



# SRM

INSTITUTE OF SCIENCE & TECHNOLOGY  
(Deemed to be University u/s 3 of UGC Act, 1956)

Department Of Computational Intelligence

18CSP109L - Major Project

## Resume Screening Application Using Data PreProcessing and NLP

Guide Name

Dr. P. G. Om Prakash

Panel Head

Dr. N. Arivazhagan

Presented By

Ganesh Kumar Tanguturi(RA2011033010118)

Amith Sai baba(RA2011026010432)



JOHN  
RESUMGO

GRAPHIC DESIGNER

Street Address  
City State ZIP Code

(123) 456-7890

email@address.com

Ut enim ad minim veniam, quis nostrud exerc. irure dolor in reprehend incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Ut enim ad minim veniam.,

### SKILLS

- HSKILL 1
- HSKILL 2
- HSKILL 3
- HSKILL 4
- HSKILL 5
- HSKILL 6
- HSKILL 7
- HSKILL 8
- HSKILL 9
- HSKILL 10
- HSKILL 11
- HSKILL 12

### EXPERIENCE

JOB TITLE - (DEC. 2012 - PRESENT)  
COMPANY NAME - City, Country

Ut enim ad minim veniam, quis nostrud exerc. irure dolor in reprehend incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

#### Major accomplishments:

- Ut enim ad minim veniam, quis nostrud exerc.
- Ut enim ad minim veniam, quis nostrud exerc.

JOB TITLE - (2006 - 2012)  
COMPANY NAME - City, Country

Ut enim ad minim veniam, quis nostrud exerc. irure dolor in reprehend incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

#### Major accomplishments:

- Ut enim ad minim veniam, quis nostrud exerc.
- Ut enim ad minim veniam, quis nostrud exerc.

### EDUCATION

DIPLOMA - (2003-2005)  
SCHOOL NAME - City, Country

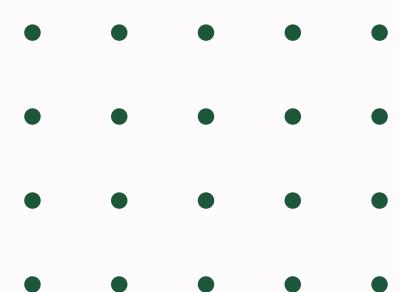
Ut enim ad minim veniam, quis nostrud exerc. irure dolor in reprehend incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam,

DIPLOMA - (2000-2003)  
SCHOOL NAME - City, Country

Ut enim ad minim veniam, quis nostrud exerc. irure dolor in reprehend incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam,

### REFERENCES

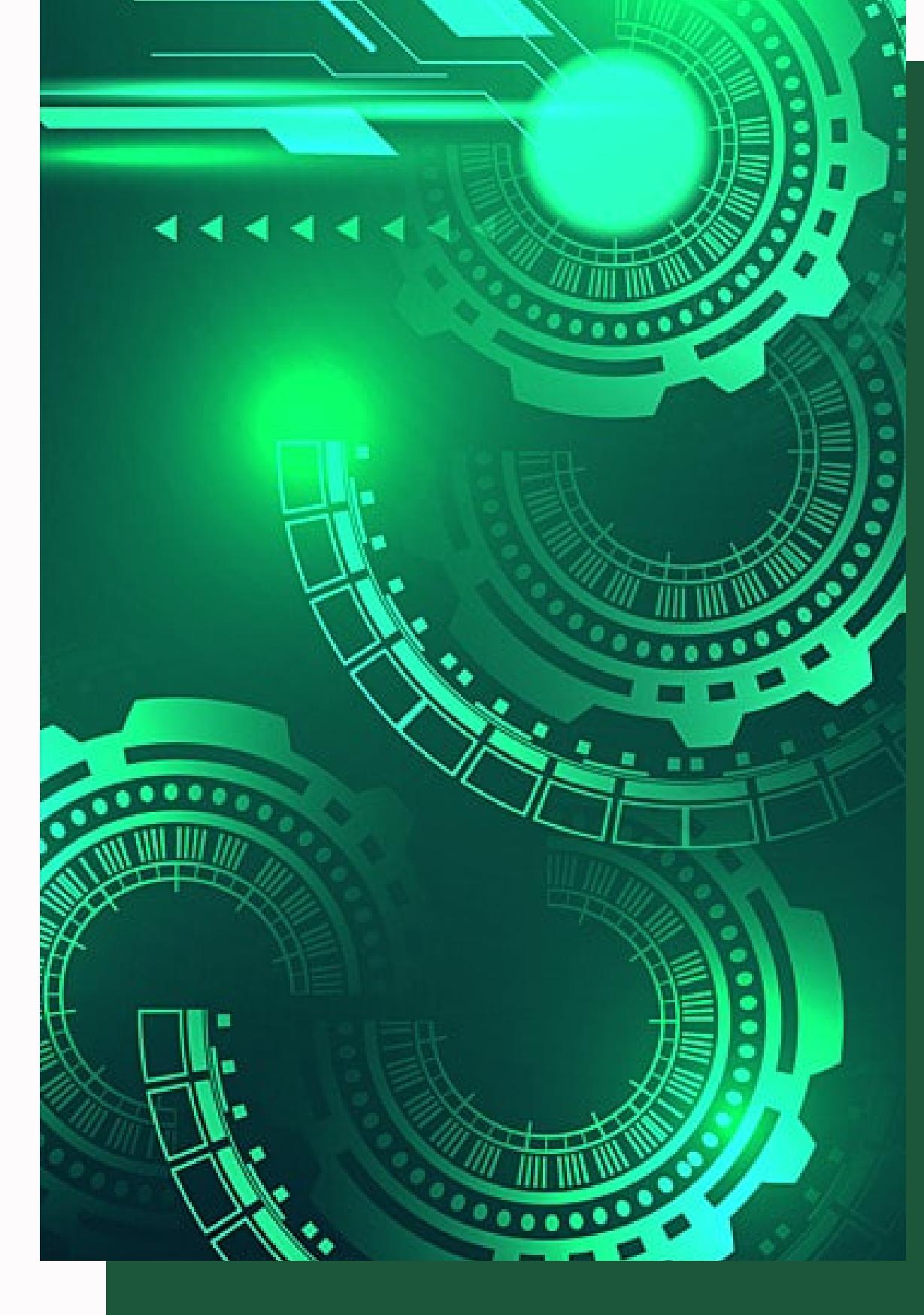
- James Smith  
(Job Title - Company Name)  
(123) 456-789
- James Smith  
(Job Title - Company Name)  
(123) 456-789





# Content

- 01** Abstract
- 02** Introduction
- 03** Motivation
- 04** Objectives
- 05** Problem Statement
- 06** Flow Diagram
- 07** Architecture Diagram





# Content

**08**

**Technologies Used**

**09**

**Existing System Vs Proposed System**

**10**

**Key Features**

**11**

**Real time Use Cases**

**12**

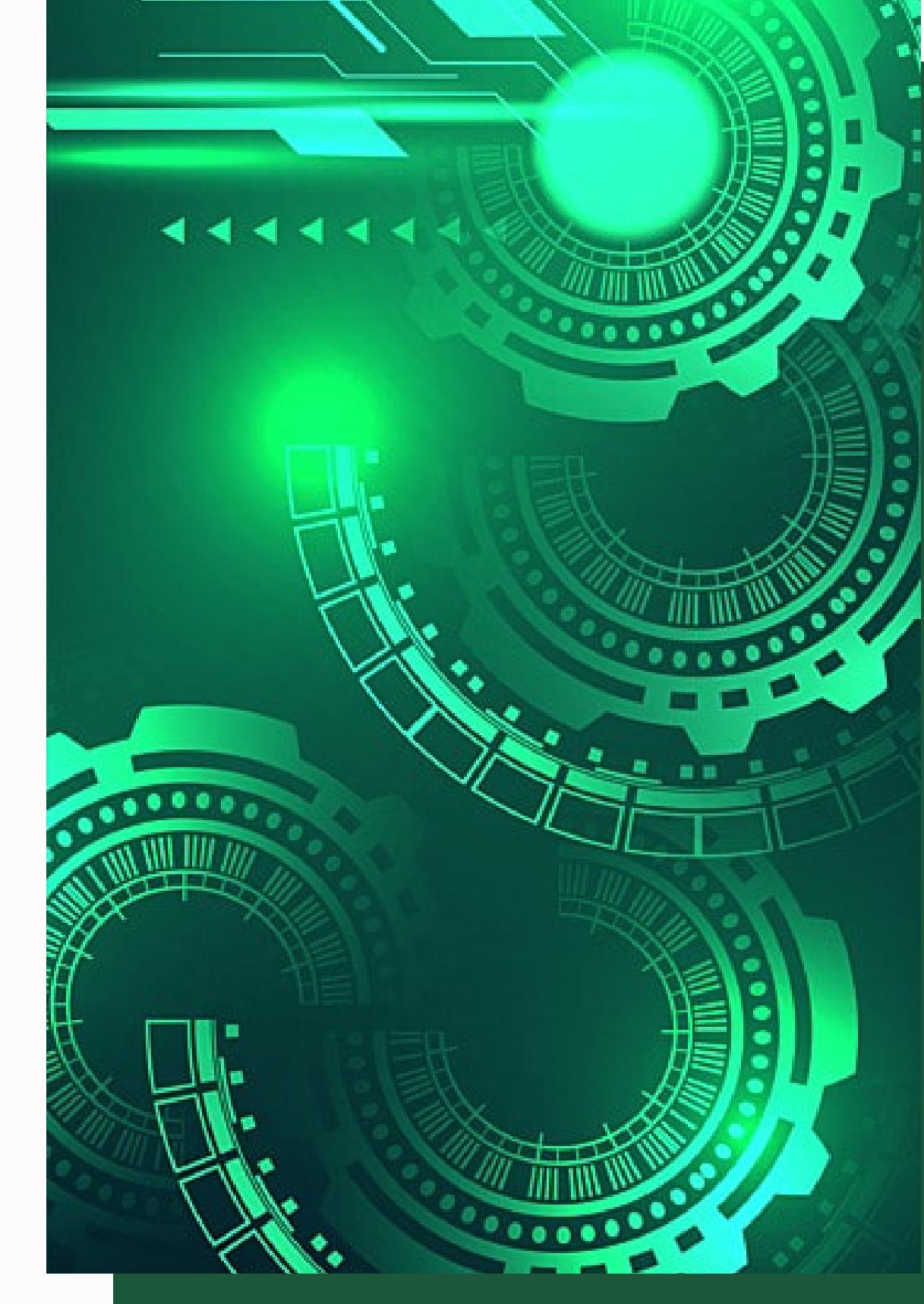
**Literature Survey**

**13**

**Algorithms Used & Advantages**

**14**

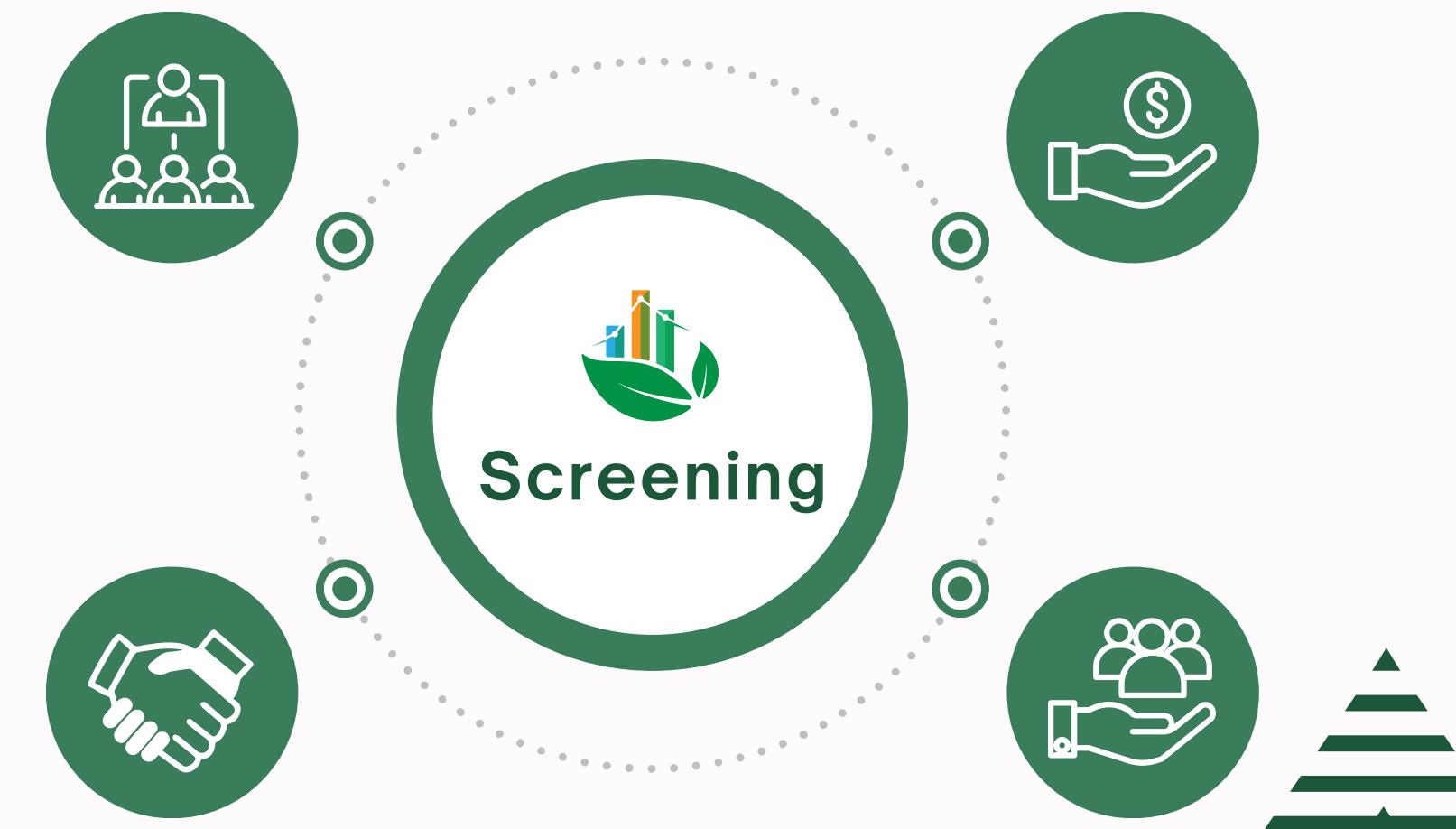
**Research Perspective**



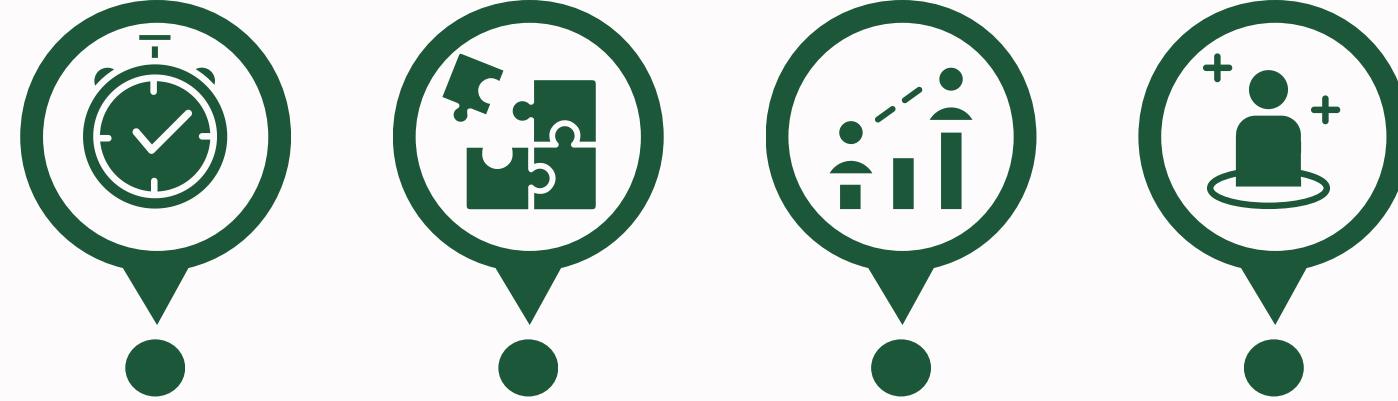


# Abstract

In the contemporary recruitment landscape, the challenge of efficiently screening a large volume of resumes and accurately matching them to job descriptions is significant. This project introduces an innovative Resume Screening and Job Matching platform, harnessing the capabilities of natural language processing (NLP), machine learning, and data visualization to transform the recruitment process. The platform is engineered to aid HR professionals and recruiters in swiftly navigating through heaps of resumes, pinpointing candidates who are most apt for specific job roles through a detailed analysis of their resumes. This paper outlines the key features and functionalities of the platform, including advanced text processing, TF-IDF analysis, a resume ranking system, and interactive data visualization, all aimed at streamlining the recruitment workflow..



# Introduction



The recruitment process, particularly the initial stages of resume screening and candidate shortlisting, is often labor-intensive and time-consuming. Traditional methods, while reliable, do not scale well with the increasing volume and complexity of job applications. To address these challenges, this project presents a cutting-edge Resume Screening and Job Matching platform.

The core of the platform is built on robust natural language processing and machine learning techniques. It begins with the parsing of resumes and job descriptions, transforming unstructured text into a structured format ready for analysis. The text data undergoes thorough cleaning and preprocessing, including tokenization, lemmatization, and the removal of stopwords, to ensure optimal data quality for analysis.

A key feature of the platform is the implementation of TF-IDF analysis, which assesses the significance of words in the resumes in relation to a corpus of documents, thereby enhancing the matching accuracy between resumes and job descriptions. The platform also introduces a sophisticated resume ranking system, which scores and ranks candidates based on the relevance of their profiles to the job description. This feature is pivotal in reducing the manual effort in candidate selection.

The user interface of the platform is developed using Streamlit, offering an intuitive and interactive experience. It allows users to easily navigate through job descriptions, select roles, and view a ranked list of candidates. Additionally, the platform incorporates advanced data visualization tools like Plotly, providing dynamic visual representations of data, such as score distributions and topic breakdowns in resumes.

Another significant aspect of the platform is the use of Latent Dirichlet Allocation (LDA) for topic modeling, which aids in clustering resumes into categories, further refining the matching process. The platform also features word cloud generation for a quick visual representation of key skills and job requirements.

In summary, this platform stands as a powerful tool for HR departments and recruitment agencies, significantly optimizing the hiring process. By automating the initial stages of resume screening, it not only conserves time but also enhances the precision of matching candidates to job roles. The platform's analytical capabilities ensure that recruiters can make data-driven decisions, thereby improving the quality of hires and contributing to the overall efficiency of the recruitment process.

# MOTIVATION

## Resume Screening



The motivation for developing a Resume Screening and Job Matching platform stems from the need to address the challenges associated with traditional resume screening methods. Manual review of resumes is a time-consuming and labor-intensive process that can lead to inconsistencies and biases in candidate evaluation. Additionally, as the volume and complexity of job applications continue to increase, traditional methods are becoming increasingly ineffective in managing the influx of resumes.



To address these challenges, a Resume Screening and Job Matching platform offers several compelling advantages:

1. Automation
2. Accuracy
3. Scalability
4. Objectivity
5. Data-driven decision-making
6. Improved hiring efficiency
7. Enhanced candidate experience



The adoption of a Resume Screening and Job Matching platform can significantly enhance the recruitment process, leading to better hiring decisions and improved efficiency for organizations.



# Objectives

The Ultimate Goal Of Resume Screening is To identify and match qualified candidates with relevant job openings efficiently and objectively.

**1st Quarter**  
Reduce Manual Effort  
and Save Time

**2nd Quarter**  
Improve matching accuracy  
and Reducing bias in candidate evaluation



**4th Quarter**

Optimize the hiring process and Gaining valuable insights into candidate profiles and job requirements



**3rd Quarter**

Enhance the candidate experience and Improving the quality of hires



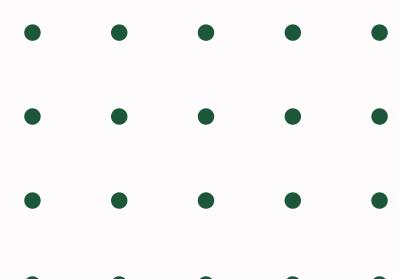


# Problem Statement

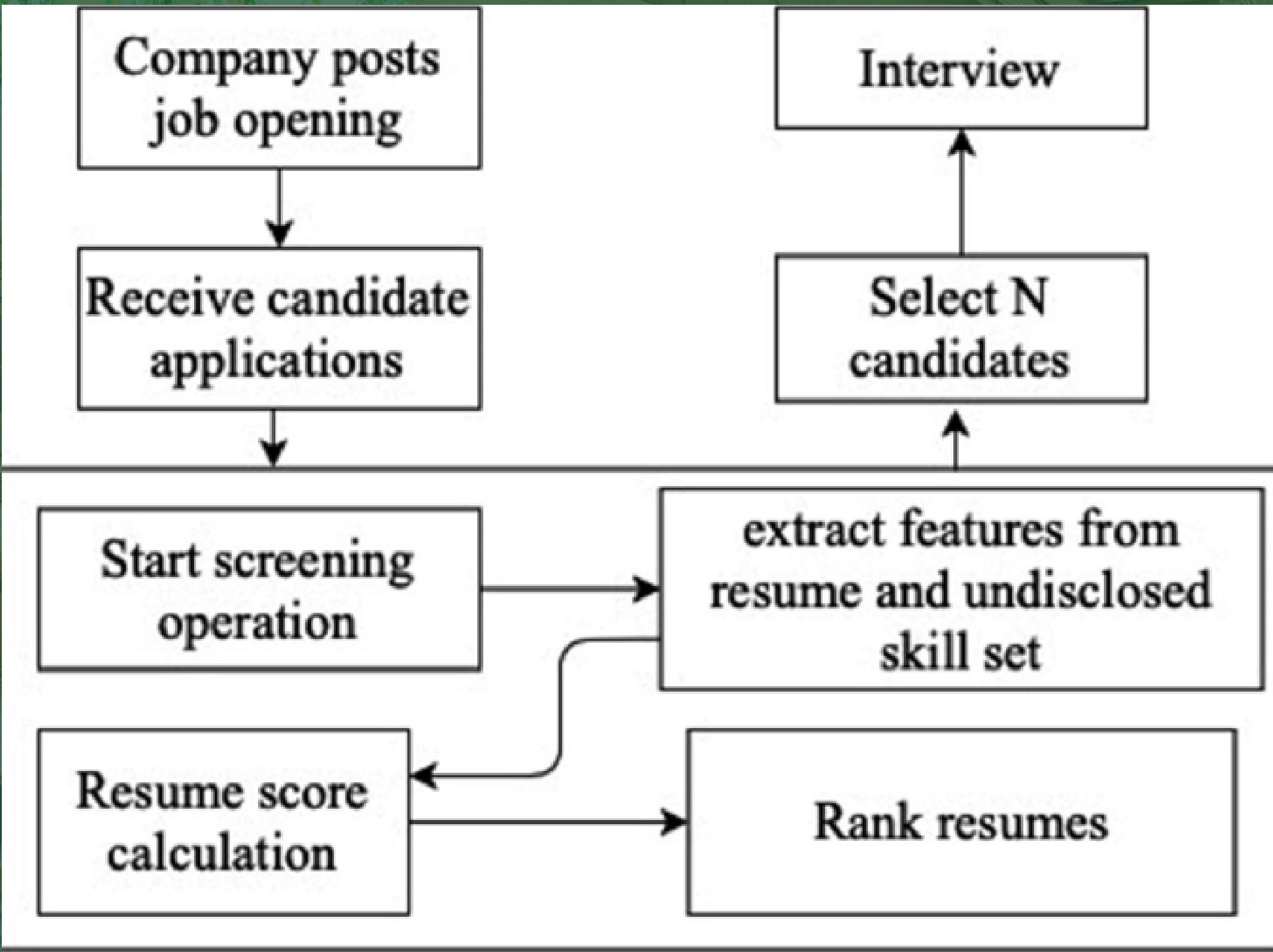
Due to the increasing volume and complexity of job applications, traditional methods of manual resume screening are becoming ineffective and time-consuming. This leads to:

- Delays in the hiring process: HR professionals are overwhelmed by the sheer volume of resumes, resulting in delays in screening and shortlisting candidates.
- Inconsistent evaluation: Manual resume screening is prone to human biases and inconsistencies, leading to unfair and inaccurate assessments of candidates.
- Overlooking qualified candidates: Relevant and qualified candidates may be missed due to the limitations of manual screening methods.
- Inefficient use of resources: HR professionals spend valuable time on repetitive tasks that could be automated, reducing their productivity and focus on strategic recruitment initiatives.

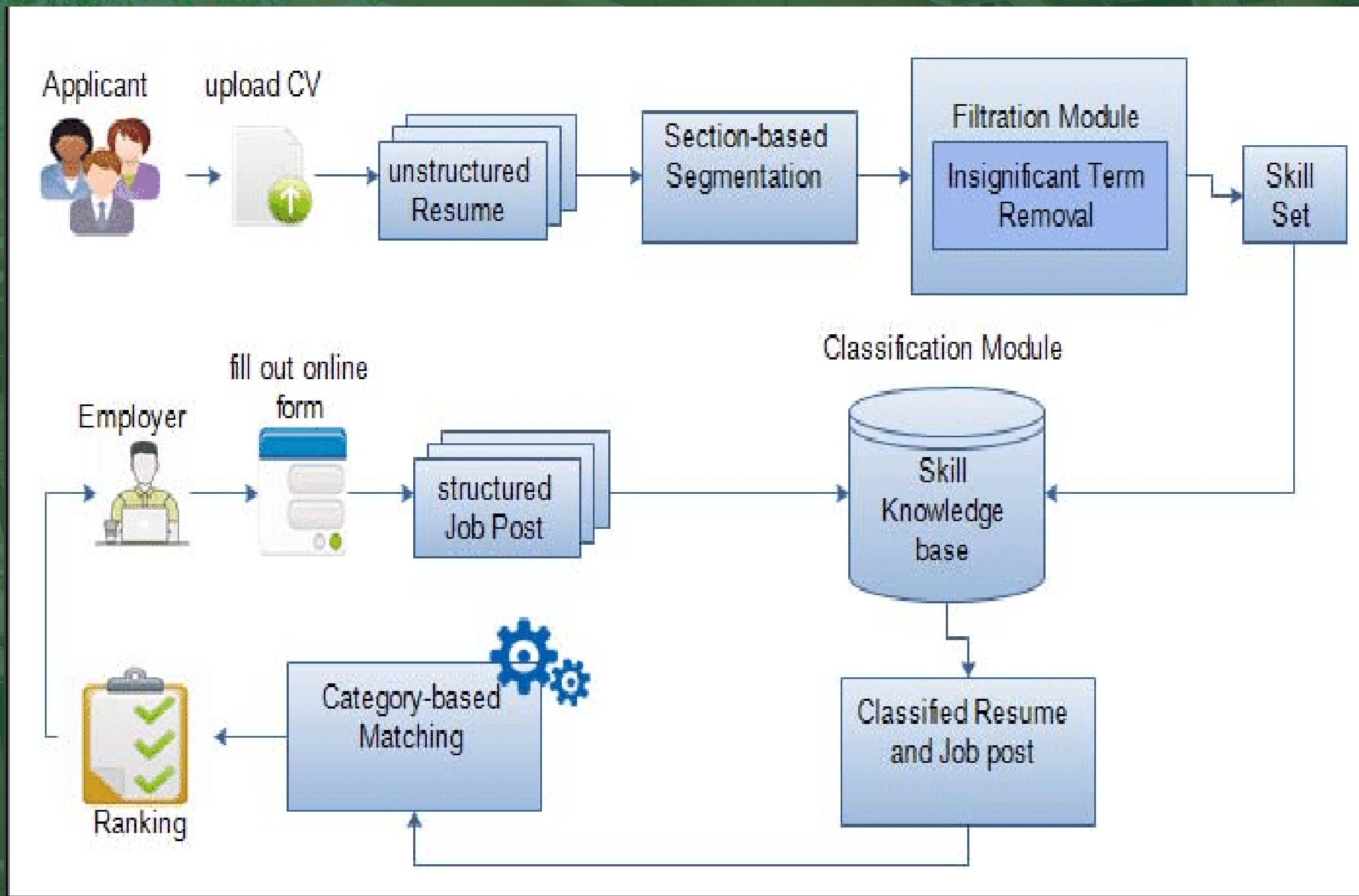
85%



# FLOW DIAGRAM



# ARCHITECTURE DIAGRAM



# Technologies Used

*Pandas*

*WorkCloud*

*Gensim*

*Python*



---

*Streamlit*



---

*Plotly*



# EXISTING SYSTEM VS PROPOSED SYSTEM

## Existing System

Manual Screening  
Time Consuming  
Subjectivity  
Limited Scalability  
Lack of Visuality

## Proposed System

Advanced Algorithms  
user Friendly  
Real time Feedback  
Visual Analysis

# Key Features

01



Resume uploading and storage in a structured format.

02



Job Description selection for tailored analysis.

03



TF-IDF and LDA-based scoring for precise matching.

04



Real-time scoring and ranking of resumes.

05



Visual representation of scores and topic distribution.

06



Word clouds for intuitive resume content overview.

# REALTIME USECASES

**HR SCREENING**

**RECRUITMENT  
AGENCIES**

**SKILL GAP  
ANALYSIS**

**AUTOMATED  
SHORTLISTING**

**CUSTOMIZED  
JOB MATCHING**

Author(s)	Year	Objective/Scope	Methodology	Key Findings
Smith et al	2015	Identify factors influencing shortlisting	Survey of HR professionals, analysis of job advertisements	Keywords and experience are top factors; 75% rely on ATS (Applicant Tracking Systems); soft skills gaining importance
Johnson and Brown	2018	Analyze biases in resume shortlisting	Experimental design, blind review of resumes	Significant bias found based on gender and ethnicity; anonymized resumes reduce bias
Gupta and Sharma	2019	Improve resume shortlisting accuracy	Machine Learning (ML) algorithms, comparative analysis	ML algorithms outperform traditional methods; reduce false positives and negatives

Author(s)	Year	Objective/Scope	Methodology	Key Findings
Chen and Wang	2020	Assess the impact of resume format	Experimental study with varied resume formats	Chronological format preferred; clear, concise, and well-organized resumes positively impact shortlisting
Kim et al.	2021	Investigate the role of social media	Content analysis of social media profiles	Positive correlation between professionally managed social media and resume shortlisting; caution with personal content
Rodriguez and Martinez	2022	Examine the role of educational background	Analysis of job applications and shortlisting outcomes	Strong correlation between relevant education and shortlisting; advanced degrees positively impact certain industries

# ALGORITHMS USED AND ITS ADVANTAGES

## TF-IDF(TERM FREQUENCT INVERSE DOCUMENT REQUENCY)

### Advantages:

- Captures the importance of terms in a document relative to the entire corpus.
- Effective in identifying key terms for matching.



## LDA (LATENT DIRICHLET ALLOCATION)

### ADVANTAGES:

- Unsupervised learning for topic modeling.
- Reveals underlying topics in a collection of documents.
- Enhances clustering and identification of similar resume

# RESEARCH PERSPECTIVE

- This project combines natural language processing and machine learning techniques to enhance the efficiency of the hiring process.
- The use of TF-IDF and LDA algorithms ensures a sophisticated level of analysis, providing a deeper understanding of resume-job compatibility.
- Visual analytics, such as word clouds and interactive plots, offer a novel way to interpret and present complex data.
- The real-time nature of the system reflects a contemporary approach to HR technology, focusing on quick and data-driven decision-making.
- The project's modular design allows for future extensions, including the integration of advanced machine learning models for continuous improvement.



Resume  
Screening

# THANK YOU

**Ganesh Kumar Tanguturi**



+91 7995252073



tk2321@srmist.edu.in

