# Abstract

# This paper proposed to designing an artificial intelligent system to extract information from unstructured resumes and transform that information into structured format and classifying those resumes based on the information extracted like skill sets, experience and educational qualification of the candidate and based on the job description of the company through a deep learning network. With that we will rank those resumes by a similarity framework and recommend top five most relevant profiles to the human resource department of the organization.

**Table of Contents**

[Abstract 1](#_Toc24536108)

[1. Introduction 3](#_Toc24536109)

[2. Background and related research 3](#_Toc24536110)

[3. Aim and Objectives 3](#_Toc24536112)

[6. Research Methodology 4](#_Toc24536113)

[7. Expected Outcomes 4](#_Toc24536114)

[8. Requirements / resources 4](#_Toc24536115)

[9. Research Plan 4](#_Toc24536116)

[References 4](#_Toc24536117)

# 1. Introduction

Among the key problems of the 21*st* century is unemployment, a multi-faceted societal and

Industrial phenomenon. Nowadays, almost every institute or department has its own job posting platform. Machine learning and Knowledge Discovery have been investigating human resource-related applications for over a decade.

It’s easy to think of resume as a summary of work experience. But a resume isn’t just about listing about professional history. Rather, it’s a strategic tool for marketing an individual brand. This phenomenon means that the structure of resume data has a great deal of uncertainty. It decreases the success rate of recommending recruits who meet most of the employer’s requirements and take up too much time of human resources to do job matching. In order to improve the efficiency of job matching, exploring an effective method to match jobs and candidates is important and with the help of AI, natural Language processing used to understand human language and it adds the ability to interpret humanly created documents.

This paper focuses on to propose a novel three-step approach i.e. First AI/ML based extraction algorithm that is proposed for resume facts extraction. Second classifying those resumes based on the information extracted like skill sets, experience and educational qualification of the candidate and based on the job description of the company through a deep learning network. Third recommending the most similar profiles.

# 2. Background and related research

To find relevant literature on e-recruiting and data mining from resumes, according to the adopted features, there are three kinds of popular methods about resume information extraction and recommendation in previous research, which can be described as follows.

The first group of methods takes keywords retrieval in consideration to screen out the

Resume which are not of use or which are not present in the given criteria. In this kind of

Scenario resumes are classified based on skill types or experience. This approach parses

Resumes for the given keyword, irrespective of what a sentence mean in the resume, once

the keyword is found that resume is recommended to the recruiter by this applicationAlthough these kinds of methods are easy to implement, the raw text content brings too many noises into the index, leading to low precision and unsatisfactory ranking results.

The second group of methods like few published studies tried to learn the information extraction rules for resumes using XML tags to identify key attributes namely email, name, street, Province, etc. In which a tag tree algorithm, in which they detected and removed the shared part among web pages with the same template, and then the main text is retained.

The third group of methods treats extracting knowledge as a semantic-based entity extraction task. The first pass is used to segment the resume. Then detailed information, such as Name and Address, are identified in certain blocks without searching globally in the whole resume. Their system integrates table analyser, CRF predictor, and content recognizer into the whole part of parsing resumes.

Keeping these things in mind and to provide a better recommendation to recruiter, This paper focuses on to propose a novel three-step approach for finding the right person with the right skills for the right role.

# 3. Aim and Objectives

To designing an artificial intelligent system that will extract information from unstructured resumes and it will classify those resumes based on the information extracted like skill sets, experience and educational qualification of the candidate and based on the job description of the company through a deep learning network. With that we will rank those resumes by a cosine similarity matrix and recommend the top five most relevant profiles to the human resource department of the organization. This paper focuses on to propose a novel three-step approach for finding the right person with the right skills for the right role i.e. first an AI/ML based information extraction. Second classifying those resumes based on the information extracted and third recommending the most similar profiles though cosine similarity.

# 4. Research Methodology

We proposed a three step approach to find the best fit candidate for the right role.The presented study is based on the corpus provided here, which includes 2000 resumes across different

1. **Information Extraction**
2. **Classifications of Resumes**

1. **Similarity Measure and Recommending**

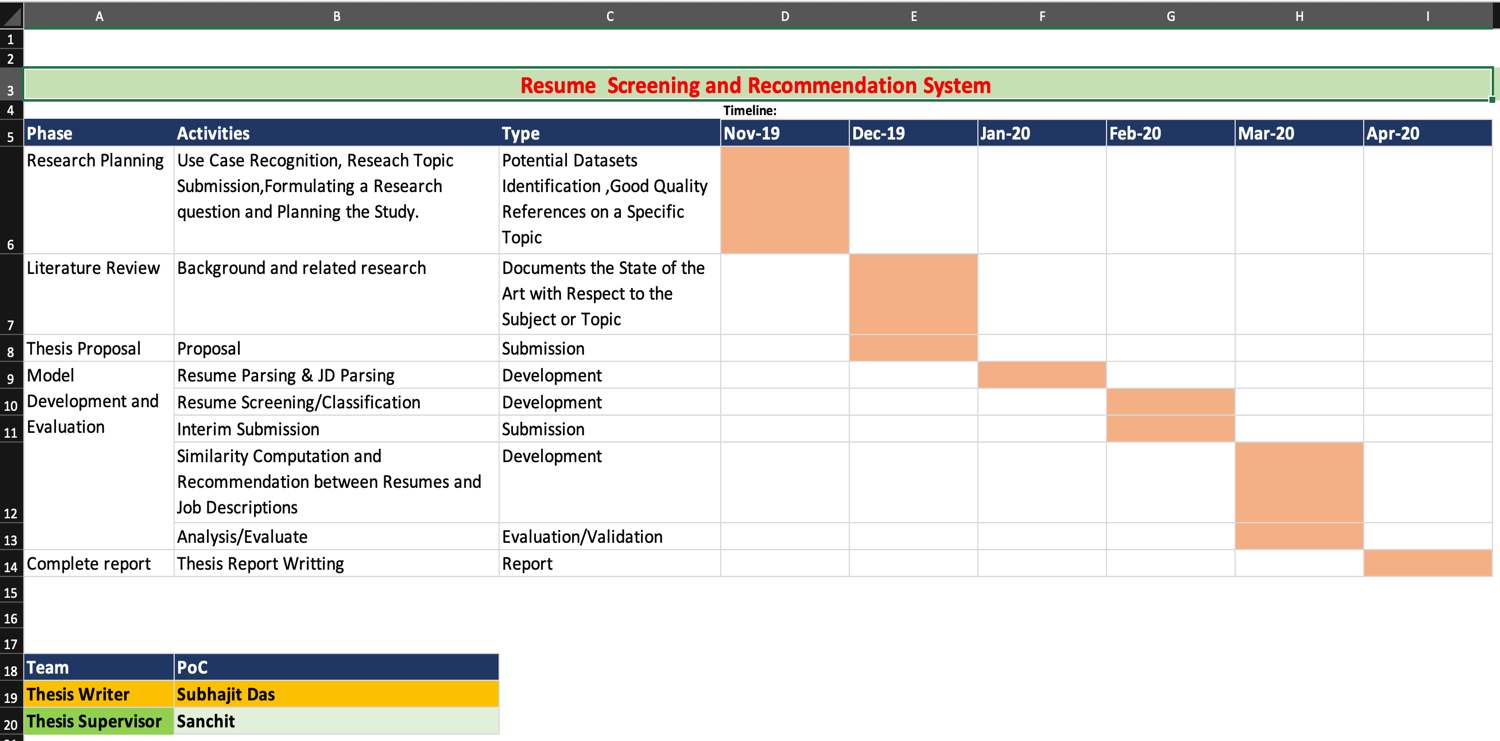
Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two word embedding vectors and determines whether two vectors are pointing in roughly the same direction. It will be used to measure document similarity. **Cosine similarity** is a measure of similarity that can be used to compare documents or, say, give a ranking of documents with respect to a given vector of query words. Let ***x*** and ***y*** be two vectors for comparison. Using the cosine measure as a similarity function, we have



# 5. Expected Outcomes

# 6. Requirements / resources

# 7. Research Plan



# References

**Refer: Harvard Referencing Guide**