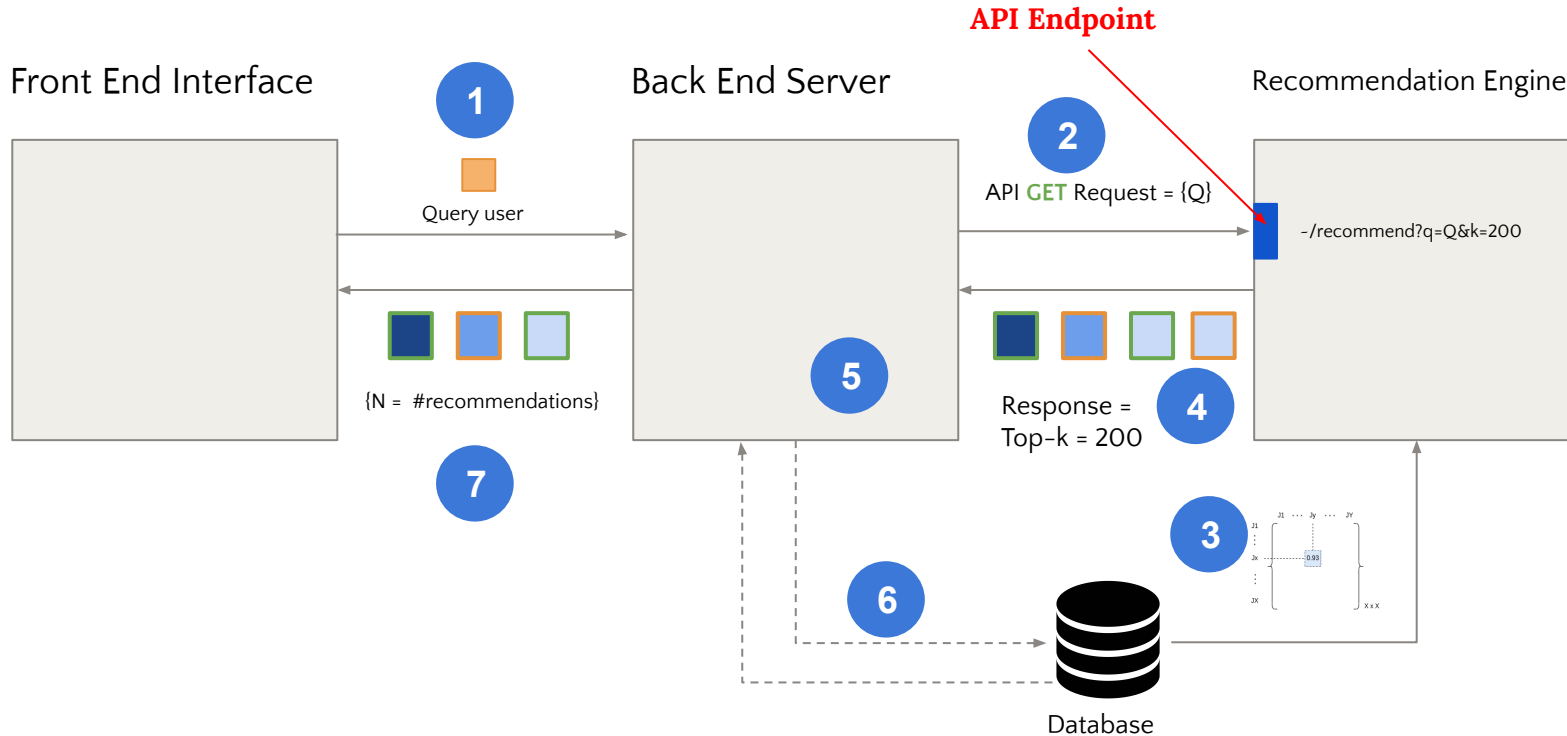


All the jobs that are in the candidate pool wont necessary be recommended because of the **expiration of the jobs**



System Integration with existing system: API Endpoints

How do we integrate the recommendation engine to the existing system?



In the coming days: evaluation of recommendations

Human Evaluation: Real user and jobs data

Coverage = $(n / N) * 100$,

n = Recommended Items

N = Total Items

Confusion Matrix

Given,

Total Jobs = X

Job candidate Pool = (JobID, Job Scores)

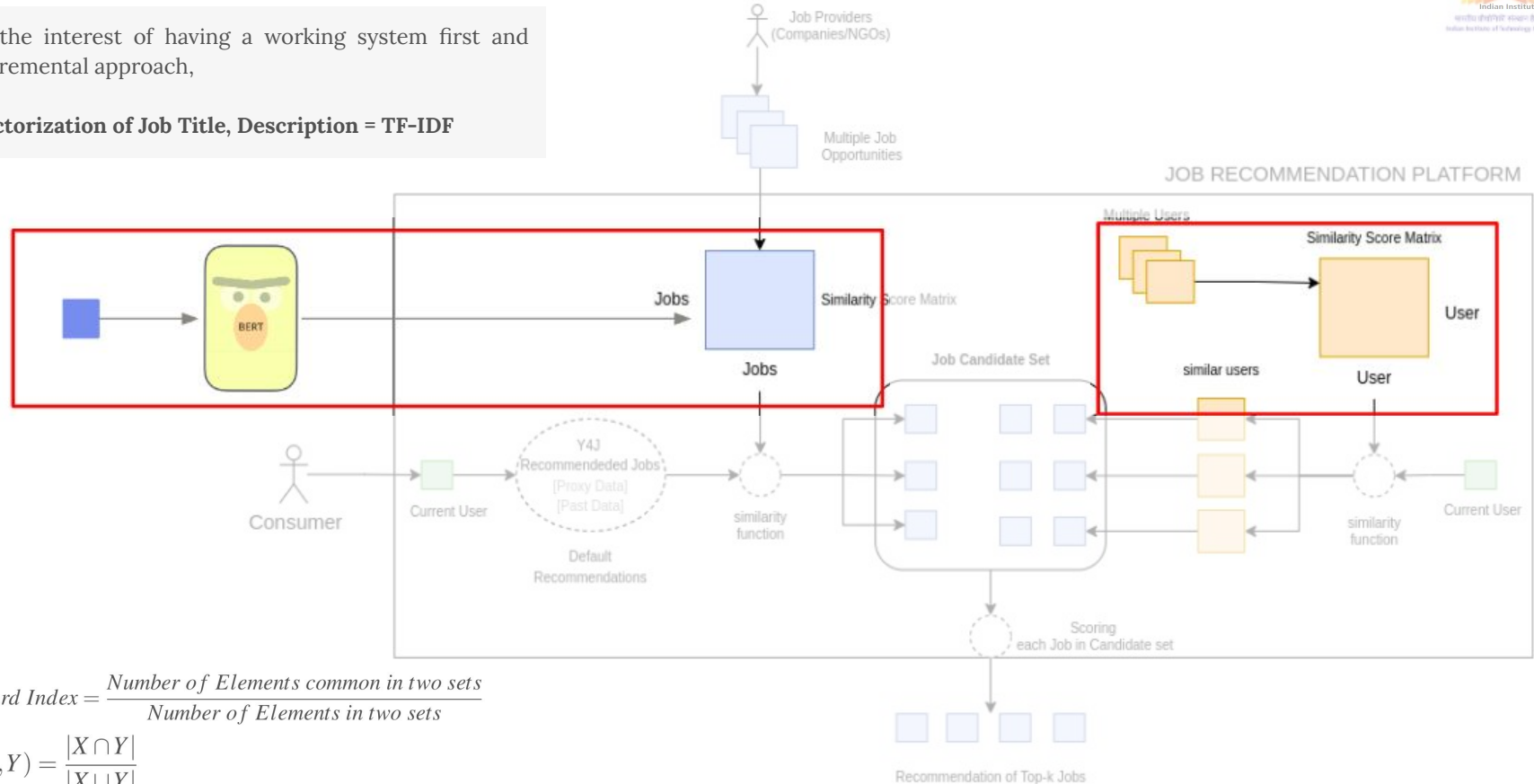
Threshold score = $\max(\text{Job Scores}) * \text{threshold value (any of 0.2, 0.3, 0.4, 0.5)}$

	Suggested to user	Not Suggested to user
Recommended by JRS	X''	$X' \text{ (Job Candidates)} - X''$
Not Recommended by JRS	0	$X - X' \text{ (Job Candidates)}$

Further Improvements: **BERT Embeddings** for better generalisations

In the interest of having a working system first and incremental approach,

Vectorization of Job Title, Description = TF-IDF



$$\text{Jaccard Index} = \frac{\text{Number of Elements common in two sets}}{\text{Number of Elements in two sets}}$$

$$J(X,Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

Further Improvements: Multilingual Aspects of platform

Amazon

SSA - Station Support Assistant

Retailing ₹ Not Specified

📍 Rajkot, Surat, Ambala, Faridabad, Gurgaon

Job Description | Roles and Responsibilities

sort, pack, day to day operational activities

Job Overview

Posted On : 10-08-2022

Location : Rajkot, Surat, Ambala, Faridabad, Gurgaon

Skills : Communication

Vacancy : 3

Job Type : Full Time

Salary : Not Specified

Application Last Date : 31-08-2023

Login & Apply Now

Common Language,
Hindi

KFC

Front Office

Food & Packaged Food ₹ Not Specified

📍 Agra, Ahmednagar, Bangalore

Job Description | Roles and Responsibilities

Participating in training activities to improve service skills and knowledge of brand promotions and products

Job Overview

Posted On : 27-08-2022

Location : Agra, Ahmednagar, Bangalore

Skills : Communication, Computer Basics, Customer Service

Vacancy : 1

Job Type : Full Time

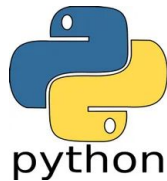
Salary : Not Specified

Application Last Date : 31-08-2023

Login & Apply Now

Local Languages

Backend Recommendation Engine Development Stack



Running tests



Logging system progress

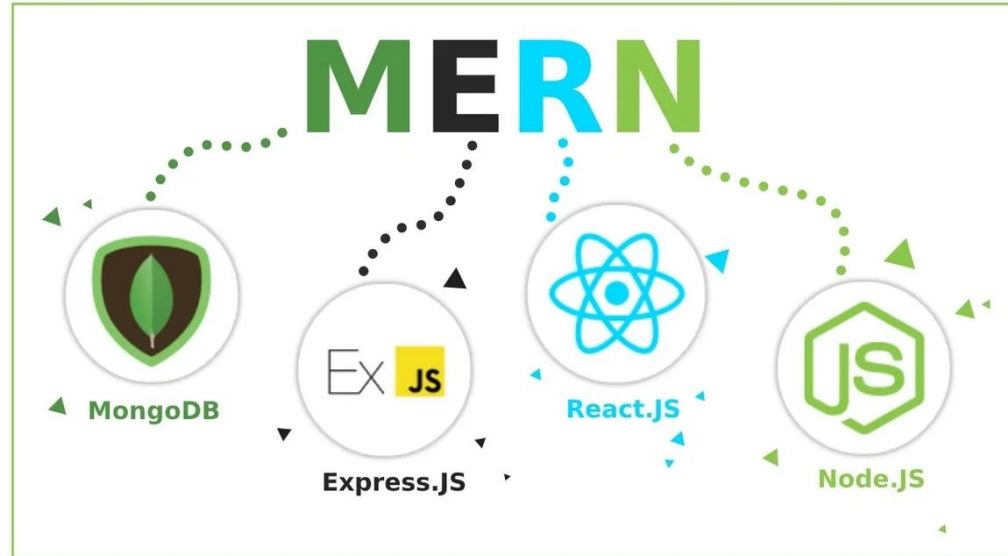


Exposing endpoints



Environment management
and deployment

Single Page Web Application (SPA) Development Stack





Thanks!

Any **questions** ?

You can find me at:

- cs21mtech16001@iith.ac.in

Academic and Professional Updates at:

- <https://shresthakamal.github.io/home>