Solution Document for Resume Screening

Author:
Jeyamaruthi Jayakumar
ID: 1001757737
Department of Computer Science & Engineering
The University of Texas, Arlington
Portfolio: https://jeyamaruthi.uta.cloud/blog/
Email ID: jeyamaruthi1995@gmail.com

Contents

1 Introduction

The objective of our project is to create a Resume Screening utilizing Natural Language Processing (NLP) and Machine Learning (ML).

Resume Screening is an act of determining whether a candidate is qualified to hire for the appropriate Job Description (JD). We accomplish that by analyzing the candidate's Education, Experience, Projects & any other information which can be extracted from the Resume.

Benefits of Resume Screening:

Generally, the hiring process commences with a Resume. For any job opportunity, there will be N number of Resumes which should be skimmed by an HR or Talent Acquisition Team, where N can vary from 1 to 1000s.

Hence, looking into 1000 Resumes is neither an easy task nor any person will be willing to do it in the 21st century. To overcome this situation, we will be utilizing NLP to screen the N number of Resumes and form a Ranking system or Plot to make human life much easier while hiring an appropriate candidate.

2 Skeleton & Technologies

2.1 Skeleton of the Design:

- Dataset Creation
- A front End
- A Back End

2.2 Technologies:

- **Front end:**
  HTML, CSS, Angular & WordPress.
- **Back end:** Jupyter Notebook – Python Version 3.7.4

  Libraries for back end:

Library Name	Usage
spaCy	To create a NLP model (*spaCy.blank*)
Pandas	To read the files
Collections	Utilizing counter from collections
PyPDF2	To read the PDF file
Pickle	To load the data to NLP model

3 Detailed Design

3.1 Dataset Creation:

- We will be collecting N number of Resumes to train the model by converting the resume into an entity dataset, which will be done manually.
  Something like:

('Govardhana K Senior Software Engineer Bengaluru, Karnataka, Karnataka - Email me on Indeed: indeed.com/r/Govardhana-K/ b2de315d95905b68 Total IT experience 5 Years 6 Months Cloud Lending Solutions INC 4 Month • Salesforce Developer Oracle 5 Years 2 Month • Core Java Developer Languages Core Java, Go Lang Oracle PL-SQL programming, Sales Force Developer with APEX. Designations & Promotions Willing to relocate: Anywhere WORK EXPERIENCE Senior Software Engineer Cloud Lending Solutions - Bangalore, Karnataka - January 2018 to Present Present Senior Consultant Oracle - Bangalore, Karnataka - November 2016 to December 2017 Staff Consultant Oracle - Bangalore, Karnataka - January 2014 to October 2016 Associate Consultant Oracle - Bangalore, Karnataka - November 2012 to December 2013 EDUCATION B.E in Computer Science Engineering Adithya Institute of Technology - Tamil Nadu September 2008 to June 2012 https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN SKILLS APEX. (Less than 1 year), Data Structures (3 years), FLEXCUBE (5 years), Oracle (5 years), Algorithms (3 years) LINKS https://www.linkedin.com/in/govardhana-k-61024944/ ADDITIONAL INFORMATION Technical Proficiency: Languages: Core Java, Go Lang, Data Structures & Algorithms, Oracle PL-SQL programming, Sales Force with APEX. Tools: RADTool, Jdeveloper, NetBeans, Eclipse, SQL developer, PL/SQL Developer, WinSCP, Putty Web Technologies: JavaScript, XML, HTML, Webservice Operating Systems: Linux, Windows Version control system SVN & Git-Hub Databases: Oracle Middleware: Web logic, OC4J Product FLEXCUBE: Oracle FLEXCUBE Versions 10.x, 11.x and 12.x https://www.linkedin.com/in/govardhana-k-61024944/', {'entities': [(1749, 1755, 'Companies worked at'), (1696, 1702, 'Companies worked at'), (1417, 1423, 'Companies worked at'), (1356, 1793, 'Skills'), (1209, 1215, 'Companies worked at'), (1136, 1248, 'Skills'), (928, 932, 'Graduation Year'), (858, 889, 'College Name'), (821, 856, 'Degree'), (787, 791, 'Graduation Year'), (744, 750, 'Companies worked at'), (722, 742, 'Designation'), (658, 664, 'Companies worked at'), (640, 656, 'Designation'), (574, 580, 'Companies worked at'), (555, 573, 'Designation'), (470, 493, 'Companies worked at'), (444, 469, 'Designation'), (308, 314, 'Companies worked at'), (234, 240, 'Companies worked at'), (175, 198, 'Companies worked at'), (93, 137, 'Email Address'), (39, 48, 'Location'), (13, 38, 'Designation'), (0, 12, 'Name')]})

- Once the dataset is manufactured, we can start loading into the model for training.

3.2 Front End/A general CSV:

We can implement either of the following:

- A single page that has a TextArea, a Browse option, and a button.

  TextArea: For adding the Job Description.
  Browse Option: To upload the Resumes.
  Button: To analyse and Rank the Resumes.

  (or)

- An old school CSV, which consists of all the necessary keywords for Job Description.
  Something like:

Machine Learning	Python Language	NLP	Front End Developer	Data Analysis
Linear Regression	NLP	NLP	Web Development	Analysing
K Means	Pandas	spaCy	HTML	Matplotlib
KNN algorithms	Numpy	NLTK	CSS	Visualizing
Random Forest	sklearn	bag of words	PHP	Plot
SVM	matplotlib	BERT	JavaScript	Graph
Decision Trees	flask	Sentiment Analysis	Wordpress	Pandas
Logistic Regression	django	Chat Bot	Laravel	Numpy
	tinker		Angular	Excel
			React	Vlookup
				SQL
				Critical Thinking
Jeyamaruthi Jayakumar				

3.3 Back End:

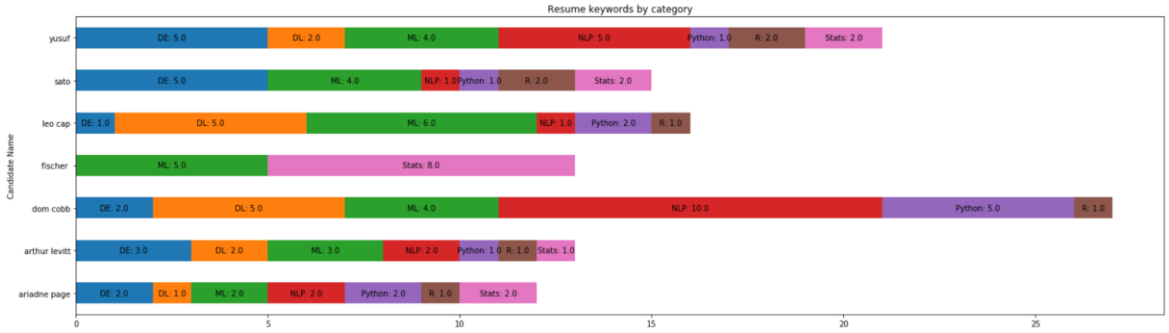
- Firstly, we will be utilizing the converted dataset for training the NLP model with the help of a library called spaCy.
- Then we will make use of Named Entity Recognition (NER) to observe the candidate's qualification. For example, from a given PDF the model will understand whether the statement talks about skills, education, or an experience.
- By doing this the model trains itself with the provided dataset and can be utilized to fetch the entity from any Resume.
- Reading the Resume i.e. for transforming the PDF into a text for the NLP model, we can utilize either PyPDF2 library or PyMuPDF library (We do not use PDFminer as it reads the document from left to right. Therefore, using it is not efficient when Resume has column separated data). The result will be like:

```
In [60]: import sys,fitz
         fname = 'Resume_Jeyamaruthi_Jayakumar.pdf'
         doc = fitz.open(fname)
         text = ""
         for page in doc:
             text = text + str(page.getText())

         tx = " ".join(text.split('\n'))
         print(tx)
```

```
JEYAMARUTHI JAYAKUMAR +1 (817)-368-9606| jeyamaruthi.jayakumar@mavs.uta.edu |https://github.com/jeyamaruthi/ |https://www.li
nkedin.com/in/jeyamaruthi-jayakumar-57864a172/ | https://jeyamaruthi.uta.cloud/blog/ PROFESSIONAL SUMMARY An energetic and de
dicated candidate offering 3 years of professional experience as a Data Analyst, who demonstrates a strong work ethic and crea
tive ability. Well-versed in development and testing for a reputed organization. SKILLS • General: Adequate Communication sk
ills, Leadership, Coordinating team projects, Interpersonal and client communications, analytical & organizational skills, Co
nceptualizing development projects. • Languages: C, C++, Python, JAVA, PHP, SAP ABAP. • Data Science: Probability & Statistic
s, Linear Algebra, R-Language, Traditional & Big Data, BI & ML, Tableau & Data Science Algorithms. • Scripting languages: HTM
L, CSS, JavaScript, Laravel. • Database: MySQL, Spatial Database, QGIS, PostGIS.. • Documentation & Modeling: MS Office, Visi
o, PowerPoint • Testing: Black box & White Box Testing (Junit, JaCoCo -Coverage). WORK HISTORY SENIOR SYSTEMS ENGINEER | 01/
2018 to 06/2019|Infosys - Chennai, Tamil Nadu • Worked closely with customers, internal staff and other stakeholders to determ
ine the planning, implementation, and integration of system-oriented projects. • Collected, Cleansed and provided modeling an
d analyses of structured and unstructured data used for major business initiatives. • Analyzed & Visualized the dataset utili
zing Excel functions such as VLOOKUP, Find/Search, Countifs etc. and Python libraries such as Numpy, Pandas, SciPy, Matplotli
b, keras etc. SYSTEMS ENGINEER | 08/2016 to 12/2017 |Infosys - Mysore, Karnataka • Handled scripting tasks for debugging and
automation using Python & SAP ABAP. • Automated tasks to reduce labor costs and increase productivity. • Re cognized the crit
ical elements of application problems. Developed and evaluated data, determined solutions & made logical recommendations. • C
onfigured, installed and upgraded databases according to the business needs. Intern|| 02/2016 to 06/2016 | Infosys - Mysore, K
arnataka • Developed an administrated file-sharing system with GUI – Backend Lang: Python • Successfully completed training
in Python & Java in Infosys Limited. PERSONAL PROJECTS | 11/2019 to Current Arlington, Texas • Projects: Portfolio, Leagu
e of Legends Tournament Database with GUI, Hypothesis Testing Project, Stock Market Prediction (using LSTM), Yahoo Stock Marke
t Visualization, Amazon Web Scrapping, Spatial Database - DFW Waze Database with QGIS, PostGIS. • Above project Description,
Document & Source Code is available in my GitHub: https://github.com/jeyamaruthi EDUCATION The University of Texas At Arling
ton - Arlington, TX | Master of Science Computer Science, Expected in 05/2021 • Majoring in Computer Science with GPA: 3.67.
```

- Meanwhile, we will either generate a Comma Separated Delimiter or we adapt a Database that fetches Job Description dynamically with the help of the front end.
- After classifying the entities, we can use the former to check with the given Job Description.
- Now, we will be left out with how many Keywords or entities are mapped with the above-collected data.
- Utilizing the data accumulated we can either **plot** a Bar Graph something like:



(or)

We can **rank** the Resume with respect to the mapped data like:

Ranking of Resumes				
S.No	**Name**	**Role**	**Email Address**	**Ph No**
1	Jeyamaruthi Jayakumar	Data Scientist	jeyamaruthi1995@gmail.com	+1 817-368-9606
2	Narmathe Sri	Technical Analyst	Narmatha@gmail.com	+91 8936786928
3	Fiaz Ahamed	Data Analyst	Fiaz94@gmail.com	+91 8145639872
4	Anantha Padmanaban	Full Stack Developer	Anantha95@gmail.com	+91 8756315297
5	Sabariish Mani	Senior Consultant	Sab.Mani1@gmail.com	+1 144-568-8971
6	Suresh Kumar	Technical Analyst	Suresh.k@gmail.com	+91 9768463218

- By doing so, it will make life easier for an HR or a Talent Acquisition team to continue for the hiring process.

4 Annotations

Label	**Description**
NLP	Natural Processing Language
ML	Machine Learning
NER	Named Entity Recognition
CSV	Comma Separated Values
PKL	Pickle file