

RESUME CLASSIFIER

*Synopsis Report submitted in partial fulfillment
of the requirement for the degree of
B. E. (Computer Engineering)*

Submitted By

BHUMI AVHAD
SAPANA SURVASE
PRATIK HALDANKAR
DIPESH BEDMUTHA

Under the Guidance of
Prof. SUJA JAYACHANDRAN

Department of Computer Engineering



(An Autonomous Institute Affiliated to University of Mumbai)

Vidyalankar Institute of Technology

University of Mumbai

2022-23

CERTIFICATE OF APPROVAL

This is to Certify that

BHUMI AVHAD

SAPANA SURVASE

PRATIK HALDANKAR

DIPESH BEDMUTHA

Have successfully carried out Project Synopsis work entitled **RESUME CLASSIFIER** in partial fulfillment of degree course in Computer Engineering as laid down by University of Mumbai during the academic year 2022-23 Under the Guidance of Prof. Suja Jayachandran

Signature of Guide

Head of Department

Examiner 1

Examiner 2

Principal

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of student	Roll No.	Signature
1. Bhumi Avhad	20102B2002	
2. Sapana Survase	20102B2005	
3. Pratik Haldankar	20102A2006	
4. Dipesh Bedmutha	20102A2009	

Date:

Acknowledgements

This Project wouldn't have been possible without the support, assistance, and guidance of a number of people whom we would like to express our gratitude to. First, we would like to convey our gratitude and regards to our mentor Prof Suja Jayachandran for guiding us with her constructive and valuable feedback and for her time and efforts. It was a great privilege to work and study under her guidance.

We would like to extend our heartfelt thanks to our Head of Department, Dr. Sachin Bojewar for overseeing this initiative which will in turn provide every Vidyalankar student a distinctive competitive edge over others.

We appreciate everyone who spared time from their busy schedules and participated in the survey. Lastly, we are extremely grateful to all those who have contributed and shared their useful insights throughout the entire process and helped us acquire the right direction during this research project

Table of Contents

Sr No	Description
1	Introduction
2	Aim and Objectives
3	Literature Survey
4	Problem Statement
5	Scope
6	Proposed System
7	Analysis
	Process Model Used for the Project
	Feasibility Study
	Timeline Chart
8	Design – Data Flow Diagram
9	Hardware and Software Requirements
10	Conclusion and Result
11	References

Abstract

One of the most important and crucial task for any company is to hire an ideal candidate for their job role. Traditional hiring practices are becoming ineffective as online recruitment grows in popularity. The traditional methods normally entail a time-consuming process of manually looking through all of the individuals who have applied, examining their resumes, and then establishing a shortlist of prospects who should be interviewed. Job seeking has grown both wiser and more accessible in our technological age. Companies receive a large number of resumes/CV's, many of which are not well-structured. There has been a great deal of effort put into the job search. The process of picking a candidate based on their resume on the other hand, has not been completely automated.

1.Introduction

Resume classifier is a process that is often used by employers to narrow down the pool of job applicants. The process can be time-consuming and arduous, especially if the employer is looking through a large number of resumes. However, machine learning algorithms can be used to help automate the process. Machine Learning (ML) / NLP could be used to automatically screen resumes for minimum qualifications, such as education level or

years of work experience. Entity recognition, for example, can be used to identify, extract key information from resumes, such as skills, qualifications and work experience (company names). This can be especially helpful when reviewing a large number of resumes, as it can help to quickly identify those that are most likely to be a good fit. This would save time for recruiters, who would otherwise have to manually review each resume. ML models can also be used to score resumes, so that the best ones are given more attention.

2.Aim & Objective

- Create a faster system for classifying resumes and reduce human effort for recruiters to Improve current resumes classification systems, which have limitations.
- Help applicants improve their resumes before applying to jobs and improve current resumes classification systems, which have limitations.
- Reduce human effort for recruiters and help applicants improve their resumes before applying to jobs

·

·

3.Literature Survey

A literature survey is the most important step in any kind of research. Before we start developing need to study the previous papers in the domain in which we are working and on the basis of the study, we can predict or generate the drawback and start working with the reference of previous papers.

1 Jagan Mohan Reddy D et al. {1} suggested joining efficient candidates before resume selection, so that the entire process can be completed in a timely and cost-effective manner. Some characteristics, such as age and salary hike, cannot be used directly for classification due to substantial variations in values that must be transformed into bins.

2. Senthil Kumaran et al. {2} used an intelligent tool for ontology called EXPERT mapping-based candidate screening to create an automated system for intelligent screening of prospects for recruitment, enhancing the precision with which candidates are matched to the requirements of the job.

3. Chirag Daryania et. al {3} proposed an Automated Resume classifier System which used Natural Language Processing and Similarity : Vector Space Model to match each CV with the job description and then suggested an approach which uses a vectorization model and cosine similarity. The calculated ranking scores could then be used to find the most suitable candidates for the job position.

4.Problem Statement

One of the most important and crucial task for any company is to hire an ideal candidate for their job role. Traditional hiring practices are becoming ineffective as online recruitment grows in popularity. The traditional methods normally entail a time-consuming process of manually looking through all of the individuals who have applied, examining their resumes, and then establishing a shortlist of prospects who should be interviewed. Job seeking has grown both wiser and more accessible in our technological age. Companies receive a large number of resumes/CV's, many of which are not well-structured. There has been a great deal of effort put into the job search. The process of picking a candidate based on their resume.

5.Scope

- Investigate more complex neural network architectures and expand dataset with more resume categories and develop models that can give feedbacks to users to improve their resume.
- Develop model that analyzes resume in detail and upscale the existing project to show the job profiles on the company career portals which match the specific keywords in the candidate's resume. List the most relevant profiles first,so that the chances of a candidate getting selected is increased significantly.

6. Proposed System

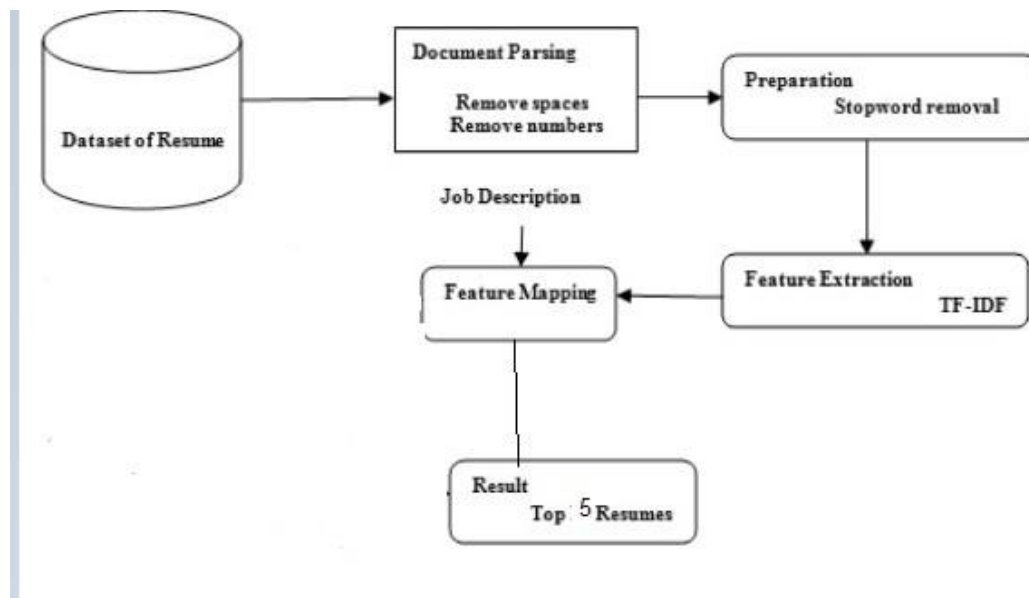


Fig 1.1 - Proposed System

- Classify resumes into categories - java Developer, Business Analyst, Data Scientist, etc.
- Used a dataset consisting of many categories
- Solution included a Machine Learning model for classification
- Display top five categories and confidence percentages
- Deploy trained model on https server
- Test on teammate's resume

• Preprocessing

The resume's provided as input would be shortlisted in this procedure to remove any special or garbage characters from the resumes. NLTK tokenizers are used to break the dataset into tokens. Stop word removal, lemmatization and vectorization are among the preprocessing operations performed on the tokenized dataset.

- Removing Stop Words:

Stop words such as and, the, was, and others appear very often in words and limit the process which determines prediction, thus they are removed.

- Lemmatization:

Lemmatization reduces derived phrases to make entirely sure that the underlying word is accurately associated with the language.

The routine phases of lemmatization are as follows: • Convert the text corpus into a list of words. • Make a corpus concordance, which includes all of the word list entries as they appear in the corpus. • Based on the concordance, link the word-forms to their lemmas.

KNN model - Is used in this model to find the resumes that are closest to the specified job description.

7. Analysis

Process Model Used for the Project:

Agile process model refers to a software development approach based on iterative development. Agile methods break tasks into smaller iterations, or parts do not directly involve long term planning. The project scope and requirements are laid down at the beginning of the development process. Plans regarding the number of iterations, the duration and the scope of each iteration are clearly defined in advance.

Each iteration is considered as a short time "frame" in the Agile process model, which typically lasts from one to four weeks. The division of the entire project into smaller parts helps to minimize the project risk and to reduce the overall project delivery time requirements. Each iteration involves a team working through a full software development life cycle including planning, requirements analysis, design, coding, and testing before a working product is demonstrated to the client.



Fig 1.2 Agile Model

Feasibility Study

A feasibility study is simply an assessment of the practicality of a proposed project plan or method. This is done by analyzing technical, economic, legal, operational and time feasibility factors. In a feasibility study, a proposed plan or project is evaluated for its practicality. As part of a feasibility study, a project or venture is evaluated for its viability in order to determine whether it will be successful.

- Operational Feasibility,
- Technical Feasibility
- Economical Feasibility.

Technical Feasibility:

The system is self-explaining and does not need any entire sophisticated training. A system has been built by concentrating on the graphical user interface concepts, the application can also be handled very easily with a novice user. The overall time that a user needs to get trained is less than 15 minutes. The system has been added with features of menu device and button interaction methods, which makes him the master as he starts working through the environment. As the software that were used as developing this application are very economical and are readily available in the market the only time that is lost by the customer is just installation time.

Operational Feasibility:

It refers to the feasibility of the product to be operational. Some products may work very well at the design and implementation but many fail in the real time environment. It introduces the study of human resources required and their technical expertise.

This product is operationally feasible as it is designed specifically for Apparel. This provides consistent and integrated data management. It also provides information at all levels of people.

Economic Feasibility:

Economic feasibility determines whether there are sufficient benefits in creating to make the cost acceptable, or is the cost of the system too high. So this signifies cost benefit analysis and savings. On the behalf of the cos benefit analysis, the proposed system is feasible and is economical regarding its presumed cost for making a system. We classified the costs of Online Shopping according to the phase in which they occur.

Timeline Chart

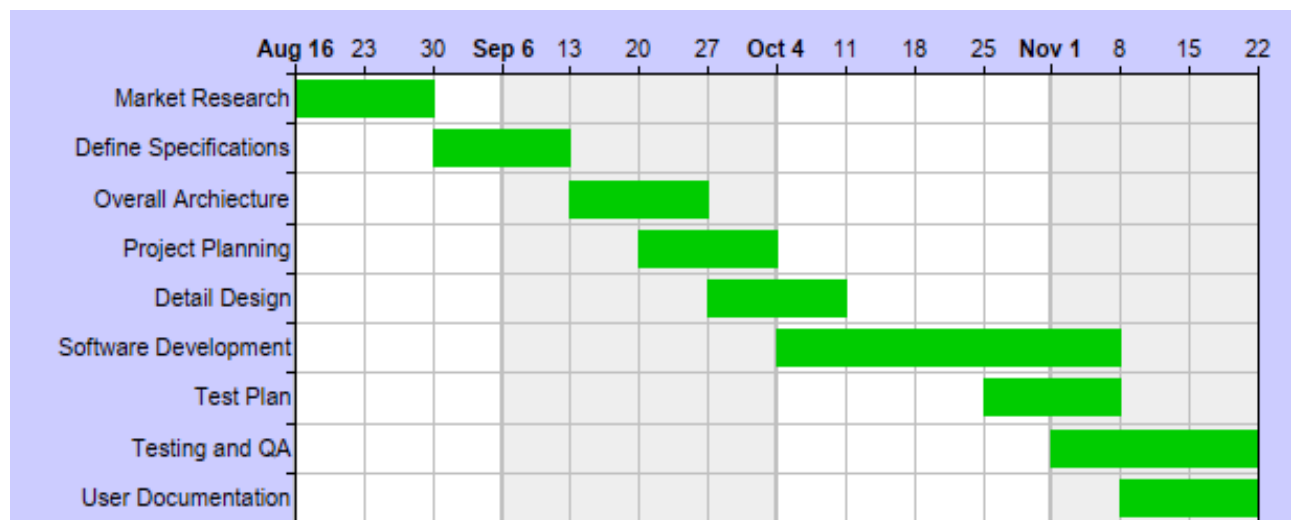


Fig 1.3 Timeline

8. Design

Data Flow Diagram

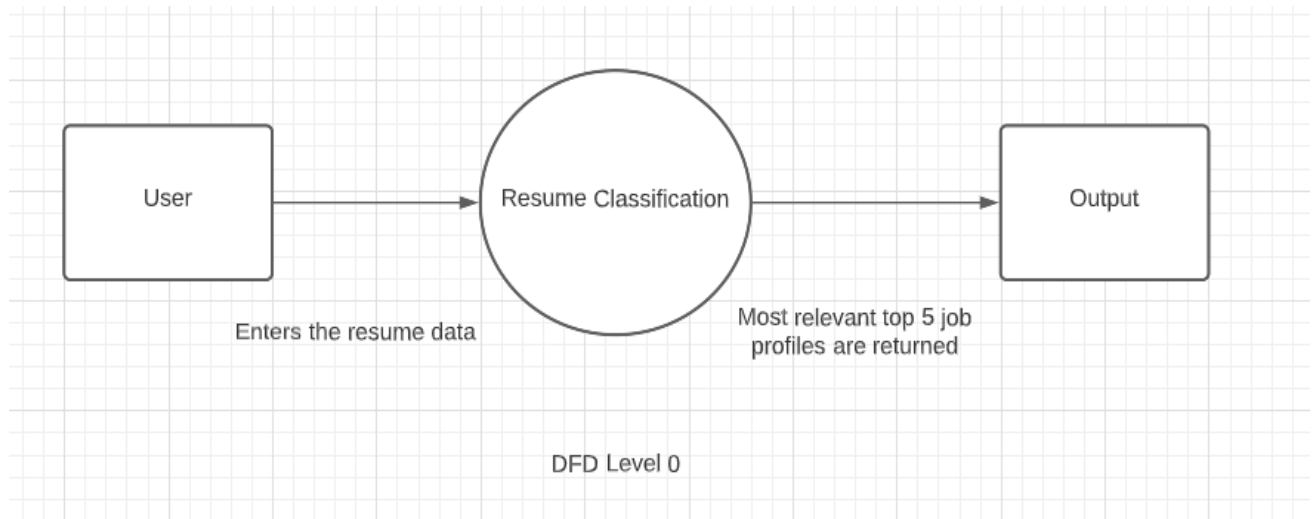


Fig 1.4 DFD_Level-0

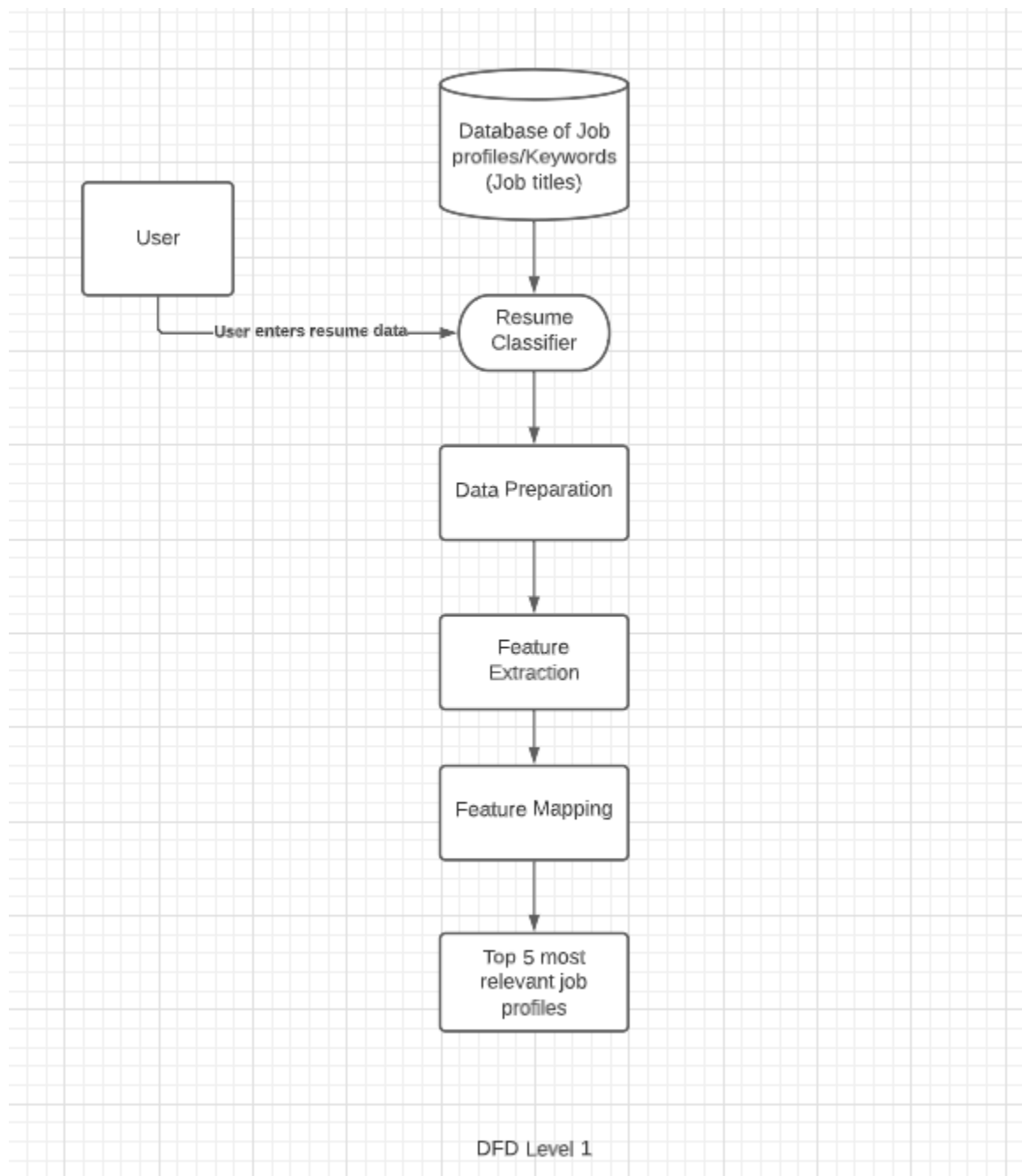


Fig 1.5 DFD_Level-1

9. Hardware & Software Requirements

Hardware Requirements:

Laptop with internet connectivity

Software Requirements:

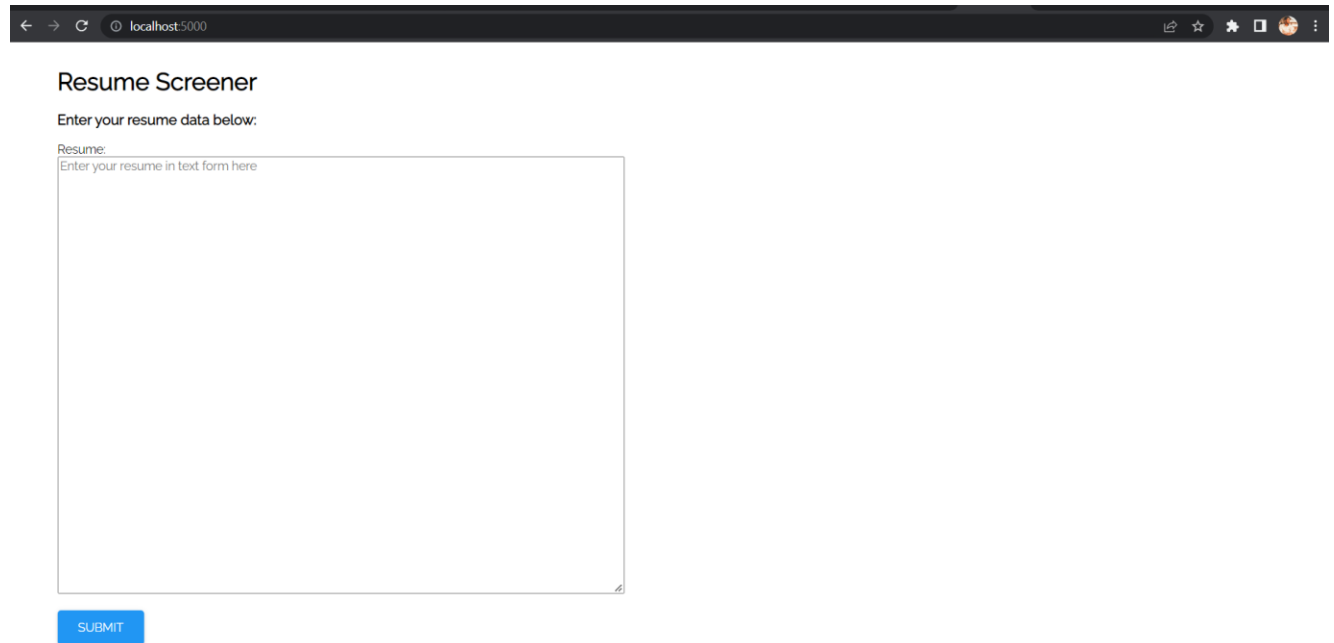
Language implemented - Python

Operating System – Windows 7/8 onwards

Python :[NLTK,matplotlib,numpy,pandas]

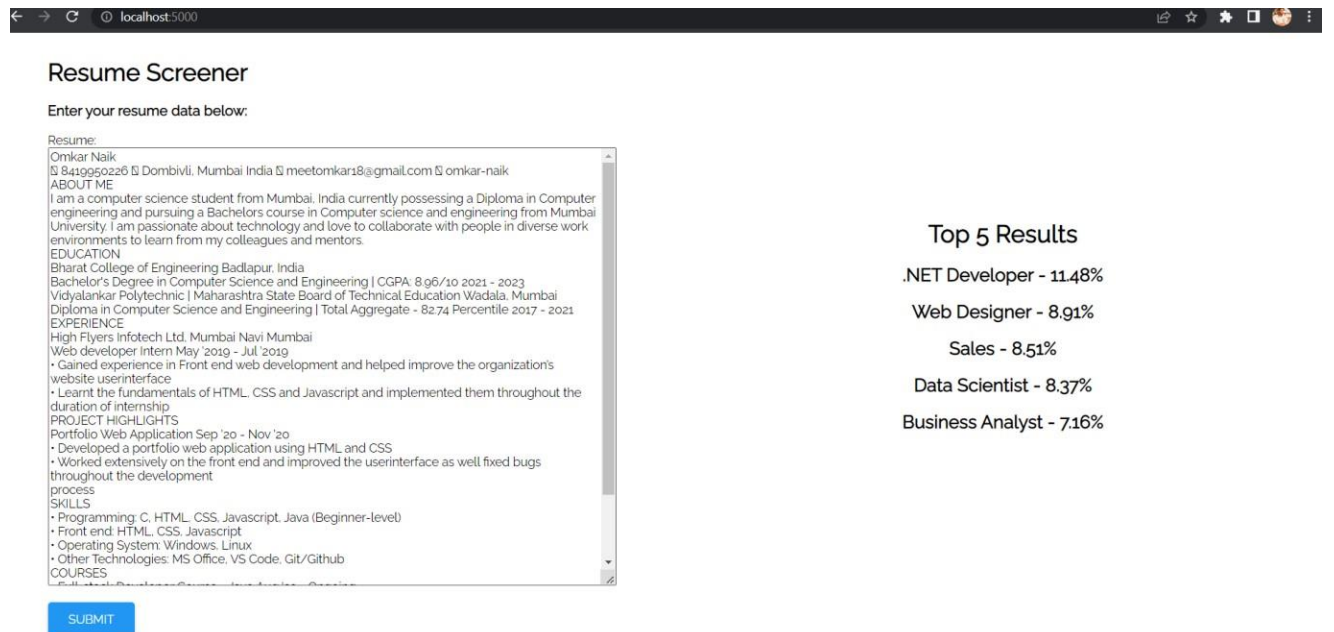
10. Conclusion & Results

Results



The screenshot shows a web browser at localhost:5000 displaying a page titled "Resume Screener". Below the title, it says "Enter your resume data below:". There is a text input area labeled "Resume:" with a placeholder "Enter your resume in text form here". A blue "SUBMIT" button is located below the input area.

Fig 1.6 Resume classifier (Before the resume data is submitted)



The screenshot shows the same "Resume Screener" page, but now with resume data entered. The data includes personal information, education, experience, project highlights, skills, and courses. To the right of the input area, the "Top 5 Results" are displayed:

- Top 5 Results
- .NET Developer - 11.48%
- Web Designer - 8.91%
- Sales - 8.51%
- Data Scientist - 8.37%
- Business Analyst - 7.16%

The blue "SUBMIT" button is still visible at the bottom of the input area.

Fig 1.7 Top 5 most relevant result shown according to the .NET developer resume

localhost:5000

Resume Screener

Enter your resume data below:

Resume:

AKSHAY VAISHNAV
C-101, Dreamland Apartment, Backbone Park, B/H Balaji Hall, Rajkot-360004, Gujarat
+91 720 393 7889 & vaishnavakshay007@gmail.com

EDUCATION

Bachelor of Engineering in Mechanical Engineering August 2012 - May 2016
Sanjaybhai Rajguru College of Engineering, Rajkot
Gujarat Technological University, CGPA: 8.38/10.00

HSC, Class XII June 2010 - April 2012
Shree S.G. Dholakiya Higher Secondary School, Rajkot, Gujarat, 57%

SSC, Class X March 2009 - March 2010
Shree S.G. Dholakiya Secondary School, Rajkot, Gujarat, 83%

SKILLS AND INTERESTS

Interests Product Development, Design, Automobile, CAD/CAE, Finite Element Analysis, Optimization, Fluid Mechanics, Robotics, Modeling and Simulation

Design Software Basic AUTOCAD, CATIA V5, ANSYS (Static Structural, Transient Structural, Static Thermal, Transient Thermal, Harmonic Response, Modal analysis, Acoustic, Fluent), OptimumLap, MATLAB

PROJECTS

Design Optimization of Hydraulic Press Plate using Finite Element Analysis January 2016 - April 2016
Major Project as a part of curriculum
An Industrial Defined Project in collaboration with Incredible Machines, Rajkot.
Designed and performed an FEA analysis of the plates of Hydraulic machine with the capacity of 250-ton
Optimization in terms of design and material reduction, leading to cost effectiveness, considering minimum deformation of plates during operation
Mathematical Modeling and Analysis of a Hydro-pneumatic Suspension
Column of a Car July 2015 - October 2015
Minor Project as a part of curriculum
Modeled a 2-DOF system considering sprung and unsprung mass of the vehicle

SUBMIT

Top 5 Results

Mechanical Engineer	- 15.41%
.NET Developer	- 12.01%
Sales	- 10.54%
Data Scientist	- 7.72%
DevOps Engineer	- 7.67%

Fig 1.8 Top 5 most relevant result shown according to the CV of an mechanical engineer

```
# Print unique categories of resumes
print(data['Category'].value_counts())
```

```
Java Developer      84
Testing             70
DevOps Engineer     55
Python Developer    48
Web Designing       45
HR                  44
Hadoop              42
Blockchain          40
Mechanical Engineer 40
Sales               40
ETL Developer       40
Operations Manager   40
Data Science        40
Arts                36
Database            33
Electrical Engineering 30
Health and fitness  30
PMO                 30
DotNet Developer    28
Business Analyst     28
Automation Testing   26
Network Security Engineer 25
Civil Engineer       24
SAP Developer        24
Advocate            20
Name: Category, dtype: int64
```

Fig 1.9 Category of job title data used for classification

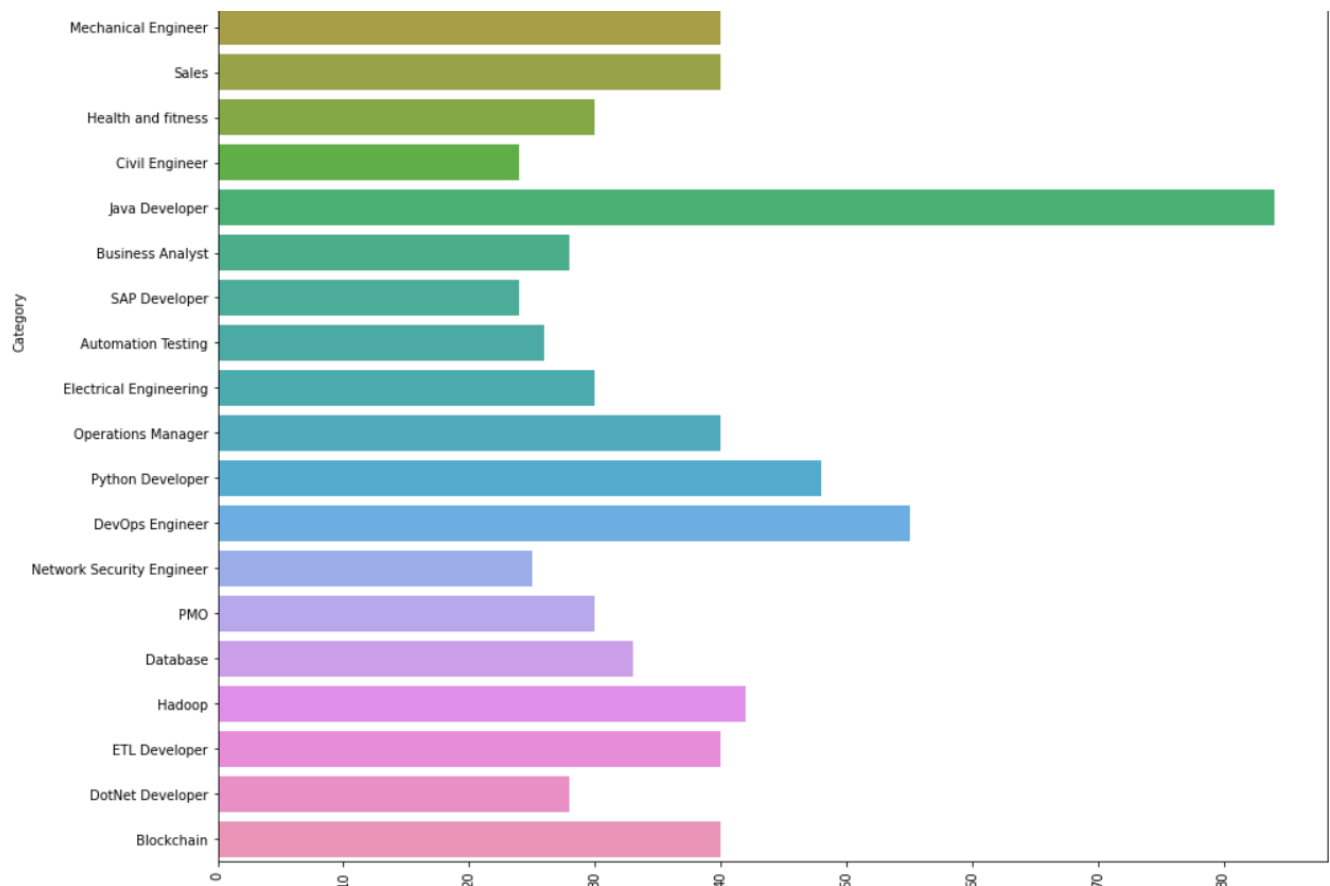


Fig 1.10 Number of job profiles available in the market according to each category of skill

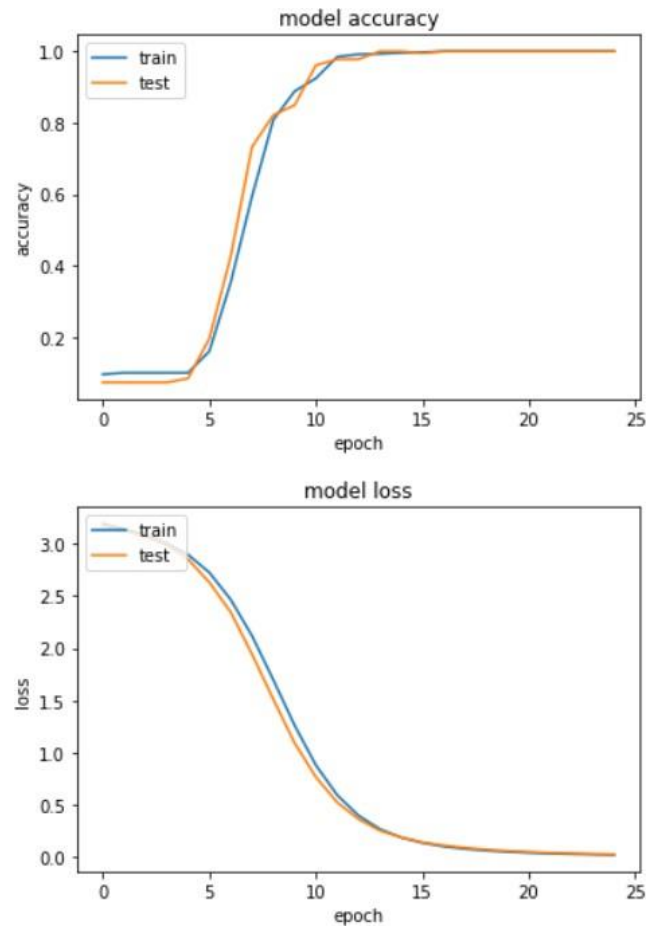


Fig 1.11 Accuracy of the model for the project

Conclusion

In this project, we learned how machine learning and Natural Language Processing can be applied to improve our day-to-day life through the example of Resume classifier. We just classified resumes in few seconds into their respective categories with a great accuracy and learning end to end model training and deployment. We have experienced collaborating with python notebooks Google collab and learned to develop TensorFlow based model.

11. References

- [1] Sankar, A. (2013). "Towards an automated system for intelligent screening of candidates for recruitment using ontology mapping (EXPERT)". *International Journal of Metadata, Semantics and Ontologies*, 8(1), 56. <https://doi.org/10.1504/ijmso.2013.054184>
- [2] Jagan Mohan Reddy D, Sirisha Regella., "Recruitment Prediction using Machine Learning", *IEEE Xplore*, 2020.
- [3] Färber,F., Weitzel, T.,Keim, T., 2003. "An automated recommendation approach to selection in personnel recruitment". *AMCIS 2003 proceedings* , 302.
- [4] Chirag Daryania, Gurneet Singh Chhabrab, Harsh Patel, Indrajeet Kaur Chhabrad, Ruchi Patel., "An Automated Resume classifierSystem using Natural Language Processing and Similarity". (2020). *Topics In Intelligent Computing And Industry Design*.
- [5] Momin Adnan, Gunduka Rakesh, Juneja Afza, Rakesh Narsayya Godavari, Gunduka and Zainul Abideen Mohd Sadiq Naseem., "Resume Ranking using NLP and Machine Learning", (2016b). *Institutional Repository of the Anjuman-I-Islam's Kalsekar Technical Campus*. <https://core.ac.uk/display/55305289>
- [6] Arvind Kumar Sinha, Ashwani Kumar, Md. Amir Khusru Akhtar., "Resume classifier using Natural Language Processing and Machine Learning A Systematic Review", (2019)., *Machine Learning and Information Processing : Proceedings of ICMLIP*.
- [7] V. V. Dixit , Trisha Patel , Nidhi Deshpande , Kamini Sonawane, "Resume Sorting using Artificial Intelligence". (2019). *International Journal of Research in Engineering, Science and Management Volume-2, Issue-4*.
- [8] Dr.K.Satheesh, A.Jahnavi, L Aishwarya, K.Ayesha, G Bhanu Shekhar, K.Hanisha, "Resume Ranking based on Job Description using SpaCy NER model". (2020). *International Research Journal of Engineering and Technology*.
- [9] Breaugh, J.A., 2009. The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review* 19, 219–231.
- [10] Zhang, L.,Fei, W. ,Wang ,L.,2015.Pj matching model of knowledge workers. *Procedia Computer Science* 60,1128–1137