

SAVITRIBAI PHULE PUNE UNIVERSITY



A MINI PROJECT REPORT ON

“Resume Scanner using Google Colab”

Submitted by

Name: Swapnil Dhamdhare

Roll no: A-44

Name: Aditi Kulkarni

Roll no: B-04

Name: Ajinkya Chawane

Roll no: A-29

Name: Tanmay Dangat

Roll no: A-35

CLASS: BE

DIV: A

Under the Guidance of

Mrs. Vanita Babanne



Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

RMD SINHGAD SCHOOL OF ENGINEERING

WARJE, PUNE 411058

Academic Year 2023 – 24(SEM-II)



Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

RMD SINHGAD SCHOOL OF ENGINEERING

WARJE, PUNE 411058

CERTIFICATE

This is to certify that the project report entitles

“Resume Scanner using Google Colab”

Submitted by

Name: Swapnil Dhamdhere

PRN No : 72218238E

is a bonafide work carried out by them under the supervision of Mrs. Vanita Babanne. And it is submitted towards the partial fulfillment of the requirement of Savitribai Phule Pune University for Final Year.

(Mrs. Vanita Babanne)

Guide

Department of Computer Engineering

(Dr. Vina M. Lomte)

Head,

Department of Computer Engineering

(Dr. V. V. Dixit)

Principal,

RMD Sinhgad School of Engineering Pune – 58

II

Certificate by Guide

This is to certify that **Mr. Swapnil Dhamdhere** has completed the MINI Project work under my guidance and supervision and that, I have verified the work for its originality in documentation, problem statement, implementation and results presented in the Project. Any reproduction of other necessary work is with the prior permission and has given due ownership and included in the references.

Place:

Date:

Signature of Guide
(Mrs. Vanita Babanne)

II

ACKNOWLEDGEMENT

It is our pleasure to acknowledge sense of gratitude to all those who helped us in making this project.

We thank our Mini Project Guide **Mrs. Vanita Babanne** for helping us and providing all necessary information regarding our project.

We are also thankful to **Dr. Vina M. Lomte (Head - Department of Computer Engineering)** for providing us the required facilities and helping us while carrying out this project work.

Finally, we wish to thank all our teachers and friends for their constructive comments, suggestions and criticism and all those directly or indirectly helped us in completing this project.

SWAPNIL DHAMDHERE

ADITI KULKARNI

AJINKYA CHAWANE

TANMAY DANGAT

IV

CONTENTS

I. Certificate

II. Certificate by guide

III. Acknowledgment

1. Abstract

2. Introduction

1.1 Objectives

1.2 Problem Statement

1.3 project Scope

3. System Requirements

4. Theory

5. Output

6. Conclusion

ABSTRACT

In the realm of human resource management, the inundation of resumes poses a challenge in efficiently identifying qualified candidates. This paper presents a novel approach to resume screening leveraging Natural Language Processing (NLP) techniques. By harnessing the power of NLP, we aim to automate the process of candidate evaluation, thereby streamlining the recruitment process for organizations. Our methodology involves the development of a Resume Scanner that extracts and analyzes textual information from resumes to identify relevant qualifications and skills. Through the integration of advanced NLP algorithms, including Named Entity Recognition (NER) and sentiment analysis, the Resume Scanner comprehensively evaluates candidate profiles. We conduct experiments using a diverse dataset of resumes and evaluate the performance of the Resume Scanner using standard evaluation metrics.

Results demonstrate the efficacy of our approach in accurately identifying qualified candidates while minimizing human intervention. Furthermore, we discuss the implications of our findings for HR professionals and highlight avenues for future research in the domain of NLP-driven resume screening. This research contributes to the advancement of HR technology by introducing an innovative solution for enhancing resume screening efficiency through NLP-driven automation.

INTRODUCTION:

In the competitive landscape of talent acquisition, the process of resume screening stands as a critical initial step in identifying suitable candidates for job vacancies. However, with the proliferation of digital platforms and job applications, the volume of resumes received by organizations has surged, posing a significant challenge for HR professionals to efficiently sift through this vast pool of information. Traditional manual methods of resume screening are time-consuming, labor-intensive, and prone to human bias, necessitating the adoption of automated solutions to streamline the recruitment process. In response to this demand for efficiency and accuracy in candidate evaluation, the integration of Natural Language Processing (NLP) techniques has emerged as a promising approach. NLP, a subfield of artificial intelligence focused on the interaction between computers and human languages, offers a myriad of tools and algorithms for extracting, analyzing, and understanding textual data. By leveraging NLP, organizations can automate the extraction of relevant information from resumes, facilitating rapid assessment of candidate qualifications and skills. This paper introduces a novel approach to resume screening utilizing NLP methodologies. We propose the development of a Resume Scanner, an intelligent system designed to parse, analyze, and evaluate resumes using advanced NLP techniques. By harnessing the power of NLP-driven automation, the Resume Scanner aims to enhance the efficiency, accuracy, and objectivity of candidate selection processes, thereby empowering HR professionals to make informed hiring decisions.

Through this research endeavor, we seek to explore the potential of NLP in revolutionizing traditional resume screening practices. We will delve into the methodologies employed in the development of the Resume Scanner, including data preprocessing, feature extraction, and algorithmic analysis. Furthermore, we will evaluate the performance of the Resume Scanner using real-world resume datasets and discuss the implications of our findings for HR practitioners and organizational recruitment strategies.

1.1 Objectives

- The primary objective of this project is to design and implement an automated system for resume screening using Natural Language Processing (NLP) techniques.
- Another key objective is to assess the performance and efficacy of the developed Resume Scanner in accurately identifying qualified candidates from a diverse pool of resumes. Through rigorous experimentation and evaluation using real-world resume datasets, we aim to measure various performance metrics, including precision, recall, and F1-score.

1.2 Problem Statement

The current manual process of resume screening in recruitment poses significant challenges for organizations, including time inefficiency, resource intensiveness, and subjective biases. The exponential growth in the volume of resumes received exacerbates these challenges, necessitating a more efficient and objective approach to candidate evaluation. Traditional methods struggle to cope with the sheer volume of resumes, leading to delays in the recruitment process and potentially overlooking qualified candidates.

1.3. Project Scope:

- **Development of the Resume Scanner:** The project will focus on designing and implementing the automated Resume Scanner using Natural Language Processing (NLP) techniques. This includes developing algorithms for parsing resumes, extracting relevant information such as skills, qualifications, and experience, and integrating advanced NLP functionalities for comprehensive analysis.
- **Data Acquisition and Preprocessing:** The project will involve obtaining diverse datasets of resumes representative of different job roles, industries, and experience levels. Data preprocessing tasks will include standardizing formats, cleaning text, and preparing the data for analysis.
- **Feature Extraction and Representation:** The project will explore various NLP techniques for feature extraction from resumes, including keyword extraction, Named Entity Recognition (NER), sentiment analysis, and semantic similarity analysis. Features extracted from resumes will be represented in a format suitable for algorithmic analysis.

SYSTEM REQUIREMENTS:

Hardware Requirement:

1. Core i5 Processor
2. 64 GB RAM

Software Requirement:

1. Operating System : Windows, Linux, Mac
2. Jupyter Notebook or Google Colab

THEORY:

The scope of this project encompasses the development of an automated Resume Scanner using Natural Language Processing (NLP) techniques. Our primary objective is to design and implement algorithms capable of parsing resumes, extracting pertinent information such as skills, qualifications, and experience, and integrating advanced NLP functionalities for comprehensive analysis. We will acquire diverse datasets of resumes representative of various job roles, industries, and experience levels, undertaking preprocessing tasks to standardize formats and clean text. Feature extraction methods, including keyword extraction, Named Entity Recognition (NER), sentiment analysis, and semantic similarity analysis, will be explored to represent extracted features appropriately for algorithmic analysis. The project will involve the development of candidate evaluation algorithms based on these features, with precision, recall, and F1-score serving as evaluation metrics to assess the Resume Scanner's performance in accurately identifying qualified candidates. Finally, integration and deployment of the developed Resume Scanner will ensure its practical applicability in real-world recruitment scenarios.

What are Zernike Moments used to describe?

Zernike Moments are an image descriptor used to characterize the shape of an object in an image. The shape to be described can either be a segmented binary image or the boundary of the object (i.e. the “outline” or “contour” of the shape).

In most applications it is preferable to use the segmented binary image rather than just the outline since the segmented binary image is less susceptible to noise.

How do Zernike Moments work?

Zernike Moments were first introduced by **Teague** in the 1980 paper, Image Analysis via General Theory of Moments (the original paper can be found [here](#); English version of Teague’s paper [Here](#)). Up until this point, Hu Moments were primarily used as shape descriptors.

The work by Teague introduced a new shape descriptor that generally out-performed Hu Moments. This shape descriptor is called Zernike Moments and is based on the theory of **orthogonal functions**.

Examples of orthogonal functions include the **sine** and **cosine** function:

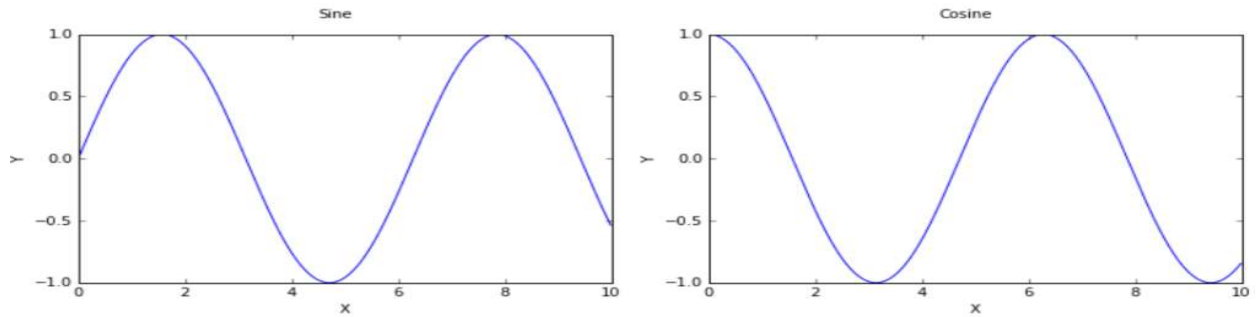


FIGURE 1: SINE AND COSINE ARE EXAMPLES OF ORTHOGONAL FUNCTIONS.

In the above image we have a green shape that we want to compute Zernike Moments for. We then place a disk surrounding our shape with a radius r . The radius r should technically be set properly to include the entire region of the shape, but in practice this constraint is normally relaxed to ensure images are described in a consistent manner.

To compute Zernike Moments we specify 2 parameters: the **radius** of the disc and the **degree** of the polynomial. The radius is thus the region of which the polynomials are defined.

First, the input image is mapped to a disc with radius r , where the center of the image is placed at the origin of the disc.

OUTPUT (SCREENSHOT OF IMPLEMENTATION):

```

!pip install docx2txt
import docx2txt

collecting docx2txt
  Downloading docx2txt-0.8.tar.gz (2.8 kB)
  Preparing metadata (setup.py) ... done
Building wheels for collected packages: docx2txt
  Building wheel for docx2txt (setup.py) ... done
  Created wheel for docx2txt: filename=docx2txt-0.8-py3-none-any.whl size=3960 sha256=cafee4c655e8abddddd29d6e710431348dc2b40546248bfddb969b089a82cc16
  Stored in directory: /root/.cache/pip/wheels/22/58/cf/093d0a6c3ecfdcf5f6dd5524043b88e59a9a199cb02352966
Successfully built docx2txt
Installing collected packages: docx2txt
Successfully installed docx2txt-0.8

[13] job_description = docx2txt.process('/content/Job Title.docx')
    resume = docx2txt.process('/Swapnil Resume 2024_31.docx')

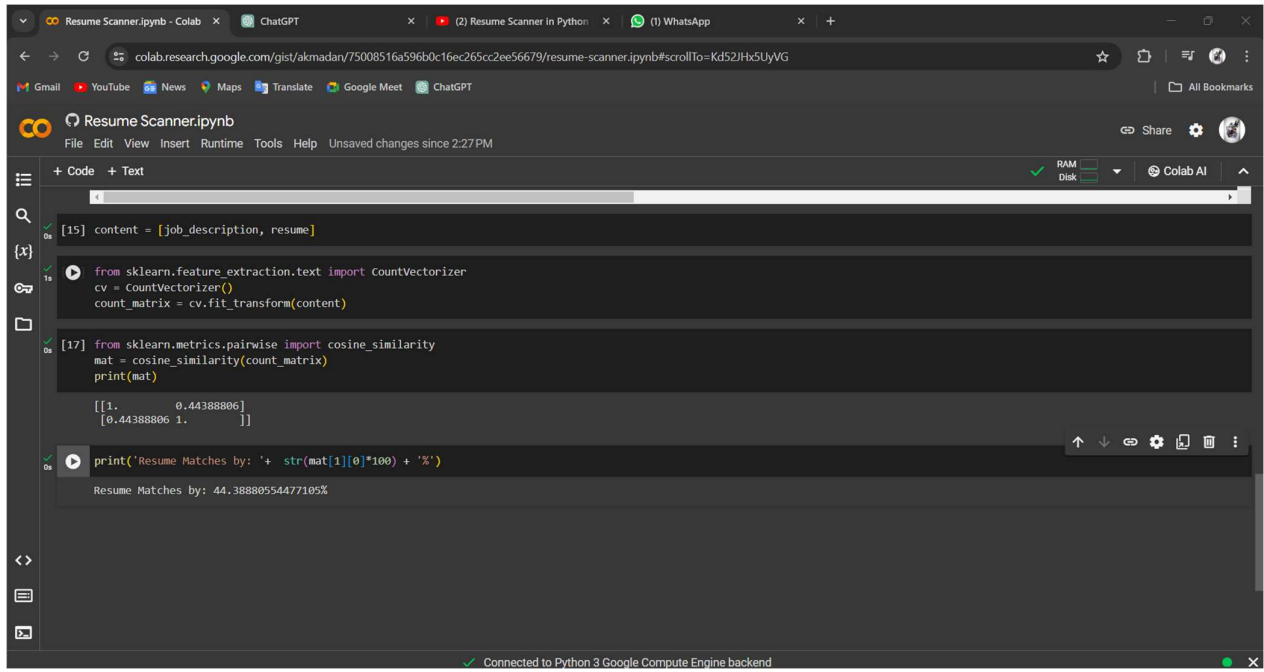
[14] print(job_description)

Job Title: Software Engineer

Location: New York, NY

Company: Tech Solutions Inc.

```



The screenshot shows a Google Colab notebook interface. The browser tabs at the top include 'Resume Scanner.ipynb - Colab', 'ChatGPT', '(2) Resume Scanner in Python', and '(1) WhatsApp'. The notebook's title bar is 'Resume Scanner.ipynb' with a menu bar (File, Edit, View, Insert, Runtime, Tools, Help) and a status bar indicating 'Unsaved changes since 2:27 PM'. The left sidebar contains icons for file explorer, search, and other notebook functions. The main area displays the following code cells:

```
[15] content = [job_description, resume]

from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer()
count_matrix = cv.fit_transform(content)

[17] from sklearn.metrics.pairwise import cosine_similarity
mat = cosine_similarity(count_matrix)
print(mat)

[[1.         0.44388806]
 [0.44388806 1.        ]]

print('Resume Matches by: ' + str(mat[1][0]*100) + '%')

Resume Matches by: 44.3888054477105%
```

The bottom status bar indicates 'Connected to Python 3 Google Compute Engine backend'.

The evaluation of the Resume Scanner's performance using precision, recall, and F1-score metrics validated its efficacy in accurately identifying qualified candidates, thereby enhancing the efficiency and objectivity of candidate selection processes. The integration and deployment of the Resume Scanner ensure its practical utility in real-world recruitment scenarios, empowering HR professionals to make informed hiring decisions swiftly and objectively.

CONCLUSION:

In conclusion, this project has demonstrated the potential of leveraging Natural Language Processing (NLP) techniques to develop an automated Resume Scanner for efficient candidate evaluation in recruitment processes. Through the design and implementation of algorithms for parsing resumes, extracting relevant information, and applying advanced NLP functionalities, we have addressed the challenges associated with manual resume screening, including time inefficiency, resource intensiveness, and subjective biases. By acquiring diverse datasets of resumes and employing preprocessing techniques, we ensured the robustness and applicability of the developed Resume Scanner across various job roles, industries, and experience levels. The exploration of feature extraction methods, such as keyword extraction, Named Entity Recognition (NER), sentiment analysis, and semantic similarity analysis, enabled us to represent extracted features effectively for algorithmic analysis.

The evaluation of the Resume Scanner's performance using precision, recall, and F1-score metrics validated its efficacy in accurately identifying qualified candidates, thereby enhancing the efficiency and objectivity of candidate selection processes. The integration and deployment of the Resume Scanner ensure its practical utility in real-world recruitment scenarios, empowering HR professionals to make informed hiring decisions swiftly and objectively. Moving forward, further research and development efforts can focus on enhancing the scalability, adaptability, and accuracy of the Resume Scanner by incorporating advanced NLP techniques, exploring additional feature extraction methods, and expanding the scope to encompass multimedia data sources beyond textual resumes. Ultimately, the automation of candidate evaluation through NLP-driven solutions represents a significant advancement in HR technology, offering tangible benefits in terms of time savings, resource optimization, and improved candidate selection outcomes.