

# **VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI**



**A PROJECT WORK PHASE-I REPORT ON**

## **RESUME SCREENING**

Submitted in partial fulfillment for the award of Degree of,

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**By**

**SHAILASHREE**

**4AL18CS077**

**SHIZA AMREEN J**

**4AL18CS078**

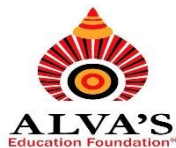
**SUJAYA**

**4AL18CS090**

**Under the Guidance of**

**Mrs. Merlyn Melita**

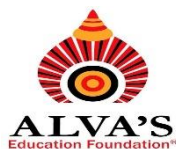
**Assistant Professor**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY  
MOODBIDRI-574225, KARNATAKA**

**2021 – 2022**

**ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY**  
**MIJAR, MOODBIDRI D.K. -574225**  
**KARNATAKA**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CERTIFICATE**

This is to certify that the Project Work Phase-I entitled “**RESUME SCREENING**” has been successfully completed by

<b>SHAILASHREE</b>	<b>4AL18CS077</b>
<b>SHIZA AMREEN J</b>	<b>4AL18CS078</b>
<b>SUJAYA</b>	<b>4AL18CS090</b>

the bonafide students of **Department of Computer Science & Engineering, Alva's Institute of Engineering and Technology** in partial fulfillment of 7<sup>th</sup> semester, BACHELOR OF ENGINEERING in **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING** of the **VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI** during the year 2021–2022. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library.

---

**Mrs. Merlyn**

**Project Guide**

---

**Mrs. Vidya**

**Project Coordinator**

---

**Dr. Manjunath Kotari**

**Head of the Department**

## ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany a successful completion of any task would be incomplete without the mention of people who made it possible, success is the epitome of hard work and perseverance, but steadfast of all is encouraging guidance.

So, with gratitude we acknowledge all those whose guidance and encouragement served as beacon of light and crowned the effort with success.

We thank our project guide **Mrs. Merlyn Melita**, Assistant Professor, in Department of Computer Science & Engineering, who has been our source of inspiration. She has been especially enthusiastic in giving her valuable guidance and critical reviews.

The selection of this project work as well as the timely completion is mainly due to the interest and persuasion of our project coordinator **Mrs. Vidya**, Assistant Professor, Department of Computer Science & Engineering. We will remember her contribution for ever.

We sincerely thank, **Dr. Manjunath Kotari**, Professor and Head, Department of Computer Science & Engineering who has been the constant driving force behind the completion of the project.

We thank our beloved Principal **Dr. Peter Fernandes**, for his constant help and support throughout.

We are indebted to **Management of Alva's Institute of Engineering and Technology, Mijar, Moodbidri** for providing an environment which helped us in completing our project.

Also, we thank all the teaching and non-teaching staff of Department of Computer Science & Engineering for the help rendered.

SHAILASHREE 4AL18CS077

SHIZA AMREEN J 4AL18CS078

SUJAYA 4AL18CS090

# **ABSTRACT**

Choosing the right people for the job is the biggest responsibility of every business since it deals with hiring the best fit candidate having relevant skills for the job profile from an immensely large pool of candidates. Recruitment is a tedious process. Wherein the first task for any recruiter is to screen resumes. Due to lack of time, companies do not have enough time to open resumes, and it is an extremely tiring task for the recruiters of company to manually go through thousands of resumes. To solve this problem, the proposed system is designed in a such a way that the recruiters can screen the resumes and the resumes submitted by the candidates are compared with the job profile requirement posted by the company by using Natural Language Processing(NLP). The extraction of data from the resume given as input by the job applicants is done by Natural Language Processing. The implementation of this project is done by using Python programming language and spaCy and NLTK(Natural Language ToolKit) an open-source libraries for Natural Language Processing in Python.

# LIST OF FIGURES

Figure No	Description	Page no
1.1	Block Diagram of Proposed System	2
4.1	System Architecture	8

# TABLE OF CONTENTS

CHAPTER NO.	DESCRIPTIONS	PAGE NO.
	ACKNOWLEDGEMENT.....	i
	ABSTRACT.....	ii
	LIST OF FIGURES.....	iii
	LIST OF TABLES.....	iv
<b>1.</b>	<b>INTRODUCTION</b>	
1.1	PROBLEM STATEMENT	1
1.2	MOTIVATION AND OBJECTIVES OF THE PROJECT	1
1.3	EXISTING SYSTEM & DRAWBACKS	2
1.4	PROPOSED SYSTEM & ADVANTAGES	2
<b>2.</b>	<b>LITERATURE SURVEY</b>	3
<b>3.</b>	<b>REQUIREMENT SPECIFICATION</b>	
3.1	FUNCTIONAL REQUIREMENTS	4
3.2	NON-FUNCTIONAL REQUIREMENTS	5
3.3	TECHNICAL REQUIREMENT (HARDWARE/ SOFTWARE)	6
3.4	SOFTWARE DESCRIPTION	6
<b>4.</b>	<b>SYSTEM ANALYSIS AND DESIGN</b>	8
	<b>REFERENCES</b>	10

# CHAPTER 1

## INTRODUCTION

Job Requirement is considered one of the major activities for humans which is a very strenuous job to find a fruitful talent. Our proposed model is basically to extract the details and statistics from the resume and ranking the resume based on the preference of the company associated and its requirements using the Natural Language Processing (NLP) techniques. Parsing and ranking the resume makes the hiring process easy and efficient. A resume contains various minute data within it and any respectable parser needs to extract out these data such as education, experience, project, address etc. So, basically, we are going to build a job portal where the employees and applicants would upload their resume for any particular job and using the NLP technique, the necessary information will be parsed and a structured resume with information will be generated and also the resumes of employee will be ranked according to the requirement of the company skill set and employees skills in the provided resume.

### 1.1 PROBLEM STATEMENT

The problem statement of this project is Due to lack of time companies and their team does not have time to read resumes and choose the best resume according to their requirements. For this task they have to take the help of any other company. For which they have to pay money. Which is a very serious problem.

### 1.2 MOTIVATION AND OBJECTIVE OF THE PROJECT

In this it gives an outlook of an ongoing project on deploying information extraction techniques in the process of resume information extraction into compact and highly-structured data. This process has been able to reduce lots of burden on the shoulder of users of recruitment agency. The Resume Parser automatically segregates information on the basis of various fields and parameters like name, phone / mobile nos. etc. and huge volume of resumes is no problem for this system and all work is done automatically without any personal or human intervention. the proposed solution involves the process of screening resumes is automated by using Natural Language Processing. The major objective of system is to automate the hiring process in order to reduce the cost of hiring and to make the hiring process more efficient.

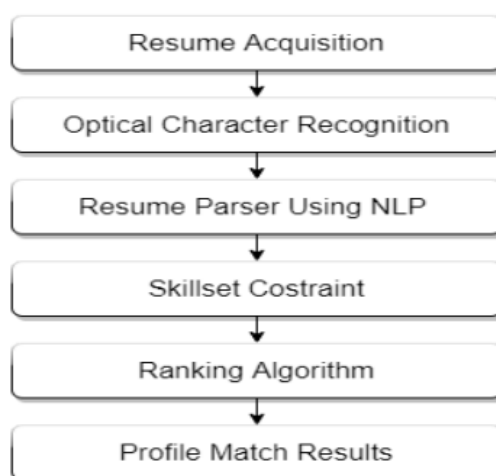
### 1.3 EXISTING SYSTEM AND DRAWBACKS

The existing system was using a manual process which makes it less efficient and more time consuming. Manual process includes Candidate to upload their resume in particular format and submit it to the agency. Then these agencies would scrutinize the candidates based on some constraints.

The drawbacks of using Manual Method Separating right candidates from the pack - India being a huge job market and with millions seeking jobs; it is humanly impossible to screen the CVs and find the right match. This makes the whole hiring process slow and inefficient costing resources to the companies. Making sense of candidate CVs - Second challenges are posed by the fact that the CVs in the market are not standard practically every resume in the market has different structure and format. HR has to manually go through the CVs to find the right match to the job description. This is resource intensive and prone to error whereby a right candidate for the job might get missed in the process.

### 1.4 PROPOSED SYSTEM AND ADVANTAGES

The proposed system In the first step, the system accepts the resume from the job applicant and performs text extraction on it. Text is then fed to the pretrained model that takes only that data which is relevant to the system. After this step, the data extracted will be displayed to the job applicant in editable format. Then it can be sent for comparison with the job description requirements which is already passed through pre-trained model and the requirements like number of year of experience, qualifications, etc. is already known. Now, the job data(after passing through pre-trained model) is compared with the resume data(after passing through pre-trained model) after which it can calculate the score based on the formula. Finally, it will display the candidate profile to the recruiter with other applicants for same profile in a rank list based on their score.



**Fig 1.1 Block Diagram of Proposed System**



## CHAPTER 2

### LITERATURE SURVEY

In this project, This technique stated parsing of the resumes with least limit and the parser works the utilization of two or three rules which train the call and address. Scout bundles use the CV parser system for the determination of resumes. As resumes are in amazing arrangements and it has different sorts of real factors like set up and unstructured estimations, meta experiences, etc. The proposed CV parser approach gives the component extraction method from the moved CV's.

E-Recruitment System Through Resume Parsing, Psychometric Test And Social Media Analysis and It follows an approach of 4 stages, the first stage was to get the data (resume) and convert them into structured format and then perform the analysis using deep learning techniques. Second step includes the psychometric test where the text mining is used to generate scores for each candidate. In the third step they perform web scraping on various social media sites to get the additional information about the candidates and recommend suitable jobs to them. In the fourth step, the system will recommend the skills and requirements in which the students are lacking and also help them to get recruited in the desired company.

An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes and In this work, a qualitative assessment of resumes on the basis of different quality parameters using a simple text analytic based approach for a resume collection was described. The resume collection was processed for two qualitative coverage, comprehensibility and the aspects; and extracted ratings are modified into a quality rating which is comprehensive. All the parameters were collectively uniformed into a combined 1 to 5 rating scale for associating a quality metric for resumes. The qualitative evaluation results obtained through the algorithmic approach were congruent to and were hence validated through the wisdom of crowds.

## CHAPTER 3

# REQUIREMENT SPECIFICATION

The System Requirement Specification (SRS) is a document, which describes completely the external behaviour of the software as well as the behaviour of the hardware. The first and foremost work of a developer is to study the system to be developed and specify the user requirements before going for the designing phase. The document show how will the system behaves and respond. The basic goal of the requirements phase is to produce the software requirements specification, which describes the complete external behaviour of the proposed system. The complete description of the behaviour of a system developed is explained below. It includes the functional requirements of the system. In addition to the functional requirements the non-functional requirements and user interface requirements are also explained. Requirement analysis is done in order to understand the problem the of human errors to solve. The problem could be eliminating an existing manual process, developing a new system, or a combination of two. For the larger system that have many feature, and that need to perform any different tasks, understanding the requirements of a system is a major task. The software should be simply and easy to understand with an interactive interface. Standard compliances specify the requirements for the standards that the system must follow. To specify requirements completely, requirement specification documents should specify the certain properties of the software. Some of them are:

- The SRS should specify all the functions of the software is to support.
- The SRS should have performance requirements.
- The SRS should specify the design constraints.

### 3.1 FUNCTIONAL REQUIREMENTS

Dataset for training is obtained from Kaggle, consist of 1000 CT scans of both large and small tumours saved in Digital Imaging and Communications in Medicine (DICOM) format. The datasets are used from Kaggle which contains the sample CT scan images of the lung and the CSV file which contains information about the patient.

#### 3.1.1 KAGGLE DATASETS

The simplest and best-supported file type available on Kaggle is the “Comma-Separated List”, or CSV, for tabular data. CSVs uploaded to Kaggle should have a header row consisting of human-readable field names. CSVs are the most common of the file formats available on Kaggle and are the best choice for tabular data.

On the Data tab of a dataset, a preview of the file’s contents is visible in the data explorer. This makes it significantly easier to understand the contents of a dataset, as it eliminates the need to open the data in a Notebook or download it locally. CSV files will also have associated column descriptions and column metadata. The column descriptions allows us to assign descriptions to individual columns of the dataset, making it easier for users to understand what each column means. Column metrics, meanwhile, present high-level metrics about individual columns in a graphic format.

## **3.2 NON-FUNCTIONAL REQUIREMENTS**

In system engineering and requirements engineering, a non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. They are contrasted with functional requirements that define specific behavior or functions. The plan for implementing functional requirements is detailed in the system design. The plan for implementing non-functional requirements is detailed in the system architecture, because they are usually architecturally significant requirements. The focus on functional requirements is because they produce tangible output. However, non-functional requirements contribute to the infrastructure rather than system behaviour. The business infrastructure, being intangible, appears negligible.

### **3.2.1 AVAILABILITY REQUIREMENT**

The system detects whether the person is suffering from lung cancer or not by taking the lung CT images. The obtained data by automatic processing, analysis and access, and this leads to more timely and cost-effective production and management effort on lung cancer detection.

### **3.2.2 FLEXIBILITY**

It provides the users to load the image easily. For that purpose we used concept flexible window, that is, the Algorithm for lung cancer detection.

### **3.2.3 MAINTAINABILITY**

Maintainability is considered, inherent to the building system design, ensuring the ease, accuracy, safety, and economy of maintenance tasks within that system. The purpose of maintainability is to improve effectiveness and efficiency of maintenance.

### **3.2.4 USABILITY REQUIREMENT**

Usability requirement specifies how easy the system must be to use. The usability requirements must be tangible so that we are able to verify them and trace them during.

### **3.2.5 SCALABILITY**

Scalability is the capability of a system or process to handle an enhanced level of operations without constraints or structural bottlenecks. Every business model gives paramount importance to business generation, which leads to higher transactional volume and its consequential surge in operational activity. Scaling up the operations to handle increased business activity is inherent and built into the system design.

### **3.3 TECHNICAL REQUIREMENTS**

In Software projects, technical requirements typically refer to how the software is built, for example: which language it's programmed in, which operating system it's created for, and which standards it must meet.

#### **3.3.1 HARDWARE REQUIREMENTS**

Hardware components that deal with basic hardware requirement to develop and to run any system. It includes processor, memory etc.

- 400 MB hard disk space.
- 4 GB/8 GB RAM.
- Intel i3 or any processor above it and 4 core CPU.
- Operating systems of windows 10 will be sufficient.
- Active internet connection and a scanner optional

#### **3.3.2 SOFTWARE REQUIREMENTS**

Software Requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application.

- Windows(x64) Operating System
- Python
- Required Datasets
- Jupyter Notebook.

### **3.4 SOFTWARE DESCRIPTION**

Software, instructions that tell a computer what to do. Software comprises the entire set of programs, procedures, and routines associated with the operation of a computer system.

#### **3.4.1 PYTHON**

Python is an interpreted high-level general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. Python is a multi-paradigm programming language.

Python uses dynamic typing and a combination of reference counting and a cycle detecting garbage collector for memory management. It also features dynamic name resolution (late binding), which binds method and variable names during program execution.

### **3.4.2 NATURAL LANGUAGE PROCESSING**

Natural Language Processing (NLP) is a process of manipulating or understanding the text or speech by any software or machine. An analogy is that humans interact and understand each other's views and respond with the appropriate answer. In NLP, this interaction, understanding, and response are made by a computer instead of a human. NLTK (Natural Language Toolkit) Library is a suite that contains libraries and programs for statistical language processing. It is one of the most powerful NLP libraries, which contains packages to make machines understand human language and reply to it with an appropriate response.

### **3.4.3 MACHINE LEARNING**

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, and computer vision. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the MNIST dataset of handwritten digits has often been used.

### **3.4.4 NATURAL LANGUAGE TOOL KIT**

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

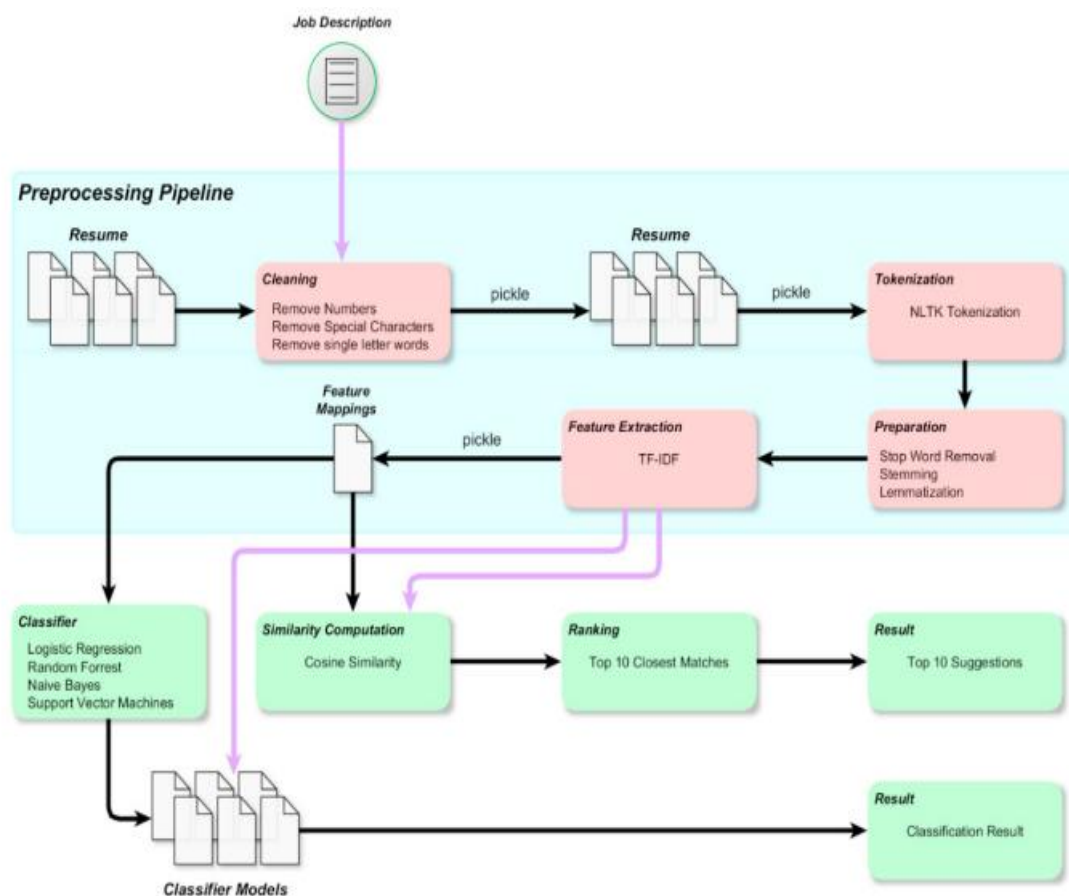
## CHAPTER 4

# SYSTEM ANALYSIS AND DESIGN

System Analysis is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components. System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem-solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose. System Design process of planning a new business system or replacing an existing system by defining its components or modules to satisfy the specific requirements.

## 4.1 ARCHITEUTURE

The System architecture is as shown in Fig 4.1



**Fig 4.1 System Architecture**

The model comprises of the phases as shown in the diagram, our entire work flow can be divided into Data Pre processing, Vectorization of text, Fig. 4.1. proposed pipeline Training and comparing different models for classification, Selection of the best model, and finally, Summarization. Preprocessing involves converting PDF to text, by making use of the well known Optical Character Recognition.

Data Cleaning involves Conversion of encoding, Censored Information ,Removal of Unnecessary separators, Removal of Punctuation and Stop Words. Vector Processing of Text in this step the classifiers and learning algorithms can not directly process the text documents in their original form, as most of them expect numerical feature vectors with a fixed size rather than raw text docs with variable length. Therefore, during this step, the text is converted to a more manageable representation. Summarization is an important part of our tool. The summarization technique that we have used here marks use of a lexical token summarizer. It is a summarizer purely based out of scoring the sentences based on the words on sentences.

# REFERENCES

- [1]. Vinaya R. Kudatarkar, ManjulaRamannavar, Dr.Nandini S. Sidnal “An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes”, 2015, IJIRAE.
- [2]. Satyaki Sanyal, Neelanjan Ghosh, SouvikHazra, Soumyashree Adhikary, “Resume Parser with Natural language Processing”, 2007,IJESC.
- [3]. Papiya Das, Manjusha Pandey, Siddharth Swarup Rautaray, “A CV parser Model using Entity Extraction Process and Big Data Tools ”, 2018,IJITCS.
- [4]. Ayishathahira and Sreejith,, “Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing”,International CET Conference on Control, Communication and Computing(IC4),2018..
- [5]. Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S,Kaveri BW, “E-Recruitment System Through Resume Parsing,Psychometric Test and Social Media Analysis”, 2019,IJARBEST.
- [6]. Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh, “NLP based Extraction of Relevant Resume using Machine Learning.