

An illustration of a resume analysis process. A hand holds a pen over a document labeled 'selected'. Another hand points to a document with a person's photo and 'CV' label. A banner across the center reads 'REZED: RESUME ANALYSIS USING NLP'. The background includes a smartphone, a pencil, a pie chart, and various papers.

REZED: RESUME ANALYSIS USING NLP



TEAM MEMBERS:

- AADHYA SHARMA
- BRINDA
DASGUPTA
- GLENNIS D'SOUZA
- JAYASREE C

123, Street
City Name
12345



PURPOSE

Resume Parsing can be very tricky as a recruiter needs to process multiple number of them. Recruiters face many dilemmas when it comes to hiring candidates as it is quite ambiguous to understand how much of exposure and practical knowledge a candidate possesses concerning a given skill. It is particularly important in understanding this as it might help a recruiter get a fair idea about the suitability of the candidate for a given project.

APPROACH

1. A dictionary or table which has various skill sets categorised is maintained. If we have words like keras, tensorflow, CNN, RNN appearing in the candidate CV, then they are merged under one column titled 'Deep Learning'.
2. We have used NLP algorithm that parses the whole resume and searches for the words mentioned in the dictionary or table.

3. The next step is to count the occurrence of the words under various category for each candidate.

4. We represent the above information in a visualized manner so that it becomes easier for the recruiter to choose the candidate. At the same time a csv file is also created which gives a score card of the different skills acquired by each candidate.

DATASET

We have considered few sample resumes for our project. All the resumes pertain to the field of Data Science and Analytics.

Jane Smith

Address Line 1 Address Line 2 City, State Zip | (123) 456-7890 | jane.smith@gmail.com

CAREER OBJECTIVE

Administrative Assistant with 6+ years of experience working directly for the President of JM Inc., a Fortune 500 company. Possesses impeccable written and verbal communication skills and excellent interpersonal skills.

CORE COMPETENCIES

- Customer Service
- Cost Efficient
- Detailed and Organized
- Supplier Relationship

PROFESSIONAL EXPERIENCE

JM INC., New York, NY

- Administrative Assistant, Apr 2008 – present
- Read and analyze incoming memos, submissions, and reports to determine their significance and plan their distribution.
- Conduct research, compile data, and prepare papers for consideration and presentation by executives, committees and boards of directors.
- Coordinate and direct office services, such as records, departmental finances, budget preparation, personnel issues, and bookkeeping, to all executives.
- Prepare invoices, reports, memos, letters, financial statements and other documents, using word processing, spreadsheet, database, or presentation software.

FLORIDA DEPARTMENT OF SOCIAL SERVICES, Orlando, FL

Rehabilitation Counselor, Aug 2004 – May 2006

- Confer with clients to discuss their options and goals so that rehabilitation programs and plans for accessing needed services can be developed.
- Prepare and maintain records and case files, including documentation such as clients' personal and eligibility information, services provided, narratives of client contacts, and relevant correspondence.
- Develop and maintain relationships with community referral sources, such as schools and community groups.
- Develop and maintain relationships with community referral sources, such as schools and community groups.
- Analyze information from interviews, educational and medical records, consultation with other professionals, and diagnostic evaluations to assess clients' abilities, needs, and eligibility for services.

EDUCATION

FLORIDA STATE UNIVERSITY, Orlando, FL

Bachelor of Art in English, May 2004

- GPA: 3.349
- Published in school's newspaper editorial
- Summer Internship for the New York Times

ADDITIONAL SKILLS

- Proficient in Microsoft Office and Adobe Illustrator CS5
- Bilingual Spanish and English
- Employee of the Month for 2 consecutive months in H&M
- Won the "Winner's Digest" 2002 Award
- Awarded an employee travel award due to "Performance Excellence" 2 years in a row through JM Inc.

JANE SMITH

Address Line 1 Address Line 2 City, State Zip | (123) 786-1011 | jane.smith@gmail.com

CAREER OBJECTIVE

Administrative Assistant with 6+ years of experience working directly for the President of JM Inc., a Fortune 500 company. Possesses impeccable written and verbal communication skills and excellent interpersonal skills.

CORE COMPETENCIES

- Customer Service
- Cost Efficient
- Detailed and Organized
- Supplier Relationship

PROFESSIONAL EXPERIENCE

JM INC.

ADMINISTRATIVE ASSISTANT 1, NEW YORK, NY | Apr 2008 – Present

- Read and analyze incoming memos, submissions, and reports to determine their significance and plan their distribution.
- Conduct research, compile data, and prepare papers for consideration and presentation by executives, committees and boards of directors.
- Coordinate and direct office services, such as records, departmental finances, budget preparation, personnel issues, and bookkeeping, to all executives.
- Prepare invoices, reports, memos, letters, financial statements and other documents, using word processing, spreadsheet, database, or presentation software.

FLORIDA DEPARTMENT OF SOCIAL SERVICES

REHABILITATION COUNSELOR 1, ORLANDO, FL | Aug 2004 – May 2006

- Confer with clients to discuss their options and goals so that rehabilitation programs and plans for accessing needed services can be developed.
- Prepare and maintain records and case files, including documentation such as clients' personal and eligibility information, services provided, narratives of client contacts, and relevant correspondence.
- Develop and maintain relationships with community referral sources, such as schools and community groups.
- Analyze information from interviews, educational and medical records, consultation with other professionals, and diagnostic evaluations to assess clients' abilities, needs, and eligibility for services.

EDUCATION

FLORIDA STATE UNIVERSITY, ORLANDO, FL

BACHELOR OF ART IN ENGLISH, May 2004

- GPA: 3.3740
- Published in school's newspaper editorial
- Summer Internship for the New York Times

ADDITIONAL SKILLS

- Proficient in Microsoft Office and Adobe Illustrator CS5
- Bilingual Spanish and English
- Certified CPR and First Aid

AWARDS AND HONORS

- Employee of the Month for 2 consecutive months in H&M
- Won the "Winner's Digest" 2002 Award
- Awarded an employee travel award due to "Performance Excellence" 2 years in a row through JM Inc.

Jane Smith

Address Line 1 Address Line 2 City, State Zip | (123) 234-5678 | jane.smith@gmail.com

CAREER OBJECTIVE

Administrative Assistant with 6+ years of experience working directly for the President of JM Inc., a Fortune 500 company. Possesses impeccable written and verbal communication skills and excellent interpersonal skills.

CORE COMPETENCIES

- Customer Service
- Cost Efficient
- Detailed and Organized
- Supplier Relationship

PROFESSIONAL EXPERIENCE

JM Inc., New York, NY

- Administrative Assistant, Apr 2008 – present
- Read and analyze incoming memos, submissions, and reports to determine their significance and plan their distribution.
- Conduct Research, compile data, and prepare papers for consideration and presentation by executives, committees and boards of directors.
- Coordinate and direct office services, such as records, departmental finances, budget preparation, personnel issues, and bookkeeping, to all executives.
- Prepare invoices, reports, memos, letters, financial statements and other documents, using word processing, spreadsheet, database, or presentation software.

Holida Department of Social Services, Orlando, FL

Rehabilitation Counselor, Aug 2004 – May 2006

- Confer with clients to discuss their options and goals so that rehabilitation programs and plans for accessing needed services can be developed.
- Prepare and maintain records and case files, including documentation such as clients' personal and eligibility information, services provided, narratives of client contacts, and relevant correspondence.
- Develop and maintain relationships with community referral sources, such as schools and community groups.
- Analyze information from interviews, educational and medical records, consultation with other professionals, and diagnostic evaluations to assess clients' abilities, needs, and eligibility for services.

EDUCATION

Holida State University, Orlando, FL

Bachelor of Art in English, May 2002

- GPA: 3.3740
- Published in school's newspaper editorial
- Summer Internship for the New York Times

ADDITIONAL SKILLS

- Proficient in Microsoft Office and Adobe Illustrator CS5
- Bilingual in Spanish and English
- Certified CPR and First Aid
- Employee of the Month for 2 consecutive months in H&M
- Won the "Winner's Digest" 2002 Award
- Awarded an employee travel award due to "Performance Excellence" 2 years in a row through JM Inc.

PACKAGES/MODULES USED

Spacy: 'Phrase Matcher' From spaCy we import PhraseMatcher. Create the rule-based PhraseMatcher . Setting a different attr to match on will change the token attributes that will be compared to determine a match.

PyPDF2 PyPDF2 is a pure-python PDF library capable of splitting, merging together, cropping, and transforming the pages of PDF files

Matplotlib, for visualization

Pandas: Counter Python Counter is a container that will hold the count of each of the elements present in the container.

Streamlit an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science.

EXECUTION

```
1 #function that does phrase matching and builds a candidate profile
2 def create_profile(file):
3     text = pdfextract(file)
4     text = str(text)
5     text = text.replace("\n", "")
6     text = text.lower()
7     #below is the csv where we have all the keywords.
8     keyword_dict = pd.read_csv('C:/Users/User/OneDrive/Documents/words_csv.csv', encoding='latin-1')
9     stats_words = [nlp(text) for text in keyword_dict['statistics'].dropna(axis = 0)]
10    NLP_words = [nlp(text) for text in keyword_dict['NLP'].dropna(axis = 0)]
11    ML_words = [nlp(text) for text in keyword_dict['Machine Learning'].dropna(axis = 0)]
12    DL_words = [nlp(text) for text in keyword_dict['Deep Learning'].dropna(axis = 0)]
13    R_words = [nlp(text) for text in keyword_dict['R Language'].dropna(axis = 0)]
14    python_words = [nlp(text) for text in keyword_dict['Python language'].dropna(axis = 0)]
15    Data_Engineering_words = [nlp(text) for text in keyword_dict['Data Engineering'].dropna(axis = 0)]
16    Other_Langs = [nlp(text) for text in keyword_dict['Other Langs'].dropna(axis = 0)]
17
18    matcher = PhraseMatcher(nlp.vocab)
19    matcher.add('Stats', None, *stats_words)
20    matcher.add('NLP', None, *NLP_words)
21    matcher.add('ML', None, *ML_words)
22    matcher.add('DL', None, *DL_words)
23    matcher.add('R', None, *R_words)
24    matcher.add('Python', None, *python_words)
25    matcher.add('DE', None, *Data_Engineering_words)
26    matcher.add('Others', None, *Other_Langs)
27    doc = nlp(text)
28
29    d = []
30    matches = matcher(doc)
31    for match_id, start, end in matches:
32        rule_id = nlp.vocab.strings[match_id] # get the unicode ID, i.e. 'COLOR'
33        span = doc[start : end] # get the matched slice of the doc
34        d.append((rule_id, span.text))
35    keywords = "\n".join(f'{{i[0]}} {{i[1]}} ({{j}})' for i,j in Counter(d).items())
36
37    # converting string of keywords to dataframe
38    df = pd.read_csv(StringIO(keywords), names = ['Keywords_List'])
39    df1 = pd.DataFrame(df.Keywords_List.str.split(' ', 1).tolist(), columns = ['Subject', 'Keyword'])
40    df2 = pd.DataFrame(df1.Keyword.str.split(' ', 1).tolist(), columns = ['Keyword', 'Count'])
41    df3 = pd.concat([df1['Subject'], df2['Keyword'], df2['Count']], axis = 1)
42    df3['Count'] = df3['Count'].apply(lambda x: x.rstrip(""))
43
44    base = os.path.basename(file)
```

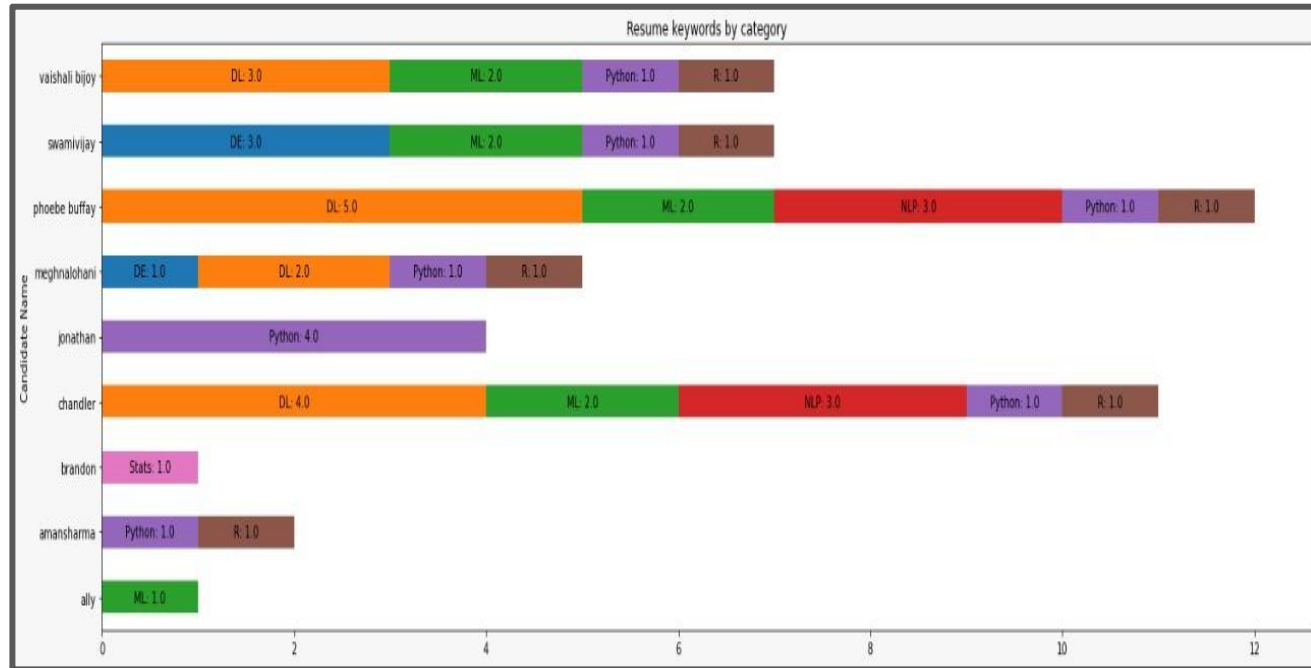
We use the a Word2Vec model.

- Where in the entire resume is parsed,
- words are converted to vectors,
- keywords (from the table) are identified and matched.
- And finally the output is given in the form of a graph.

CONCLUSION

- **Candidate names refers to the list of the candidates whose cv has been selected for the given job profile.**
- **DE,DL,ML etc refer to the different skill set that the recruiter is looking for a given job profile.**
- **Given aside is a snapshot of the scorecard where the numbers represent the individual scores of each candidate in a given skill.**

	Candidate Name	DE	DL	ML	NLP	Python	R	Stats
0	ally	0.0	0.0	1.0	0.0	0.0	0.0	0.0
1	amansharma	0.0	0.0	0.0	0.0	1.0	1.0	0.0
2	brandon	0.0	0.0	0.0	0.0	0.0	0.0	1.0
3	chandler	0.0	4.0	2.0	3.0	1.0	1.0	0.0
4	jonathan	0.0	0.0	0.0	0.0	4.0	0.0	0.0
5	meghnalohani	1.0	2.0	0.0	0.0	1.0	1.0	0.0
6	phoebe buffay	0.0	5.0	2.0	3.0	1.0	1.0	0.0
7	swamivijay	3.0	0.0	2.0	0.0	1.0	1.0	0.0
8	vaishali bijoy	0.0	3.0	2.0	0.0	1.0	1.0	0.0



The bar plot shown above gives the visualisation of the skillset score card.

A spiral-bound notebook with a light blue cover is shown against a solid light blue background. The notebook is open to a blank white page. A thin blue rectangular border is drawn on the page, framing the text. The words "THANK YOU" are printed in a bold, blue, sans-serif font, centered horizontally and vertically within the page.

THANK YOU