CSCE 5290: Natural Language Processing

Group 3 Project Increment 1

Job Description Based Resume Matcher Using Text Summarization, Keyword Extraction

1. Project Title and Team Members

Job Description Based Resume Matcher Using Text Summarization, Keyword Extraction

Suhas Siddarajgari Tellatakula	11626111
Vamshi Telukuntla	11618397
Srikanth Mamillapalli	11551359
Sairohithvarma Kantem	11606743

GitHub link: https://github.com/Vams98/NLP-project#nlp-project

2. Goals and Objectives:

- ### Motivation:

  Now-a-days, candidates applying to jobs online by uploading resumes is increasing rapidly. But unfortunately, only few of them manage to secure dream jobs in respective dream companies/firms. This is because most of the candidates' resumes will not get picked in the first place. Students frequently apply to numerous jobs with single resume irrespective of job description. There is need to be aware of creating job description tailored resumes. So, we are trying to build a solution which can analyse and understand the job descriptions by generating the sentiments and a short summary for every job description. Also, it generates the top keywords from the resumes dataset.

▪ Significance:

As companies receive plethora of applications for one job opening, it will be very difficult for the companies to go through each resume and filter the candidates. Hence, they employ artificial intelligencebased resume pickers which look for specific keywords in the resumes. Creating job description based tailored resumes is one of the crucial steps in applying for a job. A recent survey indicates that 34% of people do not know how to prepare unique resumes based on job description. Hence, this tool can be extremely helpful for candidates to check their resume relevancy to the applied job and increase their chances of landing dream job.

▪ Objectives:

With this project, we are aiming to solve above problem by taking advantage of the Natural Language Processing methodologies employing text summarization, label classification and keyword extraction techniques. The objectives of the project are as follows:

- ➢ To summarize multiple paragraph job/role descriptions into a simple yet fully understandable output by executing and finalizing best out of different text summarization techniques like pyTextRank, Spacy pipeline.
- ➢ To classify the job description based on the tone using lexicon based semantic analyzer.
- ➢ To classify the jobs/roles under few tags/groups based on the above generated short summary employing different concept/tags classification mechanisms.
- ➢ To identify and filter out the important keywords from a document by trying out and choosing best one out of different keyword extraction methods using TfIdfVectorizer.
- ➢ To provide the relevancy or matching score between job descriptions and resumes based on a machine learning model built on the generated summary and keywords extracted from resumes.

▪ Features:

Once the project is fully developed based on the above objectives, the solution is expected to have following features:

- ➢ Tool provides a brief summary from a posted job description/role requirement.
- ➢ Tool provides a sentiment analysis for the job descriptions .
- ➢ Tool can extract important relevant keywords from a resume document.

➢ Tool suggests best fit candidates' resumes to a given job description or generates matching score between job description and resume.

3. Related Work (Background):

Title 1: Automated Resume Evaluation System using NLP

Link: https://ieeexplore.ieee.org/abstract/document/9036842

Summary:

For most businesses, the hiring process is essential, and traditional hiring practices have lost their effectiveness as online hiring grows in popularity. The traditional methods of manually screening applications and creating a shortlist of qualified applicants for interviews take a long time and a lot of work. Job hunting has improved in both accessibility and sophistication in the age of technology. The selection of a candidate based solely on their resume has not, however, been totally automated. This study suggests a model that ranks resumes in accordance with the preferences and requirements of the employer by extracting useful information from them.

To accomplish the desired result, the suggested model has three segments. Using NLP approaches, the first section transforms unstructured resumes into structured data. The extraction phase, which makes up the second section, is when pertinent data from the resume is taken out and given an identifying value. In the final section, the resumes are ranked in accordance with the values assigned. The model streamlines the hiring procedure and saves time and effort by segmenting the entire process into four parts.

The suggested technique provides an automated way for recruiters to find qualified candidates by gleaning important data from their resumes. The model transforms unstructured data into structured data using NLP techniques, making it simpler to retrieve pertinent information. The model then gives the information that was retrieved an identifying value and ranks the resumes according to the values given. This strategy enables a quicker and more effective hiring procedure, which is advantageous to both the business and job seekers.

Title 2 : Ontology Based Job and Resume Matcher

Link:
https://www.cmpe.boun.edu.tr/~gungort/papers/Ontology%20Based%20Job%20and%20Resume%20Matcher.pdf

Summary:

An innovative method for enhancing the job matching process is presented in this paper in the form of an ontology-based job and resume matcher system. The authors suggest a system that uses an ontology-based model for matching, pulls qualifications and experience from both resumes and job postings. Using ontologies allows for a more accurate and efficient matching process by giving the knowledge domain a structured representation. The method provides a better user experience, requires little input from job seekers, and improves the match between the candidate and the job.

The authors provide a thorough analysis of the system's performance, proving that their strategy outperforms current practices. The system's 87% precision rating shows that the matching procedure was highly accurate. The system's architecture and the numerous parts that make up the system are also fully described by the authors. Researchers and developers interested in creating comparable systems or researching the usage of ontologies in the context of job matching will find this information to be useful.

The study offers an innovative method that uses ontologies to increase the speed and accuracy of the matching process, making a significant addition to the field of job matching and recruitment overall. The evaluation of the system gives solid proof of its efficacy, and the publication is a useful resource for academics and developers in the field due to the authors' in-depth description of the system's design and components. The essay makes a strong argument for the use of ontology-based systems in future job matching systems by illustrating their potential in the field of recruitment.

Title 3: An Intelligent framework for E-Recruitment System Based on Text Categorization and Semantic Analysis

Link: https://ieeexplore.ieee.org/abstract/document/9544102

Summary:

An new strategy is to employ NLP technology to create an autonomous text classification system for job and resume categorization in the online hiring process. Recruiters and job searchers can use the suggested system to identify appropriate job offers and resumes, respectively. Both parties depend on the system's correctness because incorrect categorization could lead to a mismatch between talents and employment requirements. To extract pertinent information from the resumes, the system uses POS tagging, tokenization, and lemmatization of the data. The Phrase Matcher is used to rate resumes based on the information provided by the recruiter, suggesting to job seekers that their skills are inadequate and giving the recruiter access to the best resumes. The system is effective in classifying resumes based on job descriptions, as shown by the results of the system evaluation, which showed better precision and recall.

The suggested method takes a novel approach to categorizing resume data on a large dataset of job descriptions by utilizing Word Order Similarity between Sentences and Domain Adaptation for the sensitive nature of resume content. The technology can offer job seekers customized job recommendations by grouping individuals based on the data in their resumes. The system's results and analyses are displayed, giving information about how well the strategy works to increase the effectiveness of the hiring process. The proposed system's capacity to spot job seekers' skill gaps and recommend them is a critical component in assisting them in improving their employability. Also, the system's capability to give the recruiter the best resumes can considerably lessen their workload so they can concentrate on the most relevant Resumes. Overall, the proposed system is a promising strategy that has the potential to change online recruitment by enhancing its precision, efficacy, and efficiency.

Title 4: A novel firefly driven scheme for resume parsing and matching based on entity linking paradigm

Link: **https://www.tandfonline.com/doi/abs/10.1080/09720529.2020.1721879**

Summary:

In order to increase the effectiveness and efficiency of the hiring process, the paper outlines a data-driven HR approach that makes use of NLP approaches. In order to effectively match resumes with job requirements, the authors specifically created a resume parser that evaluates critical recruitment parameters and incorporates a pie chart presentation into the algorithmic structure. The firefly ranking algorithm was used to assess the effectiveness and accuracy of the suggested approach, which produced an overall accuracy of 94.19%.

The paper's contribution is the creation of a strong resume parser that effectively matches resumes with job requirements, overcoming the drawbacks of traditional hiring practices. The parser can now more accurately extract crucial data from unstructured resumes thanks to the application of NLP techniques. Also, by supplying a useful tool for matching resumes with job requirements, the inclusion of a pie chart display in the algorithmic framework improves the parser's efficiency. The findings indicate that the proposed strategy has a great deal of potential to increase the effectiveness and efficiency of the hiring process.

Title 5: Resume Parsing Framework for E-recruitment

Link: **https://ieeexplore.ieee.org/abstract/document/9721762**

Summary:

In order to accurately extract information from resumes for hiring reasons, the article suggests a paradigm for resume parsing. Due to the wide range of formats used in resumes and the requirement for substantial annotated data, the existing methods, such as rule-based, supervised, and semantics-based methods, have their limits. In the suggested framework, resumes' raw content is extracted, and blocks are divided using text block categorization. Afterwards, named entity recognition is used to extract the entities, and an ontology is then used to enrich them. The suggested approach accurately extracts information from resumes that aids in the decision-making process and addresses the shortcomings of existing methodologies.

The suggested approach makes a substantial contribution to the development of recommender systems for e-recruitment. A more effective and efficient hiring procedure may result from the precise information extraction from resumes utilizing the suggested methodology. The framework is a useful advancement in the resume parsing field since it can manage the constraints imposed by current approaches. The retrieved entities are enriched using ontology, which improves the framework's accuracy. The suggested framework has the potential to considerably enhance the hiring procedure overall.

Title 6: Learning-Based Matched Representation System for Job Recommendation

Link: https://www.mdpi.com/1943792

Summary:

In this paper, a recommender system that helps job searchers locate relevant positions based on their resumes is proposed. When faced with multiple job offers, the conventional job recommendation algorithms were unable to suggest positions that appropriately matched the job seekers' profiles. The suggested method, however, uses content-based filtering to analyze and compare the similarity between the job seeker's talents and explicit elements of the job description, and then recommends the top-n positions to the job seekers. In order to match CV skills to advertised positions, the system first obtained first-hand information by extracting job descriptions from Indeed from key Saudi Arabian cities. It then examined the top talents demanded in job offers. Using decision support tools, the success and error rates of recommendations were compared to actual results.

This study is essential because it offers a solution to the issue that job searchers confront when trying to find openings that match their qualifications and expertise. The recommender system's content-based filtering method circumvents the drawback of conventional job suggestion techniques. By guaranteeing that job offers and resumes are appropriately matched, the proposed method benefits both recruiters and job seekers by saving time and money. Also, a more accurate evaluation of the effectiveness of the system can be obtained by using the decision support metrics when comparing the results of the system to reality. Considering the Saudi Arabian job market in particular, this paper offers a significant contribution to the field of job recommendation systems.

Title 7: A survey of job recommender systems

Link: https://academicjournals.org/journal/IJPS/article-full-text-pdf/B19DCA416592.pdf

Summary:

In order to create individualized recommender systems for candidates and job matching, the article presents an overview of e-recruiting procedures and current recommendation techniques. Traditional information retrieval methods are no longer sufficient due to the rise of internet-based recruiting platforms, which causes many candidates to lose out on job possibilities. The recommender system technology, which addresses concerns with information overload, has proved successful in e-commerce applications. Consequently, to enhance the functionality of e-recruiting, the essay emphasizes the significance of recommender systems.

The article provides a thorough analysis of current recommendation methods and e-recruiting procedures. It offers insights into the many methods used in personalized recommender systems, such as content-based, collaborative, hybrid, and knowledge-based strategies. It is an educational read for scholars and practitioners in the field because the writers also cover the benefits and limits of each technique. Overall, this research advances knowledge of recommender systems' function in e-recruiting and the potential advantages they may offer to both job seekers and recruiters.

Title 8: Matching Resumes to Jobs via Deep Siamese Network

Link: https://dl.acm.org/doi/abs/10.1145/3184558.3186942

Summary:

The difficult problem of proposing positions to job-seeking people by matching their resumes to job descriptions is the main topic of this study. The authors suggest a siamese adaption of convolutional neural networks to address this issue since it can accurately capture the underlying semantics and project similar job descriptions and resumes closer to one another in a semantic space. The method is evaluated on a sizable dataset of more than 5 million pairs of resumes and job descriptions, and the experimental findings show that the proposed method outperforms the state-of-the-art approaches.

This study is useful because it addresses a crucial problem in job searching by putting forth a cutting-edge method that successfully matches resumes to job descriptions. A promising method that can capture the underlying semantics of the data and increase the accuracy of the job recommendation system is the use of siamese adaption of convolutional neural networks. Strong proof that the suggested strategy outperforms current methods and might be used for real-world job recommendation systems is shown by the authors' extensive experiment.

4. Dataset:

1. As part of project increment-1, we are employing "job description.csv" for sentiment analysis and text summarization; a detailed description of the dataset is provided below.

2. The sample "resume.csv" is used to extract keywords from the "resume.csv" dataset in order to best match the candidate resume to the specified job description, a detailed explanation of the dataset is provided below.

3. “stopwords.txt” file which is used for removing the stop words as a part of preprocessing process.

1. Detailed explanation about “Job description.csv” dataset.

- **Country**: Depicts the country name for a particular given “Job description”. Data type of this column is “String”.Eg: {“United states of America”}.
- **Country code**: Depicts the country _ code for a particular given “job description”. Data type of this column is “string” of characters. Eg: {“US”}.
- **Date added:** Depicts on which date the “Job” was posted.
- **Has Expired:** Displays whether the job has expired or not expired. Data type of this column is “string” of characters. Eg:{“Yes”,”No”}.
- **Job board:** Describes on which job board, job has been posted. Eg: {“Monster”,”Indeed”,”Linkedin}. Date type of this column is “string” of characters.
- **Job Description:** Describes the Job description for a particular job that has been posted on the platform. Data type of this column is “string” of characters.
- **Job Title:** Describes the Job Title for a particular job that has been posted on the platform. Data type of this column is “string” of characters. Eg: {“It support Technician”, “Cybercoders” etc}.
- **Job Type:** Describes the type of job such as “Full Time” or “Contract”. Data Type of this column is “String” of characters.
- **Location:** Describes the location of job for a particular job that has been posted. Data Type of this column is “String” of characters.
- **Organization:** Describes the name of the organization. Data Type of this column is “String” of characters.

- **Page url:** Describes the “url” for the particular job that has been posted on the board. Data Type of this column is “String” of characters.
- **Salary:** Describes the “salary” for the job that has been posted. Data Type of this column is “String” of characters
- **Sector:** Describes the “sector” for the job that has been posted. Data Type of this column is “String” of characters
- **Uniq id:** Describes the “uniq id” for the job that has been posted. Data Type of this column is “String” of characters

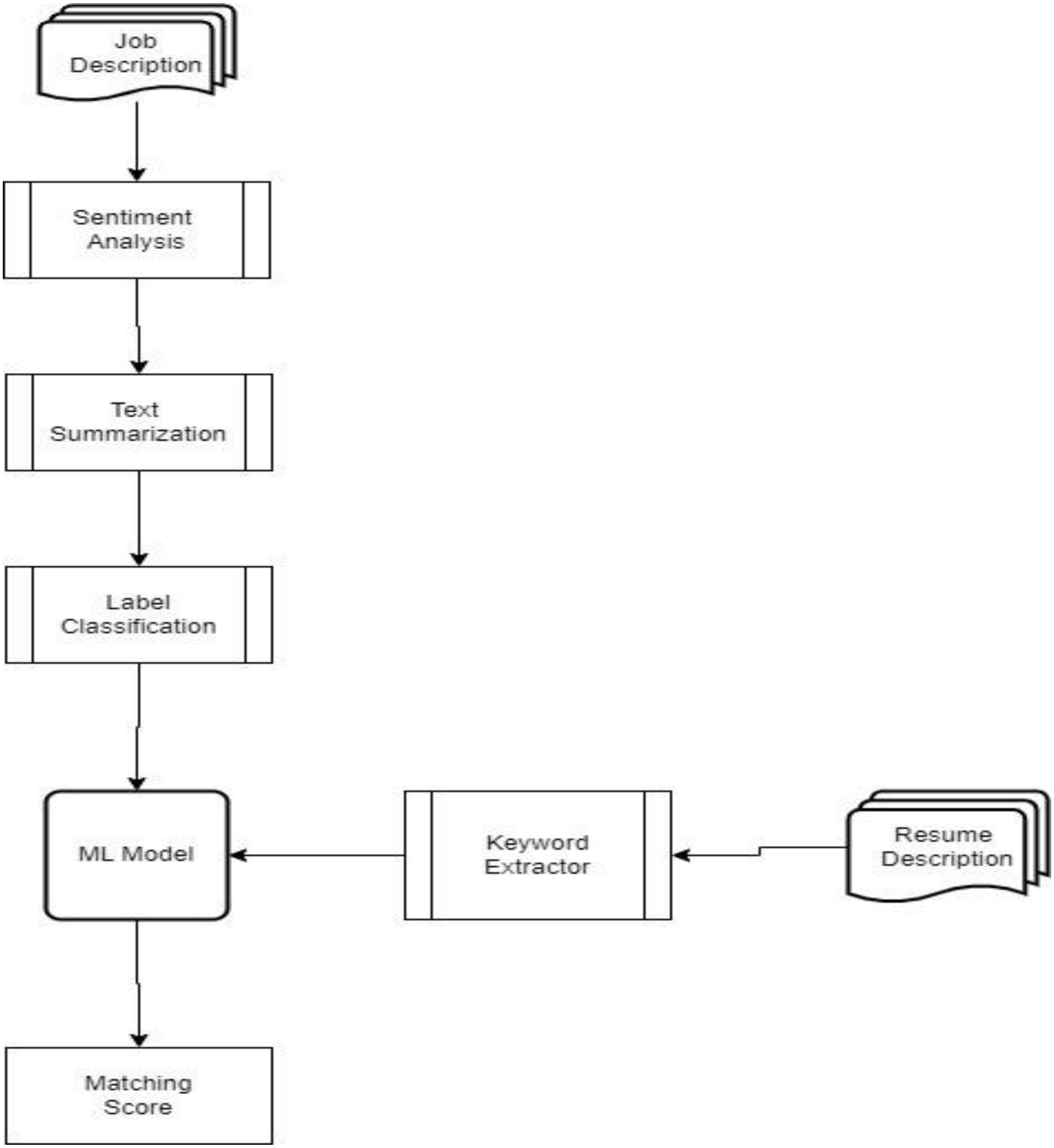
2. Detailed explanation about “Resume.csv” dataset.

- **Id:** Describes the “Id” for the job that has been posted. Data Type of this column is “String” of characters
- **Resume Str:** Describes the “Resume_str” for the job that has been posted. Data Type of this column is “String” of characters
- **Resume html:** Describes the “Resume html” for the job that has been posted. Data Type of this column is “String” of characters
- **Category:** Describes the “Category” for the job that has been posted. Data Type of this column is “String” of characters. Eg{“HR”, ”TR”}.

5. Implementation and Detail design of Features:

Detailed Workflow

According to the project’s workflow, we are performing 70% of the project in the project increment 1. First, we downloaded and read the job descriptions dataset. And perform the pre-processing before we work on the dataset. We drop the unwanted rows and columns and rows having null values. We added sentiment analyzer based on the comments from Project proposal. We performed 2 kinds of sentiment analyzer namely VaderSentiment and textblob techniques. Then we implemented the spacy pipeline and added textrank to extract brief summary for every job description. We implemented the TfIdf Vectorizer on resume dataset to extract the top keywords from resume descriptions.



Sentiment Analyzer

A sentiment analyzer that uses pre-built dictionaries or lexicons of words and their corresponding sentiment scores is known as a lexicon-based sentiment analyzer. Using the sentiment scores of the words it includes, this method seeks to determine the sentiment of a given passage of text. The working involves following steps:

- The first step is to preprocess the text data by tokenizing the text into individual words, removing stop words, and converting the text to lowercase.
- The next step is to develop a vocabulary or lexicon of terms and the sentiment scores assigned to them. The lexicon can be manually curated, crowdsourced sentiment annotations can be used, or pre-built lexicons like the AFINN lexicon or the NRC Emotion Lexicon can be used to accomplish this.
- Each word in the text is given a sentiment score based on its inclusion in the lexicon after the lexicon has been developed. For instance, the sentiment score for the word "happy" might be +1 whereas the sentiment score for the word "sad" might be -1.
- The final step is to aggregate the sentiment scores of all the words in the text to arrive at an overall sentiment score for the text. This can be done using various methods, such as taking the average of all the scores or using more sophisticated techniques such as weighted scoring.

Text Summarization

In extractive summarizing, a subset of the most crucial phrases or sentences from the original text are chosen to construct the summary. This is often accomplished by locating passages that contain words, named entities, or other significant elements that are indicative of the text's overall content. The summary is created by joining these selected sentences together.

In abstractive summarizing, the summary is produced by employing natural language generation techniques to produce new sentences that effectively convey the main ideas of the original text. This necessitates a deeper comprehension of the text's subject matter as well as the capacity to create meaningful sentences that express the same ideas as the original text.

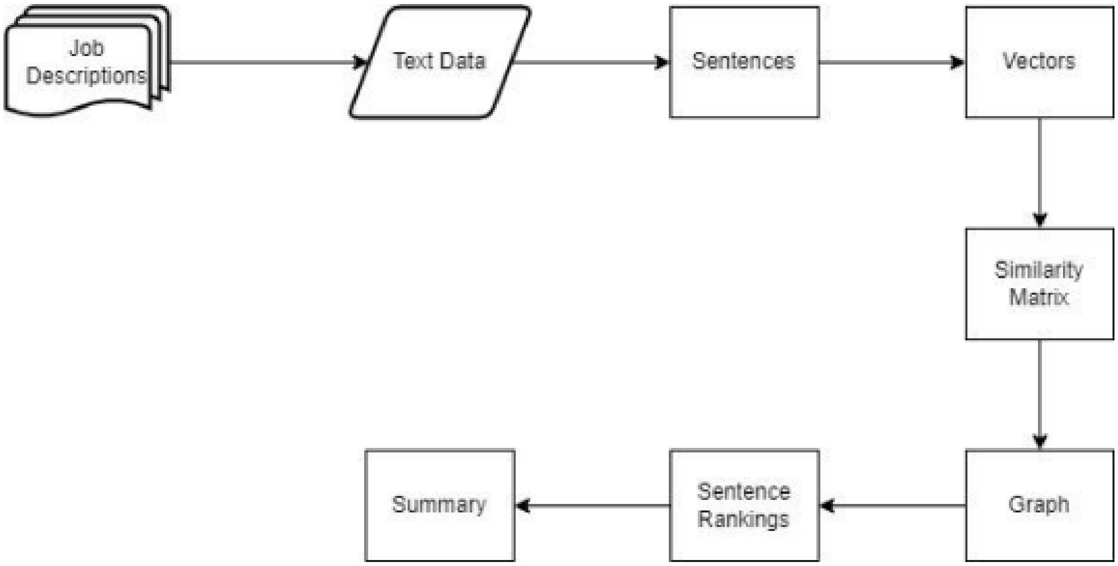


Fig 2. Text Summarization Process

Keyword Extraction

The open-source NLP library Spacy offers a number of tools for keyword extraction. The following steps are commonly involved in keyword extraction using Spacy:

- Preprocessing the text data involves breaking it down into individual words, getting rid of unnecessary words, and changing the text's case to lowercase. The Spacy tokenizer and pre-made language models can be used for this.
- The text is then subjected to part-of-speech (POS) tagging. POS tagging includes classifying each word in the text according to its syntactic type, such as a noun, verb,

adjective, or adverb. The POS tagger from Spacy can be used for this.

- Spacy can be used for entity recognition in addition to POS tagging. This entails locating named characters, groups, places, dates, or other entities in the text. These named entities might be seen as crucial textual keywords.
- After POS tagging and entity recognition have been used to preprocess and evaluate the text, the last step is to extract the most significant keywords. The built-in features of Spacy, like as phrase matching and noun chunking, can be used for this, as well as more sophisticated methods like TF-IDF or TextRank.

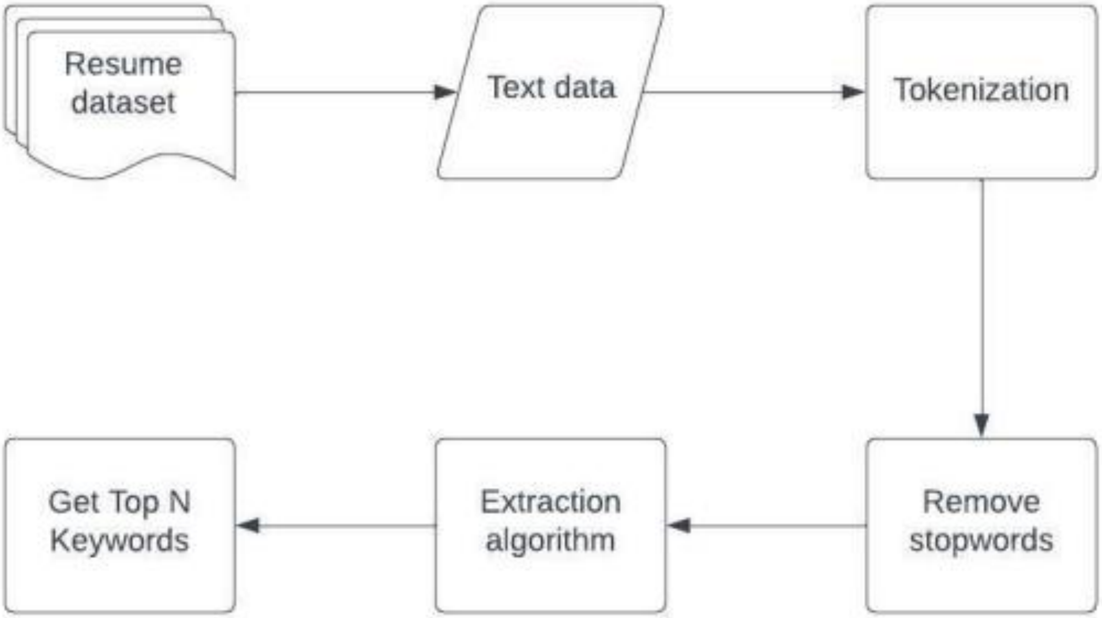
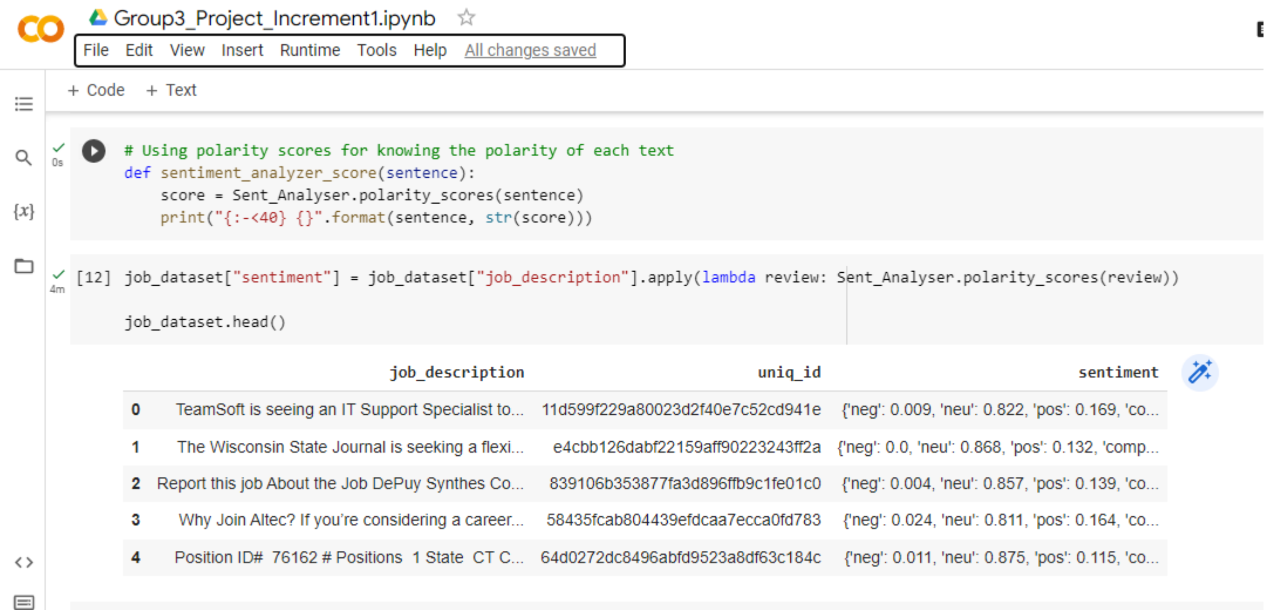


Fig 3. Keyword extraction Process

6. Analysis and Preliminary Results:

In the project increment 1, we mainly performed 3 NLP techniques namely sentiment analysis, text summarization and keyword extraction.

We first performed the VaderSentimentAnalyzer on the job descriptions in job data set and the results are following:

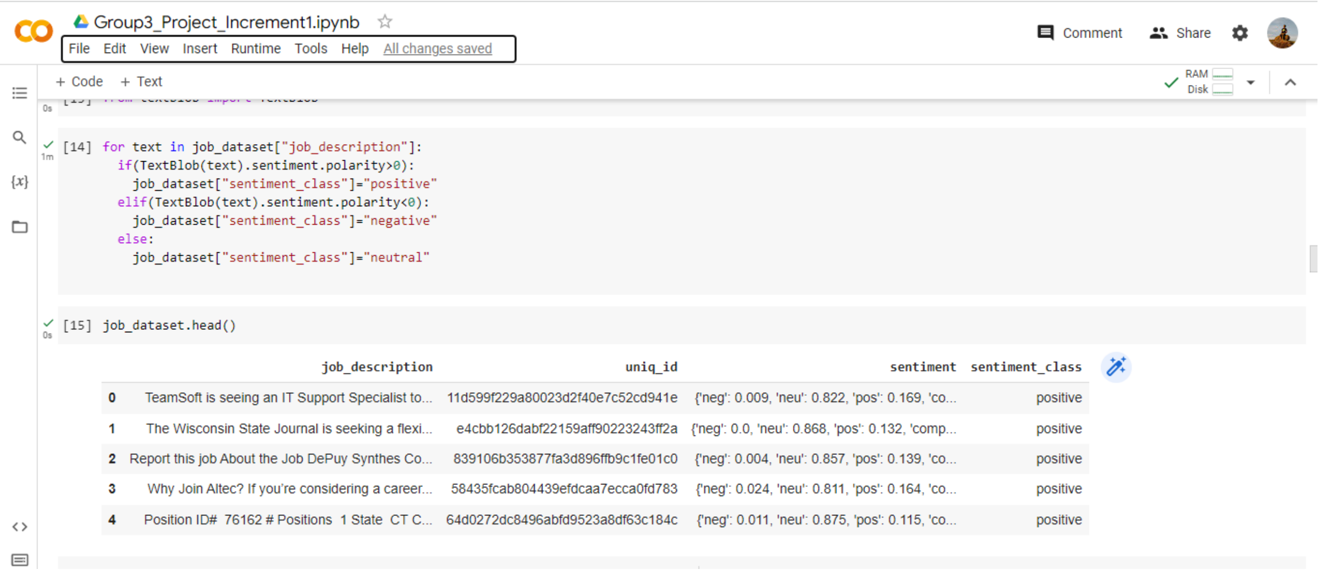

Group3_Project_Increment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text

```
# Using polarity scores for knowing the polarity of each text
def sentiment_analyzer_score(sentence):
    score = Sent_Analyser.polarity_scores(sentence)
    print("{:-<40} {}".format(sentence, str(score)))
```

```
[12] job_dataset["sentiment"] = job_dataset["job_description"].apply(lambda review: Sent_Analyser.polarity_scores(review))

     job_dataset.head()
```

	job_description	uniq_id	sentiment
0	TeamSoft is seeing an IT Support Specialist to...	11d599f229a80023d2f40e7c52cd941e	{'neg': 0.009, 'neu': 0.822, 'pos': 0.169, 'co...
1	The Wisconsin State Journal is seeking a flexi...	e4cbb126dabf22159aff90223243ff2a	{'neg': 0.0, 'neu': 0.868, 'pos': 0.132, 'comp...
2	Report this job About the Job DePuy Synthes Co...	839106b353877fa3d896ffb9c1fe01c0	{'neg': 0.004, 'neu': 0.857, 'pos': 0.139, 'co...
3	Why Join Altec? If you’re considering a career...	58435fcab804439efdcaa7ecca0fd783	{'neg': 0.024, 'neu': 0.811, 'pos': 0.164, 'co...
4	Position ID# 76162 # Positions 1 State CT C...	64d0272dc8496abfd9523a8df63c184c	{'neg': 0.011, 'neu': 0.875, 'pos': 0.115, 'co...

We also performed sentiment analysis using textblob and the resulted classification is as follows:

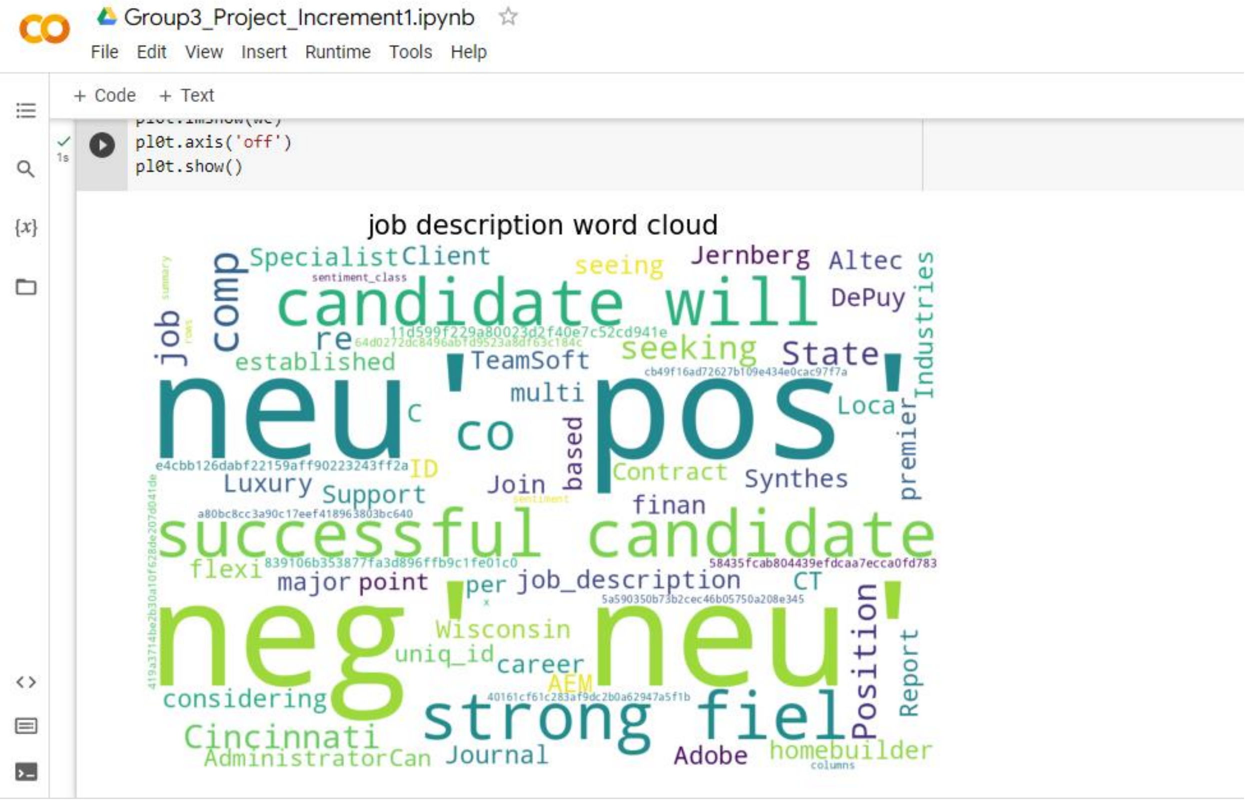

Group3_Project_Increment1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk

```
[14] for text in job_dataset["job_description"]:
       if(TextBlob(text).sentiment.polarity>0):
         job_dataset["sentiment_class"]="positive"
       elif(TextBlob(text).sentiment.polarity<0):
         job_dataset["sentiment_class"]="negative"
       else:
         job_dataset["sentiment_class"]="neutral"
```

```
[15] job_dataset.head()
```

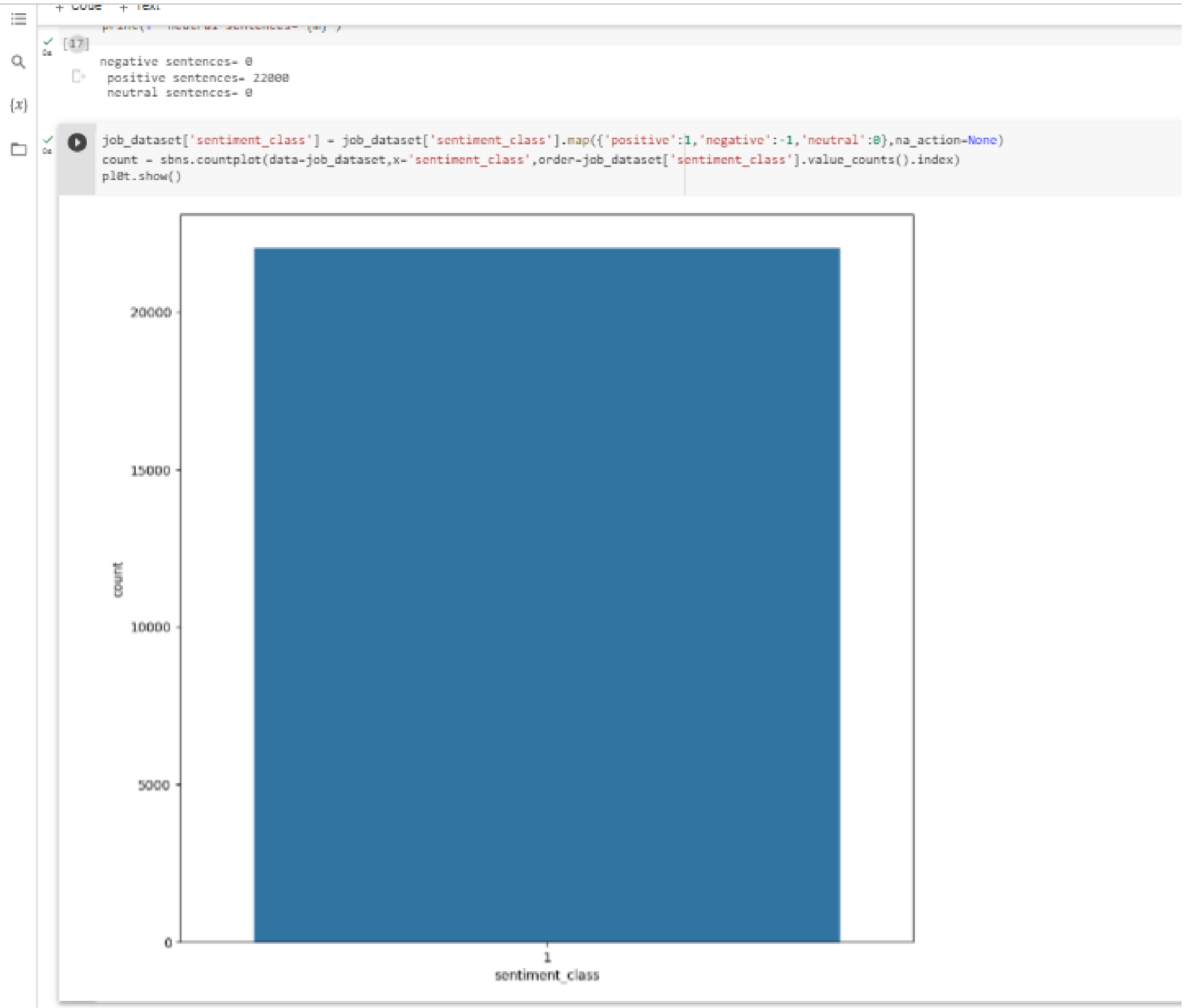
	job_description	uniq_id	sentiment	sentiment_class
0	TeamSoft is seeing an IT Support Specialist to...	11d599f229a80023d2f40e7c52cd941e	{'neg': 0.009, 'neu': 0.822, 'pos': 0.169, 'co...	positive
1	The Wisconsin State Journal is seeking a flexi...	e4cbb126dabf22159aff90223243ff2a	{'neg': 0.0, 'neu': 0.868, 'pos': 0.132, 'comp...	positive
2	Report this job About the Job DePuy Synthes Co...	839106b353877fa3d896ffb9c1fe01c0	{'neg': 0.004, 'neu': 0.857, 'pos': 0.139, 'co...	positive
3	Why Join Altec? If you’re considering a career...	58435fcab804439efdcaa7ecca0fd783	{'neg': 0.024, 'neu': 0.811, 'pos': 0.164, 'co...	positive
4	Position ID# 76162 # Positions 1 State CT C...	64d0272dc8496abfd9523a8df63c184c	{'neg': 0.011, 'neu': 0.875, 'pos': 0.115, 'co...	positive

The word cloud resulted from the job description is:



```
plt.imshow(wc)
pl0t.axis('off')
pl0t.show()
```

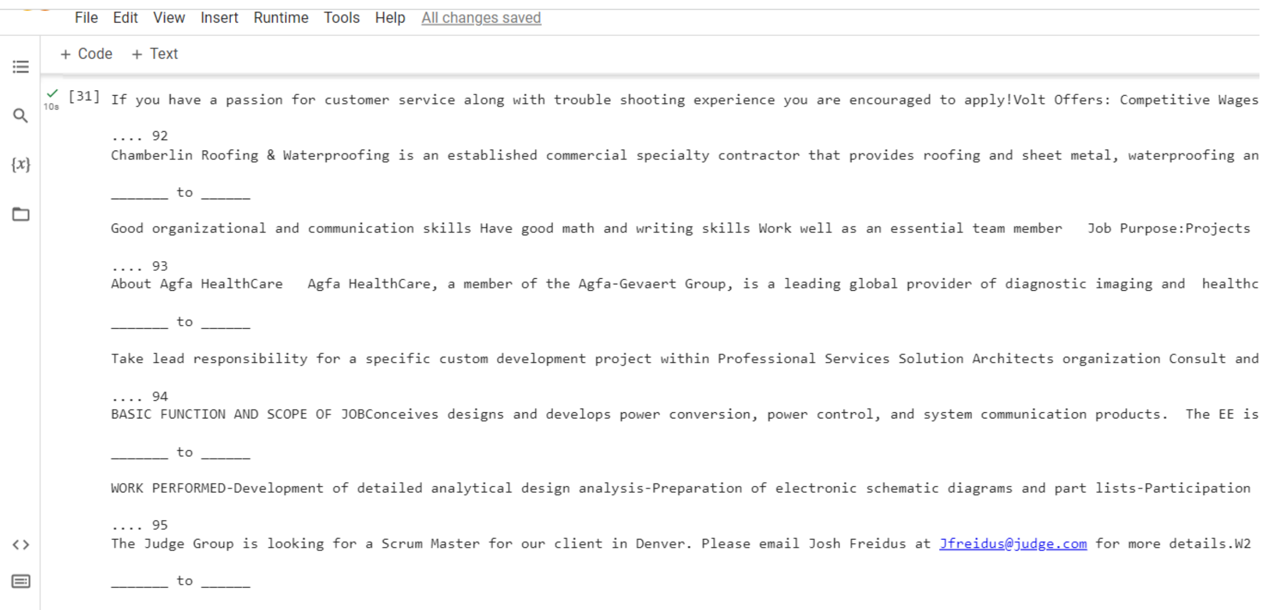
Also, we can see that almost all job descriptions are positive toned only:



```
negative sentences= 0
 positive sentences= 22000
 neutral sentences= 0
```

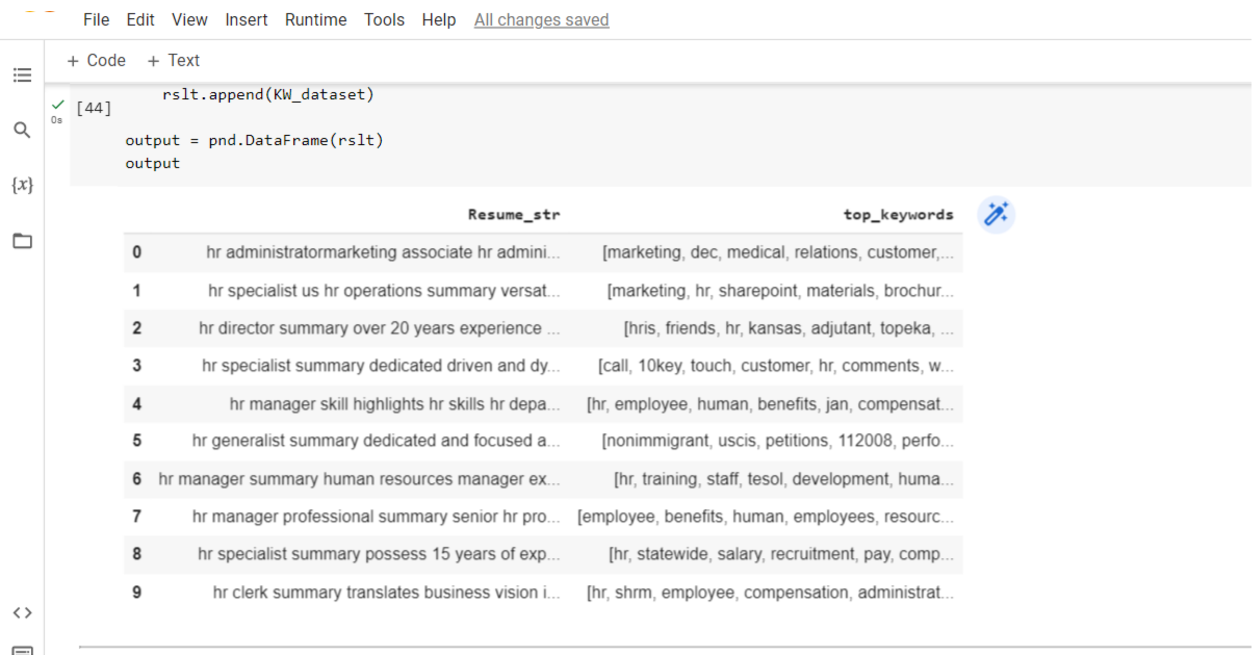
```
job_dataset['sentiment_class'] = job_dataset['sentiment_class'].map({'positive':1,'negative':-1,'neutral':0},na_action=None)
count = sbns.countplot(data=job_dataset,x='sentiment_class',order=job_dataset['sentiment_class'].value_counts().index)
pl0t.show()
```

The results from text summarization on the first 100 job descriptions is as shown below:



```
[31] If you have a passion for customer service along with trouble shooting experience you are encouraged to apply!Volt Offers: Competitive Wages

     .... 92
     Chamberlin Roofing & Waterproofing is an established commercial specialty contractor that provides roofing and sheet metal, waterproofing an

     ______ to ______

     Good organizational and communication skills Have good math and writing skills Work well as an essential team member   Job Purpose:Projects

     .... 93
     About Agfa HealthCare   Agfa HealthCare, a member of the Agfa-Gevaert Group, is a leading global provider of diagnostic imaging and  healthc

     ______ to ______

     Take lead responsibility for a specific custom development project within Professional Services Solution Architects organization Consult and

     .... 94
     BASIC FUNCTION AND SCOPE OF JOBConceives designs and develops power conversion, power control, and system communication products.  The EE is

     ______ to ______

     WORK PERFORMED-Development of detailed analytical design analysis-Preparation of electronic schematic diagrams and part lists-Participation

     .... 95
     The Judge Group is looking for a Scrum Master for our client in Denver. Please email Josh Freidus at Jfreidus@judge.com for more details.W2

     ______ to ______
```

Also, the output from the keyword extraction on the resume dataset is as follows:



```
        rslt.append(KW_dataset)
[44]
    output = pnd.DataFrame(rslt)
    output
```

	Resume_str	top_keywords
0	hr administratormarketing associate hr admini...	[marketing, dec, medical, relations, customer,...
1	hr specialist us hr operations summary versat...	[marketing, hr, sharepoint, materials, brochur...
2	hr director summary over 20 years experience ...	[hris, friends, hr, kansas, adjutant, topeka, ...
3	hr specialist summary dedicated driven and dy...	[call, 10key, touch, customer, hr, comments, w...
4	hr manager skill highlights hr skills hr depa...	[hr, employee, human, benefits, jan, compensat...
5	hr generalist summary dedicated and focused a...	[nonimmigrant, uscis, petitions, 112008, perfo...
6	hr manager summary human resources manager ex...	[hr, training, staff, tesol, development, huma...
7	hr manager professional summary senior hr pro...	[employee, benefits, human, employees, resourc...
8	hr specialist summary possess 15 years of exp...	[hr, statewide, salary, recruitment, pay, comp...
9	hr clerk summary translates business vision i...	[hr, shrm, employee, compensation, administrat...

7. Project Management:

Work completed

- ## Description

  In this increment, we worked on the major parts of the project which includes the tasks like Sentiment Analysis, Text Summarization and Keyword Extraction. We first invested time on finding relevant datasets for the project. We then performed the pre-processing of these data and did a little exploratory data analysis as well. We then executed the sentiment analysis on the job descriptions and extracted short summaries from them. We also performed a bit of pre processing on the resume dataset and extracted top keywords from the resume descriptions.

- ## Responsibility

Person	Task
Suhas Siddarajgari Tellatakula	Sentiment Analysis, Text Summarization, Keyword Extraction, Report
Vamshi Telukuntla	Sentiment Analysis, Text Summarization, Keyword Extraction, Report
Srikanth Mamillapalli	Sentiment Analysis, Text Summarization, Keyword Extraction, Report
Sairohithvarma Kantem	Sentiment Analysis, Text Summarization, Keyword Extraction, Report

- ## Issues/Concerns

  Here are few issues we faced during the course of project increment 1:

- Finding relevant datasets from plethora of data from Kaggle.
- Having neutral job descriptions.
- Having null entries in datasets.
- Having improper/ unidentified notations or symbols in the datasets.

Work to be completed

- **Description**

  In the next increment, we aim to achieve the remaining parts of the project which includes label classification on job descriptions, building a machine learning model to suggest resumes relevant to the job descriptions. We will try out different techniques to find the best possible model to generate best accurate results.

- **Responsibility**

Person	Task
Suhas Siddarajgari Tellatakula	Label classification, Machine Learning models, report
Vamshi Telukuntla	Label classification, Machine Learning models, report
Srikanth Mamillapalli	Label classification, Machine Learning models, report
Sairohithvarma Kantem	Label classification, Machine Learning models, report

- **Issues/Concerns**

  Here are few issues which we can expect as of now:

  - Considering various Machine learning models.

- Choosing the model which gives best accuracy.
- Not having enough relevant data.

8.References/Bibliography:

- https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset
- https://www.kaggle.com/code/residentmario/exploring-monster-com-job-postings
- https://towardsdatascience.com/keyword-extraction-process-in-python-with-natural-language-processing-nlp-d769a9069d5c
- https://www.topcoder.com/thrive/articles/text-summarization-in-nlp
- https://www.analyticsvidhya.com/blog/2020/11/words-that-matter-a-simple-guide-to-keyword-extraction-in-python/
- https://www.kaggle.com/code/apoorvgupta25/sentiment-analysis-using-logistics-regression/notebook

- https://www.kaggle.com/code/ahmed121ashraf131/nlp-summerization/notebook
- https://www.kaggle.com/code/akhatova/extract-keywords/notebook
- https://www.kaggle.com/code/vicely07/sentiment-analysis-of-city-of-la-job-postings
- https://www.kaggle.com/code/nezarabdilahprakasa/matching-cv-to-job-description-using-python
- https://www.kaggle.com/code/sanabdriss/nlp-extract-skills-from-job-descriptions