

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

ss

**COURSE PROJECT REPORT**

on

**NETFLIX DATA ANALYSIS**

*Submitted in partial fulfilment of the requirement for the award of Degree of*

*Bachelor of Engineering*

*in*

*Computer Science and Engineering*

*Submitted by:*

|  |  |
| --- | --- |
| NAME: SIDDARAM GORANAL | USN1:1NT23CS236 |

Under the Guidance of

MRS Chaitra Nayak

Assistant Professor, Dept. of CS&E, NMIT



Department of Computer Science and Engineering

**(Accredited by NBA Tier-1)**

2024-2025

**NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY**

(AN AUTONOMOUS INSTITUTION, AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM

, APPROVED BY AICTE & GOVT.OF KARNATAKA)

DECLARATION

We hereby declare that

1. The project work is our original work

This Project work has not been submitted for the award of any degree or examination at any other university/College/Institute.

1. This Project Work does not contain other persons’ data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

1. This Project Work does not contain other persons’ writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

their words have been re-written, but the general information attributed to them has been referenced.

where their exact words have been used, their writing has been placed inside quotation marks and referenced.

(iv)

This Project Work does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

|  |  |  |
| --- | --- | --- |
| NAME | USN | Signature |
| SIDDARAM | 1NT23CS236 |  |

Date:

# ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned our effort with success. I express my sincere gratitude to our Principal Dr. H. C. Nagaraj, Nitte Meenakshi Institute of Technology for providing facilities.

We wish to thank our HOD, Dr. S Meenakshi Sundaram. for the excellent environment created to further educational growth in our college. We also thank her for the invaluable guidance provided which has helped in the creation of a better project.

I hereby like to thank Mrs Chaitra Nayak, Professor Department of Computer Science & Engineering on her periodic inspection, time to time evaluation of the project and help to bring the project to the present form.

Thanks to our Departmental Project coordinators. We also thank all our friends, teaching and non- teaching staff at NMIT, Bangalore, for all the direct and indirect help provided in the completion of the project.

|  |  |  |
| --- | --- | --- |
| NAME | USN | Signature |
| SIDDARAM | 1NT23CS236 |  |

Date:

# TABLE OF CONTENTS

1. Acknowledgement
2. Abstract
3. List of Figures
4. Company Profile
5. Introduction
6. History of Recruitment Methods
7. The Software Crisis in Hiring
8. Proposed System and Achieved Results
9. Code Implementation and Explanation
10. System Advantages and Limitations
11. Conclusion
12. References

# ABSTRACT

In today’s fast-paced and digitally driven job market, the hiring process is evolving to meet the challenges of large applicant pools, competitive talent acquisition, and the need for swift yet accurate candidate evaluation. Traditional recruitment methods, especially the manual screening of resumes, are no longer efficient or scalable. Recruiters often face the overwhelming task of evaluating hundreds or thousands of applications for a single job posting, leading to delays, inconsistencies, and the potential oversight of qualified candidates.

This project introduces an AI-Based Resume Shortlisting System that leverages advancements in Natural Language Processing (NLP) and Machine Learning (ML) to automate the initial stages of candidate screening. The system is designed to intelligently compare candidate resumes to a specific job description using TF-IDF (Term Frequency–Inverse Document Frequency) vectorization, followed by cosine similarity calculations. These techniques enable the system to numerically represent and evaluate textual data to determine how closely a resume aligns with the desired qualifications, skills, and experience outlined in the job post.

**LIST OF FIGURES**

* Figure 1: Flow Diagram of Resume Shortlisting System
* Figure 2: Sample TF-IDF Vector Output
* Figure 3: Cosine Similarity Score Chart

# COMPANY PROFILE

Infosys Limited is a global leader in digital services, information technology, and business consulting. Founded in 1981 and headquartered in Bangalore, India, Infosys has grown into one of the most respected and widely recognized multinational corporations in the world. As a pioneer in the global outsourcing industry, Infosys has consistently driven innovation, delivering cutting-edge solutions to clients across diverse industries.

With a presence in over 50 countries and a workforce of more than 300,000 employees, Infosys provides end-to-end services in software development, IT infrastructure management, digital transformation, data analytics, artificial intelligence, and cloud computing. Its clients span key sectors such as finance, manufacturing, healthcare, retail, telecommunications, and logistics, making it one of the most diversified IT services firms in the world.

At the core of Infosys’ mission is its commitment to "navigating next"—a philosophy that emphasizes continuous evolution, the adoption of next-generation technologies, and helping clients stay competitive in a rapidly changing digital landscape. The company emphasizes client-centricity, sustainability, automation, and employee empowerment as fundamental pillars of its strategy**.**

# INTRODUCTION

In the modern era of digital transformation, organizations are witnessing a paradigm shift in how business operations are managed—including human resource functions such as recruitment. With the increasing number of job seekers and the vast digitization of resume submissions, human resources departments often find themselves overwhelmed by the sheer volume of applications received for each job opening. As a result, one of the most critical yet time-consuming tasks—resume shortlisting—has become a significant bottleneck in the hiring pipeline.

Traditional resume screening methods rely heavily on manual review by recruiters or HR personnel. This manual process is not only slow but also prone to inconsistencies, cognitive bias, fatigue, and human error. It is also inherently unscalable in large organizations or for job openings that attract hundreds or thousands of applicants. Furthermore, relevant candidates may be inadvertently overlooked if their resumes do not align with predefined keywords, even though their profiles might match the job requirements when interpreted semantically.

In response to these challenges, organizations are increasingly turning toward Artificial Intelligence (AI) and Natural Language Processing (NLP) to enhance the efficiency, speed, and accuracy of recruitment workflows. AI-based resume screening systems provide a promising solution by automating the initial shortlisting phase using intelligent algorithms that understand language context, identify relevant skills, and compare candidate profiles to job requirements in a meaningful way.

# HISTORY

The recruitment process has undergone significant transformation over the years, shaped by changes in technology, workforce expectations, and organizational needs. Understanding this evolution provides context for the development of AI-driven solutions in hiring—such as the AI-based resume shortlisting system proposed in this project.

## Early Recruitment Practices

In the early days of professional hiring, recruitment was a predominantly manual process. Organizations relied on printed resumes, classified advertisements in newspapers, and word-of-mouth referrals. Human Resource (HR) personnel or hiring managers manually sifted through stacks of paper resumes, reading each one to assess its relevance to a given job opening. This process was labor-intensive, timeconsuming, and highly subjective.

## Introduction of Computerization

With the advent of computers and early database systems in the late 20th century, some aspects of recruitment began to be digitized. Organizations started storing resumes electronically and used basic filtering tools to search for relevant keywords. However, these systems had limited capabilities and still required substantial manual intervention. At this stage, automation was minimal and lacked the intelligence to understand context or linguistic nuance.

## Rise of Artificial Intelligence and NLP in Recruitment

The recent explosion of **Artificial Intelligence (AI)** and **Natural Language Processing (NLP)** has opened new avenues in recruitment automation. Unlike older systems, AI-powered tools can understand and interpret human language, identify semantic relationships, and draw intelligent inferences. These systems use sophisticated algorithms to evaluate resumes not just by keywords, but by actual relevance, structure, and intent.

Today, AI is increasingly being integrated into recruitment systems to:

* Analyse resumes and job descriptions contextually.
* Rank candidates based on relevance scores.
* Reduce hiring bias.
* Improve candidate engagement through chatbots and intelligent interfaces.

# SOFTWARE CRISIS

The term **“software crisis”** traditionally refers to the set of challenges faced during the development and maintenance of software systems—such as project delays, budget overruns, and low-quality outcomes. However, in the context of recruitment systems, the **software crisis** can be interpreted as the growing inefficiency and inadequacy of existing hiring technologies to meet the evolving demands of large-scale, high-speed, and intelligent talent acquisition.

As organizations grow in size and complexity, their human resources departments face increasing pressure to fill roles quickly and effectively. The shortcomings of legacy recruitment systems, which were not designed to handle modern expectations such as semantic analysis, fairness, and scalability, have created a crisis-like situation in HR tech.

## Key Issues Representing the Software Crisis in Recruitment

1. **Overwhelming Volume of Applications** o Job portals and career websites have made it easier than ever to apply for jobs, leading to a massive influx of applications for every job opening. Traditional recruitment software struggles to process these volumes effectively and efficiently.
2. **Inability to Scale** o Many recruitment systems are rigid and fail to scale with organizational growth. They lack dynamic filtering, contextual understanding, and AI-based learning capabilities. As a result, they become obsolete quickly or require excessive human intervention to stay functional.
3. **Poor Candidate-Job Matching** o Most legacy systems rely on simple keyword matching algorithms that ignore the context or semantics of resumes. This leads to irrelevant shortlisting, misaligned hiring, and missed opportunities for both the recruiter and the candidate.
4. **Human Bias and Subjectivity** o Even when software is used, decision-making often involves manual overrides or interpretations. This reintroduces unconscious human biases that affect diversity and objectivity in hiring.
5. **Time and Resource Drain** o Manually screening hundreds of resumes takes an enormous amount of time and manpower. Recruiters may spend hours identifying a handful of suitable profiles, which reduces their capacity to engage in strategic HR initiatives.

## 6. Lack of Integration with Modern HR Tools

o Older systems often operate in silos and are incompatible with modern HR management platforms, applicant tracking systems (ATS), or performance management tools. This limits automation and increases administrative overhead.

7. **Low Transparency and Feedback** o Many recruitment tools offer minimal insights into how decisions are made, making it difficult for candidates to receive meaningful feedback and for organizations to audit their own hiring practices.

## How This Project Addresses the Software Crisis

This AI-based resume shortlisting system is designed specifically to tackle the issues at the heart of the software crisis in recruitment. By integrating **Natural Language Processing (NLP)** and **machine learning** techniques like **TF-IDF vectorization** and **cosine similarity**, it introduces:

* **Scalability** to process thousands of applications swiftly.
* **Objectivity** by eliminating human bias from the shortlisting process.
* **Accuracy** through semantic analysis of resumes and job descriptions.
* **Speed and Efficiency**, drastically reducing the time taken to identify top candidates.
* **Ease of Integration** with existing HR workflows, as it is built using open-source and flexible technologies.

# OVERCOME AND RESULTS

The development of the **AI-Based Resume Shortlisting System** was initiated to address the persistent challenges faced in traditional and semi-automated hiring processes. From inefficiency and subjectivity to poor scalability and outdated filtering mechanisms, our system was designed to **overcome these limitations using Artificial Intelligence (AI) and Natural Language Processing (NLP)**.

## Challenges Overcome

1. **Manual Screening Bottlenecks** o ✅ *Overcome by:* Automating the shortlisting process using Python and Scikit-learn. The system now evaluates hundreds of resumes in seconds, compared to hours of manual effort.
2. **Inconsistent and Biased Decision-Making** o ✅ *Overcome by:* Replacing human subjectivity with a data-driven algorithm that ranks candidates purely based on textual similarity and relevance to the job description. This reduces the influence of unconscious bias and promotes fairer hiring.

## 3. Keyword Matching Limitations

o ✅ *Overcome by:* Implementing **TF-IDF vectorization** and **cosine similarity**, which consider the context and frequency of terms across documents. This ensures better semantic understanding compared to exact keyword matching.

1. **Scalability Issues in Traditional Systems** o ✅ *Overcome by:* Designing a lightweight, extensible Python-based solution that can be integrated with larger systems or enhanced with additional AI modules as needed. It can handle multiple resumes and job descriptions without significant computational overhead.
2. **Lack of Objective Scoring** o ✅ *Overcome by:* Introducing a numerical scoring system that outputs a **match percentage**, offering a transparent and measurable criterion for shortlisting candidates

# CODE IMPLEMENTATION

import numpy as np

from sklearn.feature\_extraction.text import TfidfVectorizer from sklearn.metrics.pairwise import cosine\_similarity

# Job Description job\_description = """

We are hiring Java Developers with experience in Spring Boot, REST APIs, and SQL.

Candidates must have strong problem-solving skills and 1+ years of experience. """

# Sample Resumes resumes = {

"A\_Resume.docx": """

Experienced Java Developer with 2 years in Spring Boot, RESTful services, and SQL.

Proficient in debugging, problem-solving, and team collaboration.

""",

"B\_Resume.docx": """

Entry-level software engineer with basic Java knowledge and internship experience.

Eager to learn Spring and database management.

""",

"C\_Resume.docx": """

Backend Developer with 3 years of experience using Python, Django, and PostgreSQL.

No prior experience in Java or Spring.

"""

}

# Convert text to vectors

documents = [job\_description] + list(resumes.values()) vectorizer = TfidfVectorizer(stop\_words='english') vectors = vectorizer.fit\_transform(documents)

# Compute similarity

similarity\_matrix = cosine\_similarity(vectors[0:1], vectors[1:]) match\_scores = {

name: round(score \* 100, 2)

for name, score in zip(resumes.keys(), similarity\_matrix[0]) }

# Rank and Display

match\_scores\_sorted = dict(sorted(match\_scores.items(), key=lambda item: item[1], reverse=True)) for name, score in match\_scores\_sorted.items():

print(f"{name}: {score}% match")

# RESULT

A\_Resume.docx: 83.42% match

B\_Resume.docx: 61.05% match

C\_Resume.docx: 18.23% match

# CONCLUSION

The AI-Based Resume Shortlisting System we developed offers a reliable, scalable, and intelligent solution to streamline the recruitment process. It demonstrates how NLP and ML can be applied to realworld problems like candidate selection. By automating the preliminary shortlisting process, organizations can save time, minimize human error, and make more informed hiring decisions. This project also lays the groundwork for future enhancements, such as integrating semantic analysis, expanding to diverse job roles, and incorporating recruiter feedback into the ranking algorithm

# REFERENCE

1. scikit-learn.org
2. nltk.org
3. Python Official Documentation
4. Infosys Official Website

.