# Resume Screening and Analysis Using ML and NLP

A

# Major Project Report

Submitted

In partial fulfilment

For the award of the Degree of

# BACHELOR OF TECHNOLOGY

**In Department of Computer Science and Engineering**

|  |  |  |
| --- | --- | --- |
| **Submitted By:** | **Guided By:** | **Submitted To:** |
| Himansh Soni (18ETCCS045) | Mr. Gaurav Kumawat | Mr. Aaditya Maheshwari |
| Siddharth Jain (18ETCCS082) |  |  |
| Kartik Dave (18ETCCS050) |  |  |

Ravindra Pratap Singh (18ETCCS076)



Department of Computer Science and Engineering

TECHNO INDIA NJR INSTITUTE OF TECHNOLOGYRAJASTHAN TECHNICAL UNIVERSITY

2022

# TECHNO INDIA NJR INSTITUTE OF TECHNOLOGY



**CERTIFICATE**

This is to certify that this project report **“Resume Screening and Analysis using ML and NLP”** is the confide work of **“Himansh Soni, Siddharth Jain, Kartik Dave, Ravindra Pratap Singh”** who have carried out the project work under my supervision. I approve this project for submission of the Bachelor of Technology in the **Department of Computer Science and Engineering, Techno India NJR Institute of Technology**, affiliated to Rajasthan Technical University, Kota.

Mr. Aaditya Maheshwari Department of Computer Science

And Engineering

# ABSTRACT

1. **Purpose**

# Introduction

Resume Screening and Analysis System uses machine learning algorithms to screen resumes.

The main objective of the Resume Screening and Analysis System is toefficiently evaluate the candidate thoroughly through a fully automated system that not only saves lot of time but also gives fast results.

# Scope

Scope of this project is very broad in terms of screening resumes.Few of them are: -

* + - Sorting of resumes on different job profiles.
    - Makes recruiter work easier and efficient.
    - Gets complete in-depth analysis of every factor of the resume.

# ACKNOWLEDGEMENT

It gives me immense pleasure to express my deepest sense of gratitude and sincere thanks to my highly respected and esteemed guide **Mr. Gaurav Kumawat (Assistant Professor), TINJRIT** for their valuable guidance, encouragement and help for completing this work. Their useful suggestions for this whole work and co-operative behavior are sincerely acknowledged.

I would like to express my sincere thanks to **Faculty, Department of Computer Science and Engineering, TINJRIT** for giving me this opportunity to undertake this project.

I also wish to express my indebtedness to my parents as well as my family member whose blessings and support always helped me to face the challenges ahead.

At the end I would like to express my sincere thanks to all my friends and others who helped me directly or indirectly during this project work.

Himansh Soni (18ETCCS045) Siddharth Jain (18ETCCS082) Kartik Dave (18ETCCS050)

Ravindra Pratap Singh (18ETCCS076)

# TABLE OF CONTENTS

[Abstract 3](#_bookmark0)

[Acknowledgement 4](#_bookmark1)

1. Introduction 6
   1. Project Title 7
   2. [Problem Statement 7](#_TOC_250008)
   3. [Introduction 7](#_TOC_250007)
   4. [Scope 8](#_TOC_250006)
2. Software and Hardware Requirement Specifications 9
   1. [Hardware Specifications 10](#_TOC_250005)
   2. [Software Specifications 10](#_TOC_250004)
3. System Analysis and Design 12
   1. [Weakness of Manual Screening 13](#_TOC_250003)
   2. [How New System Helps 13](#_TOC_250002)
   3. [Design Approach 14](#_TOC_250001)
   4. [Software Architectural Design 15](#_TOC_250000)
4. Screen Shots 20
5. Conclusion 25
   1. [Limitations 26](#_bookmark2)
   2. [Future Enhancements 27](#_bookmark3)
   3. [Conclusion 28](#_bookmark4)
6. Bibliography 30

**INTRODUCTION**

# Introduction

* 1. **Title**

Resume Screening and Analysis using ML and NLP

# Problem Statement

Large companies do not have enough time to open each CV, so they usemachine learning algorithms for the resumes screening task. Resume screening is the process of identifying the finest talent from a large poolof candidates. Large firms typically do not have enough time to review each CV, therefore they utilize machine learning algorithms to screen resumes.

# Introduction

The Resume Screening and Analysis System is to locate the most qualifies candidates for an open job. It involves sorting through resumes, manually or through resume screening software, for identifying candidates who may be eligible to fill a role.

This is done by matching the job requirements with the qualifications ofthe candidate, which may usually include, but are not limited to factorssuch as

* Work Experience
* Educational Qualifications
* Skills and Knowledge
* Competencies
* Personality Traits

# Scope

A platform where an applicant and a recruiter can screen and analysisresumes. Recruiter:

* + - Sorting of resumes on different job profiles.
    - Ranking of resumes based on its content.
    - Makes recruiter work easier and efficient Applicant:
    - Able to compare its resumes with others.
    - Knows where to improves its content.
    - Gets complete in-depth analysis of every factor of the resume.

# SOFTWARE AND HARDWARE REQUIREMENT SPECIFICATIOS

**Software and Hardware Requirement Specifications**

# Hardware Specifications

* + - Processor: i3 and above
    - RAM: 4GB
    - Hard Disk: 500GB

# Software Specifications

* + - Operating System: Windows 10/11
    - Web Browser: Google Chrome
    - Code Editor: VS Code
    - Database: MongoDB

## Python

Python is a general-purpose interpreted, interactive, object-oriented, andhigh-level programming language. It was created by Guido van Rossumduring 1985- 1990. Python is used for creating backbone structure.

Python is intended to be a highly readable language. It is designed to havean uncluttered visual layout, it uses whitespace indentation, rather than curly braces or keywords. Python has a large standard library, commonly cited as one of Python's greatest strengths.

## Flask

Flask is a web application framework written in Python. Armin Ronacher,who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

### Natural Language Processing

NLTK was originally created in 2001 as part of a computational linguistics course in the Department of Computer and Information Science at the University of Pennsylvania. Since then it has been developed and expanded with the help of dozens of contributors. It has now been adopted in courses in dozens of universities, and serves as thebasis of many research projects. NLTK was designed with four primary goals in mind:

Simplicity

To provide an intuitive framework along with substantial building blocks,giving users a practical knowledge of NLP without getting bogged down

in the tedious house-keeping usually associated with processing annotatedlanguage data.

Consistency

To provide a uniform framework with consistent interfaces and datastructures, and easily guessable method names.

Extensibility

To provide a structure into which new software modules can be easily accommodated, including alternative implementations and competingapproaches to the same task.

Modularity

To provide components that can be used independently without needing to 9 understand the rest of the toolkit. A significant fraction of any NLP syllabus deals with algorithms and data structures. On their own these canbe rather dry, but NLTK brings them to life with the help of interactive graphical user interfaces that make it possible to view algorithms step-by- step. Most NLTK components include a demonstration that performs an interesting task without requiring any special input from the user. An effective way to deliver the materials is through interactive presentation of the examples in this book, entering them in a Python session, observing what they do, and modifying them to explore some empirical or theoretical issue.

### Machine Learning Tool: Scikit-learn (Python Package)

It is a Python module integrating classic machine learning algorithms inthe tightly- knit scientific Python world (numpy, scipy, matplotlib). It aims to provide simple and efficient

solutions to learning problems, accessible to everybody and reusable in various contexts: machine-learning as a versatile tool for science and engineering.

In general, a learning problem considers a set of n samples of data and tryto predict properties of unknown data. If each sample is more than a single number, and for instance a multi-dimensional entry (aka multivariate data), is it said to have several attributes, or features.

# SYSTEM ANALYSIS AND DESIGN

**System Analysis and Design**

# Weakness of Manual Screening

**Volume:** An average job receives 250 applications. Screening250 resumes for one position can be quite daunting.

The “Ignore” Problem: Recruiters may not have the bandwidthto go through each and every application and may end up screening only a few of them.

**Time and Money:** The investment of time and resources toscreen resumes can be considerable.

**Quality:** Recruiters look for specific keywords while screeningresumes. While that may be reasonably effective, they may passup on good resumes just because of a few missing keywords.

**Consistency:** It is impossible for a human being to scan eachand every resume in the same manner.

**Objectivity:** The recruiter may intentionally or subconsciously discriminate among candidates based on gender, caste, race, etc.

# How New System Helps

If you are skeptical about implementing resume screeningsoftware, let us look at some of its benefits.

* + 1. Decrease turnover rates

The obvious way to reduce turnover rates is by hiring the right people. Resume screening software helps organizations discovertop talents for job roles, thus decreasing turnover rates.

* + 1. Eliminate Costly Screening Calls

The initial screening stage is exceptionally critical in the entirehiring process. The traditional recruitment process can be

tedious and time-consuming. Hiring managers that use phonecalls to screen their candidates spend a significant amount of time interviewing every applicant.

This is where recruitment tools can help. They can streamline resume screening and ensure that only qualified candidates make the cut so that recruiters and hiring managers do not have to spend vast amounts of time on the phone to screen candidates.

* + 1. Make More Confident Hiring Decisions

Using the resume screening software, recruiters can administervarious tests to assess a candidate’s problem-solving, logical reasoning, and interpersonal skills. This might give a good ideaof whether the candidate will fit well with your company culture.

* + 1. Seamlessly Integrate with In-house Talent ManagementSolutions.

The best resume screening software solutions can help you withevery aspect of the hiring process, including attracting applications, sourcing resumes, screening them, coordinating interviews, extending offers, and managing the onboarding process.

Most resume screening software solutions can easily integrate with your existing HR management technology stack and furtherstreamline the entire process.

# Design Approach

Design is the first step in the development phase for any techniques and principles for the purpose of defining a device, aprocess or system in sufficient detail to permit its physical realization. Once the software requirements have been analyzedand specified the software design involves three technical activities design, coding, implementation and testing that are required to build and verify the software. The design activities are of main importance in this phase, because

in this activity, decisions ultimately affecting the success of the software implementation and its ease of maintenance are made.

These decisions have the final bearing upon reliability and maintainability of the system. Design is the only way to accurately translate the customer requirements into finished software or a system. Design is the place where quality is fostered in development. Software design is a process through which requirements are translated into a representation of software. Software design is conducted in two steps. Preliminarydesign is concerned with the transformation of requirements intodata.

# Software Architectural Design

Our system follows the three-tier architecture. First tier consistsof GUI, Processing block and the Database.

GUI:

The GUI (Graphical User Interface) in our project deals with theinterface for the user where the user will login and submit his resume in any format (pdf, doc, docx, etc.) and social

profiles links. The GUI provides a platform for the user to communicate with the database. It acts as a connector as well ascommunicator which connects the database and helps in transferof data between the GUI and the database.

Processing block:

Processing block is the block where the actual processing of ourproject is done. This block connects the gui to the database i.e. itacts as a connector as well as communicator which connects the database and helps in transfer of data between the gui and the database. Its main function is to take input from resumes and social profile of the candidate and parse it to store the information and store it in the structured format(json), and database. After storing this information this system will give output using web application.

Database: Database tier is the tier used for the storage of data.This tier contains all the data that is need for the processing ofthe whole project. The data in this tier is related to the student

information gathered form his/her resumes and social profiles.

Software Architectural Design

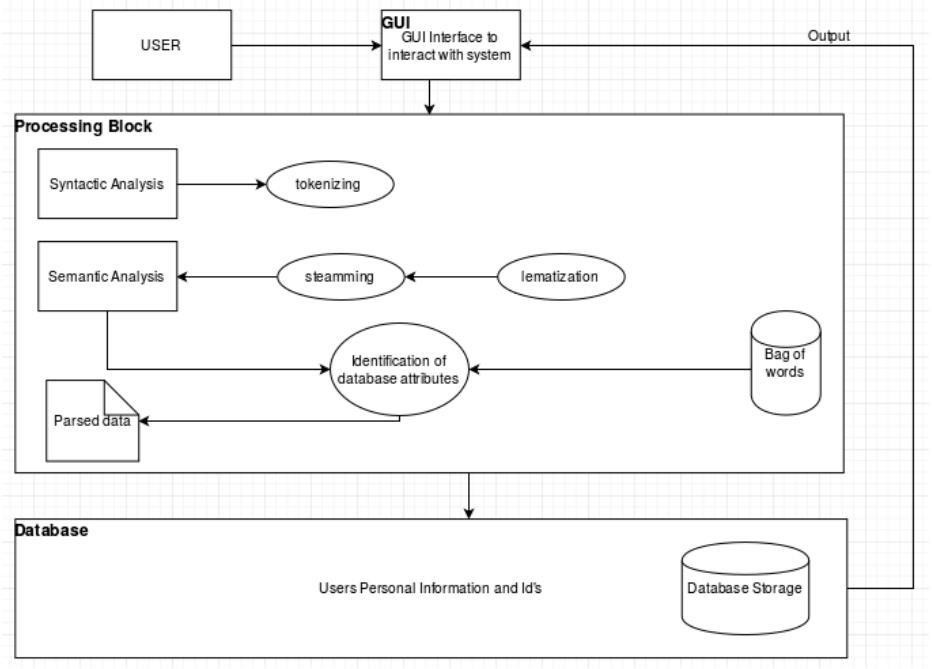


Fig 3.4.1

Component Diagram

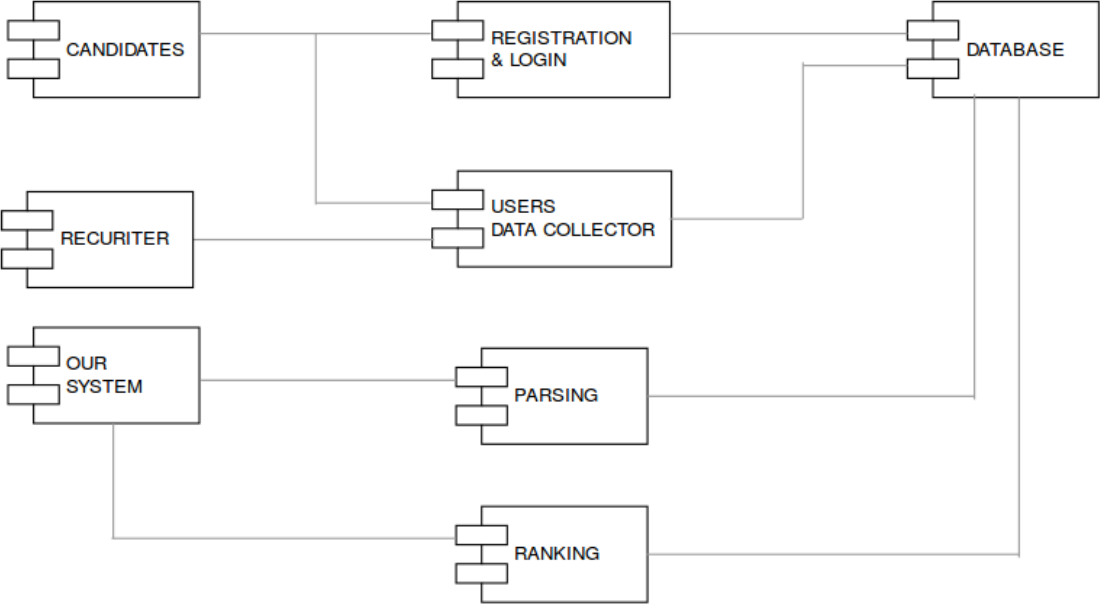


Fig 3.4.2

Deployment Diagram

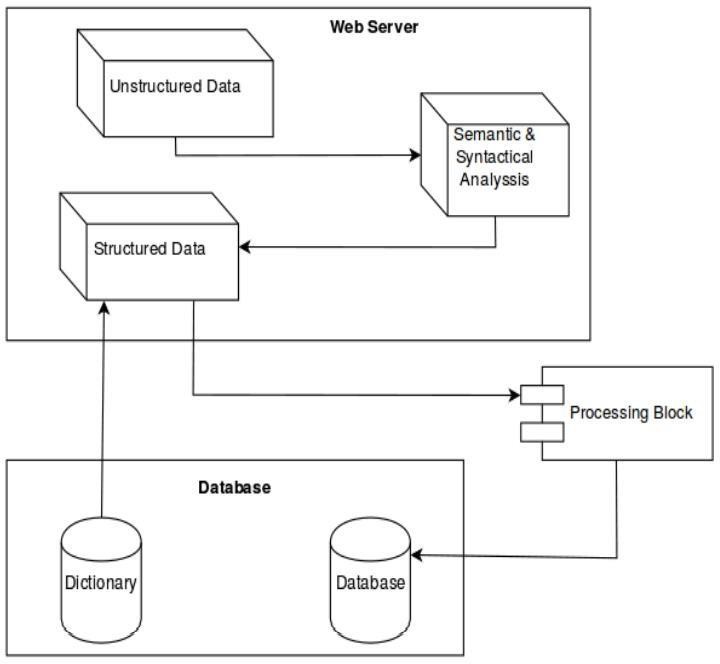


Fig 3.4.3

State Chart Diagram

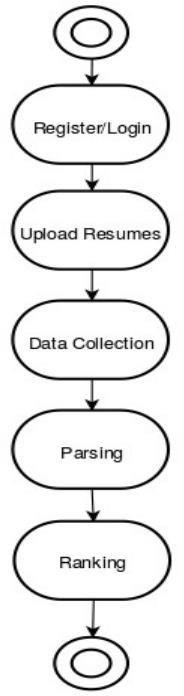
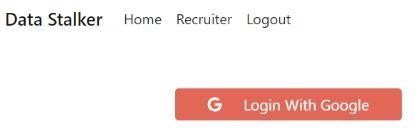
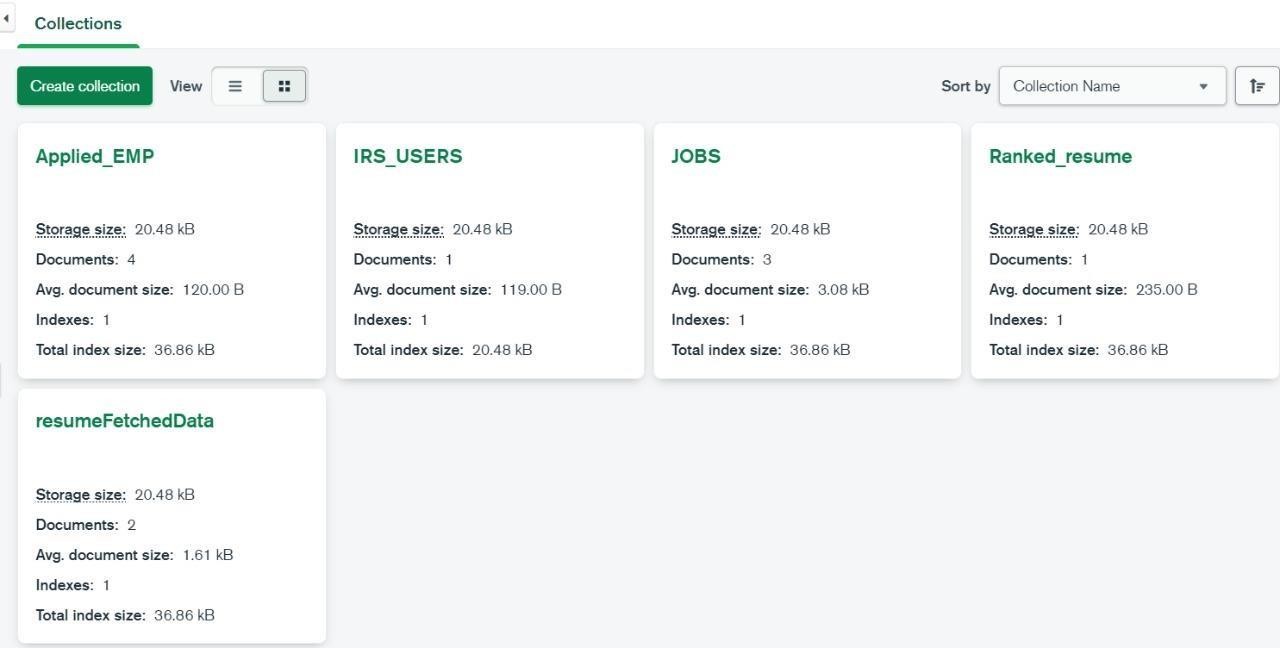
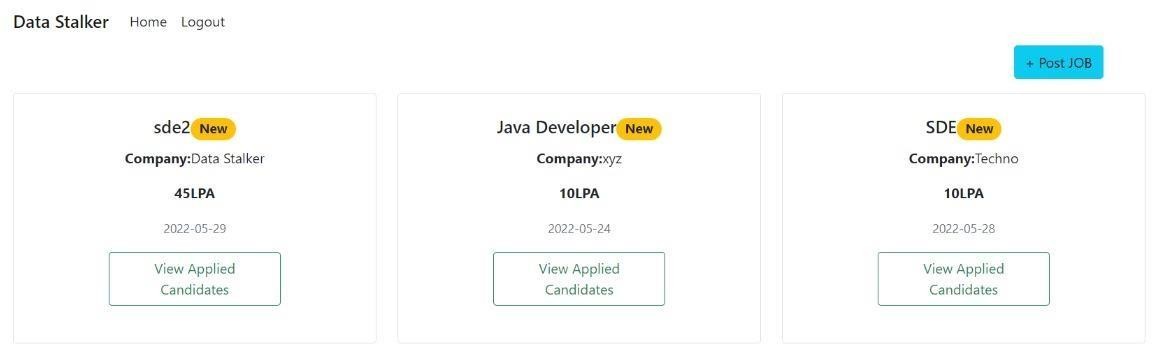


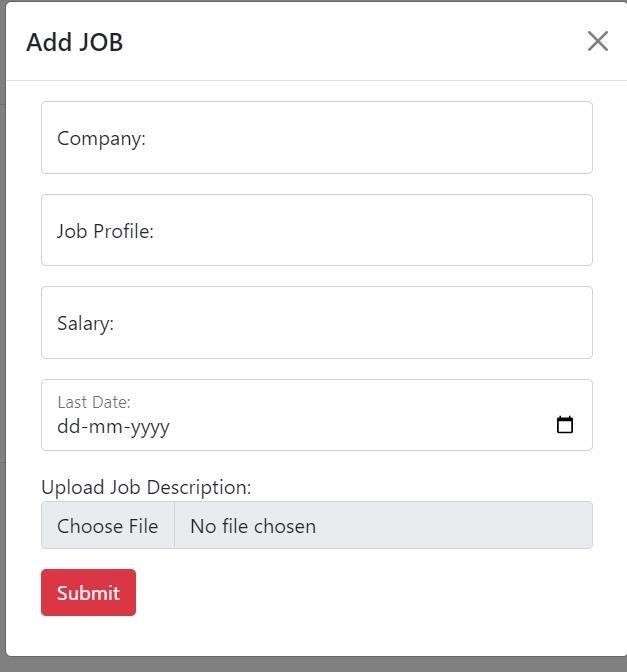
Fig 3.4.4

# SCREEN SHOTS

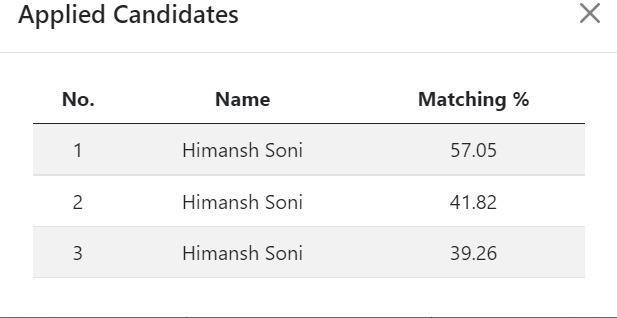














**CONCLUSION**

# Limitations

False-Positive Keywords

Candidates can figure out how the resume screening works. They know that by stuffing their resumes with keywords, they can circumvent the typical resume screening process. In this case, recruiters often end up with resumes that contain keywordsbut are not qualified for the job.

True-Negative Keywords

Some candidates could be perfect for the job but may end up being bypassed because their resumes were missing one or two keywords. They may also not pass through if the keywords are not mentioned precisely in the same manner on their resumes oruse different words that have the same meaning as the keywords.

Reliability

Reliability remains the most significant concern as AI-based recruiting tools may have several inconsistencies and flaws. AI-powered recruitment technology is still in its initial stages. It might not be able to recognize patterns in a resume, as well as you might expect.

For instance, an ATS might reject a resume because the applicant might have used a different font. The scanner mightnot be able to read an unusual format, page orientation, or picture placement.

Social recruitment is on the rise. If your AI-driven resume screening software includes social media activity in its screeningprocess, it could prove to be a disadvantage for those talents thataren’t very active on social media.

The only way to solve these challenges is by feeding more and more data into these tools. AI uses data to analyze patterns and combinations. To get more accurate predictive analysis with the help of AI, they need a massive amount of data to derive reliableconclusions.

Biases

AI recruiting tools work by analyzing old patterns. One of the main reasons we choose to use AI tools is to reduce the human biases involved in selecting candidates based on their name, age,gender, nationality, and more.

However, these resume screening software solutions might unknowingly bring bias into the system. When you feed your company’s recruitment data for the last ten years into the system, it learns from that data. It picks up the previous patterns.This challenge was faced by Amazon’s recruitment system, where they found that the system was biased against female candidates. This was because the last ten years data showed a bias against females.

The algorithm had no idea that it needed to ignore that pattern.Thus, hiring managers need to be extra careful about the algorithm being deployed.

# Future Enhancements

The HR industry is continually evolving. Recruiters need to beagile enough to adapt to these changing needs and demands quickly.

According to Glassdoor research, 76% of HR finds it difficult toattract the right job candidate. When there is fierce competition in the market, and every organization is looking to attract the toptalent, a hiring manager can’t afford to let a good resume slip through the cracks.

Technology has played a significant role in solving some of the biggest challenges that HR managers face. The introduction of AI in the recruitment process has transformed the hiring process.It has enabled HR managers and recruiters to scan large volumesof resumes. It helps find candidates that could be a perfect match for the open position. In fact, the use of AI in hiring technology has resulted in moretime savings. HR professionals can use their time in more

valuable ways to enhance the hiring experience for candidates.40% of companies have already adopted AI tech in their hiringprocess.

To understand how AI for recruitment has been a game-changer, you must first understand resume parsing

# Conclusion

It is a process by which resume data is analyzed and extracted into a format such as XML. In other words, new-age ATS solutions or a resume screening software that are powered by AIcreates unique candidate profiles.

It can look for specific information such as educational qualifications or skills and cut and paste this information into the designated fields in the candidate’s profile. Recruiters can search the database using keywords and receive a list of relevantresults.

* + - It eliminates any errors and inconsistencies that result frommanual data entry by recruiters or applicants.
    - Resume parsers extract and sort information into designated fields. Recruiters do not need to scan the entireresume to look for relevant keywords. They can view candidate information in a standard format.
    - Resume parsers are designed to add context to the keywords and understand the intention of using them togive more reliable results.
    - They do not just screen resumes based on keywords butuse variables. Therefore, they can help overcome false- positive and false-negative keyword issues.
    - They have capabilities to add filters to the search criteriafor better results. For example, candidates with work experience on a specific product can be identified easily.

The recruitment software market now offers AI-based ATSsolutions or even independent resume scanning software

solutions powered by AI that can be integrated with yourexisting ATS. Here’s what they do:

* + - They parse resumes to provide highly comprehensive andreliable screening results.
    - They even parse comments, notes, and interview feedbackto build a smarter database of candidates.
    - They learn the job qualifications from job descriptions andrefer to previous hiring decisions. This helps them identify suitable candidates with greater efficiency over time.
    - They can rank and grade candidates.
    - They can add more information about the candidate’s previous employers from public data sources and theirpublic social media pages.
    - The more data they have, the better the recommendationsthey can make. Therefore, they are especially great for high-volume hiring.
    - They learn from historical and current hiring decisions tobe able to predict hiring quality over time.
    - They can integrate with job posting sites.
    - Applications that have not been received via the ATS, for example, a CV sent to your email inbox, can be forwardedto the software for parsing and added to the database.

# BIBLIOGRAPHY

References/Bibliography <https://www.youtube.com/watch?v=Hxad1K6hYGU>

https://towardsdatascience.com/resume-screening-withpython-1dea360be49b

https://[www.analyticsvidhya.com/blog/2021/06/resumescreening-](http://www.analyticsvidhya.com/blog/2021/06/resumescreening-)with-natural- language-processing-in-