



50%

CV ANALYSIS AND OPTIMIZING THE RECRUITMENT PROCESS IN THE IT INDUSTRY USING MACHINE LEARNING

PROGRESS PRESENTATION I



Supervisor: Dr. Anuradha Karunasena
Co-Supervisor: Dr. Lakmini Abeywardhana

TEAM MEMBERS

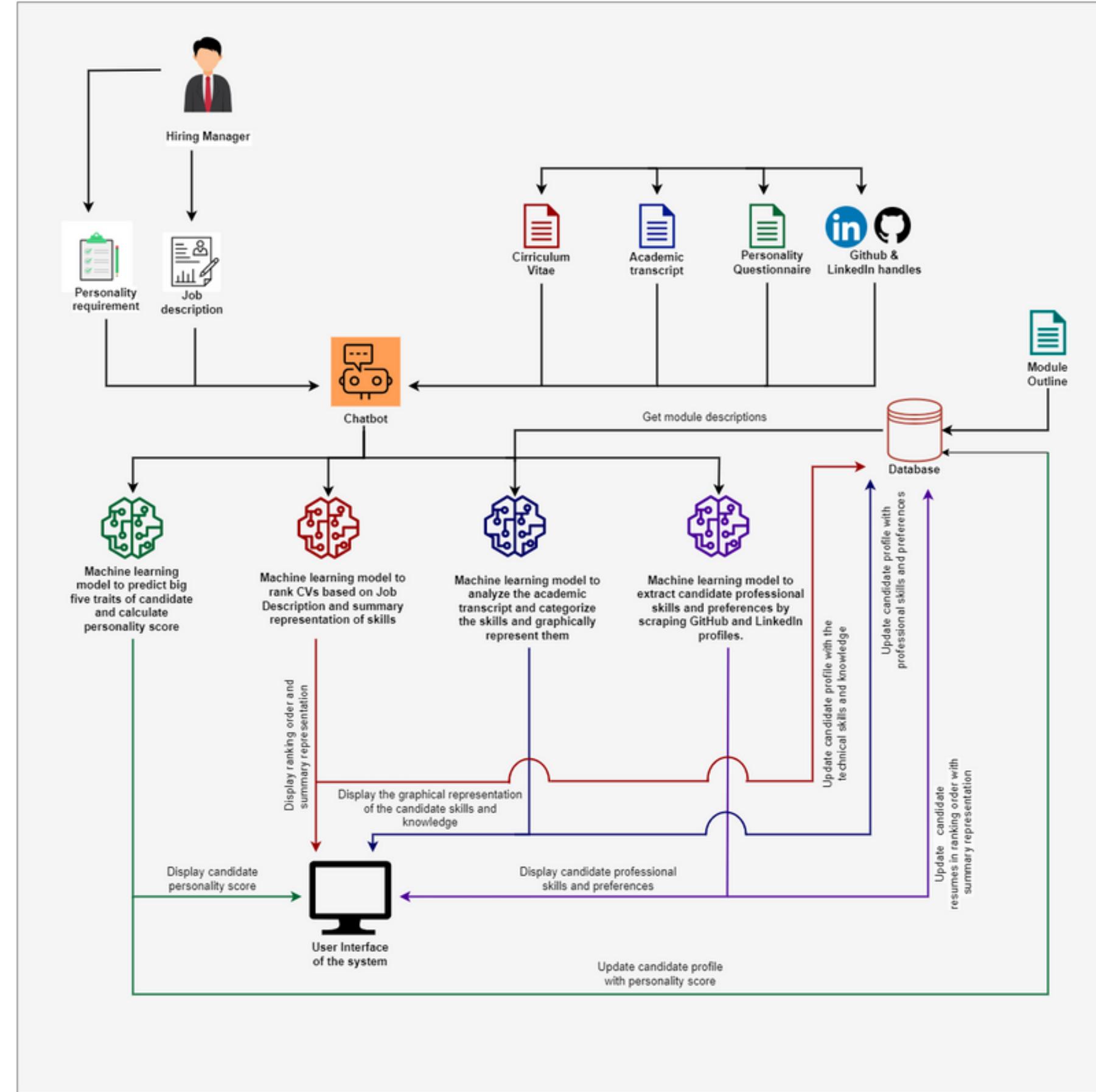
- | | |
|---|--|
| 1 | IT20207854 De Silva M. (Leader) |
| 2 | IT20231200 Zoysa E.S. |
| 3 | IT20203726 Maldeniya M.M.D. |
| 4 | IT20216900 De Silva S.R. |

INTRODUCTION

- Recruitment is a critical process for any organization that aims to hire the right talent to achieve its goals.
- Traditional recruitment process is often time-consuming, inefficient, and costly.
- Many organizations have started adopting automated recruitment systems that use machine learning, data extraction, and natural language processing techniques.
- These automated systems aim to make the recruitment process faster, more accurate, and cost-effective.
- In this research, we aim to develop an automated recruitment system that evaluates a candidate in all key areas and streamlines the recruitment process



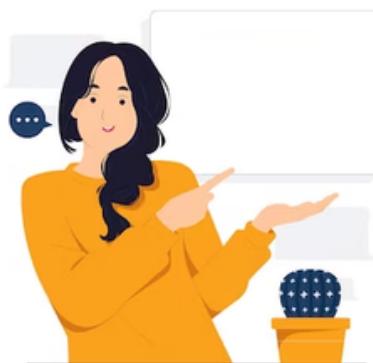
SYSTEM OVERVIEW DIAGRAM



ABILITY OF COMMERCIALIZATION

What makes 'Intellihire' special?

- Intellihire gives the opportunity for organizations to find the best candidate with a high influx of candidate applications for a specific job requirement.
- Analysing professional media platforms and reference documents is less time consuming and productive when comparing with the existing traditional methods.
- In contrast to existing tools, **Intellihire** offers a unique feature, allowing hiring managers to gauge the suitability of candidates for specific roles based on their personality traits.
- Analaysing Academic Transcript is tailored specifically for each university, ensuring accurate and relevant analysis of candidates' skills and knowledge.





IT20203726 | Maldeniya M.M.D.

Bachelor of Science (Hons) in Information Technology Specializing in
Data Science

OPTIMIZING CANDIDATE SELECTION THROUGH CV AND JOB DESCRIPTION MATCHING TECHNIQUES

BACKGROUND STUDY

- Recruitment is a crucial process for any growing company as it directly impacts the organization's success and growth
- In the recruitment process, the first task for any recruiter is to screen the CVs of all the job applicants
- Matching cv with job description helps to identify applicants who possess the required skills, qualifications, and experience necessary to excel in the role
- Ensures that only the most qualified candidates are invited for an interview



PROBLEM DEFINITION

- **The traditional manual resume screening process is laborious and time-consuming, particularly for large organizations handling a high influx of applications.**
- **The existing system lacks the capability to effectively match CVs with well-generated job descriptions,**
- **The unavailability of a suitable approach for providing a summary representation of skill proficiency of a job candidate**



PROOF OF CONCEPT



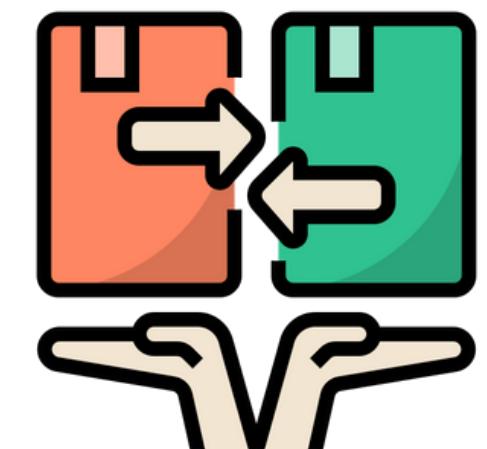
The current implementation allows for the generation of a job description by interacting with a chatbot to gather HR requirements and specifications



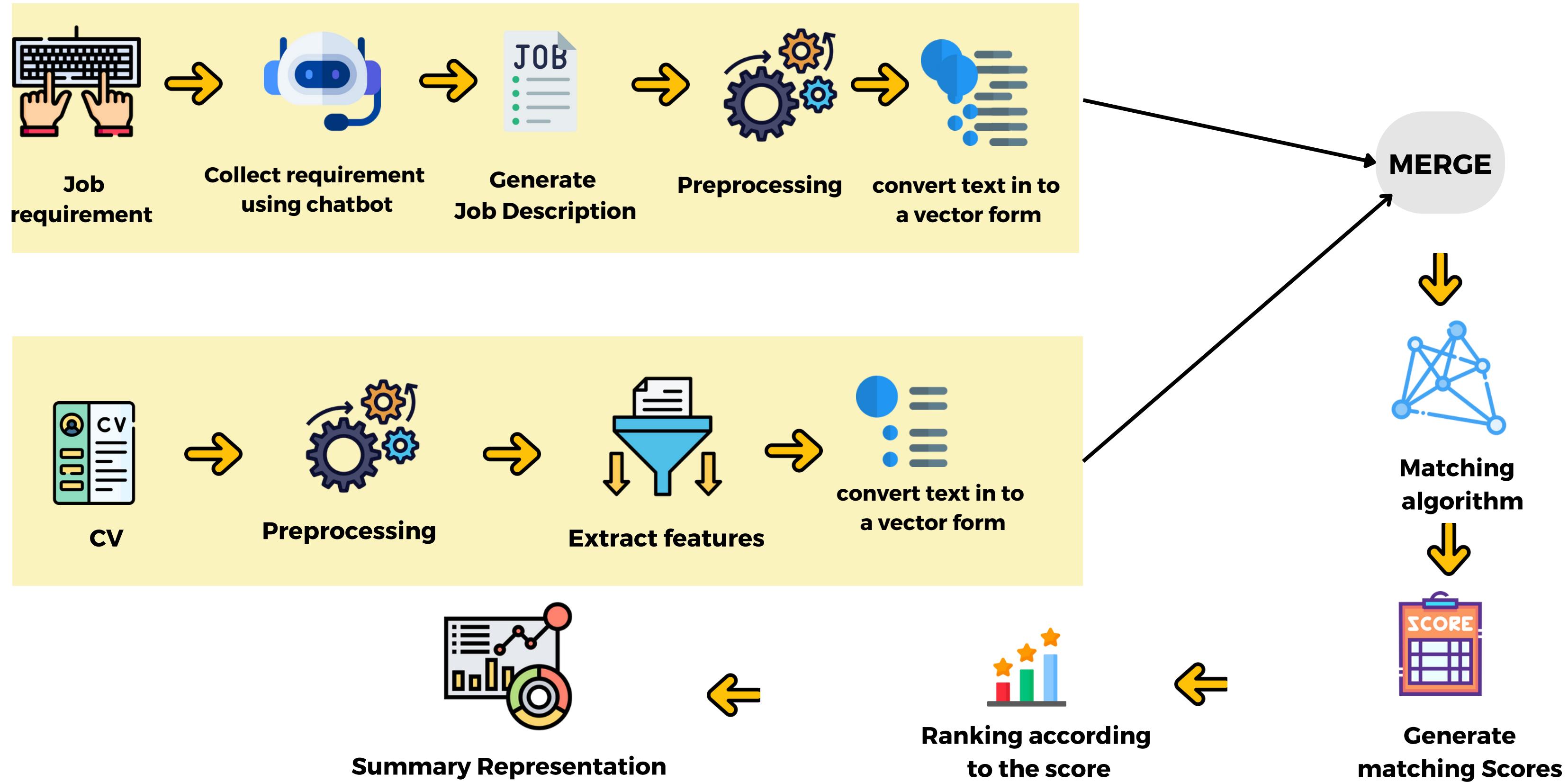
The current implementation effectively calculates the matching percentage between a resume and a job description, presenting the results in a clear and organized manner.



Current implementation will help organizations to compare candidate's skills using skill representation for further analysis



SYSTEM OVERVIEW DIAGRAM



OBJECTIVE

The objective of this research component aims to develop a system that can evaluate and rank job applicants' CVs efficiently based on the job description provided by recruiters.



SUB OBJECTIVE

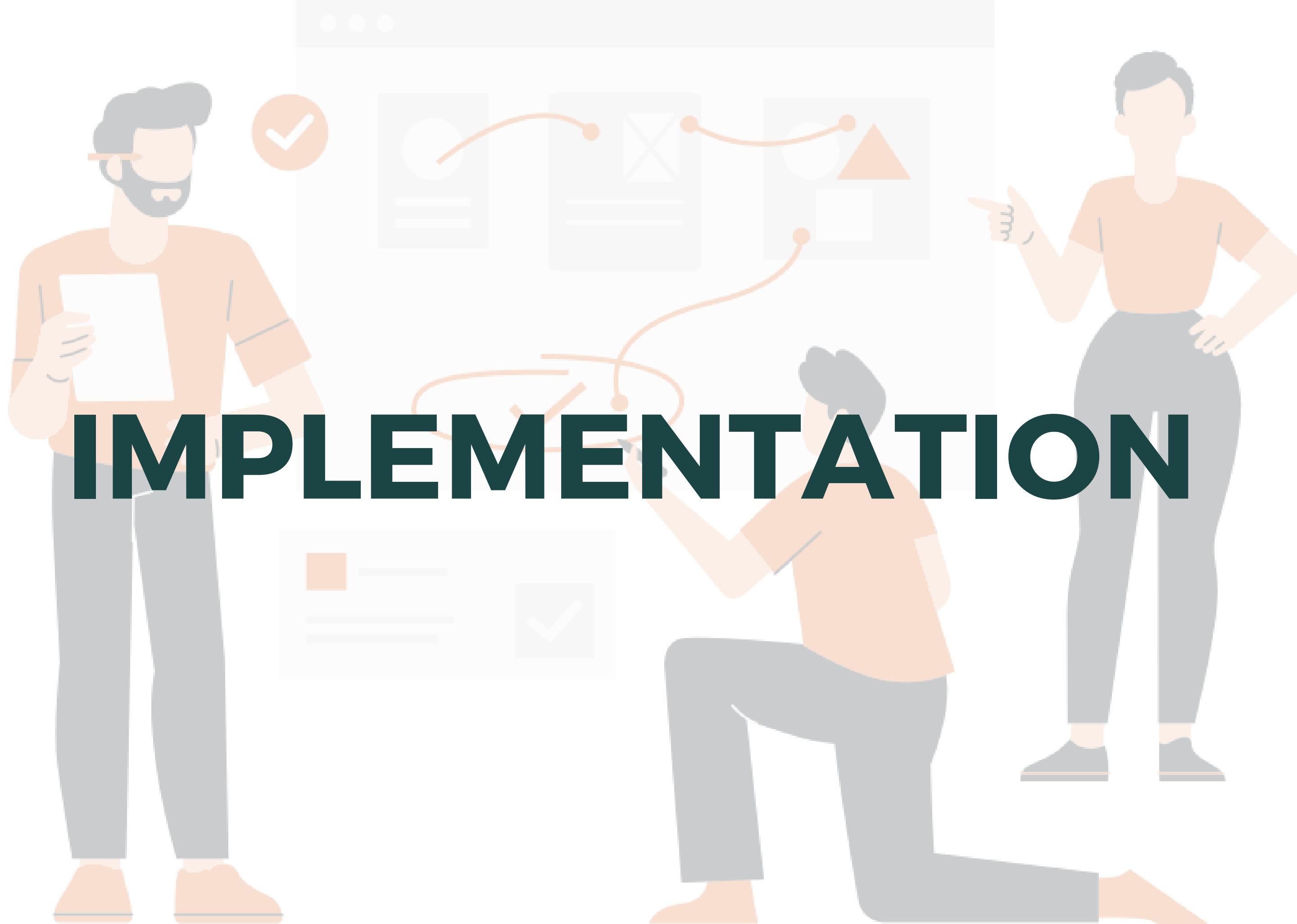
- Generate a proper structure for a job description
- Develop an algorithm to rank candidate CVs according to the Job description in an orderly manner.
- Implement the summary representation of candidates' skill proficiency



PROGRESS



IMPLEMENTATION



IMPLEMENTATION



Created a Rasa chatbot that leverages HR inputs to generate accurate job descriptions in a structure.



Machine learning models are employed to determine the matching percentage between a CV and a job description

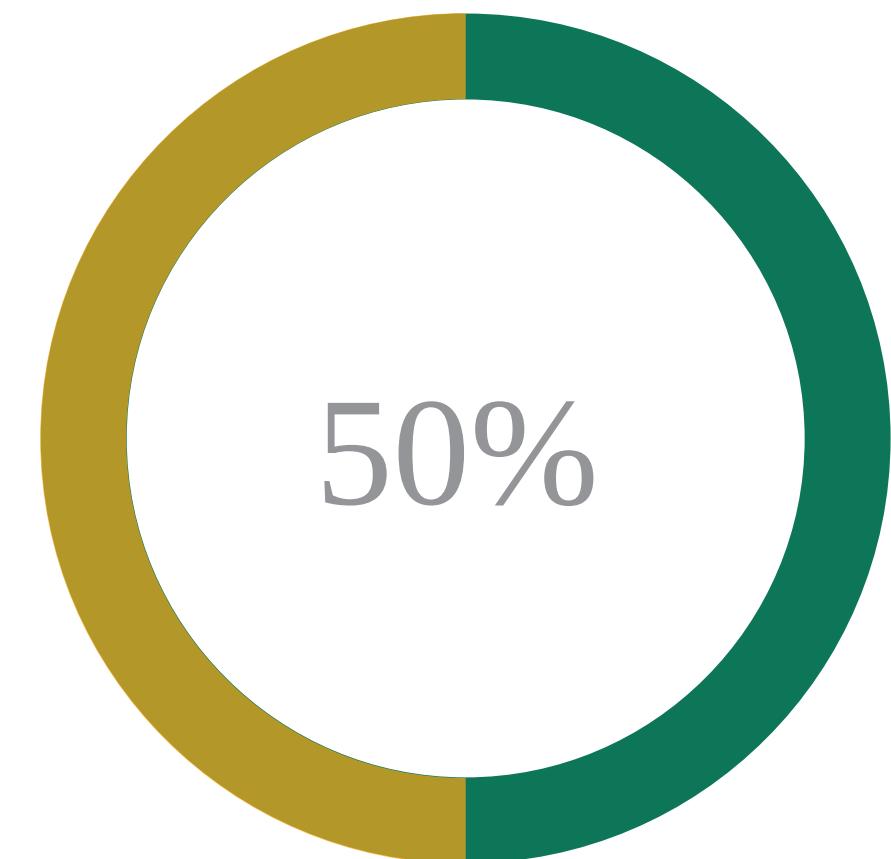
- Used TF-IDF as Text Vectorizer
- Used BoW as Text Vectorizer



Constructed a stacked ensemble model that predicts the matching percentage between a CV and a job description.



Organized resumes in an ordered manner based on their alignment with the job description



FUTURE WORKS



Generate a summery representation of candidate's skills along with the job requirements



Develop RASA chatbot to generate a proper job description using a natural language generation model



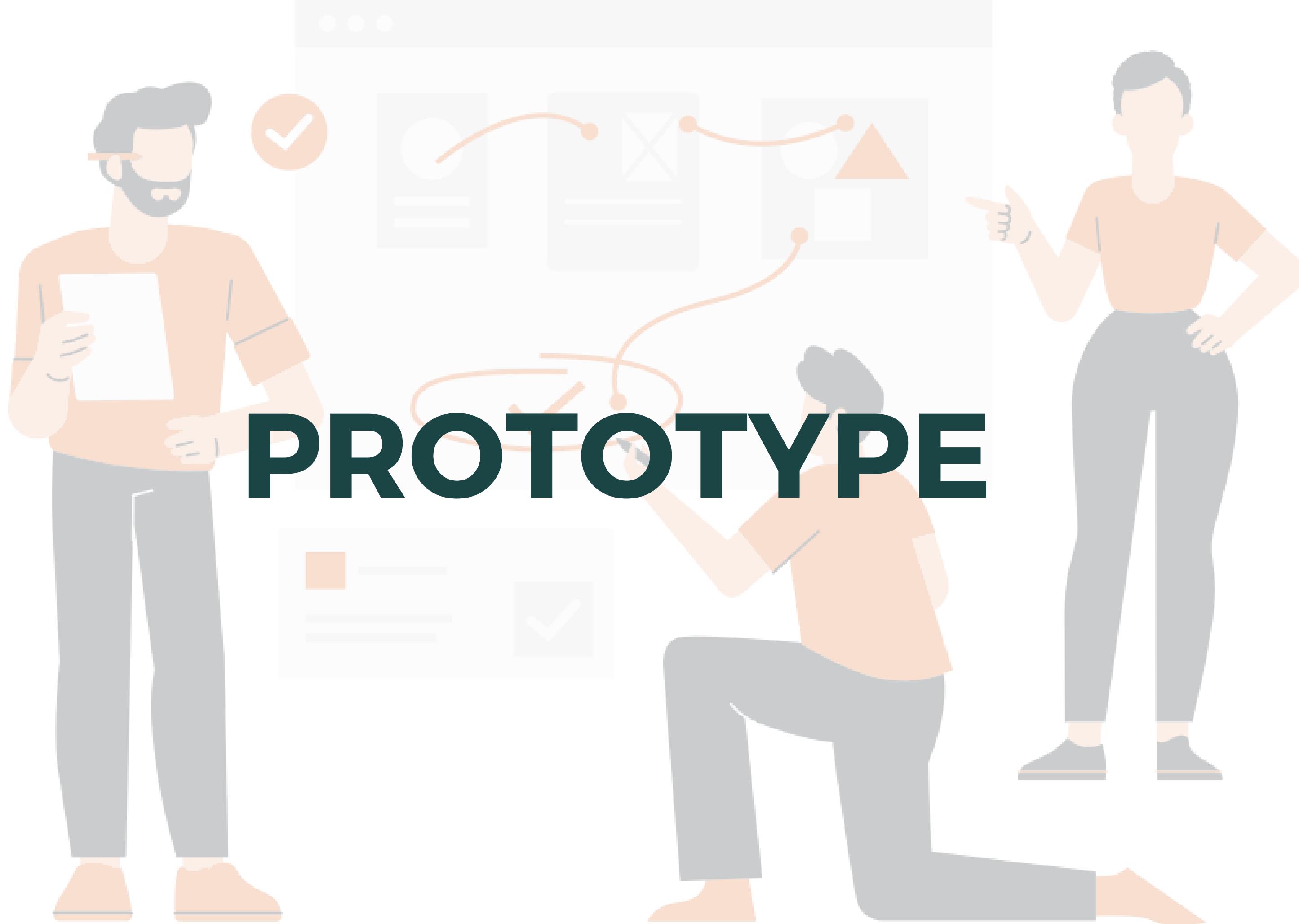
React Front-end design



Testing for functionality and possible improvements



PROTOTYPE



PROTOTYPES

The screenshot shows a web-based prototype for an AI hiring tool. At the top, a dark navigation bar features the brand name "INTELLIHIRE" in white, followed by five menu items: "Home" (highlighted in green), "CV Analysis", "Professional Skills", "Academic Transcript", and "Personality Prediction". Below the navigation is a large, stylized purple background graphic with abstract shapes and a central search icon labeled "search". Overlaid on this graphic is a white rectangular chat interface. The conversation starts with two "Hi" messages, one from each participant. The user then asks, "I want to create a job description". The bot responds with, "what is the job title". The user replies, "Software engineer". At the bottom of the chat window is a text input field containing "Type Here!" and a blue send button with a white arrow icon. In the bottom right corner of the page, there is a small watermark or logo for "Nexcent".

PROTOTYPES

The screenshot shows a prototype of an IntelliHire application. At the top, there is a dark navigation bar with the brand name "INTELLIHIRE" on the left and five menu items: "Home", "CV Analysis", "Professional Skills", "Academic Transcript", and "Personality Prediction". Below the navigation bar is a light purple header section containing the title "Matching Score". The main content area is enclosed in a blue-bordered box. It features two input fields: "Job Description" and "Resume", each with a corresponding "Upload" button. To the right of these fields is a large, rounded rectangular button labeled "Calculate". At the bottom of the page, there is a dark footer bar with links for "Company", "Support", and "Stay up to date". The "Company" section includes links for "About us", "Blog", "Contact us", "Pricing", and "Testimonials". The "Support" section includes links for "Help center", "Terms of service", "Legal", "Privacy policy", and "Status". The "Stay up to date" section features a text input field for "Your email address" with a small "X" icon to its right.

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Matching Score

Job Description

Upload

Resume

Upload

Calculate

Company

About us

Blog

Contact us

Pricing

Testimonials

Support

Help center

Terms of service

Legal

Privacy policy

Status

Stay up to date

Your email address

Copyright © 2023 Landify UI Kit.
All rights reserved.

SLIIT FACULTY OF COMPUTING

PROTOTYPES

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Maching percentage

Resume ID	Candidate Name	Percentage
RE001	Nisal denkotage	70%
RE002	Natasha Perera	69%
RE003	Piyumika Silva	55%
RE004	Lili anthony	51%
RE005	Sachin de Silva	45%
RE006	Dilshan Ranasinghe	25%

Job description

Job Title: Customer Support Representative

Job Description: We are seeking a dedicated and friendly Customer Support Representative to join our team. As a Customer Support Representative, your primary responsibility will be to provide exceptional service and assistance to our customers. You will serve as the first point of contact for inquiries, complaints, and technical support.

Responsibilities:

1. Respond promptly and professionally to customer inquiries via phone, email, or chat.
2. Provide accurate and relevant information about our products and services.
3. Assist customers in troubleshooting and resolving technical issues.
4. Handle customer complaints and ensure timely resolution.
5. Maintain a high level of product knowledge to effectively address customer queries.
6. Document customer interactions, issues, and solutions in the CRM system.

Intellihire
Copyright © 2020 Landify UI Kit.
All rights reserved

Company Support Stay up to date
About us Help center Your email address
Blog Terms of service
Contact us Legal
Pricing Privacy policy
Testimonials Status

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Resume macthing Percentage

Resume ID - RE001
Candidate Name - Natasha Silva



1242 x 546

Company Support Stay up to date
About us Help center Your email address
Blog Terms of service
Contact us Legal
Pricing **ABEBC6** Privacy policy
Testimonials Status

CV MATCHING STACKED MODEL

TF-IDF-based stacked model Accuracy and Matching Percentage

```
[59] ✓ 55.4s
...
Mean Squared Error: 190.54427381057695
Model Score (Training): 0.5460491689235174
Model Score (Test): 0.34396149933579623
```

```

Matching Percentage Prediction: [18.25741104]
Matching Percentage Prediction: [18.41348427]
Matching Percentage Prediction: [14.47569354]
```

BoW-based stacked model Accuracy and Matching Percentage

```

...
Mean Squared Error: 127.70211641100448
Model Score (Training): 0.8085903393991631
Model Score (Test): 0.6259048609615556
```

```

-
Matching Percentage Prediction: [26.95535279]
Matching Percentage Prediction: [34.72938952]
Matching Percentage Prediction: [39.77751138]
```

RASA CHATBOT

```
utter_goodbye:
- text: Bye
utter_iamabot:
- text: I am a bot, powered by Rasa.
utter_ask_first_name:
- text: What is your first name?
utter_ask_last_name:
- text: So {first_name}, what is your last name?
utter_ask_job_title:
- text: What is the job title or position for which you are creating the job description?
utter_ask_job_summary:
- text: Could you please provide a brief summary of the job?
utter_ask_job_responsibilities:
- text: What are the main responsibilities and duties associated with this role?
utter_ask_job_qualifications:
- text: What qualifications and skills are required for this position?
utter_ask_job_education:
- text: What level of education is required? (e.g., Bachelor's degree, Master's degree)
utter_ask_job_employment_type:
- text: Is this a full-time or part-time position?
utter_ask_job_work_schedule:
- text: What will be the typical work schedule or working hours for this role?
```

```
def validate_job_Qualifications(
    self,
    slot_value: Any,
    dispatcher: CollectingDispatcher,
    tracker: Tracker,
    domain: DomainDict,
) -> Dict[Text, Any]:
    """Validate `job_Qualifications` value."""

    print(f"job_Qualifications given = {slot_value} length = {len(slot_value)}")
    if len(slot_value) <= 2:
        dispatcher.utter_message(text=f"That's a very short job_Qualifications. I'm assuming you missed.")
        return {"job_Qualifications": None}
    else:
        return {"job_Qualifications": slot_value}
```

```
# Define CSV file path
csv_file = "slot_values.csv"

# Write slot values to CSV
with open(csv_file, 'a', newline='') as file:
    writer = csv.writer(file)
    writer.writerow([first_name, last_name, job_title, job_summary, jobResponsibilities, jobQualifications, jobEducation, jobEmploymentType, jobWorkSchedule])
return []
```

```
(install_demo3) PS D:\RB> cd y
(install_demo3) PS D:\RB\y> rasa run actions
C:\Users\PC\anaconda3\envs\install_demo3\lib\site-packages\rasa\core\tracker_store.py:1048: MovedIn20Warning:
  SQLAlchemy 2.0. To prevent incompatible upgrades prior to updating applications, ensure requirements files a
1 to show all deprecation warnings. Set environment variable SQLALCHEMY_SILENCE_UBER_WARNING=1 to silence th
  Base: DeclarativeMeta = declarative_base()
C:\Users\PC\anaconda3\envs\install_demo3\lib\site-packages\sanic_cors\extension.py:39: DeprecationWarning: di
  SANIC_VERSION = LooseVersion(sanic_version)
2023-05-24 04:42:56 INFO    rasa_sdk.endpoint - Starting action endpoint server...
2023-05-24 04:42:56 INFO    rasa_sdk.executor - Registered function for 'action_print_slots'.
2023-05-24 04:42:56 INFO    rasa_sdk.executor - Registered function for 'validate_name_form'.
2023-05-24 04:42:56 INFO    rasa_sdk.endpoint - Action endpoint is up and running on http://0.0.0.0:5055
```

```
Bot loaded. Type a message and press enter (use '/stop' to exit):
Your input -> hi
Hey! How are you?
Your input -> i want to create a job description
What is your first name?
Your input -> Manushika
So Manushika, what is your last name?
Your input -> a
That's a very short name. I'm assuming you mis-spelled.
So Manushika, what is your last name?
Your input -> 
```

REFERENCE

Dataset- <https://www.kaggle.com/datasets/mukund23/a-perfect-fit>

- [1] Muntaha Mehboob, M.Saad Ali, Saif ul Islam, Syed Sarmad Ali, "Evaluating Automatic CV Shortlisting Tool For Job," in Mohammad Ali Jinnah University International Conference on Computing (MAJICC), China, 2022.
- [2]Dr.K.Satheesh(Professor), A.Jahnavi, L.Iswarya, K.Ayesha, G.Bhanusekhar, K.Hanisha, "Resume Ranking based on Job Description using SpaCy NER model," in International Research Journal of Engineering and Technology (IRJET), Ganguru, India, 2020.
- [3]Rasika Ransing, Akshaya Mohan ,Kailas Mahavarkar, Nikita Bhrugumaharshi Emberi, "Screening and Ranking Resumes using Stacked," 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), p. 6, December 2021.
- [4]YanWang, Yacine Allouache,Christian Joubert, "Analysing CV Corpus for Finding Suitable Candidates," in The Thirteenth International Conference on Advances in Databases, Knowledge, and Data Applications, Valencia, 2021
- [5]"Sri Lanka aiming 200,000 ICT workforce by 2022," Information and Communication Technology Agency of Sri Lanka, Sri Lanka, 2022.

IT20231200 | Zoysa E.S.

**Bachelor of Science (Hons) in Information Technology Specializing in
Data Science**





PROFESSIONAL USER PROFILING USING DIGITAL FOOTPRINTS

BACKGROUND STUDY

- In today's digital age, companies are increasingly relying on automated online job recruitment processes to screen potential candidates quickly and efficiently
- As part of this process, candidate profiles on social media and professional networking sites are being analyzed.
- In order to effectively address the needs and expectations of employers, it is crucial to implement a more efficient and accurate method for evaluating a candidate's skills and abilities.

PROBLEM DEFINITION

- Are CV's reliable? Can we go beyond the CV and asses a candidate?
- Can we proofread canidate's professional preferences before interviewing?
- Can we asses candidate's professional skills in point of view of others?



PROOF OF CONCPET

- Current implementation is able analyse a candidate through **Linkedin, Github** and **Letter of recommendations**.
- Current implementation is able to predict **candidate's desired job category** by analysing Linkedin added skills.
- Current implementation is able to give **sentiment comments** on Letter of recommendations/ service letters/ reference letters



OBJECTIVE

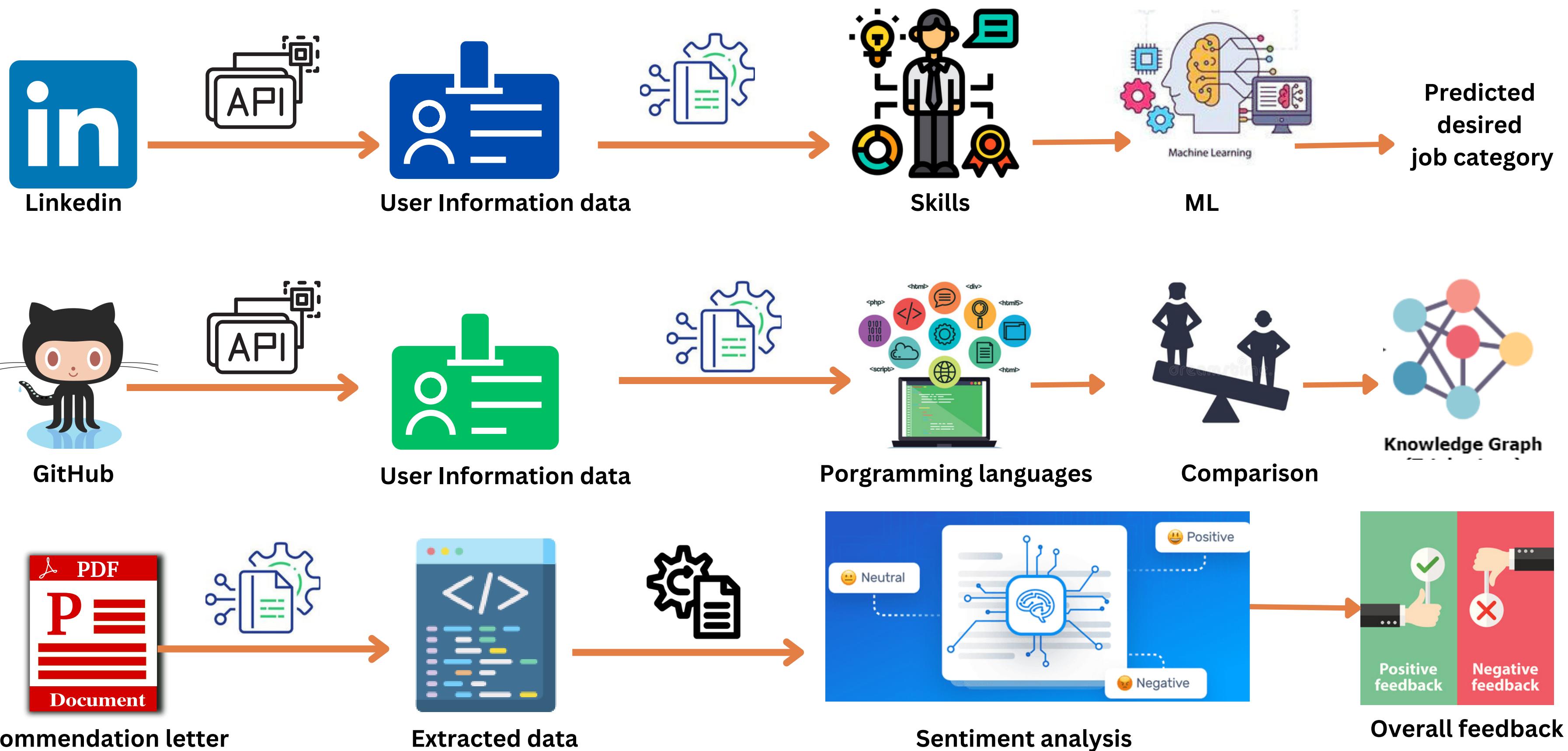
The main objective of this component is to give the opportunity for a HR system to evaluating and confirm candidate's user profile through professional social media platforms like GitHub and LinkedIn user profile and recommendation letter.

SUB OBJECTIVES

- Extracting the content of the candidate's LinkedIn and GitHub user profile and letter of recommendations.
- Analyse programming language proficiency and compare with a peer candidate
- Identify candidate's professional preferences.



SYSTEM OVERVIEW DIAGRAM



IMPLEMENTATION



PROGRESS



CURRENT PROGRESS



Extract Github user profile data from GitHub REST API and extract language proficiency and compare with another user using a knowledge graph



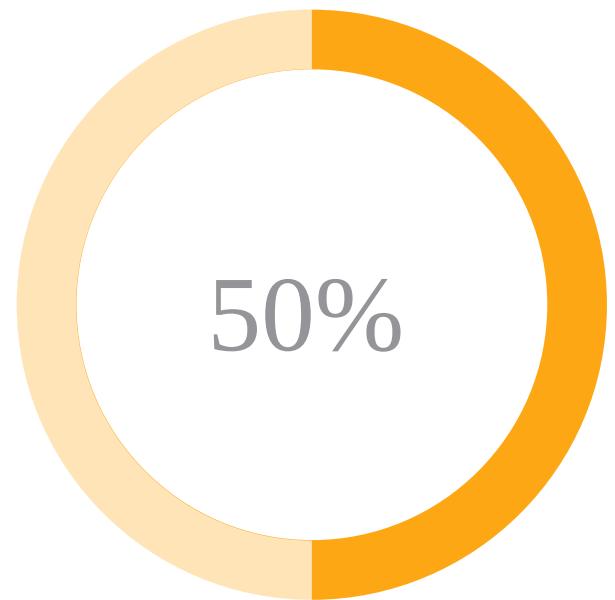
Extract LinkedIn user profile data using API and extract skills



Train a machine learning model to predict candidate's desired job category applying skills extracted from LinkedIn



Train a machine learning model to sentiment analysis on candidate's reference/recommendation letter



EXPECTED FUTURE PROGRESS



Develop sentiment analysis model for letter into a advanced level using new technologies which will deliver more reliable output.



Develop classification model



Frontend Design with React.js



Testing for functionality and possible improvements



Discover Language Proficiency

```
Extract User Profile Data from GitHub passing the Username

from github import Github

# Authenticate with GitHub
access_token = "ghp_TLOmXnjewFn0fGakhPyJKeF2wC0L2Uqjmc"
g = Github(access_token)

# Fetch user information
username = "Maldeniyaa99"
user = g.get_user(username)

# Extract user data
```

Input

```
Overall Language Proficiency (Percentage):
- Java: 76.55%
- JavaScript: 5.38%
- HTML: 10.61%
- CSS: 7.45%
```

Current user language proficiency

```
from github import Github

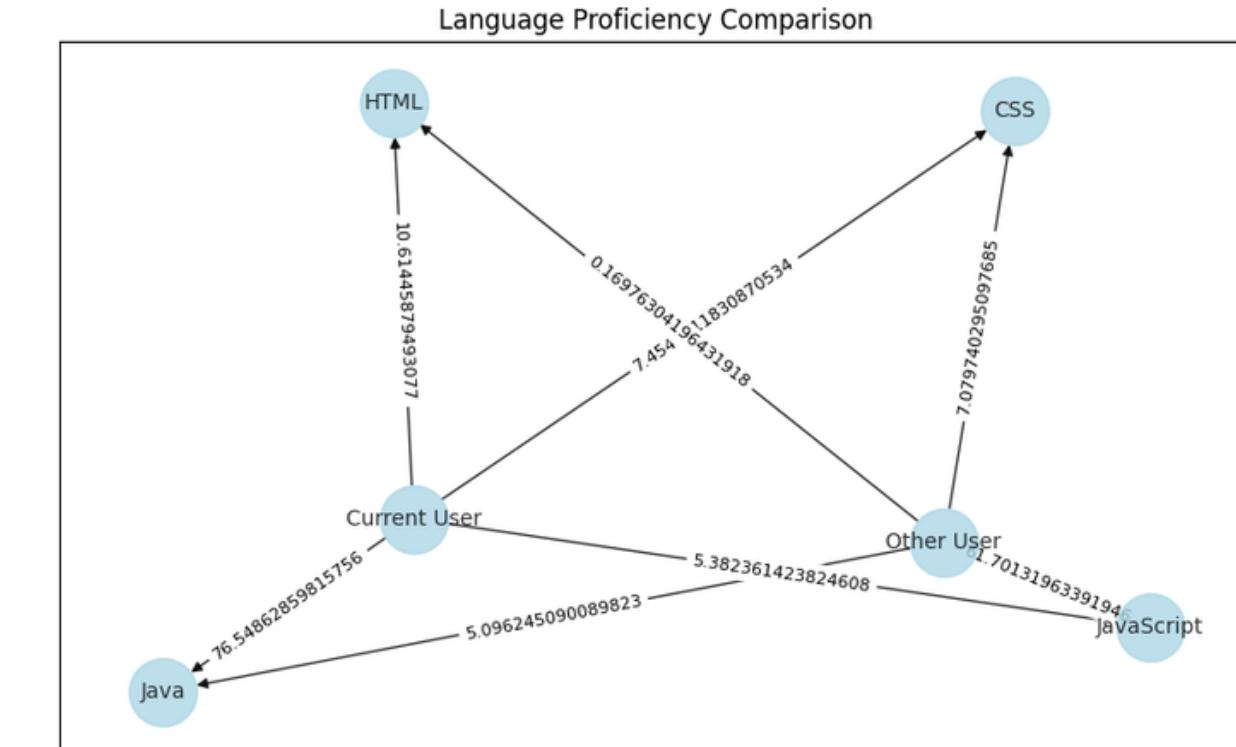
# Authenticate with GitHub
access_token = "ghp_TLOmXnjewFn0fGakhPyJKeF2wC0L2Uqjmc"
g = Github(access_token)

# Fetch user information
current_username = "Maldeniyaa99"
current_user = g.get_user(current_username)
```

Add another user

```
[28]: ...
... Language Proficiency Comparison:
- CSS:
  - Current User: 7.45%
  - Other User: 7.08%
- JavaScript:
  - Current User: 5.38%
  - Other User: 61.76%
- HTML:
  - Current User: 10.61%
  - Other User: 0.17%
- Java:
  - Current User: 76.55%
  - Other User: 5.18%
```

Comparison



Knowledge graph visualization

Predict Desired Job Category



Chanux Bro · 2nd
Software Engineer, Social Media Consultant, Most Popular Socialmedia personality in LK UX awarded

Talks about #teck, #youtube, #srilanka, #chanuxbro, and #education

Sri Lanka · Contact info

33,272 followers · 500+ connections

Sanjula Ranasinghe, Wishwa Rathnaweera, and 7 other mutual connections

+ Follow Message More

Connect if you know each other Connect

LinkedIn user profile

```
import requests
api_endpoint = 'https://nubela.co/proxycurl/api/v2/linkedin'
linkedin_profile_url = 'https://www.linkedin.com/in/chanuxbro/'
api_key = 'wL50ibTkp7AagYs0MzcUGA'
headers = {'Authorization': 'Bearer ' + api_key}

response = requests.get(api_endpoint,
                        params={'url': linkedin_profile_url, 'skills': 'include'},
                        headers=headers)
```

Input

```
Skills = profile_data['skills']
Skills
[0]
['Linux',
 'HTML',
 'Image Processing',
 'Artificial Intelligence',
 'Big Data Analytics',
 'Project Management',
 'E-commerce SEO',
 'Entrepreneurship',
 'Social Entrepreneurship',
 'Mobile Application Development',
 'Business Strategy',
 'Consulting',
 'Cyber-security',
 'Information Security Consultancy',
 'Online Marketing Analysis',
 'PHP',
 'IT Consulting',
 'Social Media Consulting',
 'E-commerce Consulting',
 'Start-up Consulting']
```

```
skills = list_string
print(model.predict(fitted_vectorizer.transform([skills])))

[0]
['Information Technology']
```

Output

```
#read csv file
df_dataset = pd.read_csv(path)

[0]
[0s]

#To view the first few rows,
df_dataset.head()

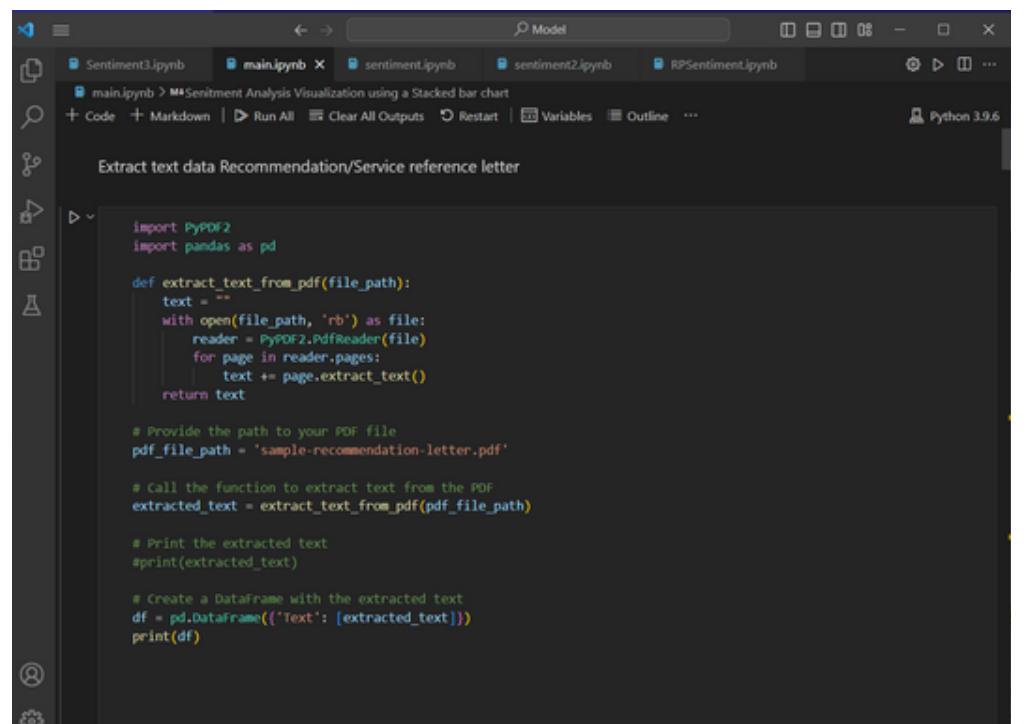
[0]
[0s]

...   index  category          clean_skills
0       1      HR  ['Performance Management', 'Human Resources', ...]
1       2      HR  ['Talent Acquisition', 'Employee Engagement', ...]
2       3      HR  ['Human Resources', 'Recruiting', 'Team Manage...
3       4      HR  []
4       5      HR  ['Team Management', 'Human Resources', 'Employ...
```

Dataset

Extracted data

Sentiment on the letter



The screenshot shows a Jupyter Notebook interface with several tabs at the top: Sentiment3.ipynb, main.ipynb (selected), sentiment.ipynb, sentiment2.ipynb, and RPSentiment.ipynb. The main content area contains the following Python code:

```
import PyPDF2
import pandas as pd

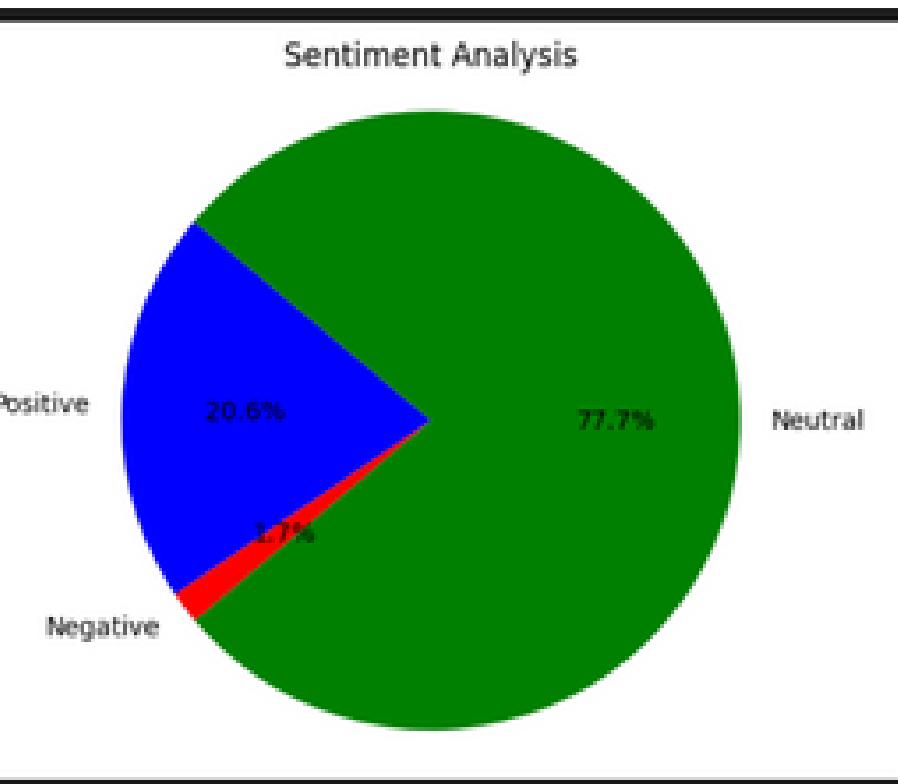
def extract_text_from_pdf(file_path):
    text = ""
    with open(file_path, 'rb') as file:
        reader = PyPDF2.PdfReader(file)
        for page in reader.pages:
            text += page.extract_text()
    return text

# Provide the path to your PDF file
pdf_file_path = 'sample-recommendation-letter.pdf'

# Call the function to extract text from the PDF
extracted_text = extract_text_from_pdf(pdf_file_path)

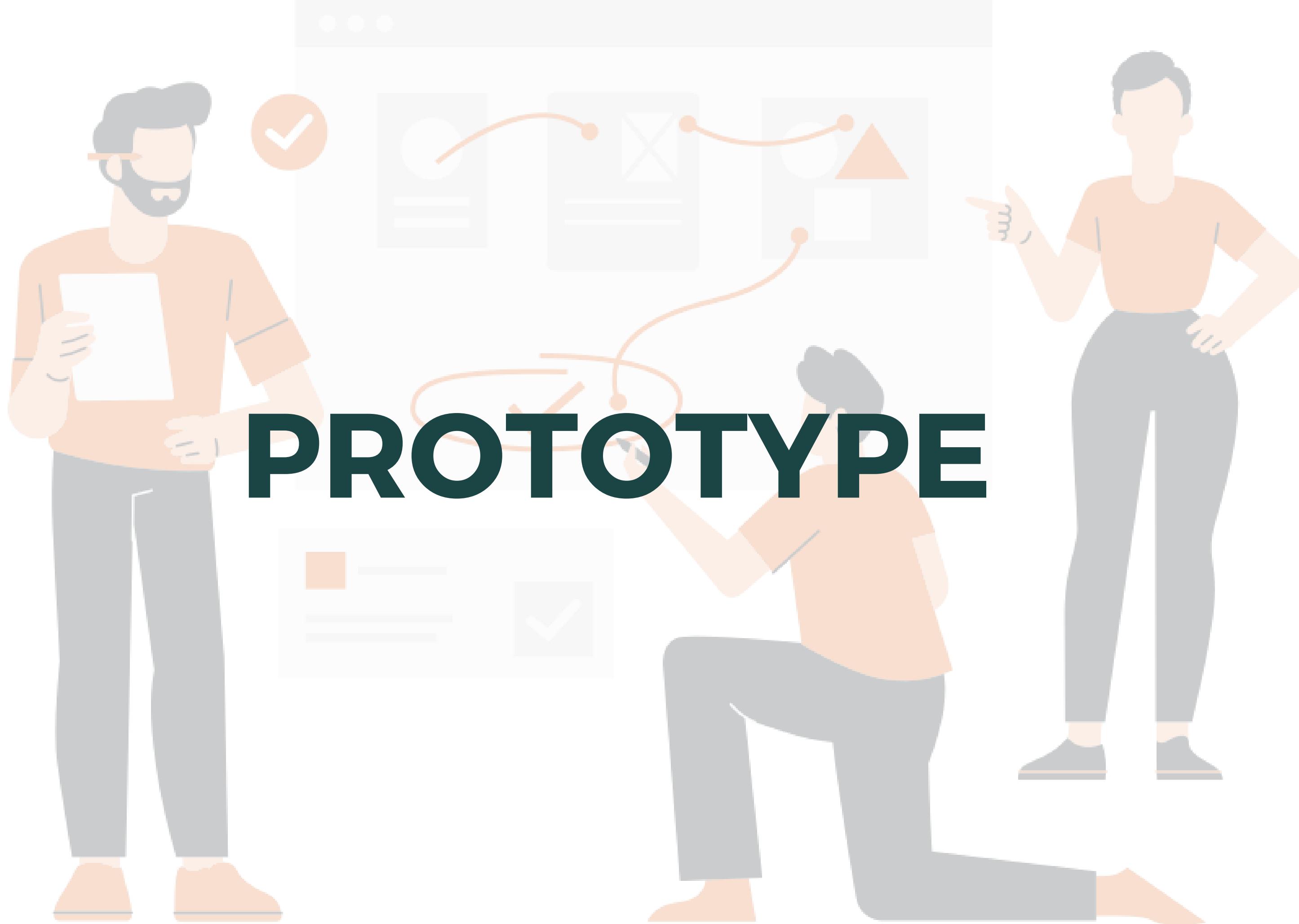
# Print the extracted text
#print(extracted_text)

# Create a DataFrame with the extracted text
df = pd.DataFrame([{'Text': [extracted_text]}])
print(df)
```



Extract content of the PDF letter

PROTOTYPE



PROTOTYPES

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Predict Desired Job Category

- Compare candidate's skills and discover desired job passion

 Go

GitHub Language Proficiency

- Discover candidate's overall programming language proficiency
- Compare two candidates over language proficiency

 Go

Read Reference Letter

- Quickly go through the Recommendations/ Reference/ Service letter and find the sentiment

 Go

Intellihire

Copyright © 2020 Landify UI Kit.
All rights reserved.

Company Support Stay up to date

About us Help center Your email address ↗
Blog Terms of service
Contact us Legal
Pricing Privacy policy
Testimonials Status



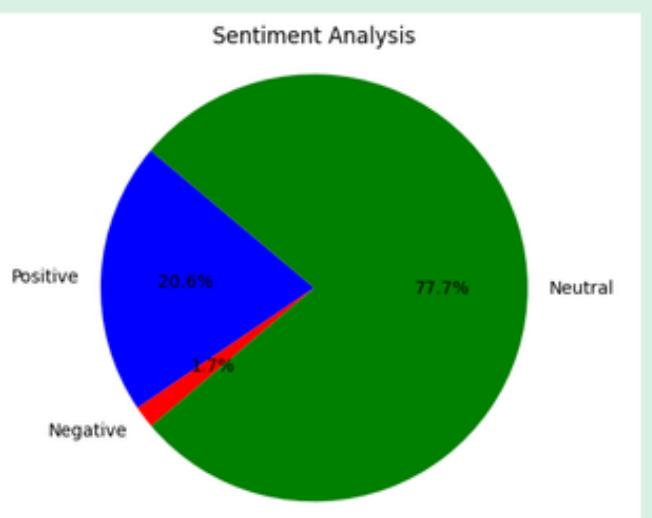
INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Select or Drop the letter here!

Compare

Sentiment Analysis



Sentiment	Percentage
Positive	20.6%
Negative	1.7%
Neutral	77.7%

Intellihire

Copyright © 2020 Landify UI Kit.
All rights reserved.

Company Support Stay up to date

About us Help center Your email address ↗
Blog Terms of service
Contact us Legal
Pricing Privacy policy
Testimonials Status



 SLIIT
FACULTY OF COMPUTING

2023-098

PROTOTYPES

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Enter candidate GitHub URL here!

Find

Compare with a peer candidate

Programming Language Proficiency

Language	Percentage
Java	55%
HTML	25%
CSS	15 %
JAVA SCRIPT	5%

Intellihire

Company Support Stay up to date

About us Help center Your email address

Blog Terms of service

Contact us Legal

Pricing AEBBC6 Privacy policy

Testimonials Status

Copyright © 2020 Landify UI Kit.
All rights reserved

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Enter peer candidate GitHub URL here!

Compare

Language	Value
Java	5
HTML	8
CSS	10
JS	18

```
graph TD; Java -- "0.568259451176" --> CurrentUser[Current User]; Java -- "5.096245090089821" --> OtherUser[Other User]; HTML -- "0.269793041283987134" --> CurrentUser; HTML -- "2.4124219318" --> OtherUser; CSS -- "7.02974079097065" --> CurrentUser; CSS -- "7.02974079097065" --> OtherUser; JavaScript -- "5.382361423824608" --> CurrentUser; JavaScript -- "7013319433913e" --> OtherUser;
```

Intellihire

Company Support Stay up to date

About us Help center Your email address

Blog Terms of service

Contact us Legal

Pricing AEBBC6 Privacy policy

Testimonials Status

Copyright © 2020 Landify UI Kit.
All rights reserved

PROTOTYPES

The screenshot shows a web application interface for 'INTELLIHIRE'. At the top, there is a dark navigation bar with the 'INTELLIHIRE' logo on the left and five menu items: 'Home' (highlighted in green), 'CV Analysis', 'Professional Skills', 'Academic Transcript', and 'Personality Prediction'. Below the navigation bar is a light blue main content area. On the left side of this area, there is a text input field with the placeholder 'Enter candidate LinkedIn URL here!' and a blue 'Predict' button below it. On the right side, there is another blue button labeled 'View more Details in LinkedIn'. The bottom of the page features a dark footer bar with the 'Intellihire' logo, copyright information ('Copyright © 2020 Landify UI Kit. All rights reserved.'), social media links (Instagram, Facebook, Twitter, YouTube), and a navigation menu with links like 'About us', 'Blog', 'Contact us', 'Pricing', 'Testimonials', 'Help center', 'Terms of service', 'Legal', 'Privacy policy', and 'Status'. There is also a 'Stay up to date' section with an input field for an email address.

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Enter candidate LinkedIn URL here!

Predict

Designer!!!

View more Details in LinkedIn

Powered By ProxyCurl

Intellihire

Copyright © 2020 Landify UI Kit.
All rights reserved.

About us
Blog
Contact us
Pricing
Testimonials
Help center
Terms of service
Legal
Privacy policy
Status

Stay up to date

Your email address

REFERENCES

- [1] S. Gupta, B. Gupta, and S. Gupta, "A Novel Method for Technical Candidate Assessment using Github Repository Inspection Automation," in 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, Jan. 2022, pp. 1-5. doi: 10.1109/ICCCI54379.2022.9740986.
- [2] R. G. U. S. Gajanayake, M. H. M. Hiras, P. I. N. Gunathunga, E. G. Janith Supun, A. Karunasenna, and P. Bandara, "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer," in 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, Dec. 2020, pp. 168-173. doi: 10.1109/ICAC51239.2020.9357279.
- [5] S. M. Patil, R. Singh, P. Patil, and N. Pathare, "Personality prediction using Digital footprints," in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, May 2021, pp. 1736-1742. doi: 10.1109/ICICCS51141.2021.9432380.
- [7] H. K. S M, S. Hegde, S. G, S. M, S. R, and S. L. N, "User Interest Prediction based on Social Network Profile with Machine Learning," in 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, Apr. 2021, pp. 1-6. doi: 10.1109/I2CT51068.2021.9418126
- [8] R. T. R. Jayasekara, K. A. N. D. Kudarachchi, K. G. S. S. K. Kariyawasam, D. Rajapaksha, S. L. Jayasinghe, and S. Thelijjagoda, "DevFlair: A Framework to Automate the Pre-screening Process of Software Engineering Job Candidates," in 2022 4th International Conference on Advancements in Computing (ICAC), Colombo, Sri Lanka: IEEE, Dec. 2022, pp. 288-293. doi: 10.1109/ICAC57685.2022.10025337.

IT20207854 | De Silva M.

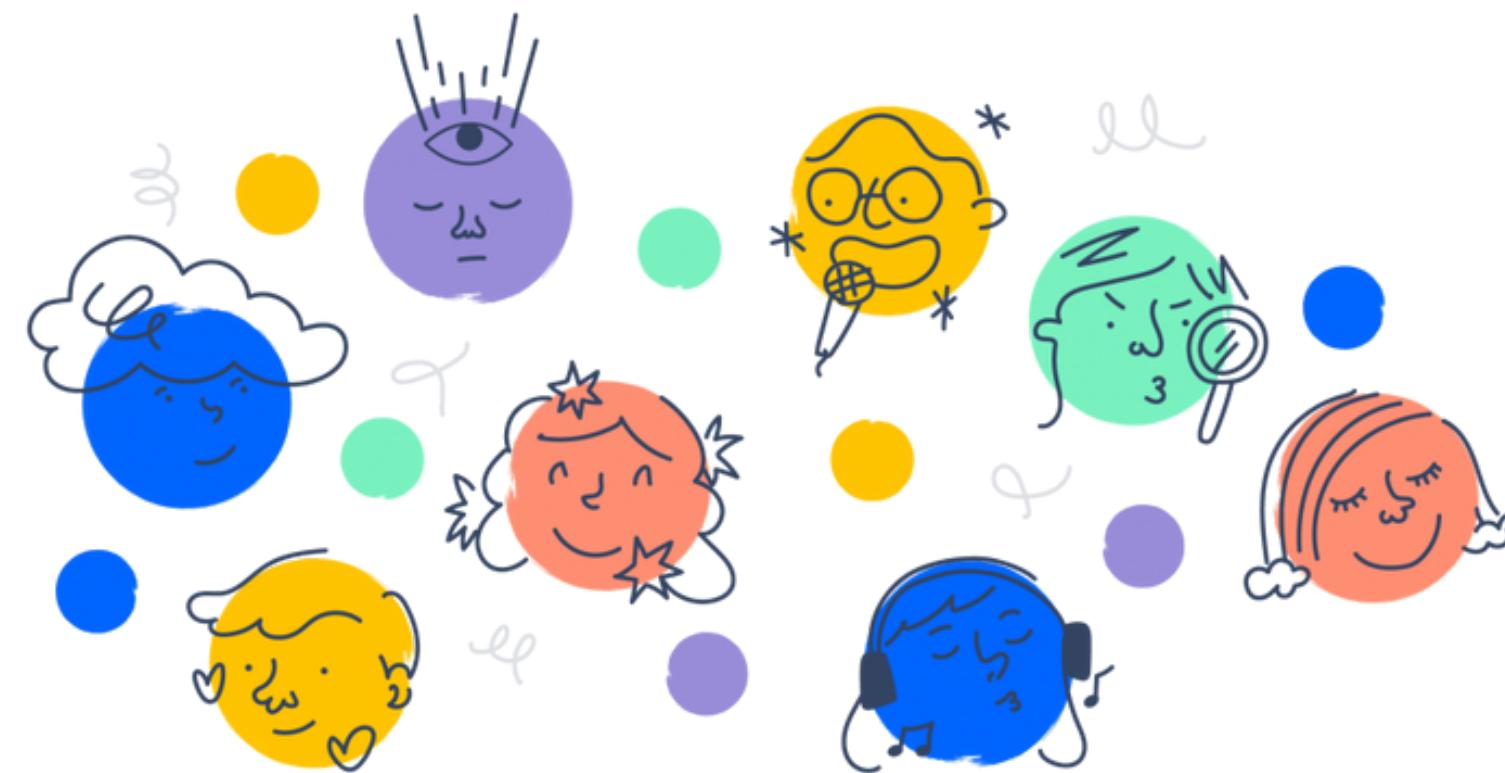
**Bachelor of Science (Hons) in Information Technology Specializing in
Data Science**



PERSONALITY PREDICTION OF CANDIDATES

PROBLEM DEFINITION

- Personality traits and characteristics play a crucial role in determining job fit, job satisfaction, and overall success, benefitting both the individual and organization as a whole.
- Existing systems do not evaluate if a candidate's personality traits match those needed for the job role.



PROPOSED SOLUTION

MAIN OBJECTIVE

- The research objective of this specific component is to develop a feature to assess the compatibility between a candidate's personality and the personality traits necessary for a particular job role.

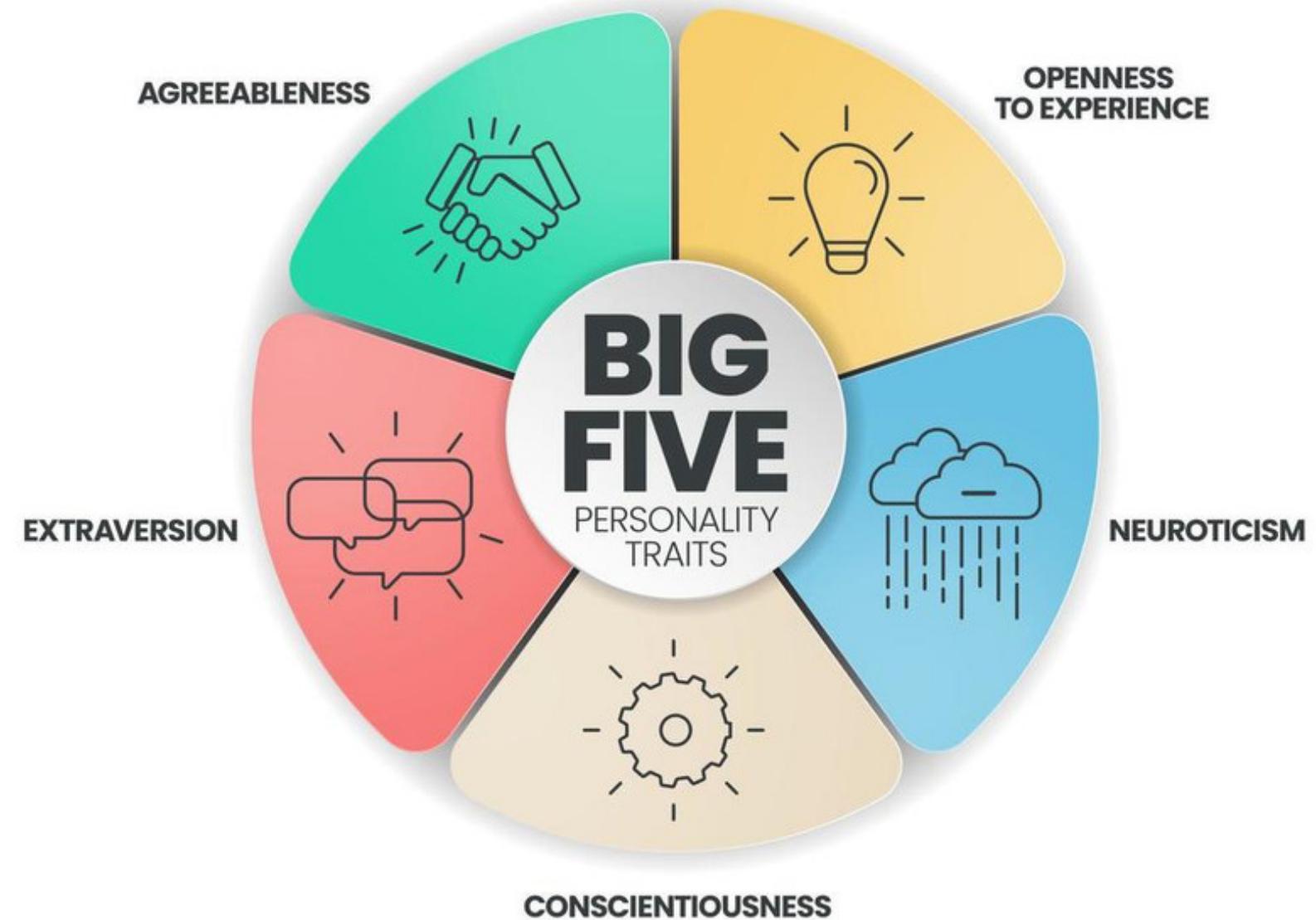
SUB OBJECTIVES

- Develop a feature to identify the presence of Big Five Personality Traits of candidates.
- Develop an algorithm to evaluate how well the personality traits of the candidate aligns with the job role.



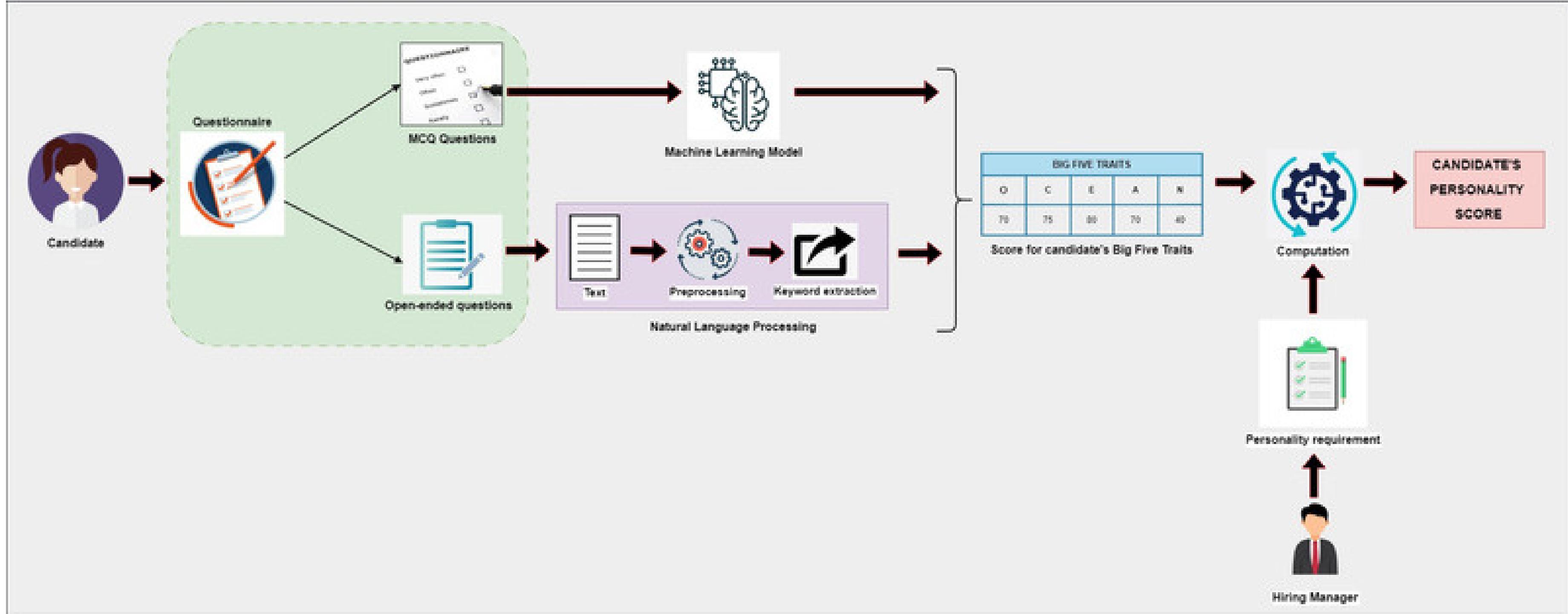
THE BIG FIVE MODEL

**WHAT ARE
THE BIG FIVE
PERSONALITY
TRAITS?**



- Big Five traits are considered to be highly relevant to job performance and organizational success, making it a popular choice for employee selection and development.

SYSTEM OVERVIEW DIAGRAM



PROGRESS



IMPLEMENTATION



CURRENT PROGRESS



Identify the characteristics of the Big Five Traits



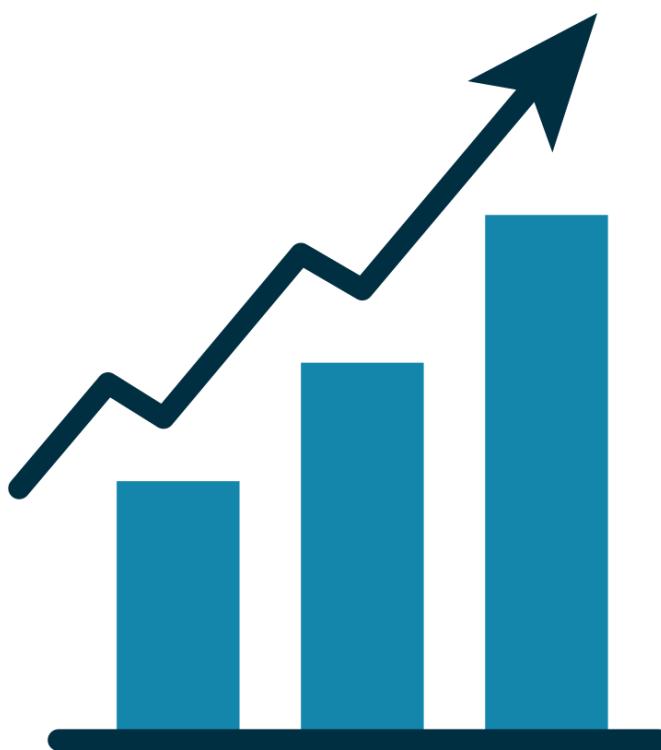
Identify the main personality clusters using K-means clustering

- ✓ Data Preprocessing
- ✓ Explanatory Data Analysis
- ✓ Elbow Visualization
- ✓ Principal Component Analysis



Identify keywords from the dataset to determine the Big Five Traits

- ✓ Extract keywords using 'Bag of Words'
- ✓ Extract keywords using 'TF-IDF'
- ✓ Extract keywords using 'KeyBERT'



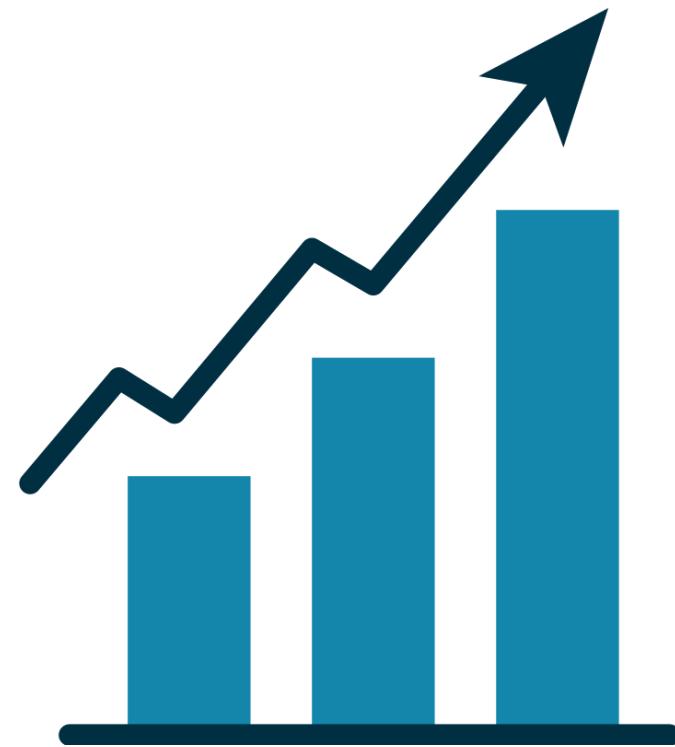
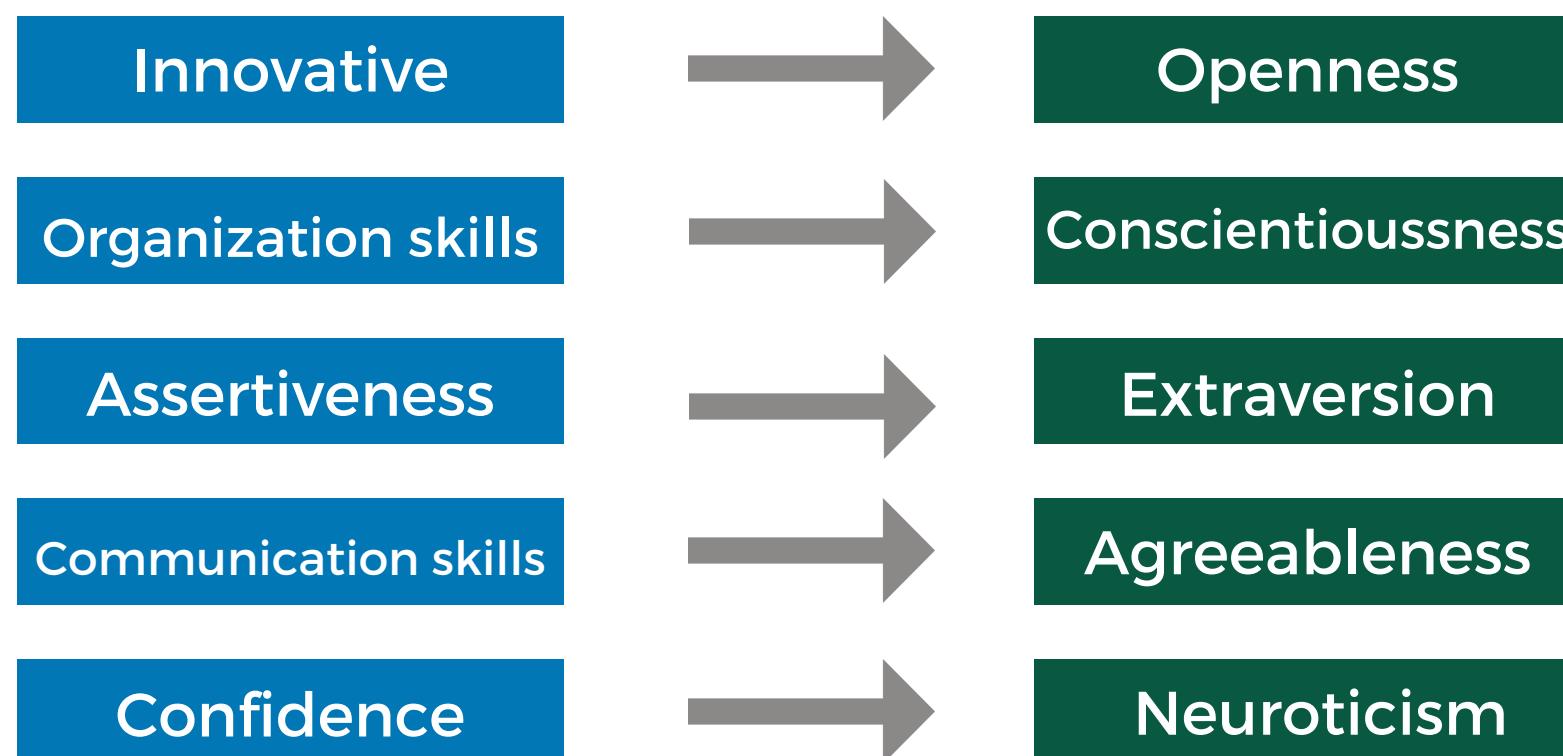
CURRENT PROGRESS



Algorithm for predicting Big Five personality trait distribution of candidate.



Algorithm for predicting Big Five personality trait distribution required for a specific job role.



CURRENT PROGRESS

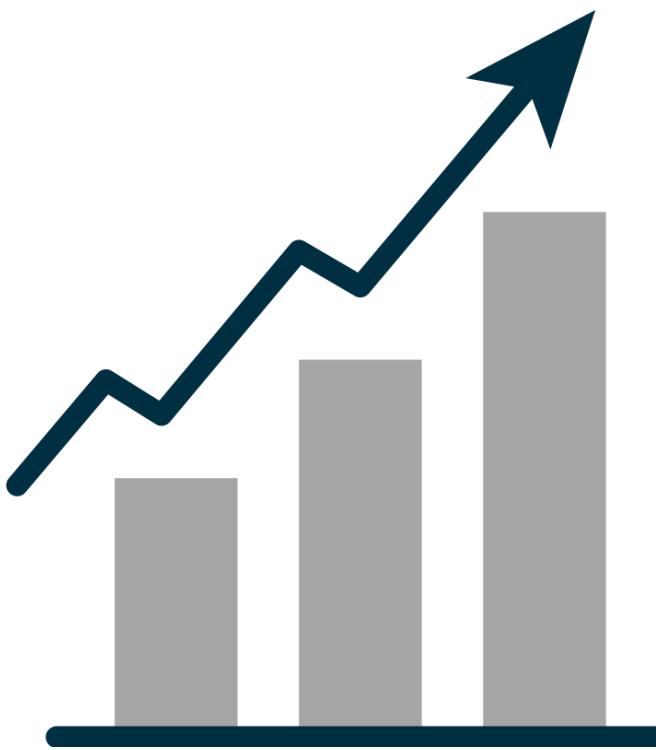
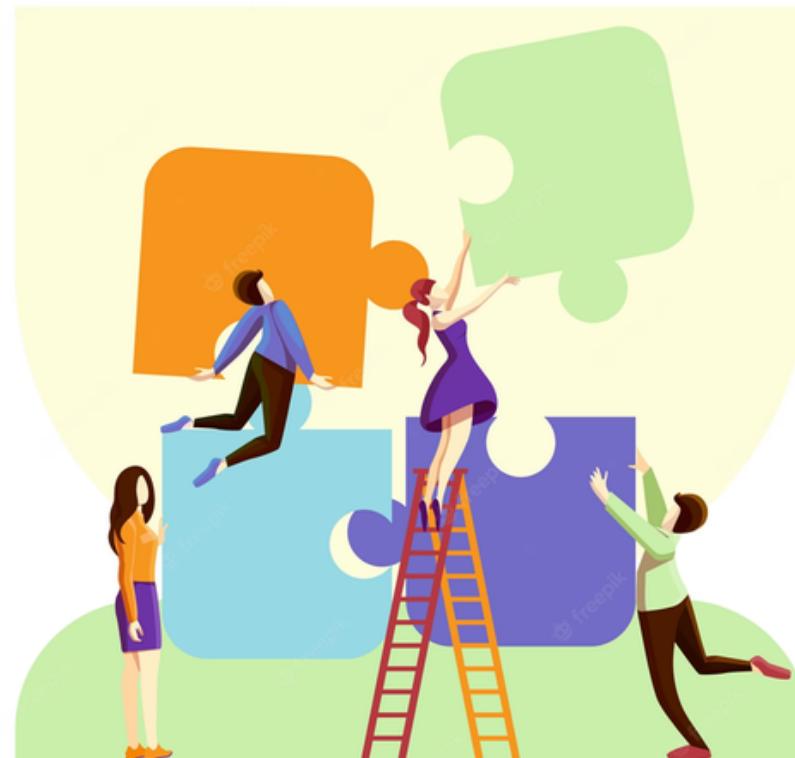
 **Algorithm to assess the suitability of a candidate for a particular job role.**

-  RandomForest
-  XGBoost
-  Naive Bayes

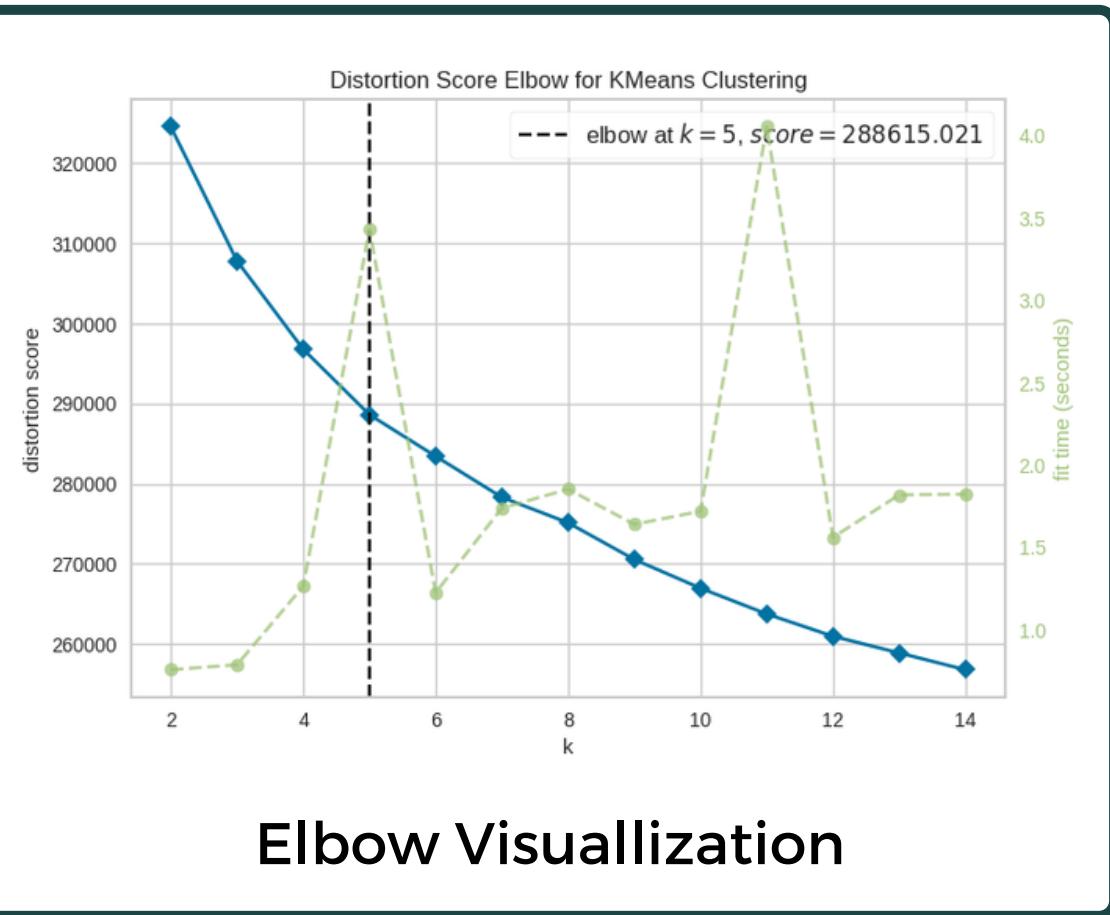


EXPECTED FUTURE PROGRESS

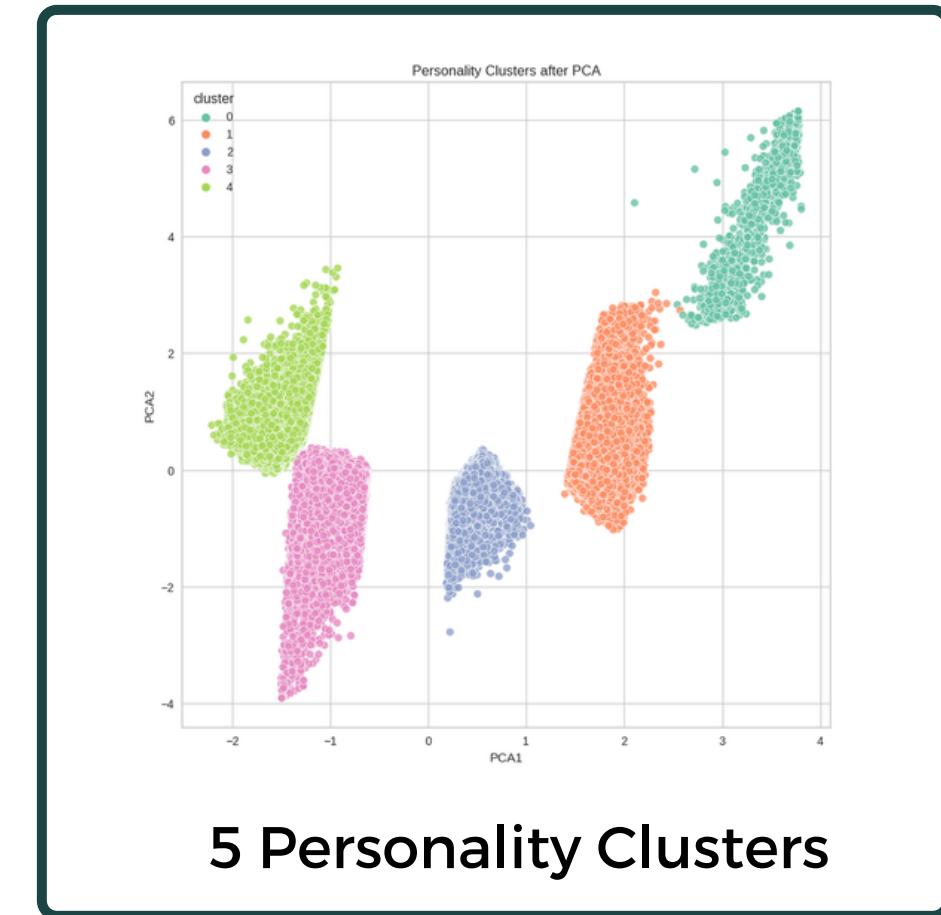
- ▶ Incorporate candidate's personality score in the weighted final score.
- ▶ React frontend design
- ▶ Testing for functionality and possible improvements



OUTPUTS



Elbow Visualizzazione



5 Personality Clusters

```
# Show the updated DataFrame
print('Candidate personality score: ', score)

Candidate personality score:  0    72.957672
dtype: float64
```

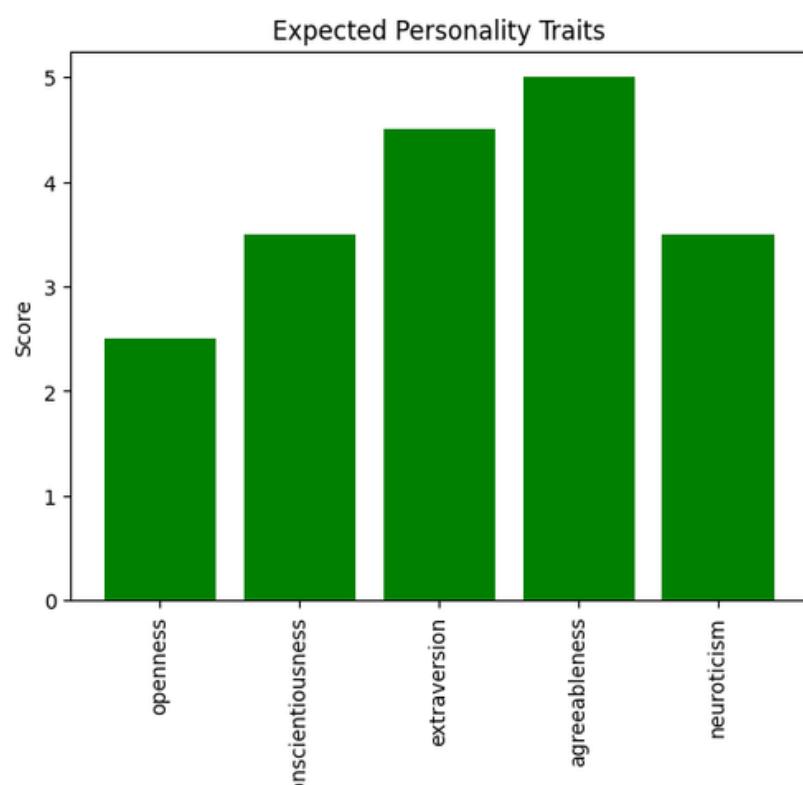
Candidate Personality Compatibility

```
# Evaluate the model's performance
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

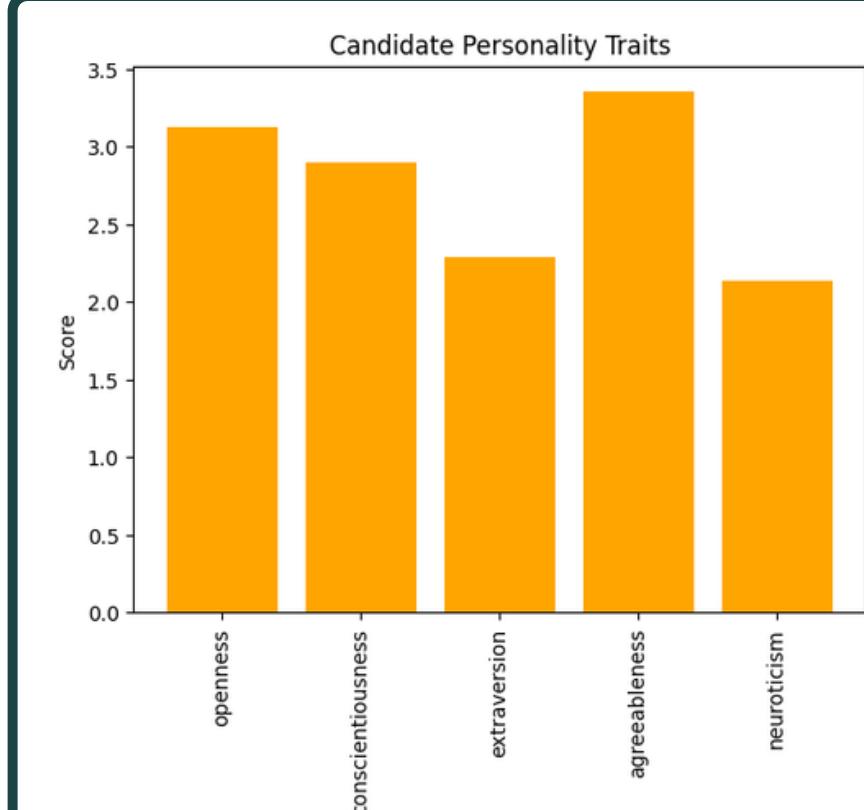
Accuracy: 0.9538769690720632
```

Accuracy of XGBoost Model

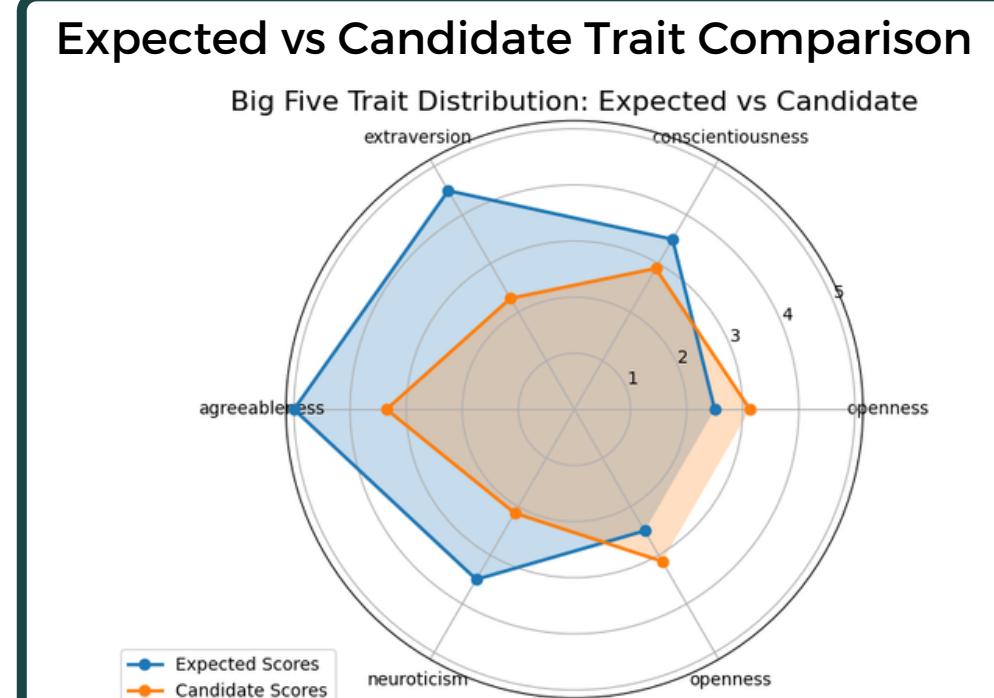
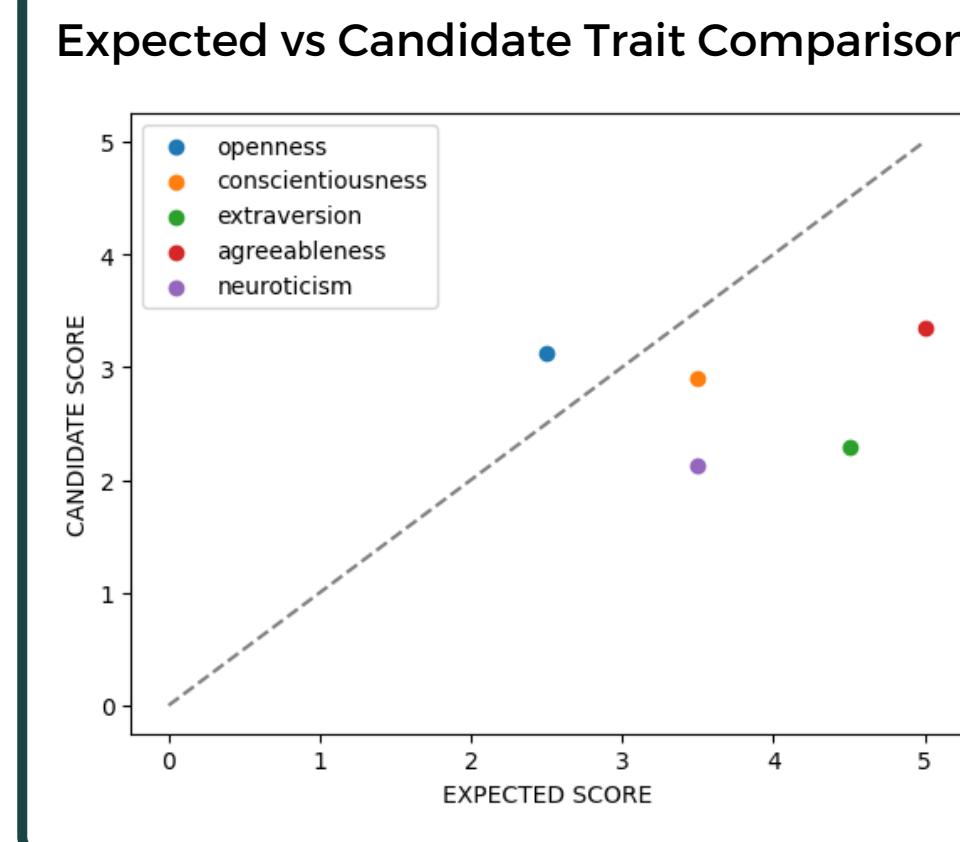
OUTPUTS



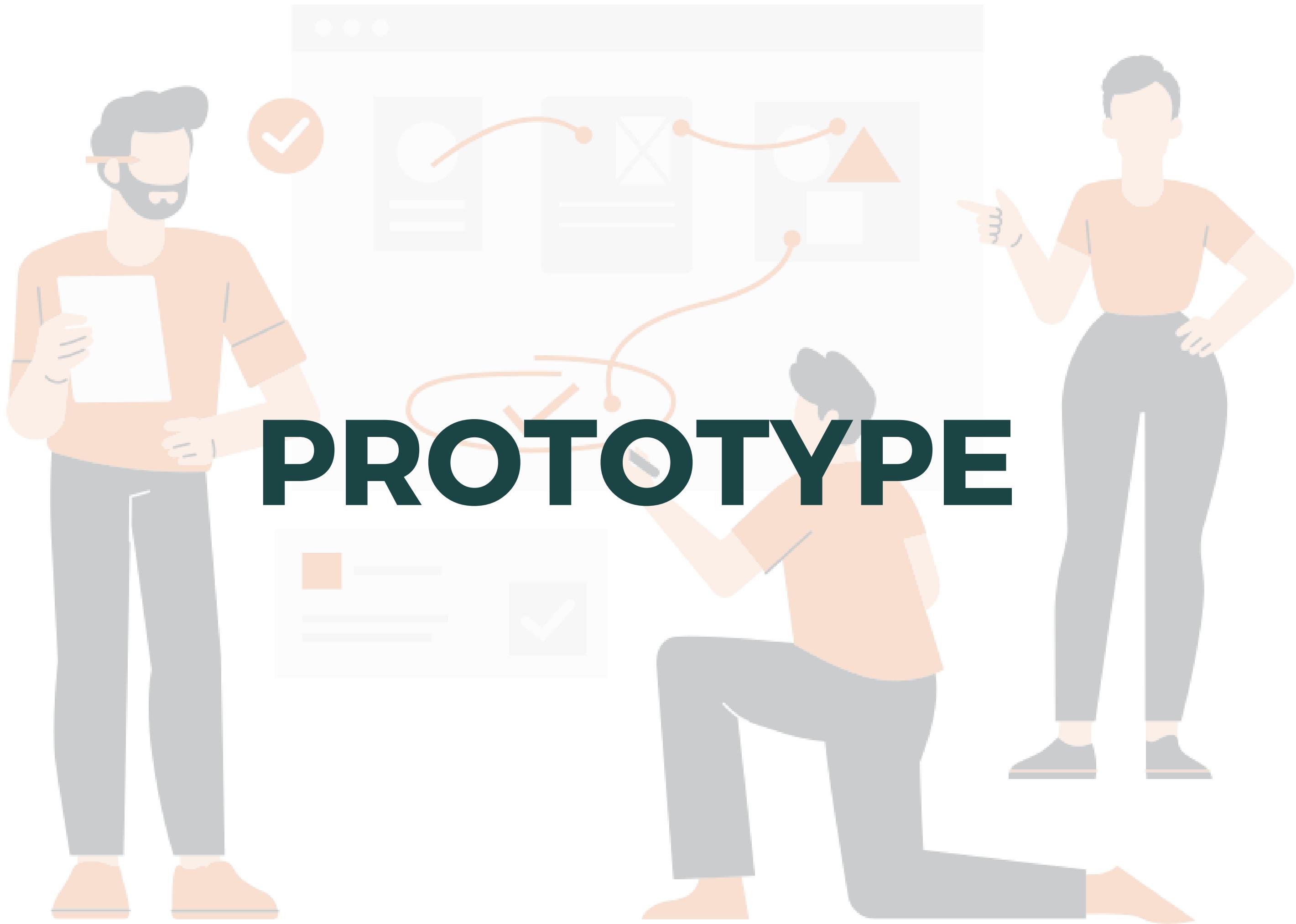
Expected Personality Traits



Candidate Personality Traits



PROTOTYPE



INTELLIHIRE - PERSONALITY PREDICTION



CANDIDATE RESPONSES

Candidate ID	Candidate Name	Personality Match
IT1021	Tharindu Herath	View response -
IT1022	Natasha Perera	View response -
IT1023	Aanya Silva	View response -
IT1024	Roshan De Silva	View response 68%
IT1025	Sachin de Silva	View response 57%
IT1026	Dilshan Ranasinghe	View response 73%

[+ Add new](#)

Candidate ID IT1021

Candidate Name Tharindu Herath

Applying position Project Manager

How often do you seek out new experiences or take risks?

I am someone who enjoys taking on new challenges and seeking out new experiences. I believe that by pushing myself out of my comfort zone, I am able to learn and grow both professionally and personally. In my previous roles, I have taken on new projects and tasks that were outside of my immediate area of expertise, and I found these experiences to be incredibly rewarding. Of course, I always make sure to assess the risks involved before taking on a new challenge, but I am not afraid to take calculated risks in order to achieve my goals. Overall, I would say that I am someone who is open to new experiences and willing to take risks when the situation calls for it.

How do you ensure that your work is completed accurately and on time?

I ensure accurate and timely completion of work by breaking down projects into manageable tasks and prioritizing based on importance and urgency. I use productivity tools and communicate regularly with my team to track progress and address issues early. Seeking feedback from colleagues and stakeholders, taking ownership of my work, and being willing to put in extra effort are also important to me in ensuring successful completion of projects.

[VIEW PERSONALITY COMPATIBILITY](#)

EXPECTED PERSONALITY TRAITS

Applying Position

Leadership skills

Analytical & Problem Solving skills

Adaptability to changes

Team player

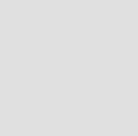
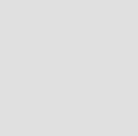
SUBMIT



COMPATIBILITY SCORE



COMPATIBILITY SCORE 71%

[Save score](#)

REFERENCES

- Dataset
- <https://www.verywellmind.com/the-big-five-personality-dimensions-2795422>
- "R. T. R. Jayasekara, K. A. N. D. Kudarachchi, K. G. S. S. K. Kariyawasam, D. Rajapaksha, S. L. Jayasinghe and S. Thelijjagoda, "DevFlair: A Framework to Automate the Pre-screening Process of Software Engineering Job Candidates,""
- M. A. Iqbal, F. A. Ammar, A. R. Aldaihani, T. K. U. Khan and A. Shah, "Building Most Effective Requirements Engineering Teams by Evaluating Their Personality Traits Using Big-Five Assessment Model,"
- "R. G. U. S. Gajanayake, M. H. M. Hiras, P. I. N. Gunathunga, E. G. Janith Supun, A. Karunasenna and P. Bandara, "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer,""
- M. Rehman, A. K. Mahmood, R. Salleh and A. Amin, "Mapping job requirements of software engineers to Big Five Personality Traits," 2012 International Conference on Computer & Information Science (ICCIS), Kuala Lumpur, Malaysia, 2012. (n.d.).
- "S. K. Nivetha, M. Geetha, R. S. Latha, K. Sneha, S. Sobika and C. Yamuna, "Personality Prediction for Online Interview," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022"

IT20216900 | De Silva S.R.

**Bachelor of Science (Hons) in Information Technology Specializing in
Data Science**



ANALYZING THE ACADEMIC TRANSCRIPT TO UNDERSTAND THE TYPES OF SKILLS AND KNOWLEGDE THAT A CANDIDATE HAS GAINED DURING THEIR DEGREE PROGRAM

PROBLEM DEFINITION

It is essential to accurately assess the skills and knowledge acquired by candidates during their degree program to ensure a suitable match with job requirements.

Existing evaluation methods often fail to provide a comprehensive understanding of the skills and knowledge gained by candidates, leading to potential mismatches and inefficiencies in the hiring process.



OBJECTIVES

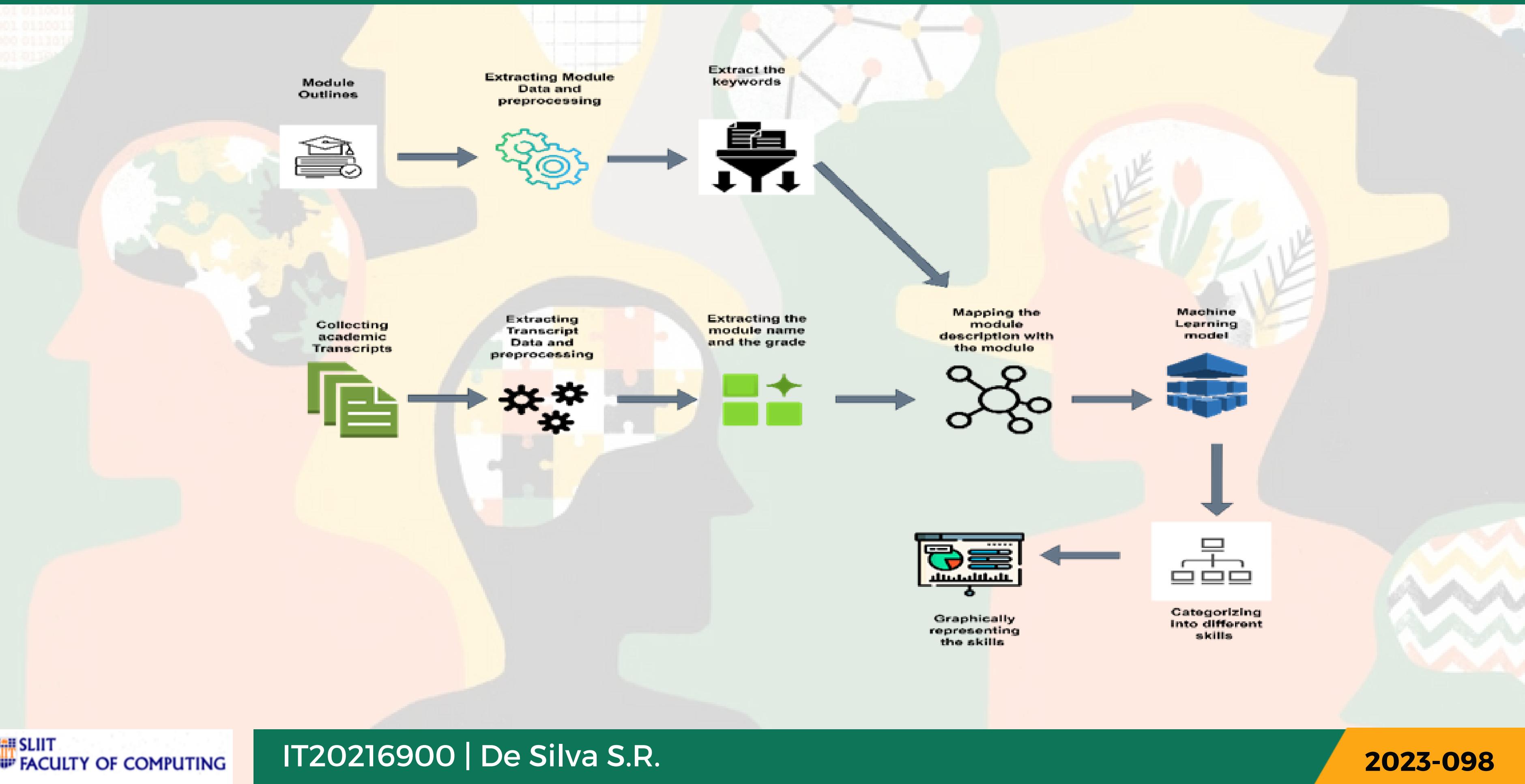
MAIN OBJECTIVE

- To analyze academic transcripts as a means to gain a comprehensive understanding of the specific skills and knowledge that candidates have acquired throughout their degree program.

SUB OBJECTIVES

- To develop an algorithm for categorizing and classifying the skills and knowledge reflected in academic transcripts..
- Create an intuitive, user-friendly graphical representation of a candidate's skilled area and knowledge to simplify hiring decisions.
- To assess the extent to which academic transcripts can serve as a reliable indicator of a candidate's skills and knowledge proficiency.

SYSTEM OVERVIEW DIAGRAM



PROGRESS



CURRENT PROGRESS

- ✓ Extracted the important keywords from the module outlines
- ✓ Built and trained a Custom Name Entity Recognition(NER) Model to Identify the Module Titles and the Grades from an Academic transcript
- ✓ Trained a model to identify the Skill Area with the accuracy of 87%
- ✓ Extracted Transcript data using Optical Character Recognition
- ✓ Categorized the the modules into different skill areas and graphically represented them



EXPECTED FUTURE PROGRESS

- Enhance the accuracy of the models
- Optimize and fine-tune the OCR and text processing: With different preprocessing techniques, OCR engines to improve text recognition performance.
- Front end Implementation
- Develop the web application
- Testing for functionalities and further improvements

100%

IMPLEMENTATION



Tools and Technologies

Tools and Technologies

- Extract Keywords from Module Outlines

n-grams and nltk library

- Custom NER Model

Spacy

- Model to Predict the Skill Area

Countvectorizer, Support Vector Machines, XGBoost, Naive Bayes

- Extract Transcript Data

Optical Character Recognition

- Tool

Google Colab

CUSTOM NER MODEL

Custom NER output

IT3030 CODE
Programming Applications and Frameworks MODULE_TITLE
1 SEMESTER
Apr - 2021 PERIOD
4 CREDITS
A GRADE
IT3050* CODE
Employability Skills Development - Seminar MODULE_TITLE
1 SEMESTER
Apr - 2021 PERIOD
1 CREDITS
C GRADE
T3041 CODE
Information Retrieval and Web MODULE_TITLE
2 SEMESTER
Oct - 2021 PERIOD
4 CREDITS
B+ GRADE
T3051 CODE
Fundamentals of Data Mining MODULE_TITLE
2 SEMESTER
Oct - 2021 PERIOD
4 CREDITS
B+ GRADE
T3061 CODE
Massive Data Processing and Cloud MODULE_TITLE
2 SEMESTER
Oct - 2021 PERIOD
4 CREDITS
B- GRADE
T3071 CODE
Machine Learning and Optimization MODULE_TITLE
2 SEMESTER
Oct - 2021 PERIOD
4 CREDITS
B- GRADE
1T3110* CODE
Industry Placement MODULE_TITLE
2 SEMESTER
Oct - 2021 PERIOD
8 CREDITS
C GRADE
174021 CODE
Internet of Things and Big Data Analytics MODULE_TITLE
1 SEMESTER
Jun - 2022 PERIOD

E	#	LOSS_TOK2VEC	LOSS_NER	ENTS_F	ENTS_P	ENTS_R	SCORE
0	0	0.00	52.29	14.40	9.24	32.65	0.14
4	200	151.16	1704.66	99.46	99.46	99.46	0.99
8	400	22.88	37.06	99.68	99.68	99.68	1.00
14	600	40.78	43.16	99.68	99.68	99.68	1.00
22	800	56.30	44.88	99.78	99.78	99.78	1.00
31	1000	117.53	45.19	99.78	99.78	99.78	1.00
42	1200	109.26	55.50	99.78	99.78	99.78	1.00
56	1400	459.08	66.02	99.78	99.78	99.78	1.00
73	1600	475.02	89.73	99.78	99.78	99.78	1.00
93	1800	1482.81	100.33	99.78	99.78	99.78	1.00
118	2000	1065.48	112.98	99.78	99.78	99.78	1.00
148	2200	2867.07	136.36	99.78	99.78	99.78	1.00
185	2400	864.50	170.98	99.78	99.78	99.78	1.00

```
    warnings.warn(  
        Losses: {'ner': 1080.9641413251313}  
        Losses: {'ner': 81.69892219707003}  
        Losses: {'ner': 26.13685030728511}  
        Losses: {'ner': 32.97153806471643}  
        Losses: {'ner': 2.39907462387285}  
        Losses: {'ner': 22.977387887101763}  
        Losses: {'ner': 4.599581692964621e-06}  
        Losses: {'ner': 1.9882955234192774e-08}  
        Losses: {'ner': 5.26225924263131e-09}  
        Losses: {'ner': 2.2673915734306062e-07}
```

Row 97 - Top N-grams: ['machine learning', 'deterministic algorithm', 'machine learning discuss', 'technology different', 'algorithm technique may', 'discuss machine learning', 'realworld problem cannot', 'discuss machine learning', 'realworld problem cannot']
Row 98 - Top N-grams: ['business analysis', 'order develop effective', 'skill essential', 'aim give student', 'aim give', 'babok focus role', 'babok focus', 'business requirement', 'business requirement analysis', 'business requirement analysis', 'business requirement analysis']
Row 99 - Top N-grams: ['role is', 'maturity hris', 'relationship hr', 'hris role', 'hris hris become', 'hris hris', 'hris hr analytics', 'hris hr', 'hris become study', 'hris become']
Row 100 - Top N-grams: ['quality assurance', 'course provider', 'participant broad awareness', 'tool achieving software', 'tool achieving', 'participant broad', 'practical knowledge variety', 'awareness many', 'quality assurance', 'course provider', 'participant broad awareness', 'tool achieving software', 'tool achieving', 'participant broad', 'practical knowledge variety', 'awareness many']
Row 101 - Top N-grams: ['knowledge management', 'reused fully', 'knowledge acquisition assessment', 'created captured represented', 'coordination codification knowledge', 'module focus know', 'knowledge management', 'reused fully', 'knowledge acquisition assessment', 'created captured represented', 'coordination codification knowledge', 'module focus know']
Row 102 - Top N-grams: ['computing device', 'study interaction', 'user computing device', 'user computing', 'behavioral science', 'often regarded', 'several field', 'occurs user', 'occurs user interface', 'inte', 'computing device', 'study interaction', 'user computing device', 'user computing', 'behavioral science', 'often regarded', 'several field', 'occurs user', 'occurs user interface', 'inte']
Row 103 - Top N-grams: ['nosql web service', 'software development implementation', 'goal software', 'allowing concentrate', 'allowing concentrate actual', 'key concept', 'key concept software', 'actual goal so', 'nosql web service', 'software development implementation', 'goal software', 'allowing concentrate', 'allowing concentrate actual', 'key concept', 'key concept software', 'actual goal so']
Row 104 - Top N-grams: ['networking sdn', 'develop good', 'underlying physical', 'simulation environment mininet', 'emerging paradigm', 'emerging paradigm data', 'current usage future', 'current usage', 'under', 'networking sdn', 'develop good', 'underlying physical', 'simulation environment mininet', 'emerging paradigm', 'emerging paradigm data', 'current usage future', 'current usage', 'under']
Row 105 - Top N-grams: ['iot ecosystem', 'iot ecosystem particular', 'focus iot ecosystem', 'phase iot ecosystem', 'analyzes unstructured data', 'analyzes unstructured', 'industry approaching it', 'ecosystem pa', 'iot ecosystem', 'iot ecosystem particular', 'focus iot ecosystem', 'phase iot ecosystem', 'analyzes unstructured data', 'analyzes unstructured', 'industry approaching it', 'ecosystem pa']
Row 106 - Top N-grams: ['cloud migration', 'cloud migration techniques', 'virtualization techniques', 'module reuse', 'virtualization methods', 'virtualization techniques finally', 'virtualization techniques finally', 'virtualization module reuse']

N-gram output

MODEL BUILDING

Accuracy of XG Booster Model

Accuracy: 0.862379421221865

Accuracy of SVM Model

```
[22]: #svm_model.score(x_test, y_test)
prediction = svm_model.predict(x_test)
accuracy_score(y_test,prediction)

0.8752411575562701
```

Accuracy of Naive Bayes Model

Accuracy: 0.7980707395498392

Categorization of the skill areas

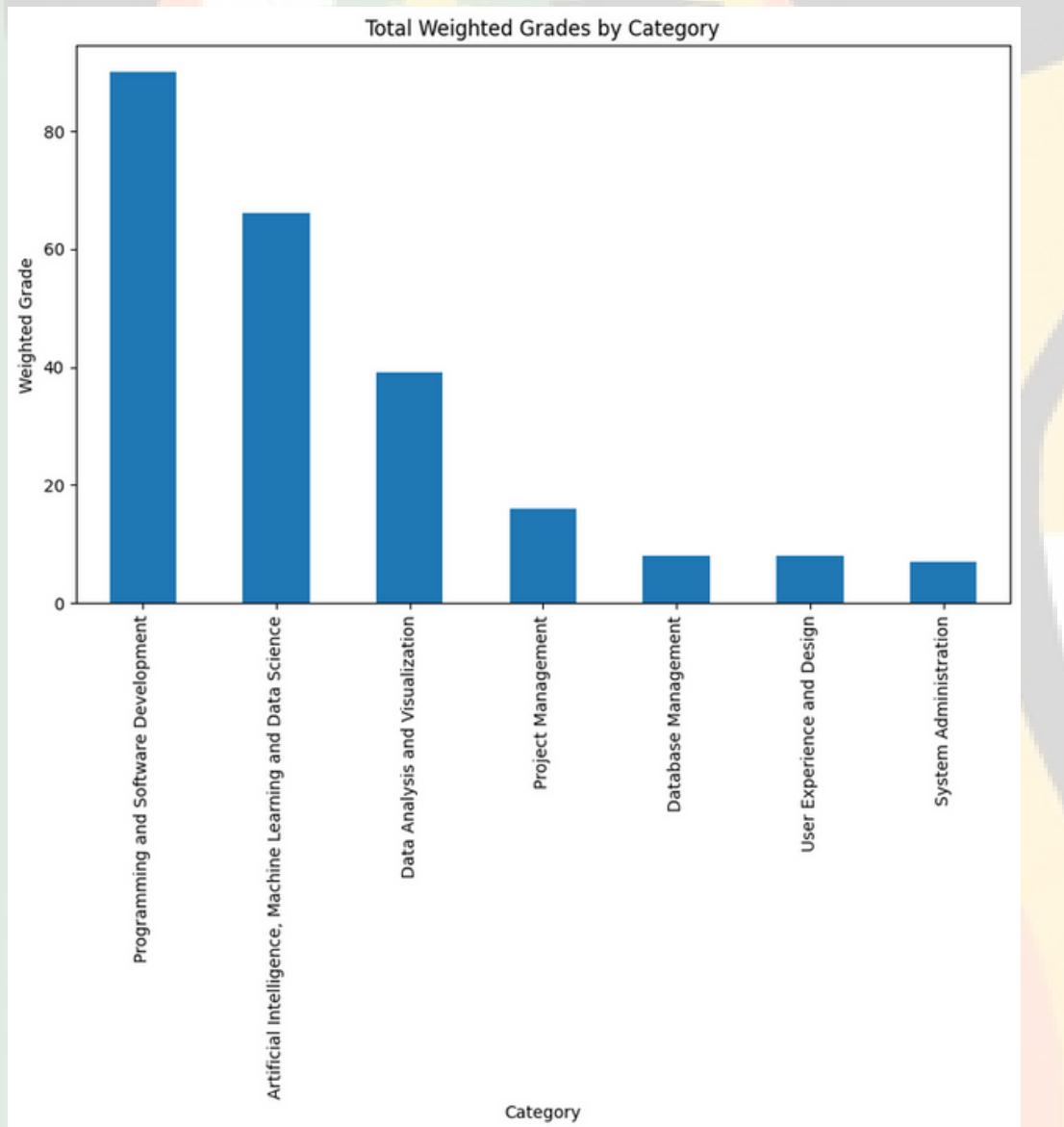
	Category	Module Title	Module Keywords	Weighted Grade
0	Programming and Software Development	introduction programming	data types, control structures, functions, pol...	9.0
1	Artificial Intelligence, Machine Learning and ...	introduction computer systems	computer systems, computer organization, combi...	4.0
2	Artificial Intelligence, Machine Learning and ...	mathematics computing	Logic Control, Number Systems, Differentiation...	6.0
4	Artificial Intelligence, Machine Learning and ...	object oriented concepts	Object Oriented Programming, design solutions,...	9.0
5	Programming and Software Development	software process modeling	Software Engineering, requirement specificatio...	9.0
7	Data Analysis and Visualization	information systems data modeling	Information Systems, Data Modeling, Business P...	9.0
11	Programming and Software Development	database management systems	Database design, Hands-on experience, Schema r...	9.0
13	System Administration	operating systems system administration	System administration, Lecture tutorial lab, K...	7.0
14	Programming and Software Development	mobile application development	Mobile application development,Mobile technolo...	7.0
15	Programming and Software Development	data structures algorithms	Data structures,Stacks,Queues,Linked lists,Tre...	9.0
16	Project Management	project	Final report writing, System testing activity,G...	9.0
18	Data Analysis and Visualization	probability statistics	Descriptive statistics, Statistical techniques...	9.0
19	Data Analysis and Visualization	theory practices statistical modelling	Statistical models,Regression,Time series,Stat...	NaN
20	Programming and Software Development	data warehousing business intelligence	Data integration,Business intelligence,Data qu...	NaN
21	Programming and Software Development	programming applications frameworks	Enterprise systems engineering,Web-based appli...	9.0
24	Data Analysis and Visualization	fundamentals data mining	Theory,Practices,Data mining applications,Data...	7.0
28	Data Analysis and Visualization	internet things big data analytics	Internet of Things (IoT),IoT architectures,Dat...	5.0
29	User Experience and Design	visual analytics user experience design	Visual analytics,User experience design,Analyt...	8.0
31	Project Management	research project	engineering industry, software engineering ind...	7.0
32	Database Management	database administration storage systems	Database administration,Storage systems,Oracle...	8.0

Total Score of Skill Areas

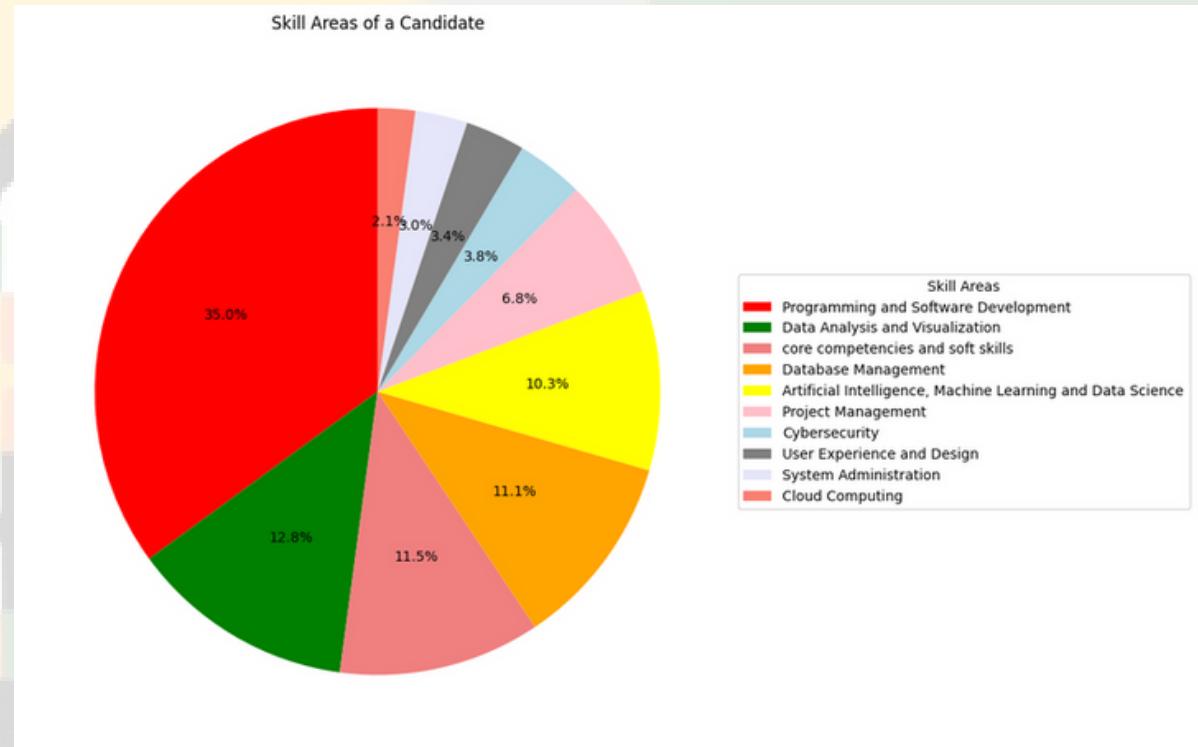
```
Category
Programming and Software Development          90.0
Artificial Intelligence, Machine Learning and Data Science 66.0
Data Analysis and Visualization                39.0
Project Management                           16.0
Database Management                         8.0
User Experience and Design                  8.0
System Administration                        7.0
Name: Weighted Grade, dtype: float64
```

GRAPHICAL REPRESENTATION

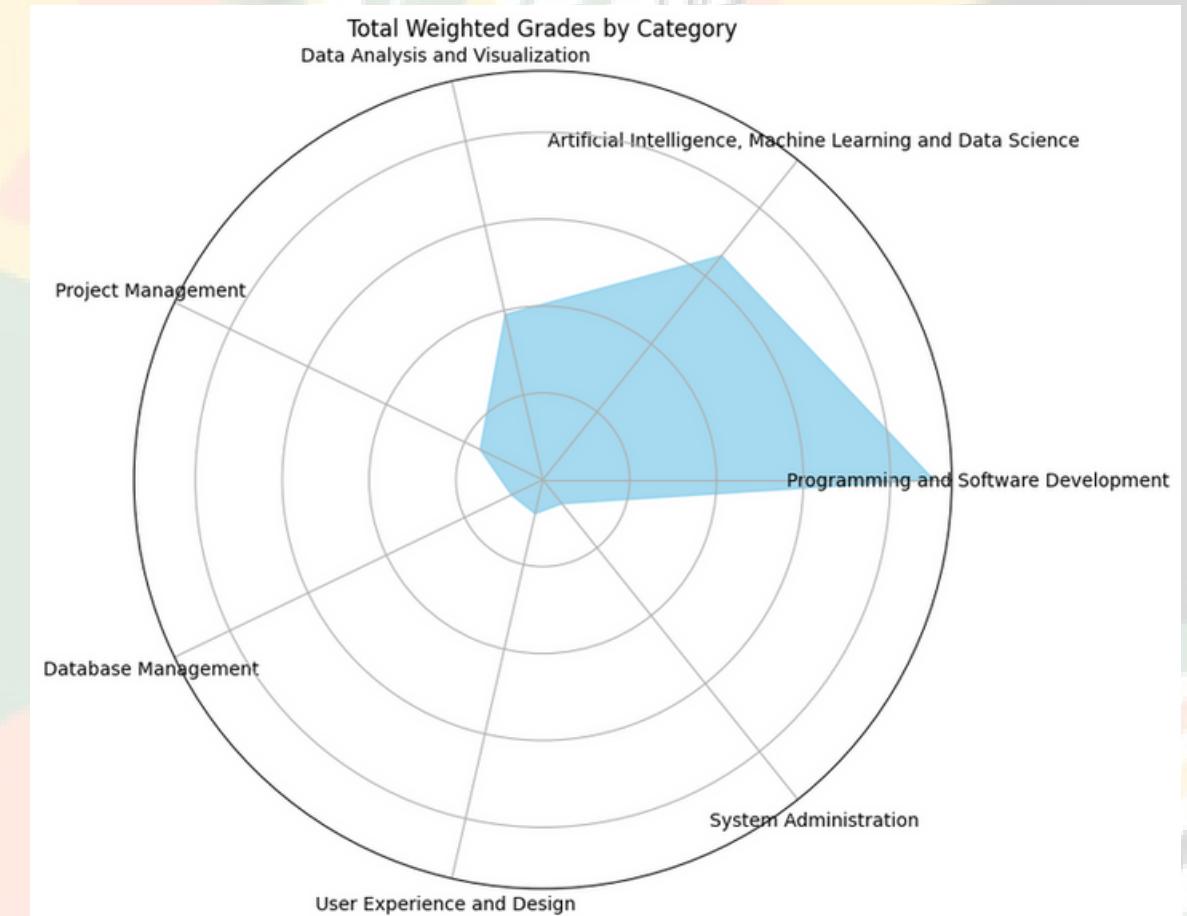
BAR CHART



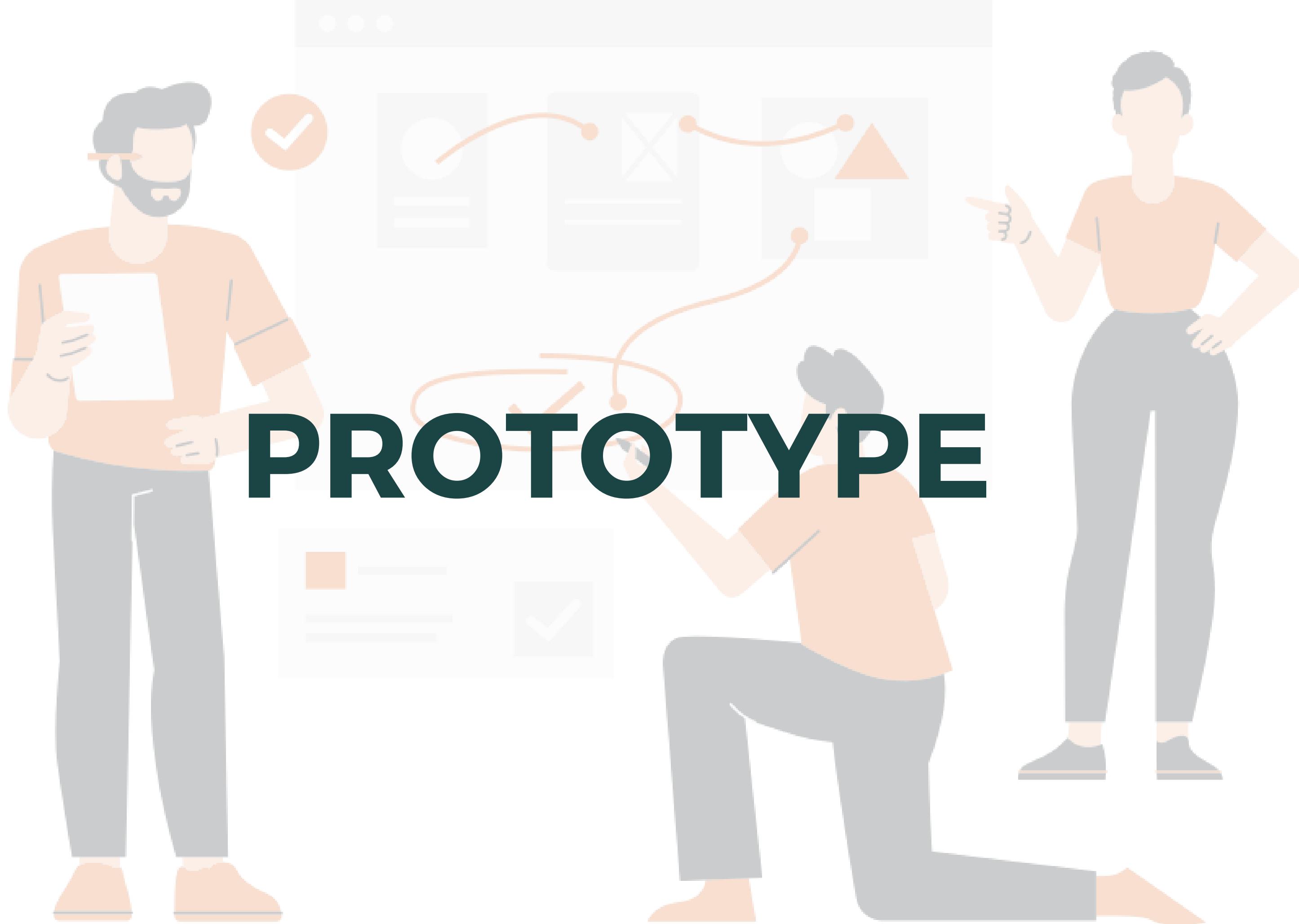
PIE CHART



RADAR GRAPH



PROTOTYPE



PROTOTYPE

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

ACADEMIC TRANSCRIPT

Candidate ID	Candidate Name	
IT1021	Tharindu Herath	View transcript insights
IT1022	Natasha Perera	View transcript insights
IT1023	Aanya Silva	View transcript insights
IT1024	Roshan De Silva	View transcript insights
IT1025	Sachin de Silva	View transcript insights
IT1026	Dilshan Ranasinghe	View transcript insights

[+ Add new](#)

Intellihire
Copyright © 2020 Landify UI Kit.
All rights reserved.

[Company](#) [Support](#) [Stay up to date](#)

[About us](#) [Help center](#) [Send](#)

[Blog](#) [Terms of service](#)

[Contact us](#) [Legal](#)

[Pricing](#) [Privacy policy](#)

[Testimonials](#) [Status](#)

PROTOTYPE

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript

ACADEMIC TRANSCRIPT

Upload the transcript

Candidate ID

Candidate Name

Upload Transcript Upload Transcript

Intellihire

Copyright © 2020 Landly UI Kit.
All rights reserved.

[About us](#) [Help center](#)
[Blog](#) [Terms of service](#)
[Contact us](#) [Legal](#)
[Pricing](#) [Privacy policy](#)
[Testimonials](#) [Status](#)

Stay up to date

Your email address

OBJECTIVES

INTELLIHIRE

Home CV Analysis Professional Skills Academic Transcript Personality Prediction

Candidate ID: IT1021

Candidate Name: Tharindu Herath

Skill Areas of a Candidate

Skill Area	Percentage
Programming and Software Development	35.0%
Data Analysis and Visualization	12.8%
core competencies and soft skills	11.5%
Database Management	11.1%
Artificial Intelligence, Machine Learning and Data Science	10.3%
Project Management	6.8%
Cybersecurity	3.8%
User Experience and Design	3.4%
System Administration	3.0%
Cloud Computing	2.1%

Intellihire

Copyright © 2020 Landify UI Kit.
All rights reserved.

Stay up to date

Your email address

Company Support

About us Help center

Blog Terms of service

Contact us Legal

Pricing Privacy policy

Testimonials Status

REFERENCES

[1]	M. Hiras, R. Gajanayake, P. Gunathunga and E. Supun, "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer," p. 6, 2020. (Research A)
[2]	T. Rahman, J. Nwokeji, . R. Matovu, S. Frezza, H. Sugnanam and A. Pisolkar, "Analyzing Competences in Software Testing: Combining Thematic Analysis with Natural Language Processing (NLP)," p. 9, 2021. (Research B)

CONCLUSION

- As SMMS, our goal is to develop a user-friendly and efficient solution for hiring managers to identify and select top candidates for their organization.
- By leveraging advanced technologies machine learning and Natural Language Processing techniques Intellihire can provide insightful candidate data and streamline the recruitment process.
- With our software, hiring managers can save time and resources while ensuring they make the best hiring decisions for their organization.
- We are confident that Intellihire will drive significant value and success for organizations looking to streamline and optimize their recruitment process.

