



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

PROGRESS PRESENTATION II



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

TEAM MEMBERS

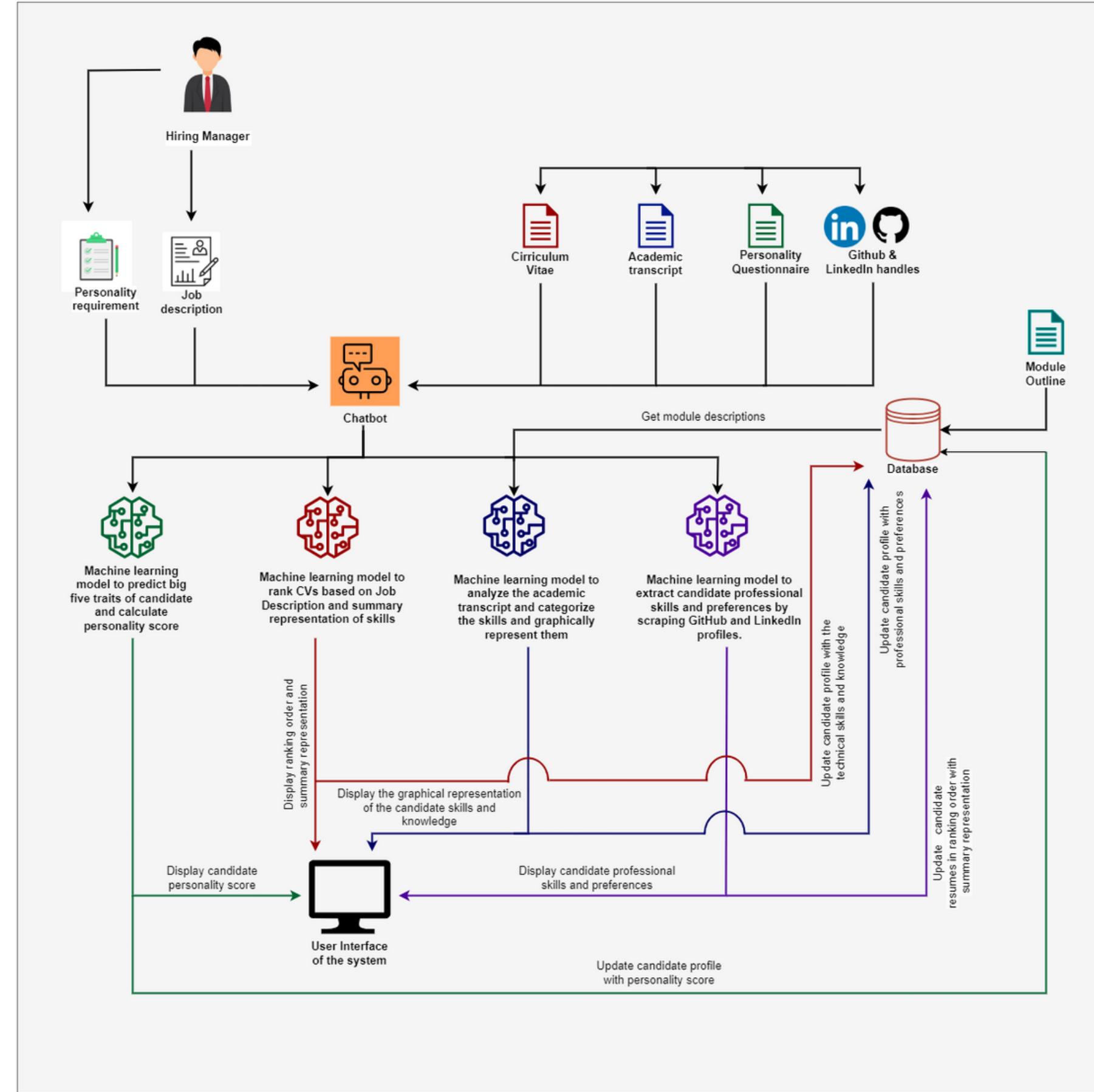
- | | |
|---|---|
| 1 | IT20207854 De Silva M. (Leader) |
| 2 | IT20216900 De Silva S.R. |
| 3 | <u>IT20231200 Zoysa E.S.</u> |
| 4 | <u>IT20203726 Maldeniya M.M.D.</u> |

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



IT20207854 | De Silva M.

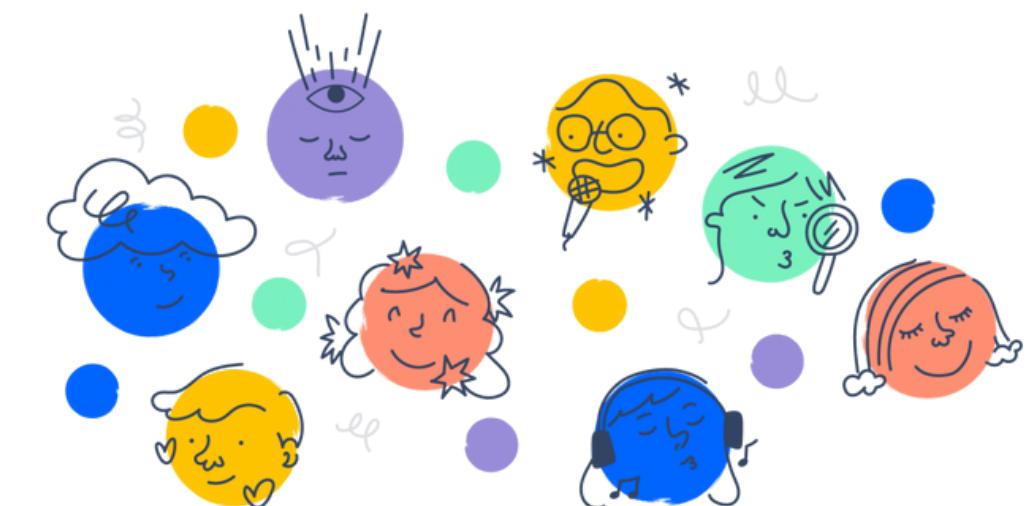
**Bachelor of Science (Hons) in Information Technology Specializing in
Data Science**



PERSONALITY ASSESSMENT OF CANDIDATES

BACKGROUND AND LITERATURE

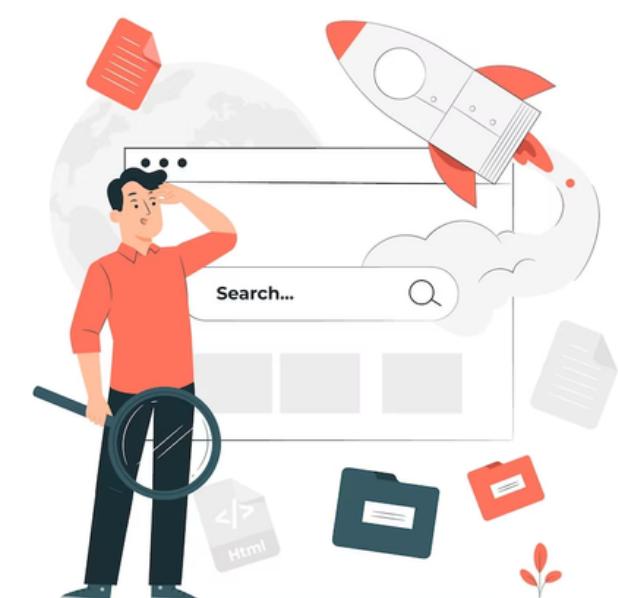
- ▶ Personality traits play a crucial role in determining job fit & job satisfaction benefitting both the individual and organization.
- ▶ Personality is more predictive of professional effectiveness than previous experience [1].
- ▶ Recruiters use personality tests as an assessment tool to help filter applicants and identify candidates with attributes that match the organizational requirements



BACKGROUND AND LITERATURE

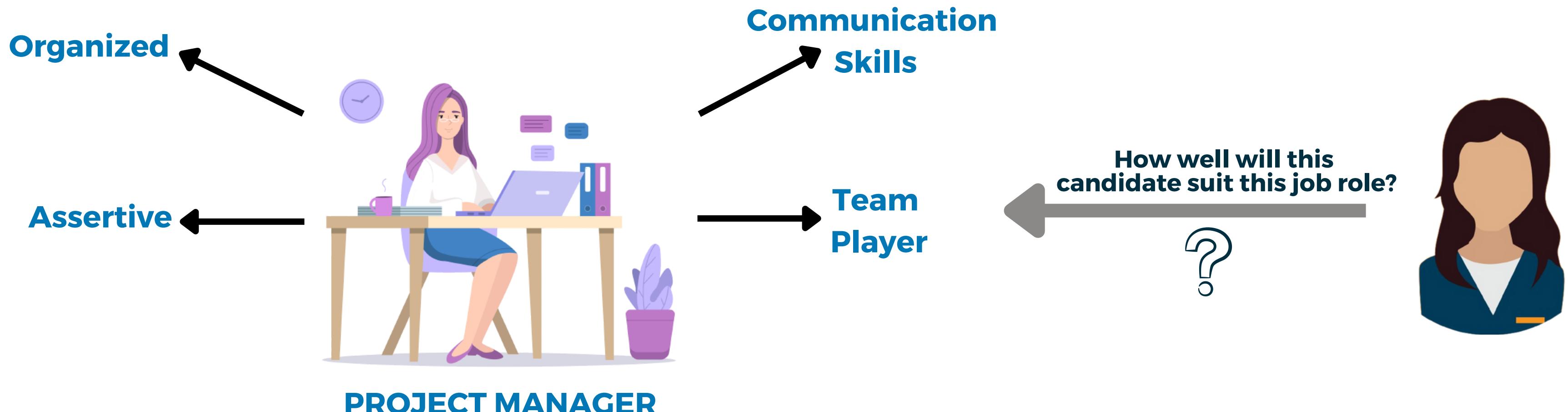
► A number of research has been done on using personality tests for recruitment.

Research	Evaluation Method	Evaluation Criteria
DevFlair: A Framework to Automate the Pre-screening Process of Software Engineering Job Candidates (2022)	Data from candidate's LinkedIn profile	Distribution of Big Five Personality Traits
A Multimodal Interviewee Evaluation Approach for Candidates Facing Video Interviews (2022)	Video interview of candidate	Distribution of Big Five Personality Traits
Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer (2020)	Phone call transcript of answers of open-ended questions	Distribution of Big Five Personality Traits



RESEARCH PROBLEM

- ▶ Existing systems do not evaluate if a candidate's personality traits match those needed for the job role.
- ▶ The purpose of this research component is to assess the candidates personality traits align with the requirements for the job role,



OBJECTIVES

MAIN OBJECTIVE

- Develop a feature to assess the compatibility between a candidate's personality and the personality traits necessary for a particular job role.

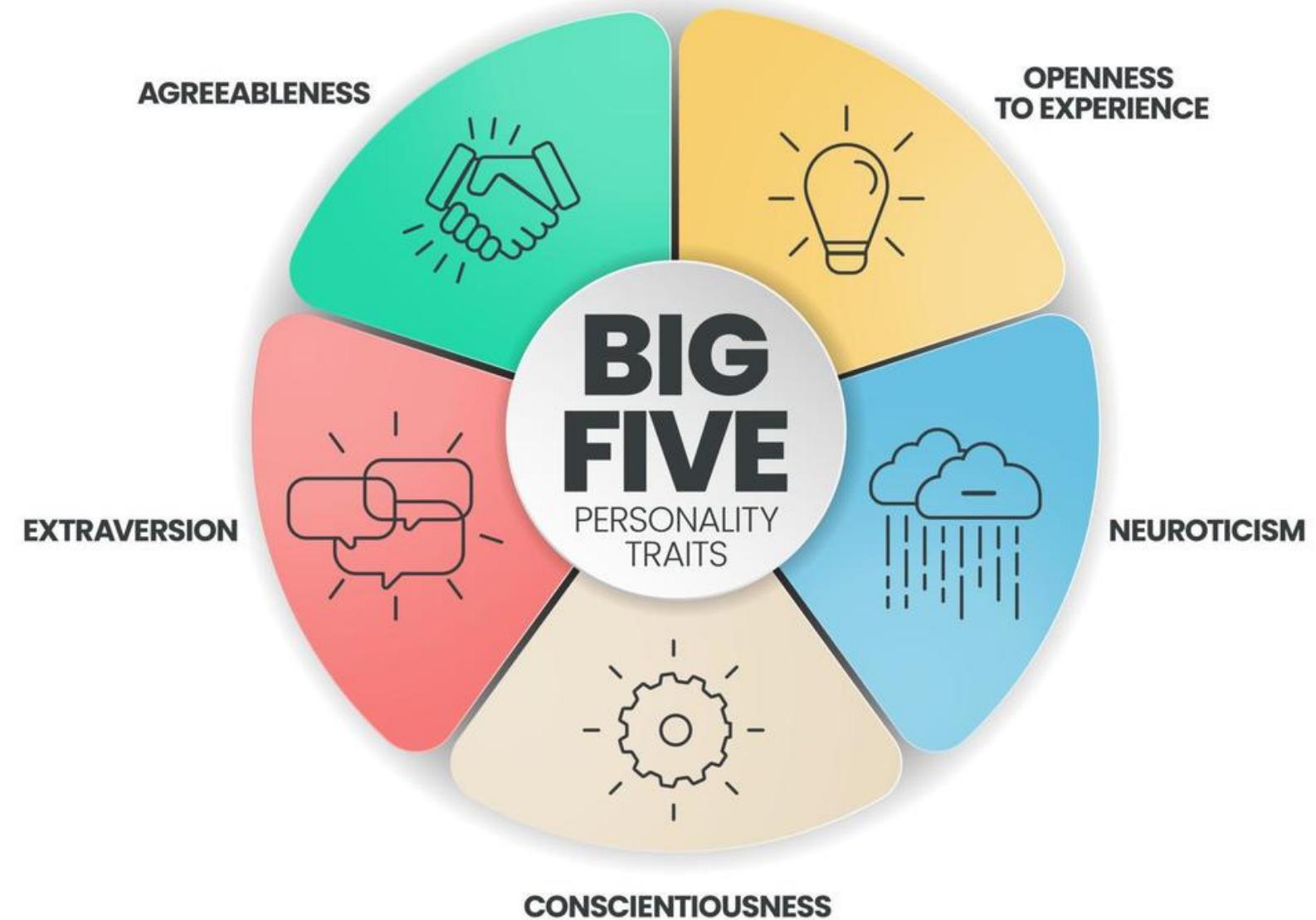
SUB OBJECTIVES

- 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.



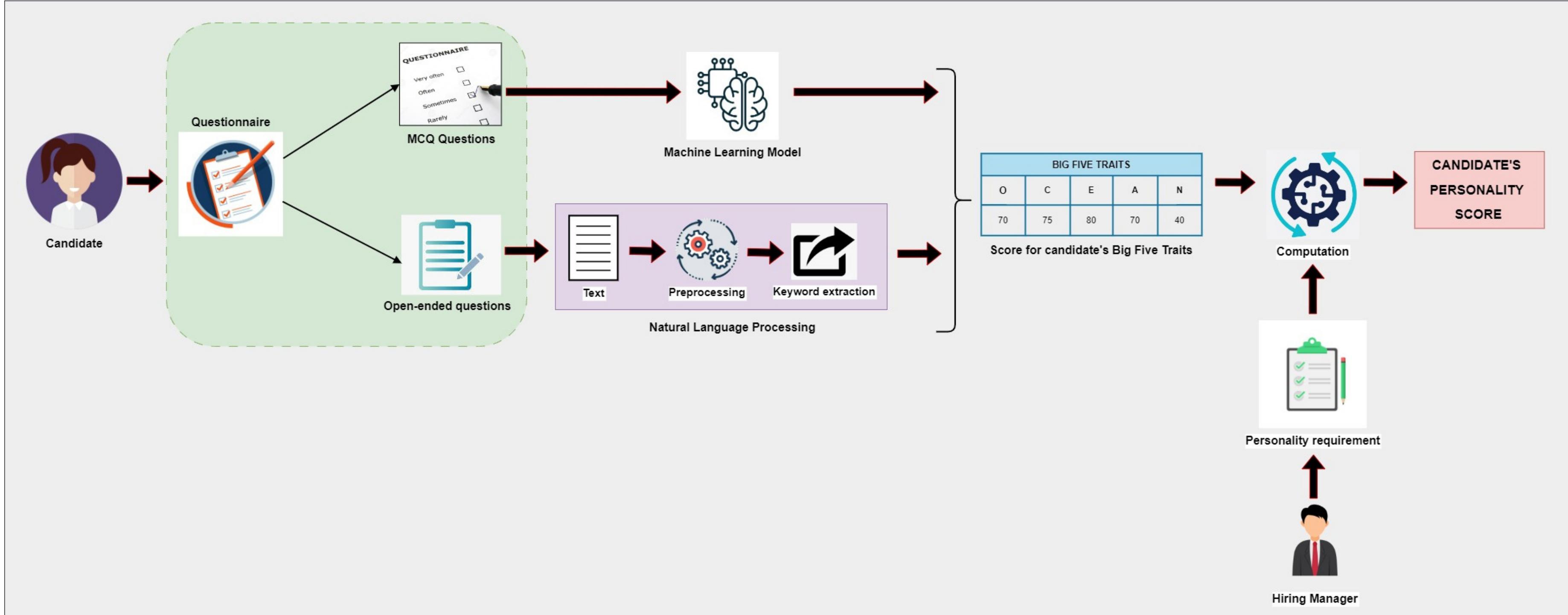
THE BIG FIVE MODEL

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



- The Big Five model stands as the most prevalent and widely-accepted trait framework in the history of personality psychology [2].

SYSTEM OVERVIEW DIAGRAM



METHODOLOGY

1 Identify the main personality clusters using K-means clustering

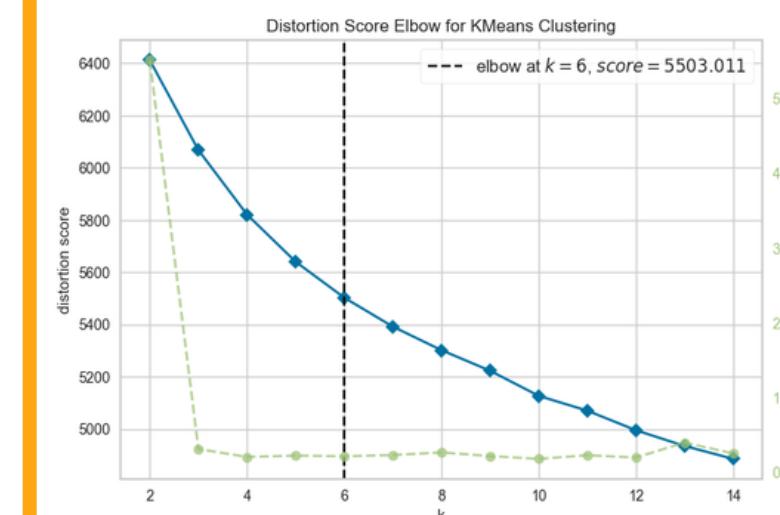
Data Preprocessing



Explanatory Data Analysis (EDA)



Elbow Visualization



Principal Component Analysis (PCA)

	PCA1	PCA2	cluster
0	-3.805313	-0.623585	2
1	1.439889	-2.491525	0
2	0.309490	-2.362485	0
3	0.539203	0.813371	3
4	-3.254717	-2.408617	2

METHODOLOGY

2 Identify keywords from the dataset to determine the Big Five Traits

Bag of Words



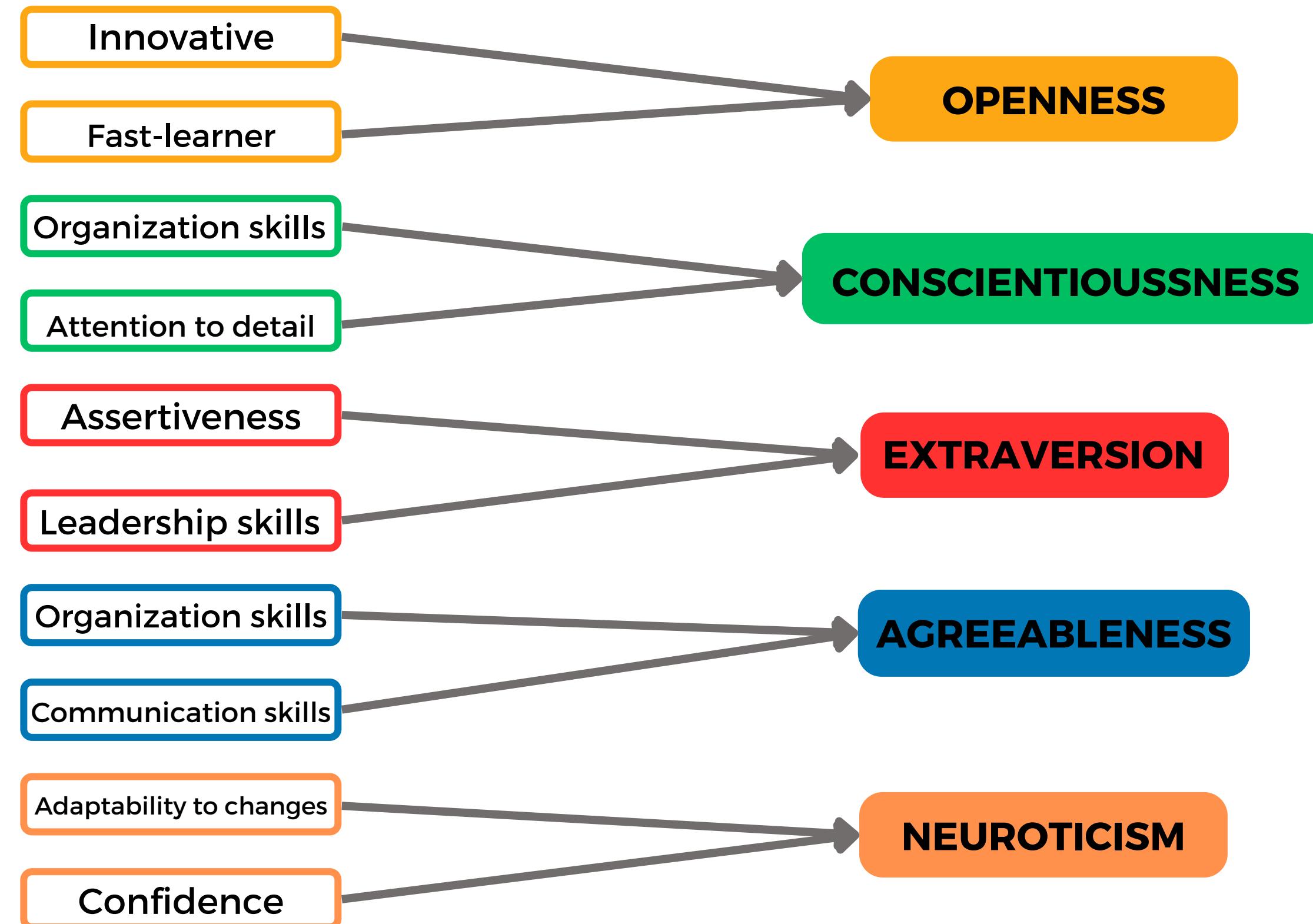
TF-IDF

keyBERT

OPENNESS	CONCIENTIOUSNESS	EXTRAVERSION	AGREEABLENESS	NEUROTICISM
new	work	connection	feedback	problem
risk	time	people	believe	stress
learn	ensure	relationship	help	difficult
try	task	social	understand	challenge
believe	complete	meeting	ask	stressful
opportunity	accurately	building	person	break
experience	prioritize	opportunity	listen	overwhelmed

METHODOLOGY

3 Mapping the personality requirements for job roles with Big Five Traits [6]



METHODOLOGY

④ Algorithm to assess the suitability of a candidate for a particular job role.

- ✓ Determine the Big Five Trait distribution of the candidate
- ✓ Determine the personality cluster(from K-means clustering) that a candidate belongs to and identify behaviours.
- ✓ Compare the difference between the candidate's personality traits and expected personality traits

$$\text{Score} = \Sigma \left(\frac{\text{trait(candidate)}}{\text{trait(expected)}} \right) * \frac{100}{5}$$



DEVELOPED SOLUTION

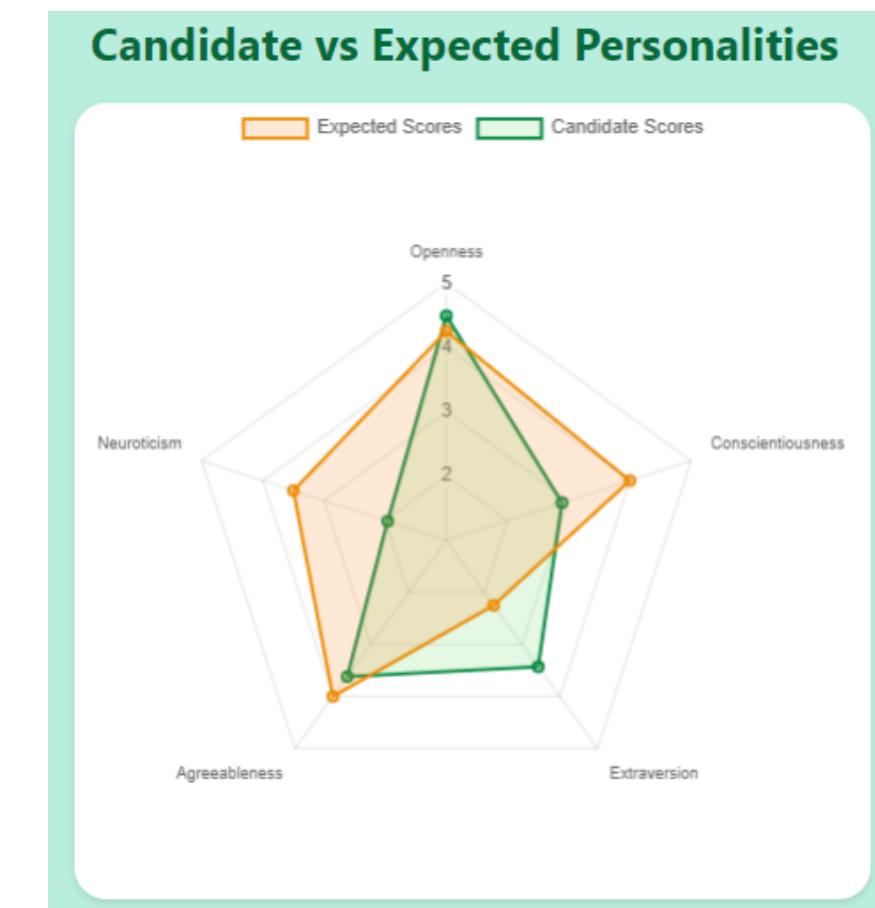
- 3 user-friendly interfaces that enable recruiters to gain a comprehensive insight into a candidate's personality effortlessly.

Candidate Name: Manushika Maldeniya

	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Final Scores	4.49	2.89	3.43	3.62	1.96
Expected Scores	4.25	4.0	2.25	4.0	3.5

Candidate personality - Job role match: 79.75%

Show details



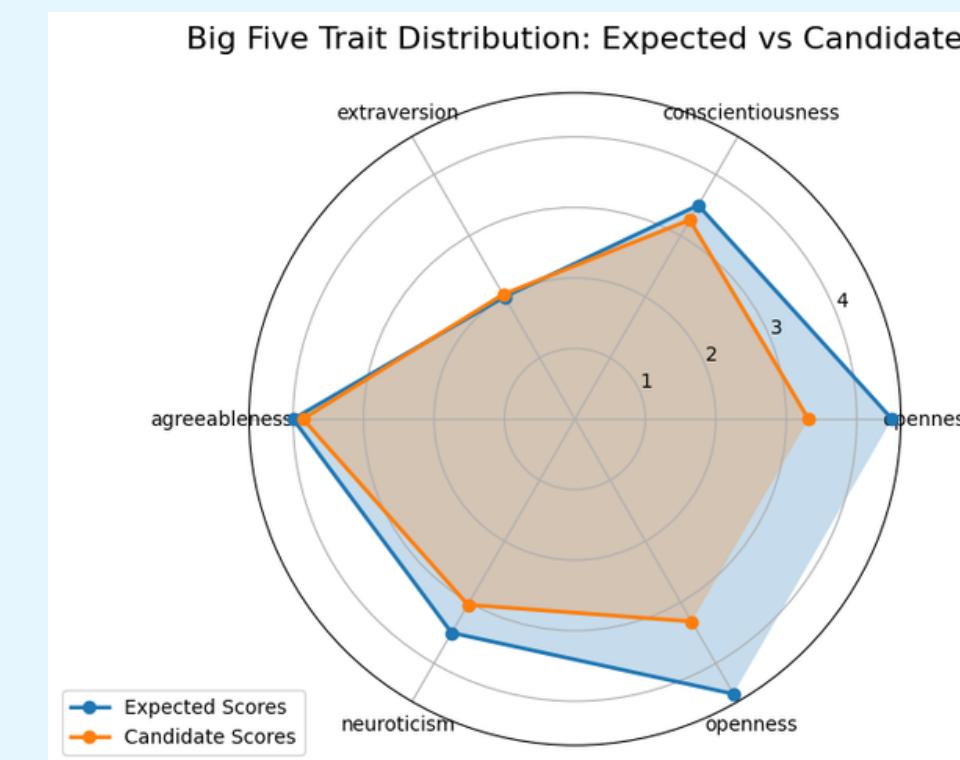
This personality suggests that the individual may be more relaxed and flexible in their approach to tasks and responsibilities. Additionally, they exhibit a high degree of agreeableness, suggesting they are generally cooperative, friendly, and considerate in their interactions with others.

[Demo Video>>](#)

RESULTS AND DISCUSSION



MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
Random Forest	0.99	0.992	0.992	0.992
Naive-Bayes	0.96	0.958	0.97	0.97
XGBoost	0.95	0.96	0.96	0.96



REFERENCES

- Dataset
- [1] A. Furnham, K. V. Petrides, C. J. Jackson, and T. Cotter, “Do personality factors predict job satisfaction?,” *Personality and Individual Differences*, vol. 33, no. 8, pp. 1325–1342, Dec. 2002, doi: 10.1016/S0191-8869(02)00016-8.
- [2] T. A. Judge and C. P. Zapata, “The Person-Situation Debate Revisited: Effect of Situation Strength and Trait Activation on the Validity of the Big Five Personality Traits in Predicting Job Performance,” *AMJ*, vol. 58, no. 4, pp. 1149–1179, Aug. 2015, doi: 10.5465/amj.2010.0837.
- [3] 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.
- [4] W. G. Y. Randika, M. T. A. R, K. L. O. G. Liyanage, A. Karunasena, and K. M. L. P. Weerasinghe, “A Multimodal Interviewee Evaluation Approach for Candidates Facing Video Interviews,” in 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India: IEEE, Oct. 2022, pp. 1–7. doi: 10.1109/ICCCNT54827.2022.9984585.
- [5] 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: IEEE, Dec. 2020, pp. 168–173. doi: 10.1109/ICAC51239.2020.9357279.
- [6] J. Darby, “HR Blog,” thomas.co, May 16, 2023. <https://www.thomas.co/resource-type/hr-blog>

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

INTRODUCTION AND BACKGROUND

Academic transcripts offer critical insights into a candidate's qualifications and skills, serving as a key component in the evaluation process.

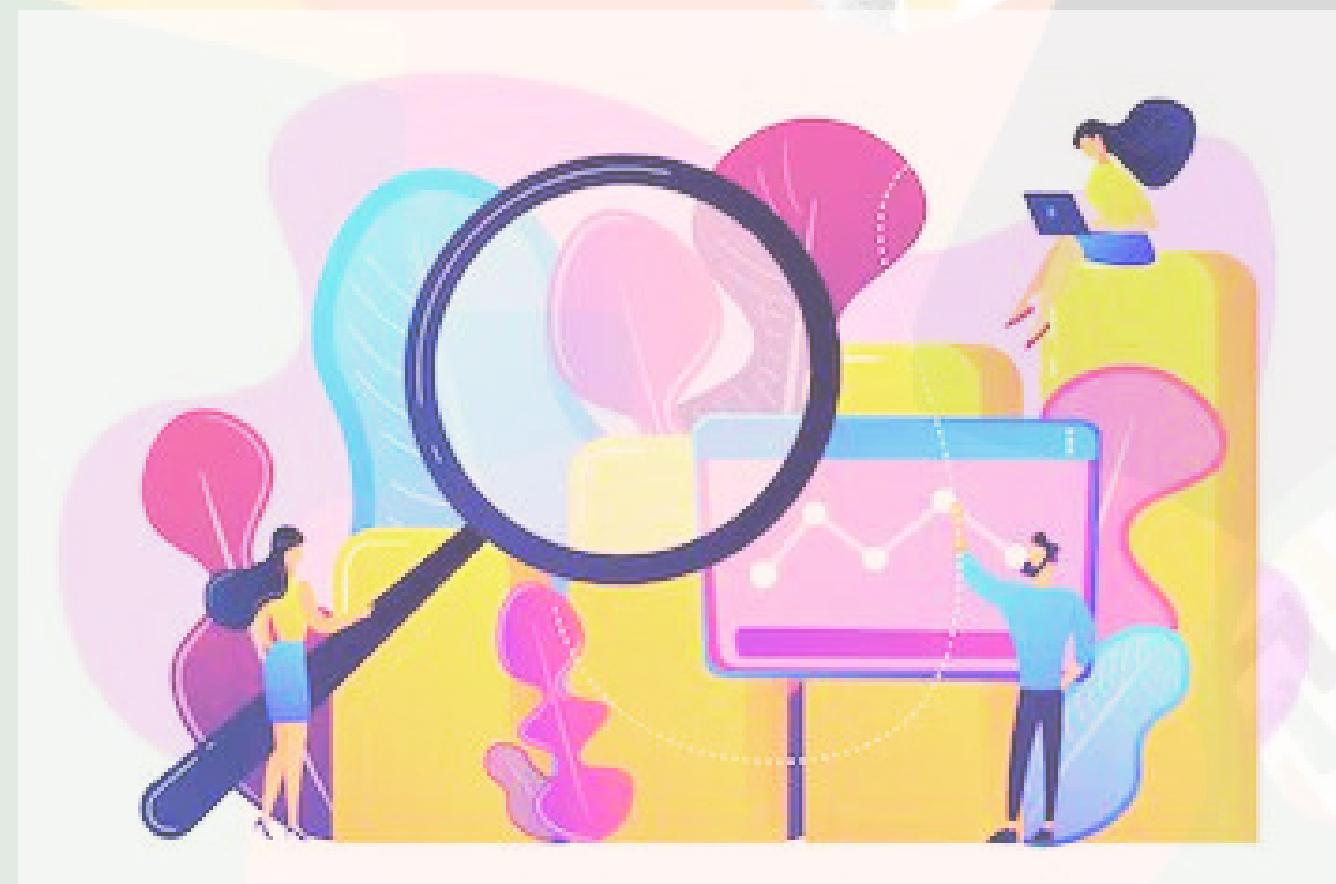
Recognizing the importance of academic transcripts for assessing technical skills helps make well-informed decisions in hiring.



PROBLEM DEFINITION

Existing systems lack precision in the analysis of academic transcripts, leaving room for incomplete evaluations of candidates.

It also fall short in providing precise assessments, especially when leveraging module outlines from universities for a more thorough evaluation.



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.

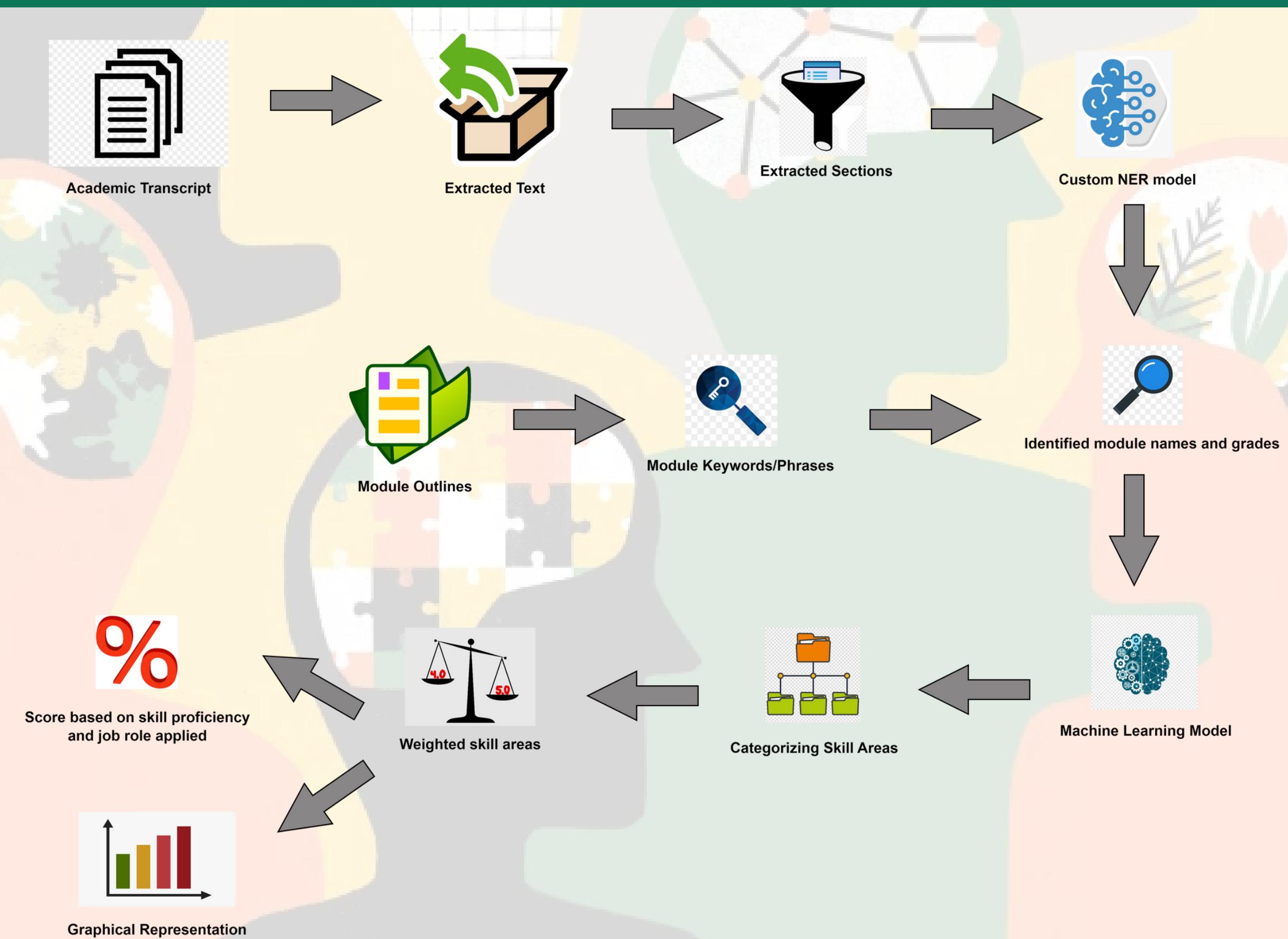
SUB OBJECTIVES

- 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.

LITERATURE REVIEW

	Candidate skill analysis using academic transcript	Combining Thematic Analysis with Natural Language Processing (NLP)	Analysing academic transcript
Use academic transcript	✓		✓
Use module outline	✓	✓	✓
Graphically represent a candidate's expertised areas			✓
Focuses on the whole IT industry	✓		✓
Based in Sri Lanka			✓

SYSTEM OVERVIEW DIAGRAM



METHODOLOGY



MODULE OUTLINE KEYWORDS

Extract the important keywords from the module outlines



Collected Module outlines from
a specific university

Data-preprocessing

n-grams



Module keywords/Phrases

ACADEMIC TRANSCRIPT EXTRACTION



Academic Transcripts

Convert to grayscale

Remove light noise

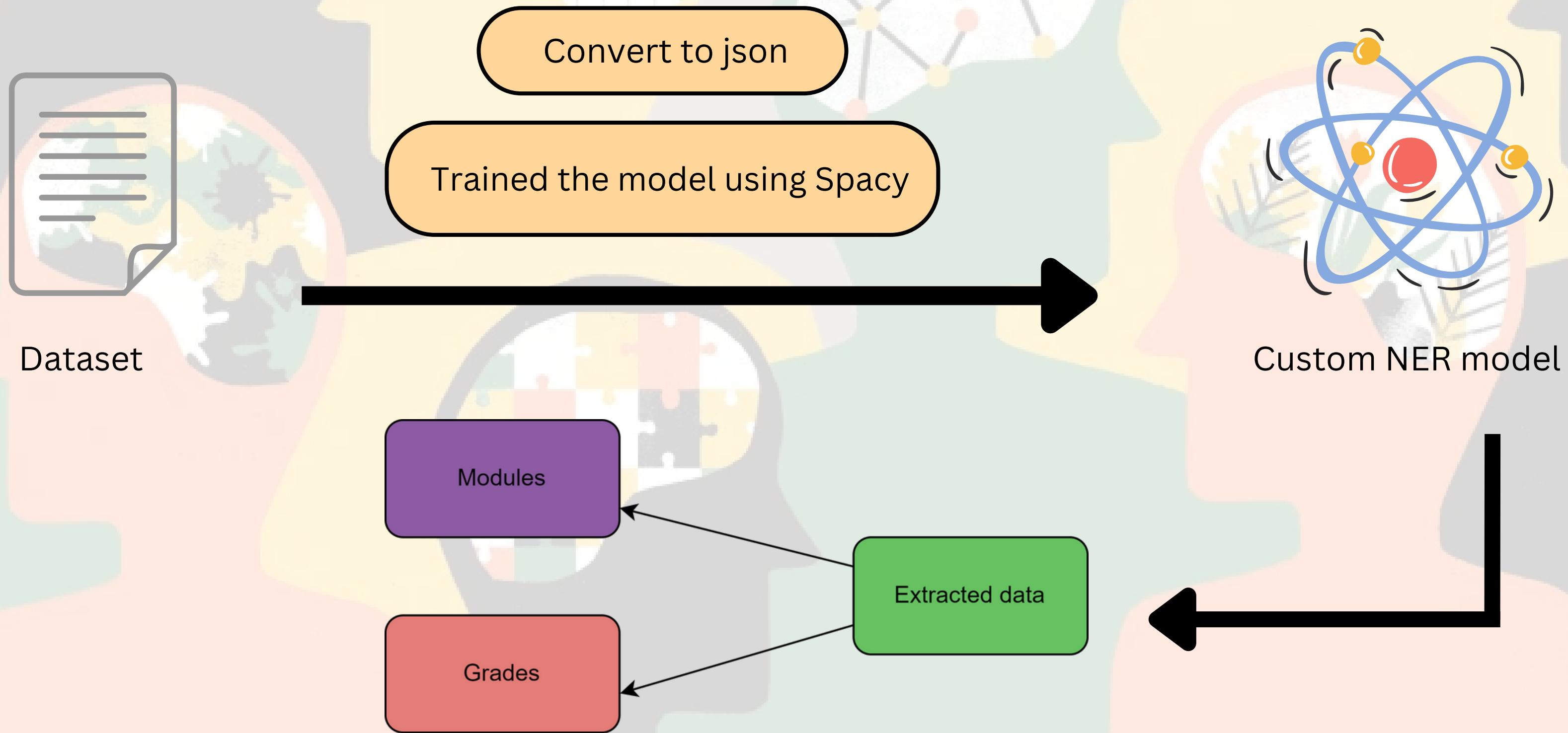
Adjust Configurations

OCR Pytesseract and Poppler

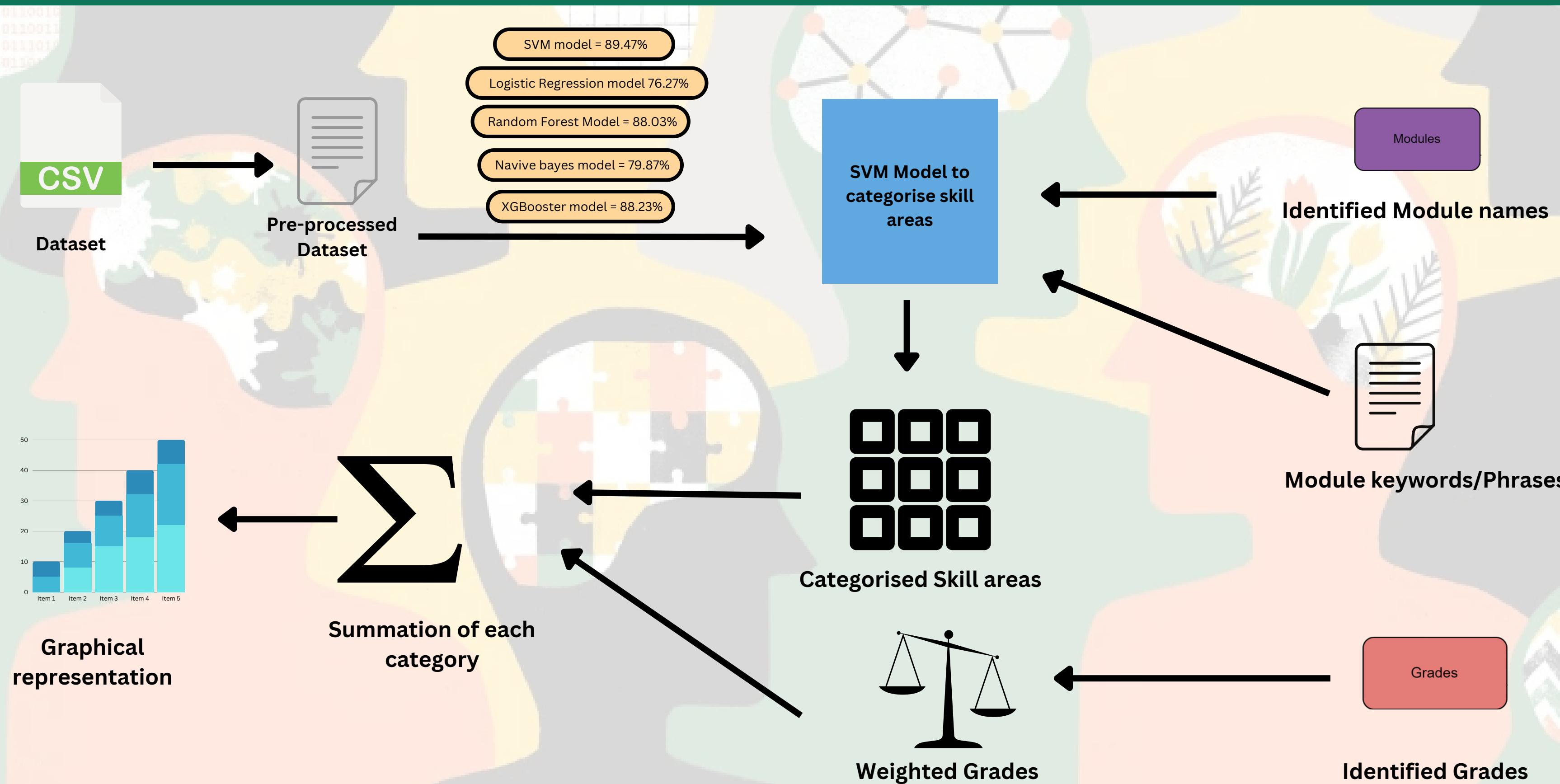


Extracted data of an Academic Transcript

CUSTOM NER MODEL



SKILL AREA PREDICTION AND CATEGORIZATION



SCORE PREDICTION MODEL

Σ

Summation of each category



Job Role

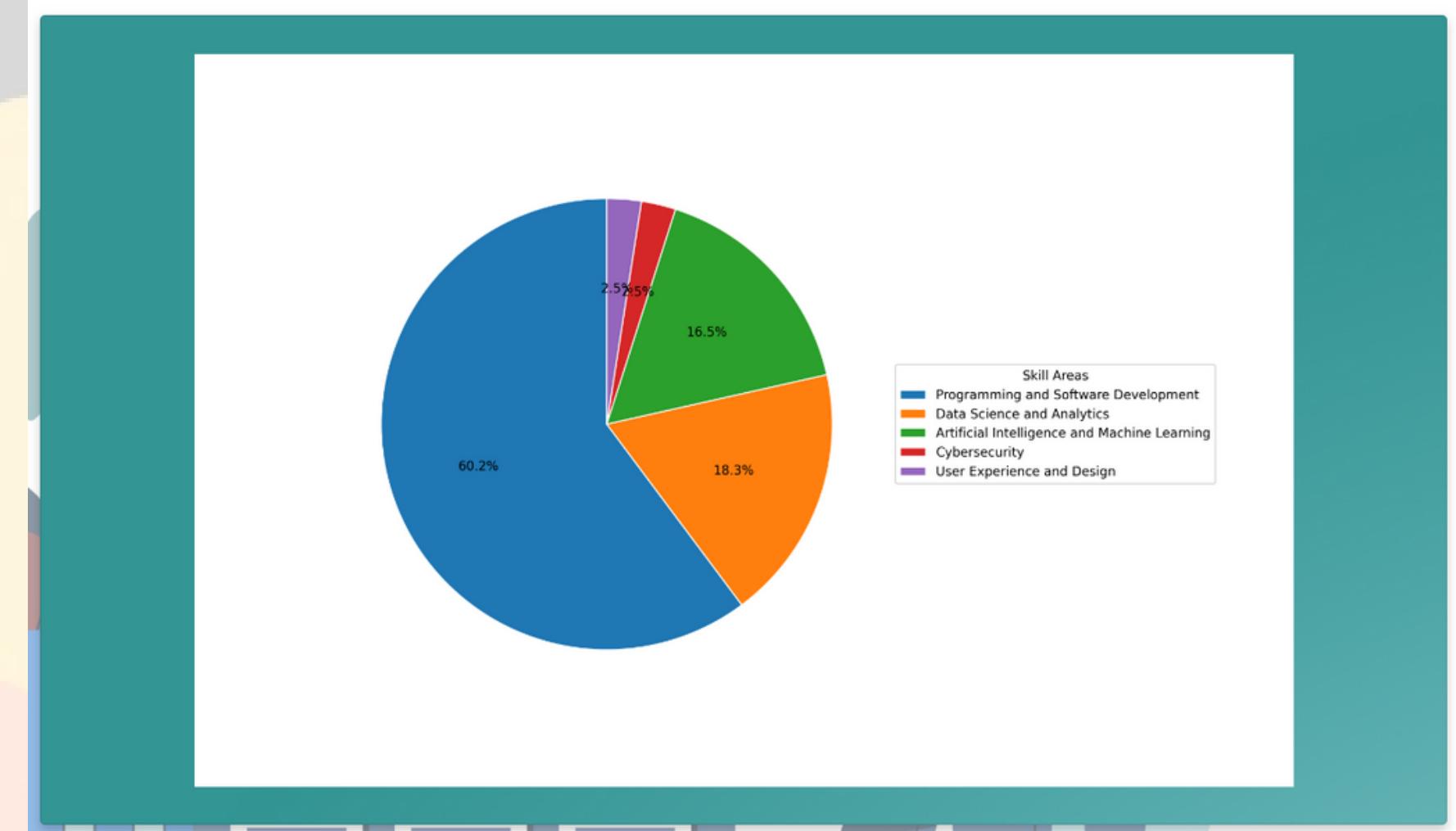
RandomForestRegressor
MSE = 75%

~~100~~

Score based on skill proficiency and job role applied

DEVELOPED SOLUTION

When an academic transcript is uploaded, the solution is able to analyse it and categorise the skill areas



PROGRESS



CURRENT PROGRESS

- ✓ Extracted the important keywords from the module outlines
- ✓ Built and trained a Custom Name Entity Recognition(NER) Model
- ✓ Trained a model to identify the Skill Area
- ✓ Extracted Transcript data using Optical Character Recognition
- ✓ Categorized the the modules into different skill areas
- ✓ Graphical representation of skill areas
- ✓ Calculated a score for a candidate to match the job role based on their skill areas



CURRENT PROGRESS

- ✓ Optimized OCR text extraction by converting the scanned images to grayscale and applying a threshold to remove light noise, and used Tesseract OCR with specific configuration settings to enhance accuracy and readability.
- ✓ Enhanced the custom NER model's accuracy
- ✓ Optimised the accuracy of skill area prediction model
- ✓ Front end implementation
- ✓ developed the web application



EXPECTED FUTURE PROGRESS

- Deploying the web application
- Fine tuning the web application
- Storage Implementation

100%

RESULTS



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
IT3110* 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}
```

N-gram output

MODEL BUILDING

Accuracy of XG Booster Model

↳ Accuracy: 0.8823151125401929

Categorization of the skill areas

	Category	Module Title	Module Keywords	Weighted Grade
0	Programming and Software Development	introduction programming	data types, control structures, functions, poi...	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

Accuracy of SVM Model

0.8947539659568061

Accuracy of Naive Bayes Model

↳ Accuracy: 0.7987138263665595

Accuracy of Logistic Regression Model

↳ Accuracy: 0.7627
/usr/local/lib/python3.16

Accuracy of Random Forest Model

↳ [nltk_data] Downloading package...
[nltk_data] Package stopword...
Accuracy: 0.8803858520900322

Accuracy of Random Forest Model

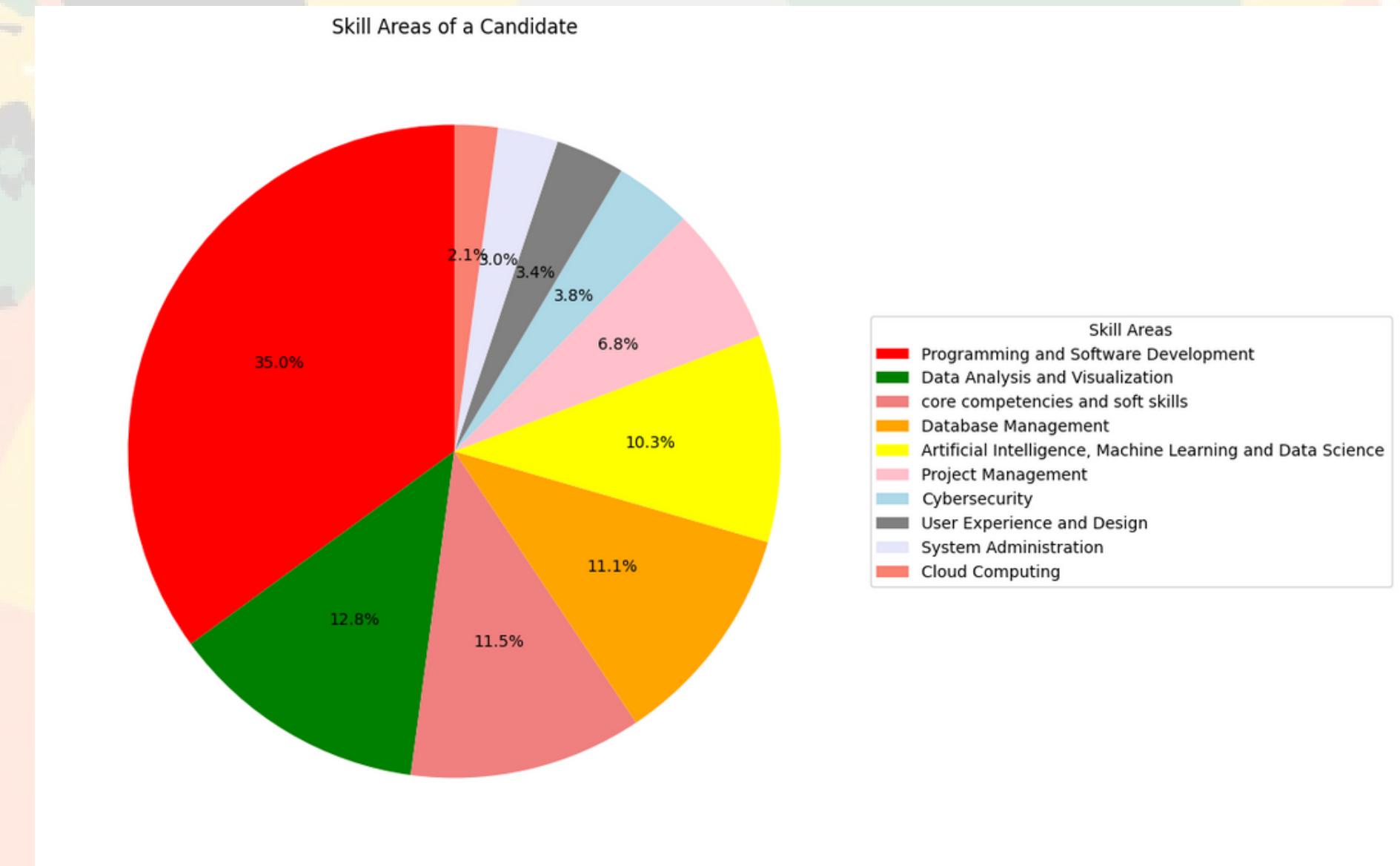
↳ Mean Squared Error: 75.54
R-squared (R2) Score: -0.08

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

PIE CHART



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)

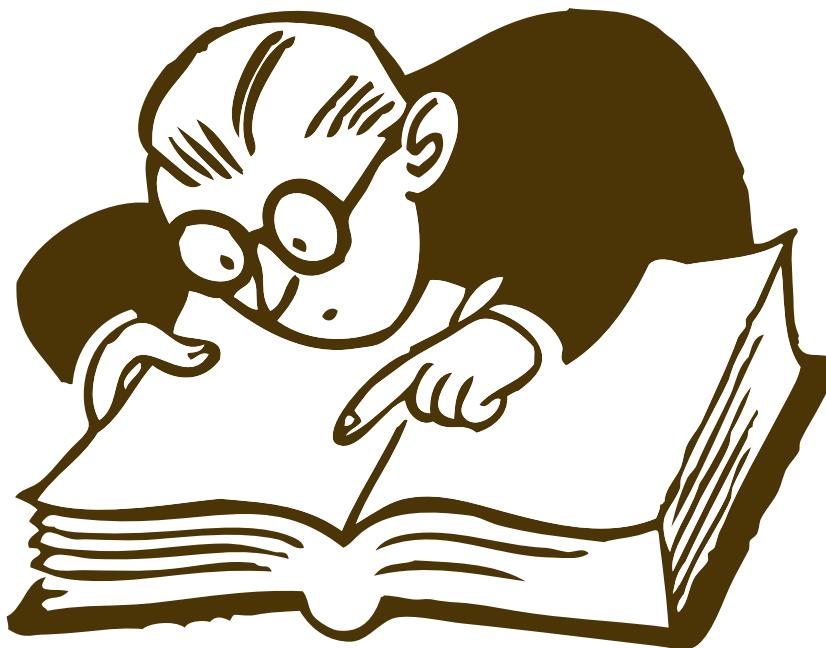
IT20231200 | Zoysa E.S.

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



PROFESSIONAL SKILL ANALYSIS USING DIGITAL FOOTPRINTS

BACKGROUND STUDY



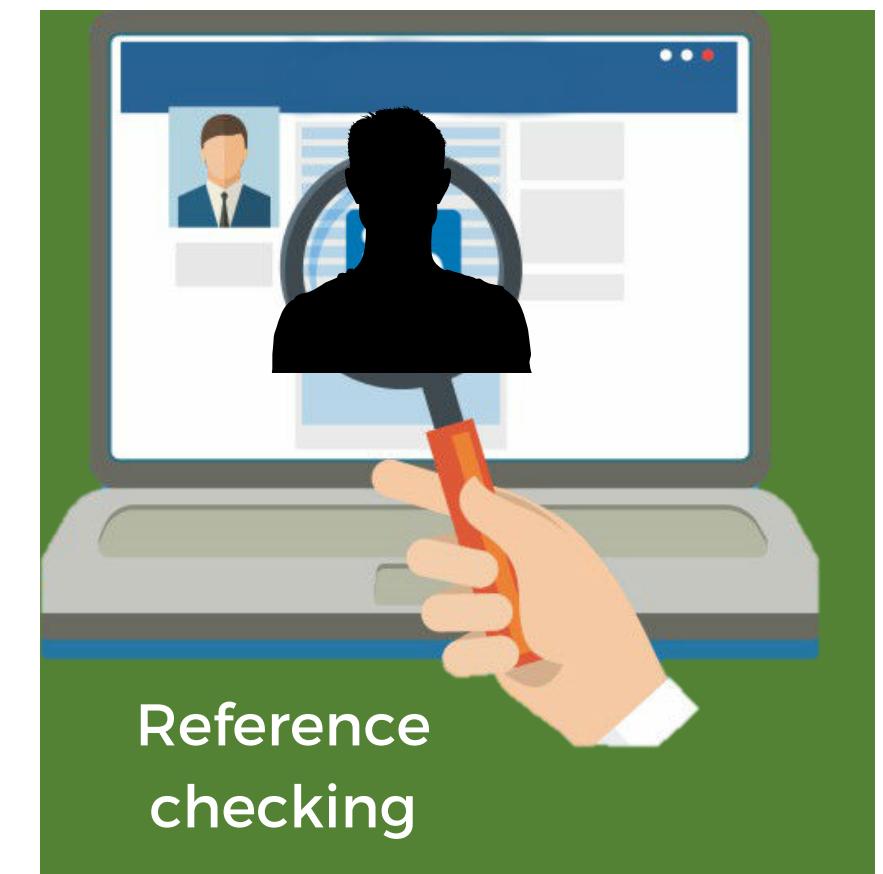
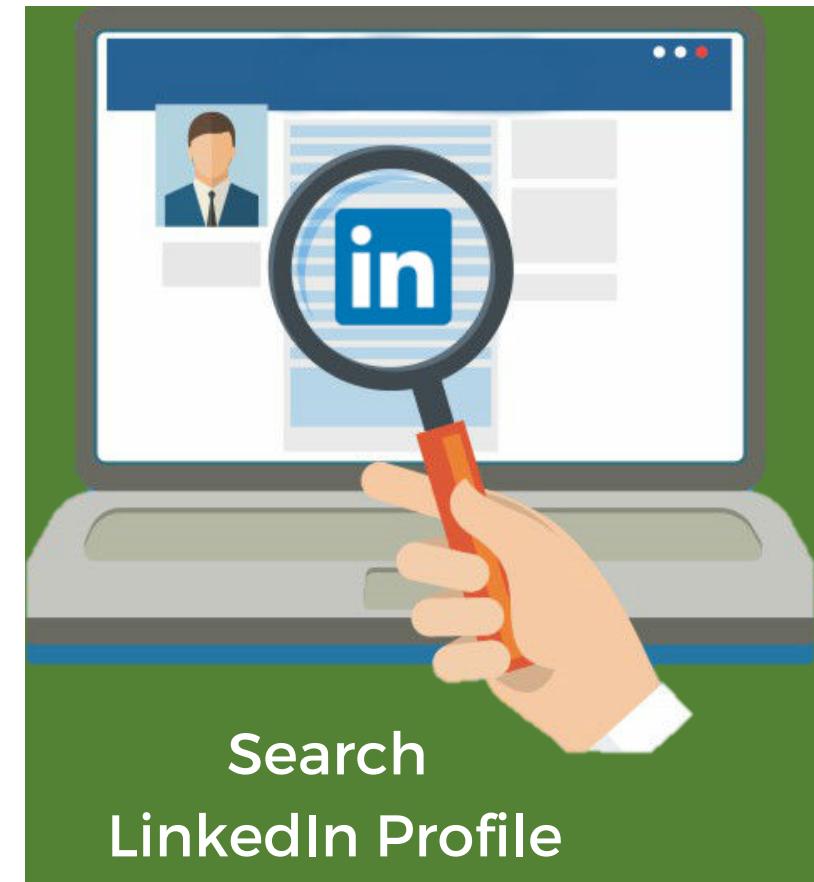
- In today's digital age, companies are increasingly relying on automated online job recruitment processes to screen potential candidates quickly and efficiently
- As part of this process, candidate profiles on social media and professional networking sites are being analyzed.
- In order to effectively 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.

BACKGROUND STUDY

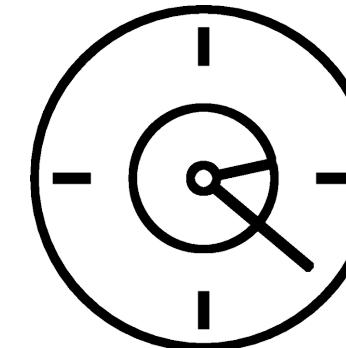
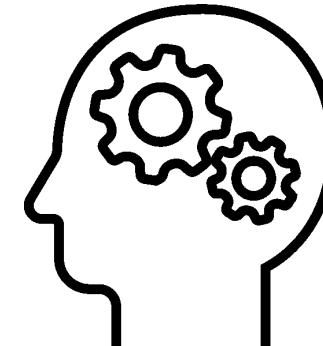
Why we need a tool ?



In a case of
CV flood



Less
Productivity



Time
consuming

Literature Review

Feature Research	Based GitHub	Based on LinkedIn	Based on both GitHub and LinkedIn	Evaluate Professional skills	Sentiments Analysis on background check	Job Category prediction	Cross Validate technical skills
Research A [7]				✓			
Research B [5]		✓		✓			
Research C [2]				✓			
Research D [1]	✓						
Research E [8]	✓	✓	✓				
Proposed system	✓	✓	✓	✓	✓	✓	✓

RESEARCH PROBLEM

- Are CV's reliable? Can we go **beyond the CV** and asses a candidate?
- Can we **optimize** the recruitment process without human interefrence?
- Can we go for better **decesion making** with the **trending technologies**?



OBJECTIVE

The main objective of this component is to give the opportunity for a HR system to evaluating and confirm candidate's professional skills through professional social media platforms like GitHub and LinkedIn user profile and employee background checking with less human interference.

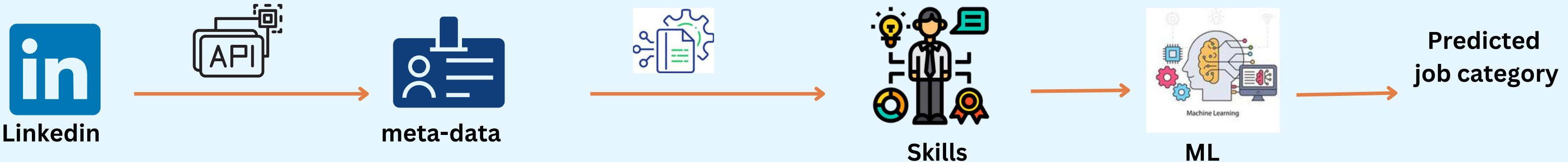
SUB OBJECTIVES

- Extracting the content of the candidate's LinkedIn and GitHub user profile and Reference feedback responses.
 - Analyse programming language proficiency
 - Matching job category based on LinkedIn mentioned skills.
 - Cross validate the skills and output
 - Summary representation of background checking

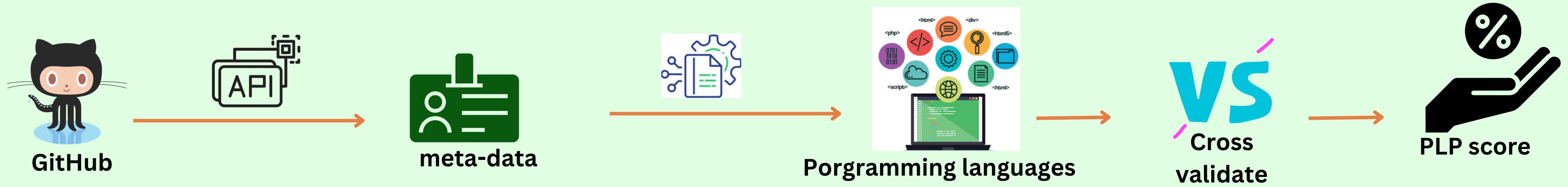


SYSTEM OVERVIEW DIAGRAM

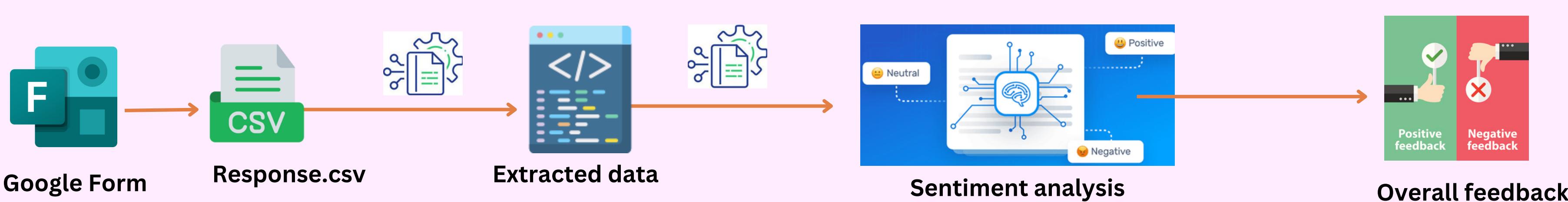
LinkedIn based job category prediction



GitHub programming language proficiency



Employee reference checking



LinkedIn job category prediction

Dataset creation using a survey

Data pre-processing

- remove missing values
- get unique categories
- text-vector transformation
- find the top 3 co-related terms

Prediction - Job Category



LinkedIn Scraping

- Authenticate with LinkedIn
- ProxyCurl API scraping
- Scrape Skills



Hyper parameter tuning for Linear SVC

- Linear SVC - 66.38%

List and Loop through list of multi class classification models

- Random forest classifier - 57.9%
- LinearSVC - 63.6%
- MultinomialNB - 55.0%
- Logistic Regression - 57.9%
- XGBClassifier - 43.7%

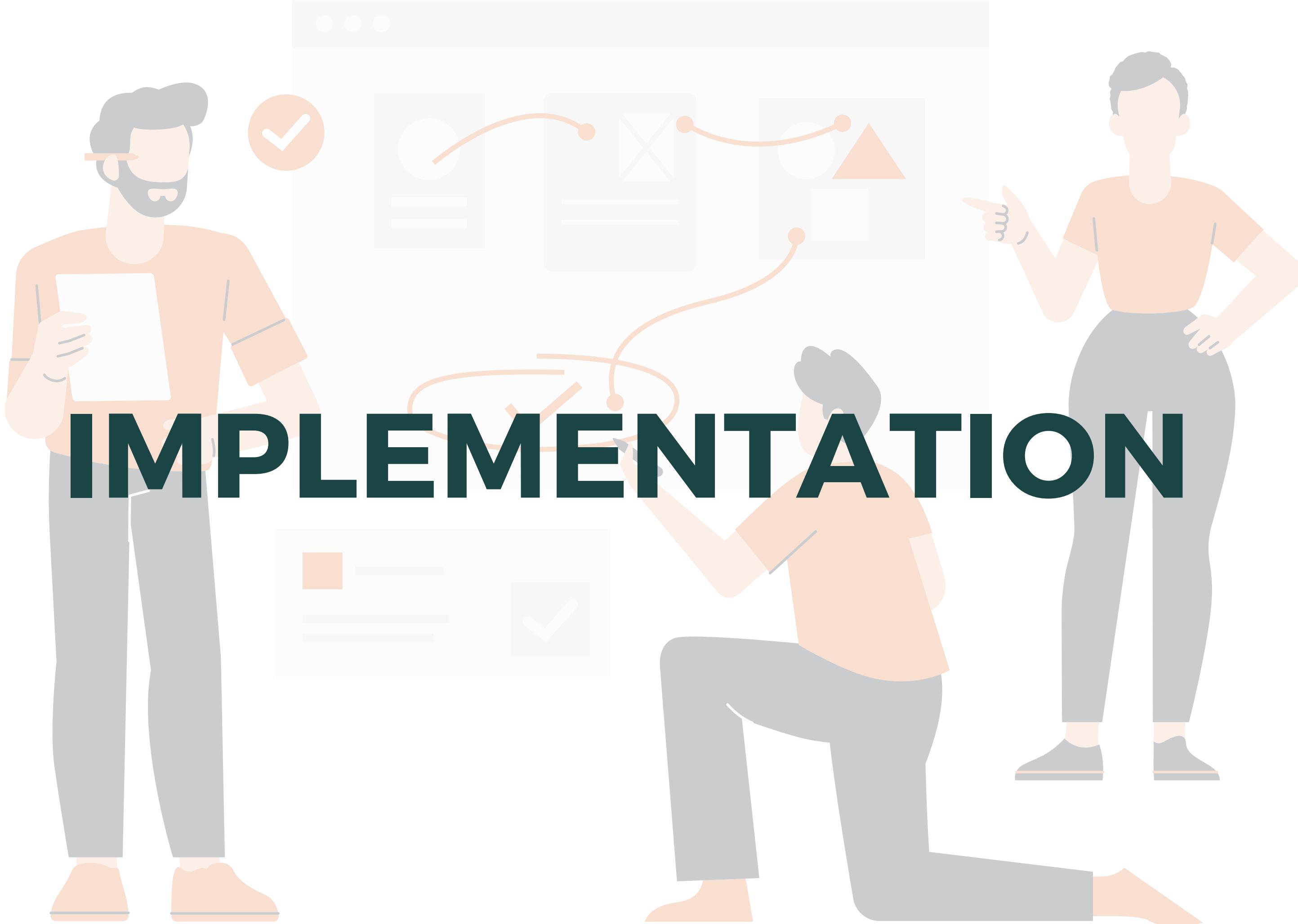
Cross validation

COMPONENT ACROSS CROSS VALIDATION

Programming language proficiency (PLP)



IMPLEMENTATION





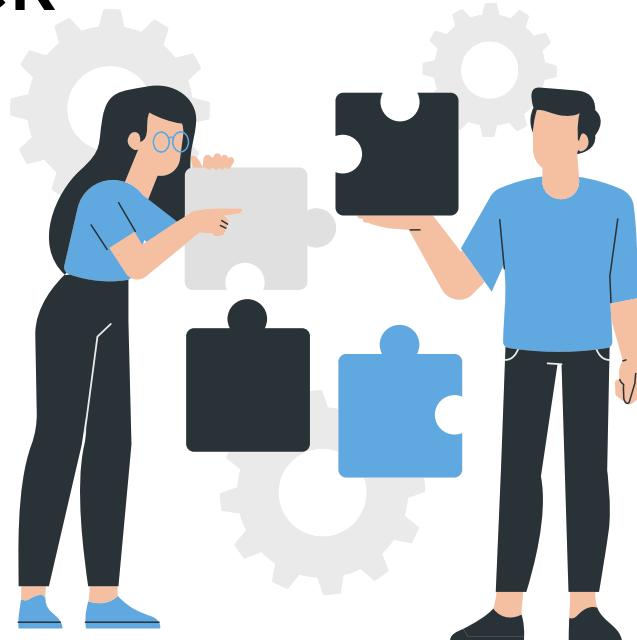
PROGRESS



Developed Solution

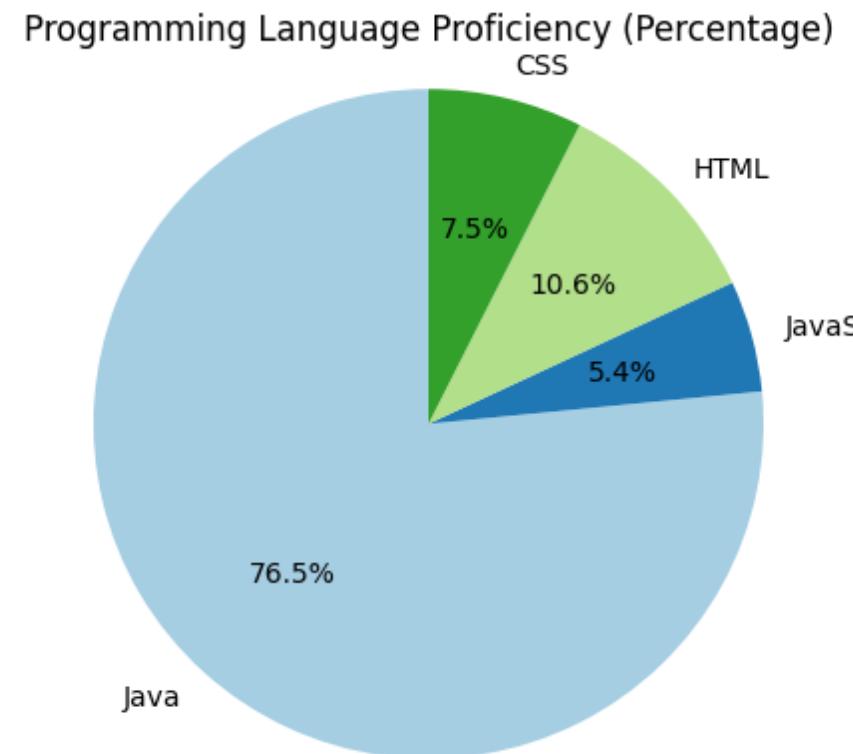
Current implementation is able to,

- “[GitHub Polygraph](#)” - Profile candidate overall programming language proficiency in the context of candidate all public repositories in GitHub user account. A proficiency comparator option analyze multiple candidates.
- “[SkillSage-LinkedIn Alchemist](#)” - Predict & find the most relatable job category based on the skills in the mention in the LinkedIn profile.
- “[PastPort -Reference Evaluator](#)” - Analyze candidate’s background check responses and visualize the sentiment.



Results and Discussion

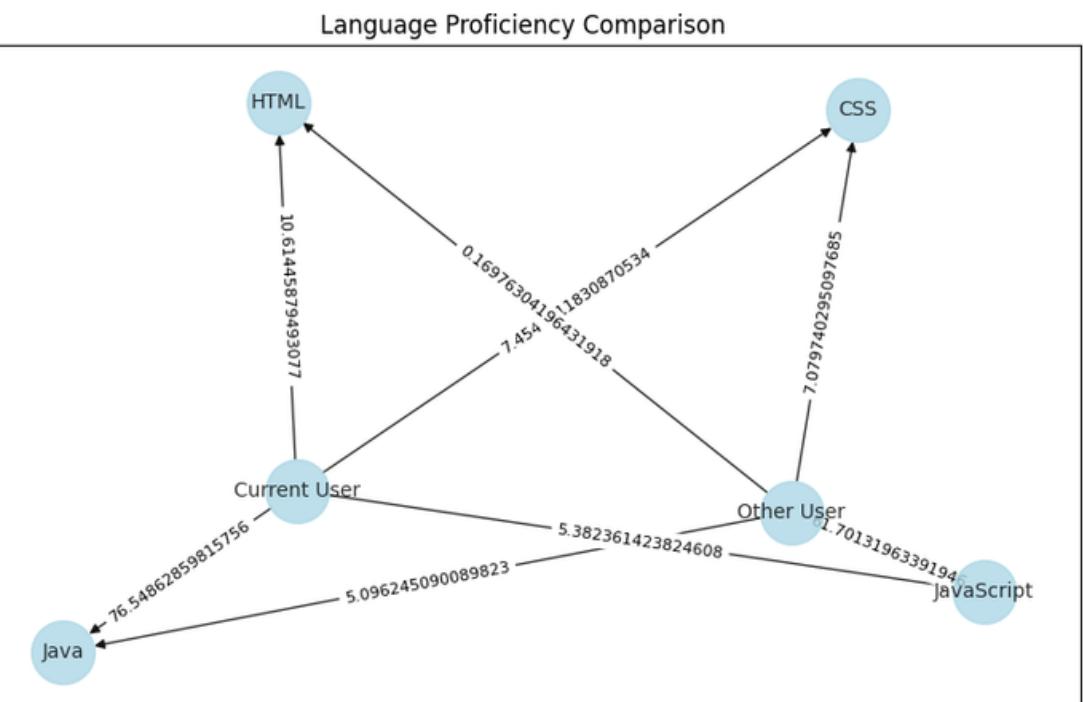
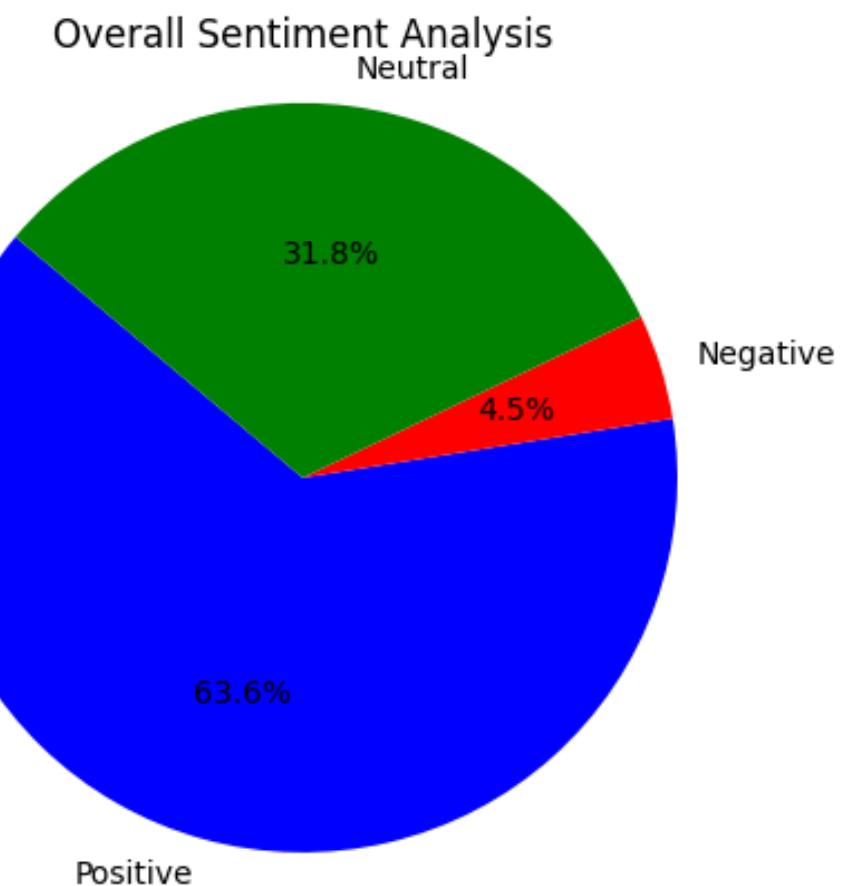
Github



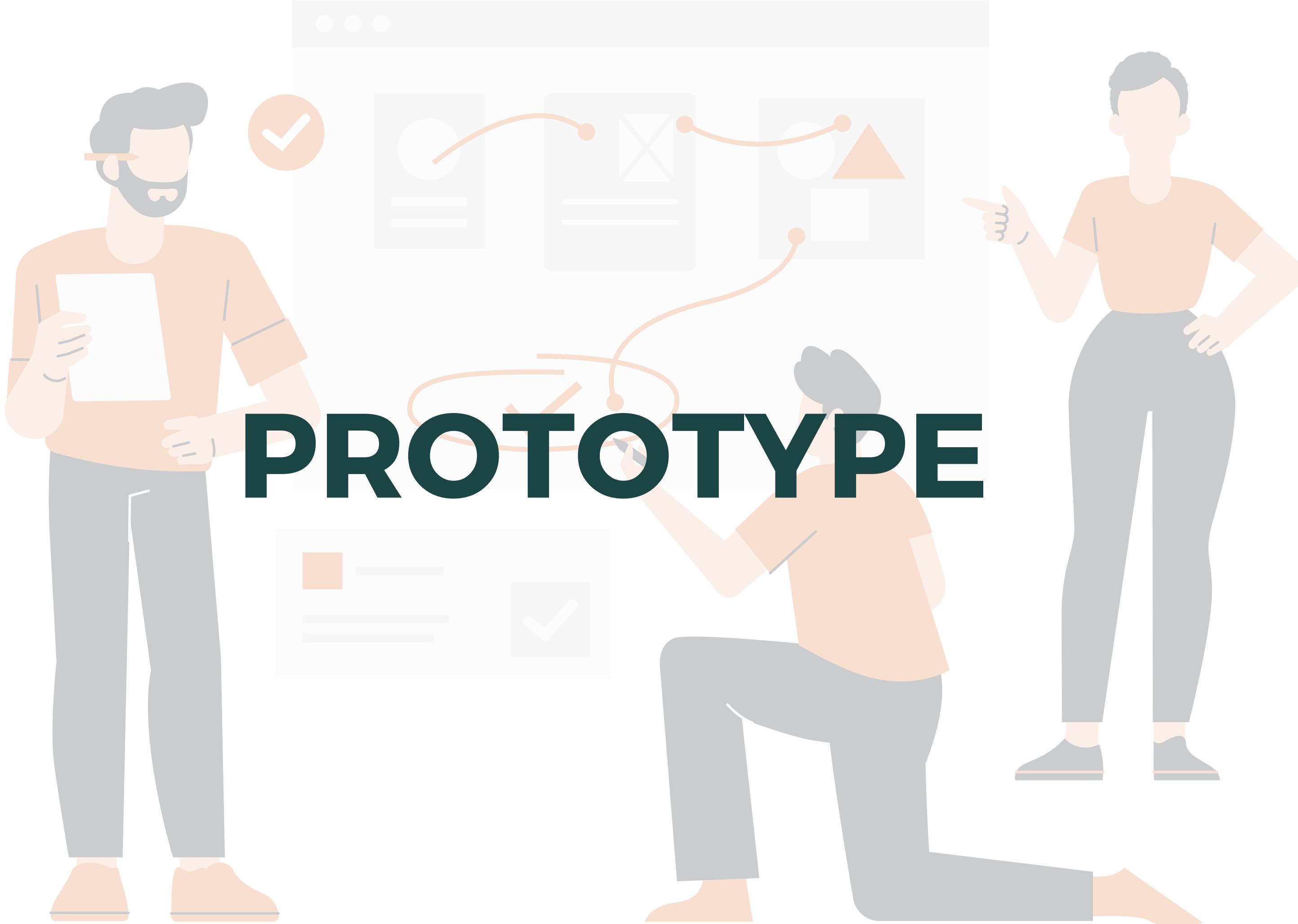
LinkedIn

LinerSVC	66.38%
LogisticRegression	57.90%
XGB Classifier	43.70%
Random Forest	57.90%
MultinomialNB	55.0%

References

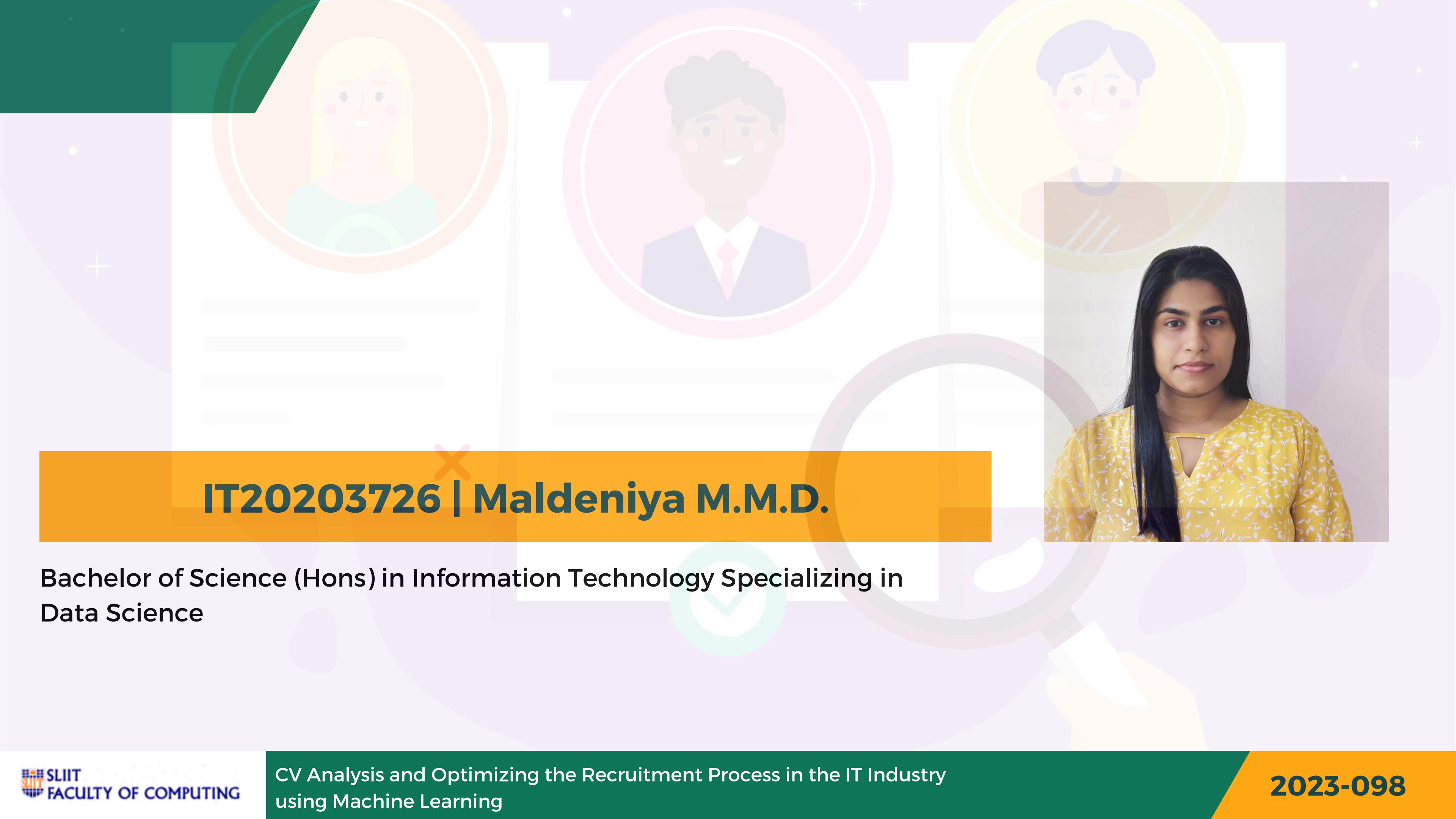


PROTOTYPE



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.



IT20203726 | Maldeniya M.M.D.

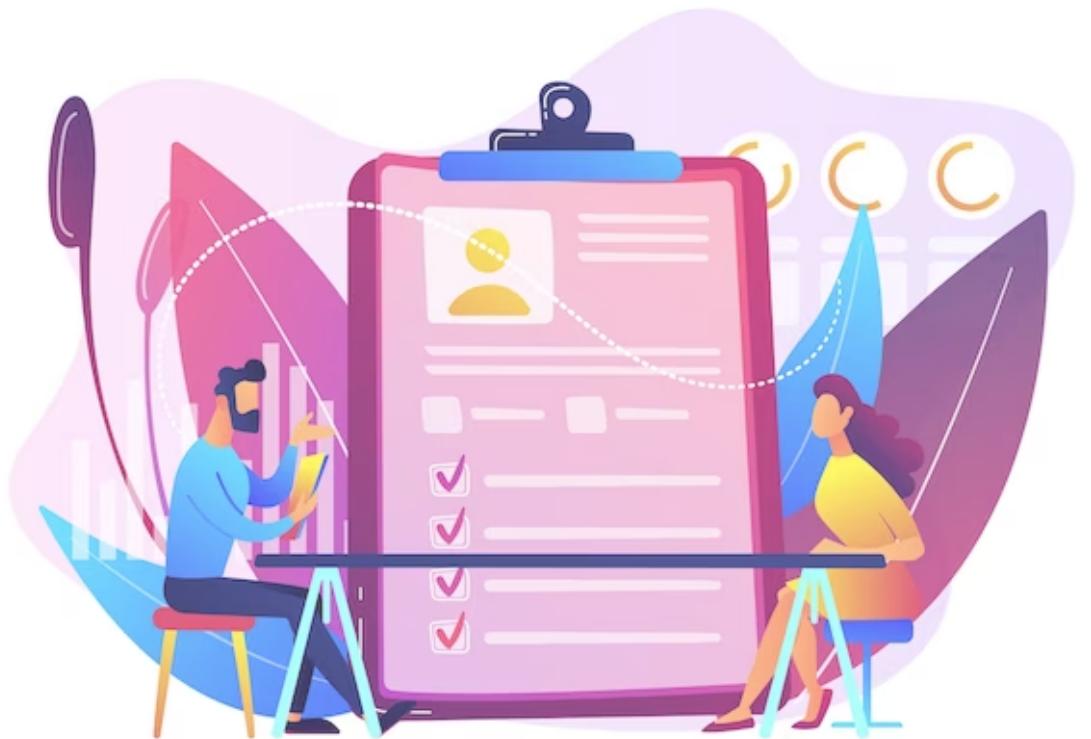
Bachelor of Science (Hons) in Information Technology Specializing in
Data Science



OPTIMIZING CANDIDATE SELECTION THROUGH CV AND JOB DESCRIPTION MATCHING TECHNIQUES

BACKGROUND AND LITERATURE

- ▶ CV ranking streamlines candidate selection in modern recruitment.
- ▶ The initial responsibility for any recruiter is to thoroughly review the CVs of all job applicants,
- ▶ Matching cv with job description helps to identify applicants with right talent.



BACKGROUND AND LITERATURE

RESEARCH

Evaluating Automatic CV Shortlisting Tool
For Job
Recruitment Based On Machine Learning
Techniques

Screening and Ranking Resumes using
Stacked Model

A Job Post and Resume Classification System
for Online Recruitment

Developed Solution

DATA COLLECTION

CV text data from Linkdein
Profiles

Resumes of candidates and
their job roles

candidate resumes in
structured format

**candidate resumes and
generated job description
in structured format**

RANKING TECHNIQUE

Cosine similarity computation
Ranking

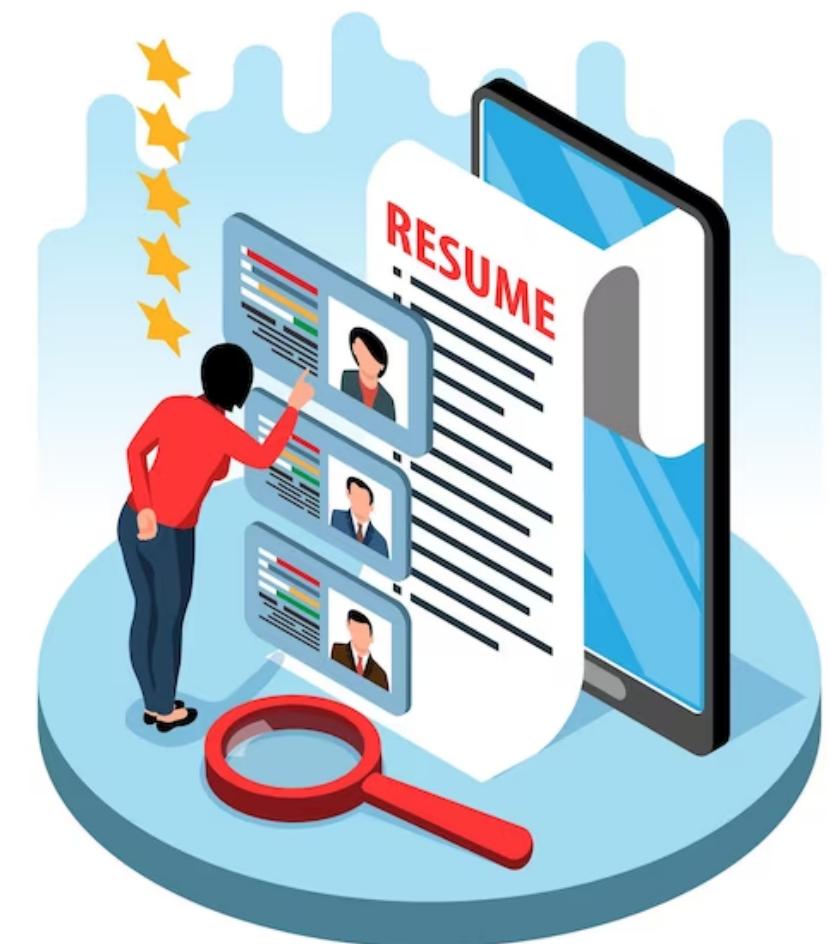
stacked classifier to predict job
role with cosine simmilarity

Skill-Based Resume
Classification Ranking

**Stacked regressor model to
predict Matching percentages**

RESEARCH PROBLEM

- ▶ Traditional resume screening is slow and tedious
- ▶ Matching CVs to job descriptions is challenging due to the use of unstructured, manually written job descriptions.
- ▶ The unavailability of a suitable approach for summarizing a job candidate's skills in a way that perfectly matches what recruiters seek



OBJECTIVE



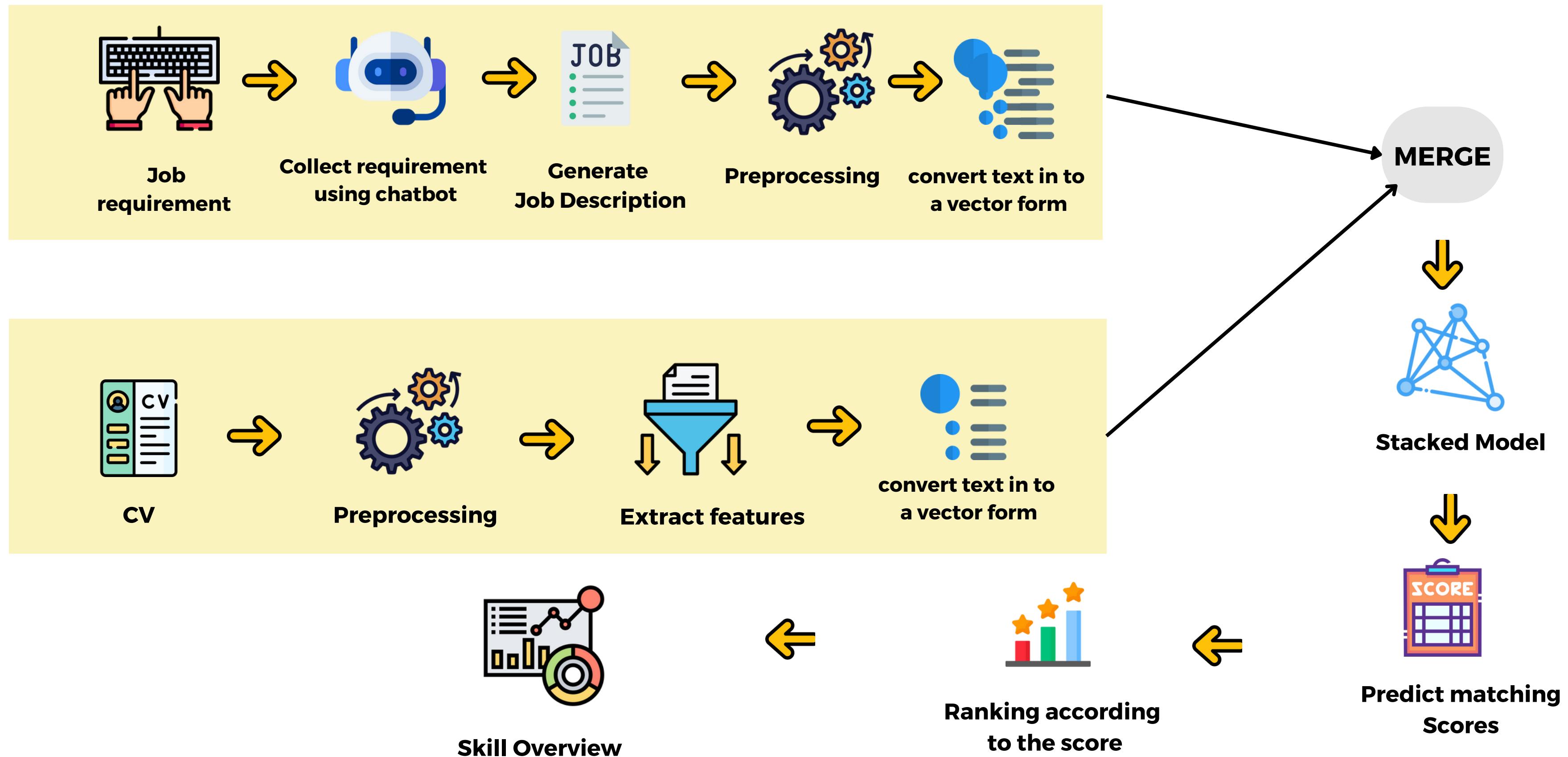
The objective of this research component aims to develop a system that can evaluate and rank job applicants' CVs efficiently based on the job description provided by recruiters.

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' technical skill proficiency



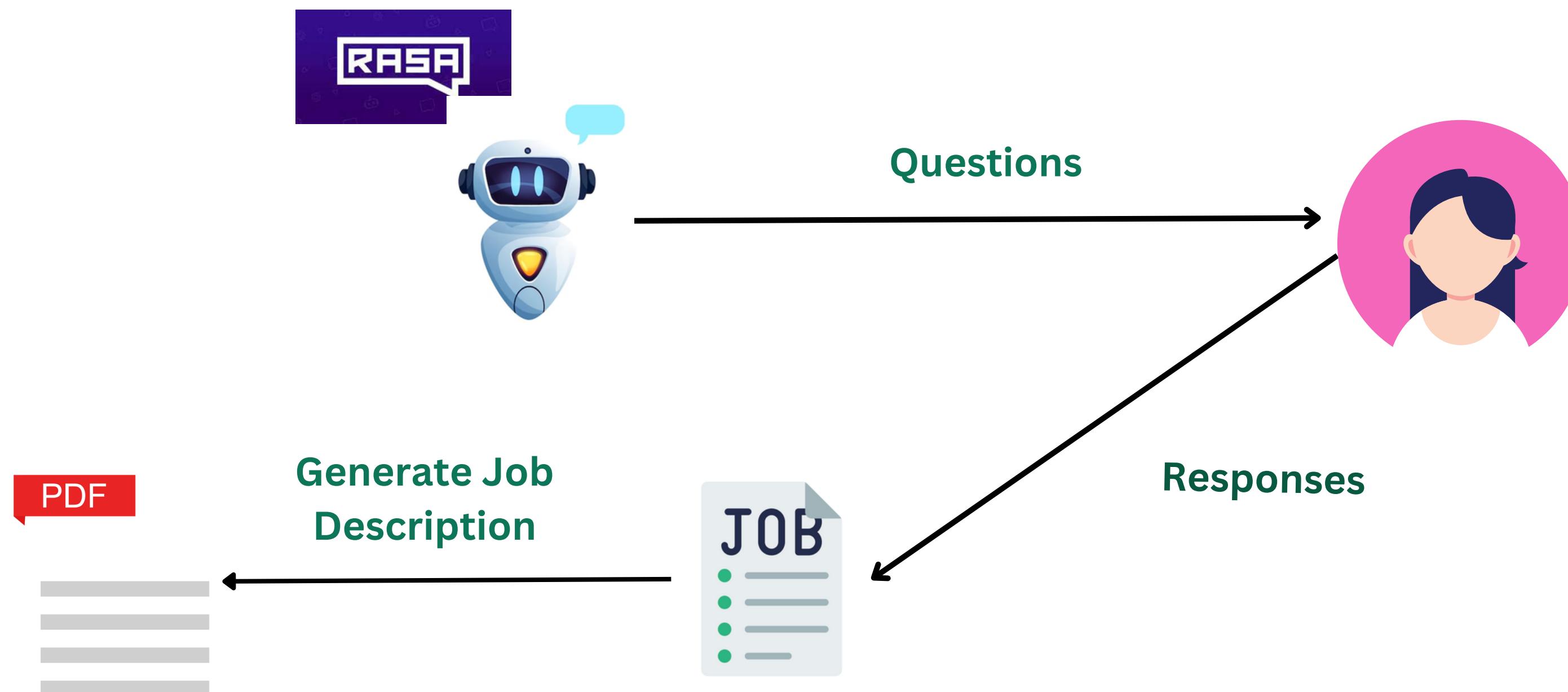
SYSTEM OVERVIEW DIAGRAM



METHODOLOGY



Automated Generation of Job Description using Rasa Framework



METHODOLOGY

► Extraction of data

Data Preprocessing

Explanatory Data Analysis

Feature Vectorization



Feature Vectorization with TF-IDF vectorizer

Feature Vectorization with Count Vectorizer

► Stacked Model Implementation

Multiple individual machine-learning models used

Hyper parameter tuning

Stacked Model



Stacked Model with TF-IDF Vectorizer

Stacked Model with Count Vectorizer

METHODOLOGY



Ranking Resumes according to the Matching percentages

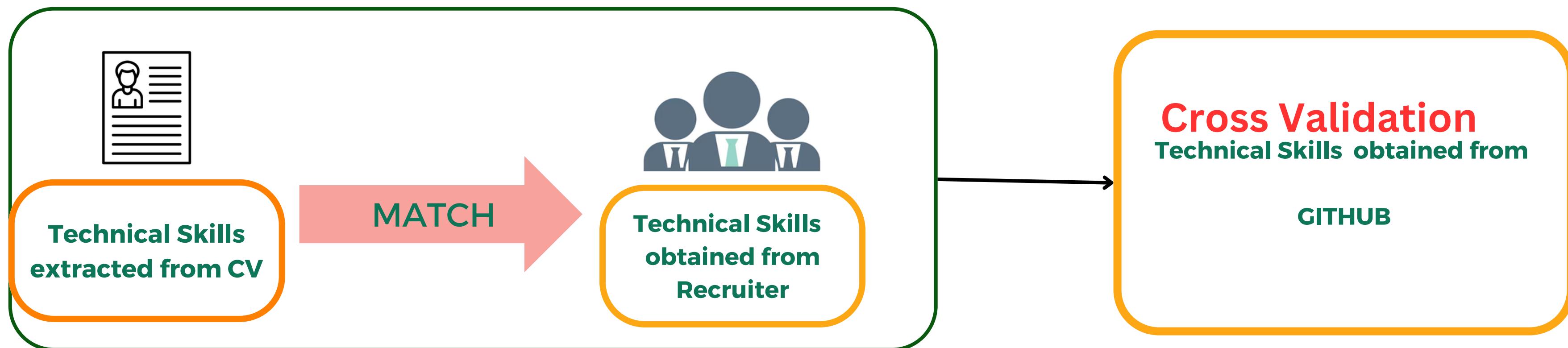
Candidate resumes are ranked according to the highest matching percentages to the lowest



Technical proficiency of candidates

Technical skills are extracted from CV

Match extracted skills with the provided skills list by the recruiter



DEVELOPED SOLUTION



Job Description Generator

The Rasa framework is used to develop a chatbot that prompts questions to the user. Upon receiving user responses, it generates a job description



Stacked Model

This approach involved combining the individual ML models that demonstrated favorable accuracies, resulting in a single unified model with significantly improved accuracy.

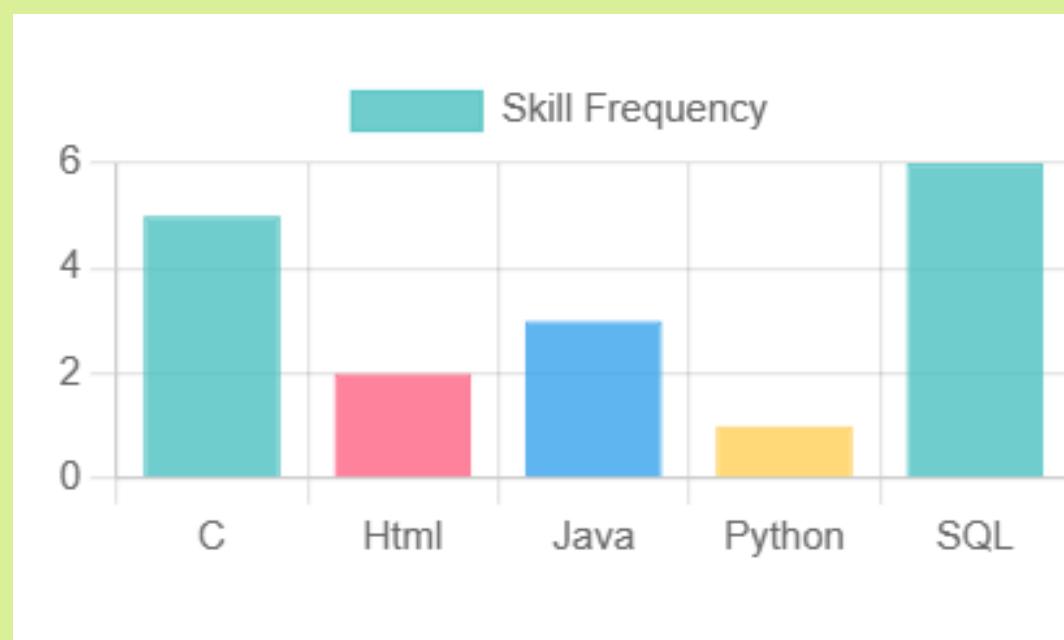


Technical Proficiency Overview

Candidate technical skills are extracted and matched using keywords with the recruiter's inputted skill set to demonstrate technical proficiency. These skills are cross-validated.

RESULTS AND DISCUSSION

```
Bot loaded. Type a message and press enter (use '/stop'  
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?
```



MODEL	Accuracy
Support Vector Regressor	64.74%
Random forest Regressor	37.64%
Decision Tree Regreeor	42.63%
XgBoost	45.31%

Model	Accuracy
The stacked model with Count vectorizer	68.10%
The stacked model with TF-IDF vectorizer	61.28%

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.

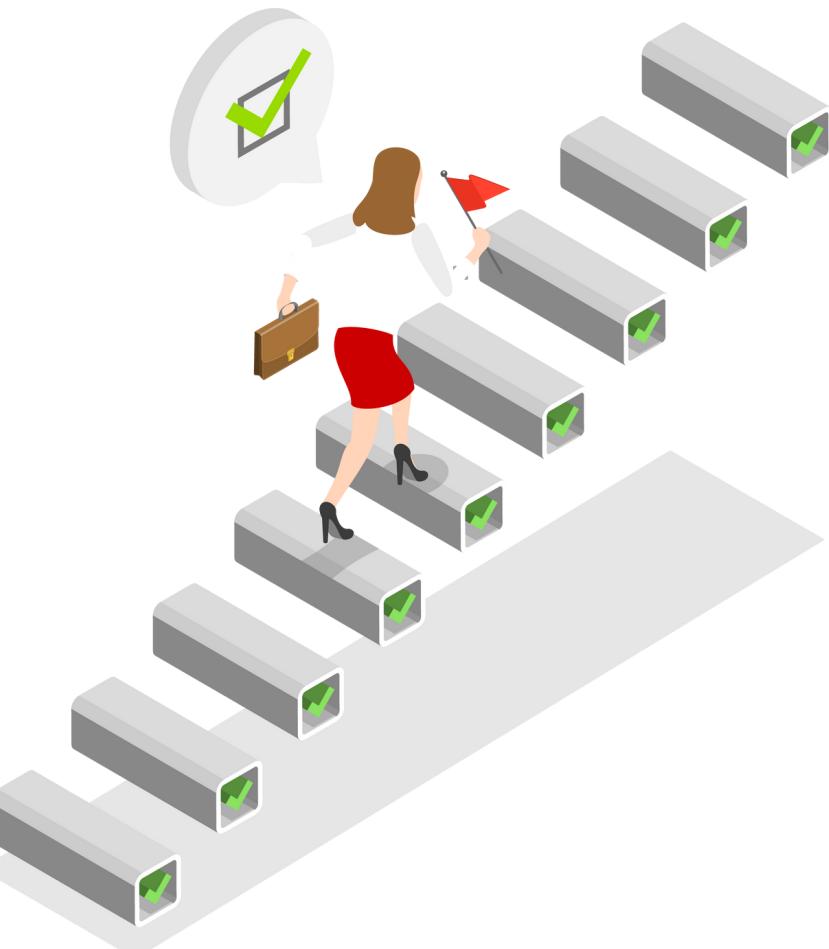
PROGRESS

COMPLETED

Integrated functional web application

TO-DO

Host the web application
Storage Implementation



CONCLUSION

- As Team Intellihire, 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.



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

De Silva M.

Zoysa E.S.

Maldeniya M.M.D.

De Silva S.R.

INTRODUCTION

- Traditional recruitment processes are often plagued by inefficiencies, consuming valuable time and resources while leaving room for human error.
- To address these challenges, a growing number of organizations are turning to automated recruitment systems powered by machine learning, data extraction, and natural language processing techniques.
- In this research, we embark on the journey to develop an innovative automated recruitment system that not only accelerates the hiring process but also provides a comprehensive evaluation of candidates in key areas, promising to reshape the landscape of talent acquisition.

RESEARCH PROBLEM

Inefficient traditional hiring processes in the IT sector require a more accurate and efficient method for evaluating candidate skills and abilities. This study proposes leveraging machine learning, data extraction, and natural language processing to streamline the hiring process for IT-related positions.

RESEARCH OBJECTIVES

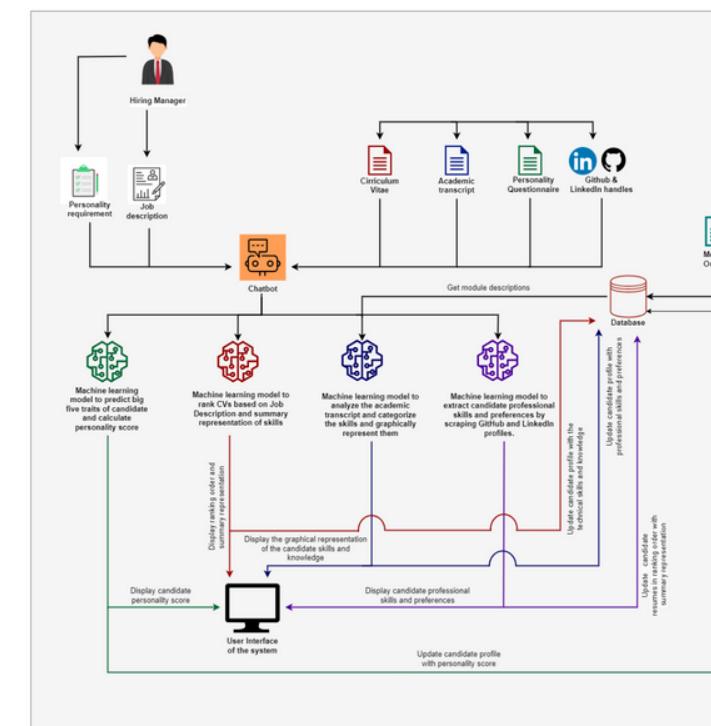
Develop a comprehensive candidate evaluation system to streamline the recruitment process.

- Evaluate and rank job applicants' CVs efficiently based on the job description.
- Analyze academic transcripts from specific universities to assess candidate expertise in relevant areas.
- Assess compatibility between a candidate's personality and job role-specific personality traits.
- Evaluate candidate user profiles by leveraging professional social media platforms such as GitHub and LinkedIn.

LITERATURE SURVEY

- [1] A comprehensive screening approach for candidate selection. They used phone call transcripts to assess personality traits, achieving high accuracy with Logistic regression. For Software Engineer positions, they proposed pre-screening methods incorporating GitHub profiles, CV data, LinkedIn skills, academic transcripts, and recommendation letters. Their evaluation categorized academic transcripts into skill areas, considering a C+ threshold for eligibility. This research highlights the importance of aligning candidate skills with job requirements, benefiting both employers and candidates.
- [2] Advanced machine learning techniques were used to systematically rank IT industry resumes based on predefined criteria, offering valuable insights for recruiters in making informed decisions during applicant evaluations.

SYSTEM DIAGRAM



SOLUTION AND METHODOLOGY

Optimizing Candidate Selection: Utilizes an AI chatbot for structured job description generation, compares job descriptions with resumes to compute matching percentages, and employs a scoring algorithm to evaluate candidate skill profiles.

Analyzing Academic Transcripts: Extracts data from academic transcripts, identifies module titles and grades using NER models, matches them with module keywords, categorizes skills, and predicts skill areas using various machine learning models based on weighted grades.

Aligning Candidate Personality with Job Roles: Utilizes the Big Five Personality Model, combining multiple-choice and open-ended questionnaires, along with keyword extraction and clustering algorithms, to determine the compatibility between a candidate's personality and job role requirements.

Evaluating Professional Skills: Utilizes GitHub data to assess programming language proficiency, a LinkedIn skill-based job title classifier with TFIDF vectorization and machine learning, and sentiment analysis on reference check forms for candidate skill validation.

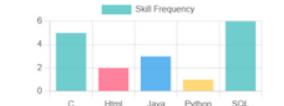
Finally, by employing all four components, a comprehensive final score is calculated for each candidate, facilitating a holistic assessment of their suitability for a specific job role.

REFERENCES

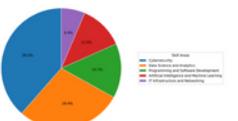
- [1] R. G. U. S. Gajanayake, M. H. M. Hiras, P. I. N. Gunathunga, E. G. Janith Supun, A. Karunasenna, and P. Bandara, "Candidate Selection for the Interview using GitHub Profile and User Analysis for the Position of Software Engineer," in 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka: IEEE, Dec. 2020, pp. 168–173. doi: 10.1109/ICAC51239.2020.9357279.
- [2] R. Ransing, A. Mohan, N. B. Emberi, and K. Mahavarkar, "Screening and Ranking Resumes using Stacked Model," in 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India: IEEE, Dec. 2021, pp. 643–648. doi: 10.1109/ICEECCOT52851.2021.9707977.

RESULTS AND DISCUSSION

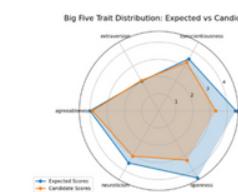
Candidate selection optimization included a chatbot with an accuracy of 1.00, a stacked model for resume matching, and visualized skill proficiency assessment.



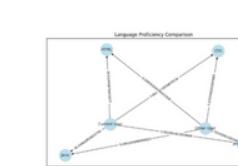
In analyzing academic transcripts for candidate skills, a custom NER model showed low loss, while the pre-trained custom NER model had a perfect score. The SVM model demonstrated the highest accuracy for skill prediction at 0.8810.



Aligning candidate personality traits with job role requirements involved using MCQ and open-ended questionnaires. K-means clustering identified 5 personality clusters, and XGBoost predicted candidate clusters accurately. The radar graph assessed compatibility with job role traits, resulting in higher accuracy using XGBoost compared to previous research.



In evaluating professional skills, a GitHub code proficiency comparator quantitatively assesses programming language skills. A LinkedIn-based job title classifier, with Linear Support Vector Machine, achieved 66% accuracy for predicting job categories based on skills, enhancing hiring efficiency.



ACKNOWLEDGEMENT

- We extend our heartfelt gratitude to Dr. Anuradha Karunasena, our supervisor, and Dr. Lakmini Abeywadharana, our co-supervisor, for their invaluable guidance throughout this research.
- We also wish to express our appreciation to the Faculty of Computing at Sri Lanka Institute of Information Technology for granting us the opportunity to undertake this research.

THANK YOU

