



CS688 MINING ON THE OF DATA SCIENTIST JOB DESCRIPTIONS USING BERT

Yuandi TANG U65674688

TOPICS

1. RESEARCH BACKGROUND

2. DATA AND PREPARATION

3. PROCESSION AND RESULTS

4. CONCLUSION





RESEARCH BACKGROUND

RESEARCH SCENARIO

Employment of data scientists is projected to grow 36 percent from 2021 to 2031, much faster than the average for all occupations. About 13,500 openings for data scientists are projected each year, on average, over the decade.

Source: US Bureau of Labor Statistics

RESEARCH SCENARIO

Data Scientist is one of the hottest jobs in the US job market, it has over 300k job openings on LinkedIn right now.

Source: LinkedIn

The screenshot shows a LinkedIn search interface with the following details:

- Search Bar:** data scientist, United States, Search button.
- Filter Buttons:** Jobs, Date posted, Experience level, Company, Job type.
- Result Summary:** Data scientist in United States, 322,946 results, Set alert toggle.
- Job Listings:**
 - Jr Data Scientist** at Apex Systems (Charlotte, NC On-site), \$50/hr - \$60/hr, 4 benefits, Actively recruiting, Promoted, Easy Apply.
 - Sr Data Scientist US Citizen/Secret Clearance Required** at Echelon Services, LLC (Tobyhanna, PA On-site), Medical, Vision, Dental, 401(k), Actively recruiting, Promoted, 7 applicants.
 - Data Scientist** at AE Studio (United States Remote), Actively recruiting, Promoted.
 - Sr. Data Scientist Methodology Pharma SQL PYTHON REMOTE US** at ICON plc.

RESEARCH SCENARIO

As of 2022, over 1,000 universities have a data science degree at all levels. In our database, there are over 980 different programs, with the Master of Data Science degree being the most popular.

Source: [DataScienceProgram.org](https://www.datascienceprogram.org)





QUESTION

- What can we learn from job descriptions of data scientists from various companies? Any common and differences between them?
- How can we learn and be more prepared before we start our careers as data scientists?



DATA & PREPARATION

DATA INTRODUCTION

- Researched 20 current entry-level data scientist hiring ads on LinkedIn
- Looked up their job descriptions
- Keep the responsibility and requirement part, ignoring the company introduction
- Copy to text and saved into 20 text files(txt).



DATA INTRODUCTION

LinkedIn search results for "data scientist" in United States:

- Staff Data Scientist** at HID Global, Austin, TX (Hybrid). Contract - Mid-Senior level. 501-1,000 employees - IT Services and IT Consulting. Promoted.
- Data Scientist in Richmond, VA (HYBRID)** at Dice, Richmond, VA. 1 alum works here. Promoted - 7 applicants.
- Data Scientist** at Systems Planning and Analysis, Inc., Arlington, VA (Hybrid). 9 alumni work here. Promoted.
- Lead Data Scientist** at Motion Recruitment, Los Angeles, CA (On-site). Medical, Vision, Dental, 401(k). Actively recruiting.

Description:
Possess Advanced Excel, Power BI, Alteryx, SQL skills with the ability to drive insights from vast quantities of data.
Will support Network build analytics with a focus on Small Cell programs and Business initiatives.
Adhoc reporting functions, database management and data mining experience
Wireless telecom experience for basic network data understanding will be a plus.
Candidates who are willing to relocate may also apply.

A red box highlights the job description, and a black arrow points from it to the responsibilities section of a detailed job posting.

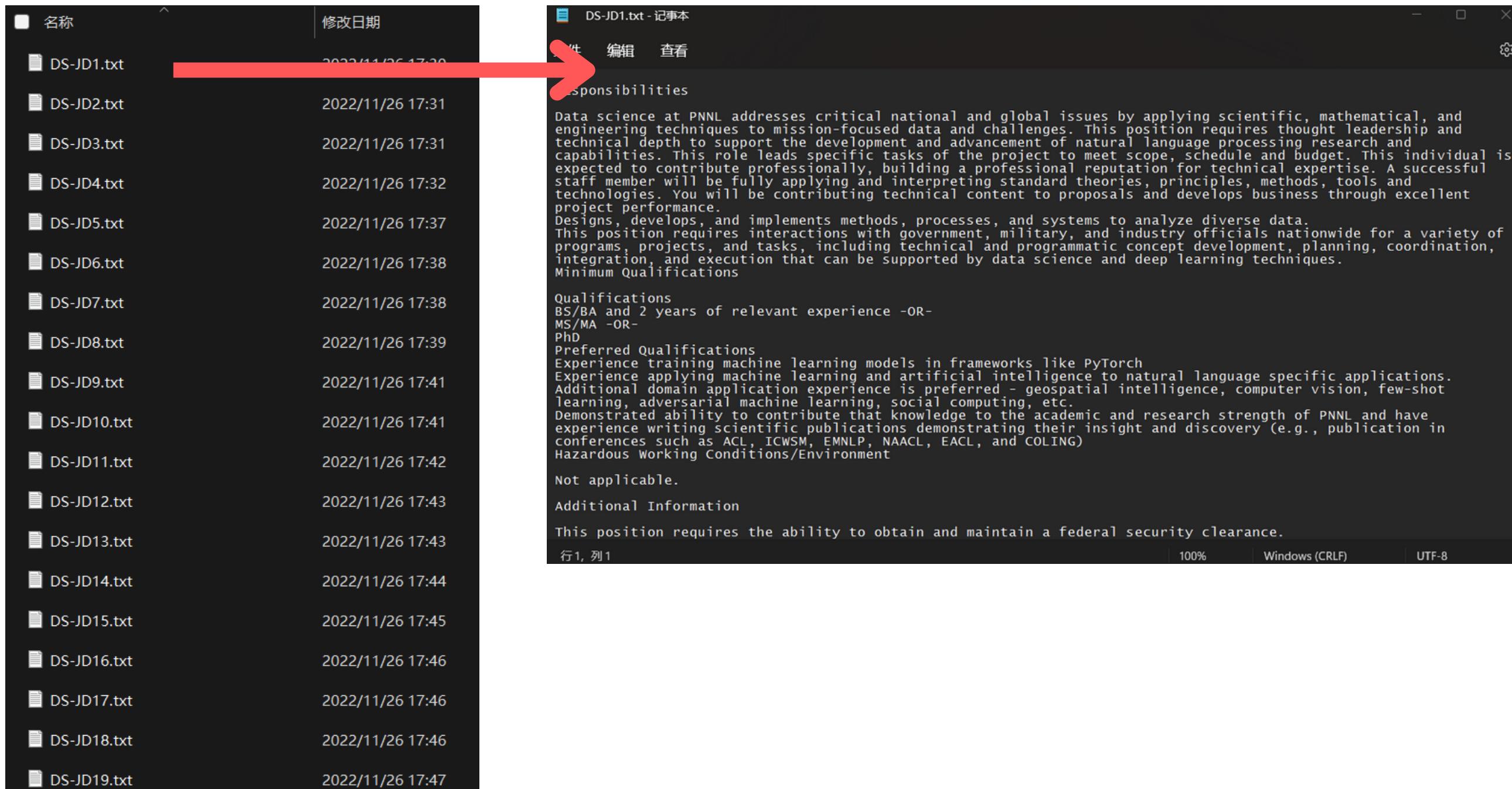
Responsibilities

Data science at PNNL addresses critical national and global issues by applying scientific, mathematical, and engineering techniques to mission-focused data and challenges. This position requires thought leadership and technical depth to support the development and advancement of natural language processing research and capabilities. This role leads specific tasks for the project to meet scope, schedule and budget. This individual is expected to contribute directly, leading, providing, replicating, facilitating, and applying technical expertise. A successful staff member will be fully applying and interpreting standard theories, principles, methods, tools and technologies. You will be contributing technical content to proposals and develops business through excellent project performance. Designs, develops, and implements methods, processes, and systems to analyze diverse data. This position requires interactions with government, military, and industry officials nationwide for a variety of programmatic projects and tasks, including technical and programmatic concept development, planning, coordination, integration, and execution that can be supported by data science and deep learning techniques.

Qualifications
BS/BA and 2 years of relevant experience -OR- MS/MA -OR- PhD
Preferred Qualifications
Experience training machine learning models in frameworks like PyTorch
Experience applying machine learning and artificial intelligence to natural language specific applications.
Additional domain application experience is preferred - geospatial intelligence, computer vision, few-shot learning, adversarial machine learning, federated learning, etc.
Demonstrated ability to contribute that knowledge to the academic and research strength of PNNL and have experience writing scientific publications demonstrating their insight and discovery (e.g., publication in conferences such as ACL, ICWSM, EMNLP, NAACL, EACL, and COLING)
Hazardous Working Conditions/Environment
Not applicable.
Additional Information
This position requires the ability to obtain and maintain a federal security clearance.



DATA INTRODUCTION



名称 修改日期

名称	修改日期
DS-JD1.txt	2022/11/26 17:47
DS-JD2.txt	2022/11/26 17:31
DS-JD3.txt	2022/11/26 17:31
DS-JD4.txt	2022/11/26 17:32
DS-JD5.txt	2022/11/26 17:37
DS-JD6.txt	2022/11/26 17:38
DS-JD7.txt	2022/11/26 17:38
DS-JD8.txt	2022/11/26 17:39
DS-JD9.txt	2022/11/26 17:41
DS-JD10.txt	2022/11/26 17:41
DS-JD11.txt	2022/11/26 17:42
DS-JD12.txt	2022/11/26 17:43
DS-JD13.txt	2022/11/26 17:43
DS-JD14.txt	2022/11/26 17:44
DS-JD15.txt	2022/11/26 17:45
DS-JD16.txt	2022/11/26 17:46
DS-JD17.txt	2022/11/26 17:46
DS-JD18.txt	2022/11/26 17:46
DS-JD19.txt	2022/11/26 17:47

DS-JD1.txt - 记事本

文件 编辑 查看

Responsibilities

Data science at PNNL addresses critical national and global issues by applying scientific, mathematical, and engineering techniques to mission-focused data and challenges. This position requires thought leadership and technical depth to support the development and advancement of natural language processing research and capabilities. This role leads specific tasks of the project to meet scope, schedule and budget. This individual is expected to contribute professionally, building a professional reputation for technical expertise. A successful staff member will be fully applying and interpreting standard theories, principles, methods, tools and technologies. You will be contributing technical content to proposals and develops business through excellent project performance.

Designs, develops, and implements methods, processes, and systems to analyze diverse data. This position requires interactions with government, military, and industry officials nationwide for a variety of programs, projects, and tasks, including technical and programmatic concept development, planning, coordination, integration, and execution that can be supported by data science and deep learning techniques.

Minimum Qualifications

BS/BA and 2 years of relevant experience -OR- MS/MA -OR- PhD

Preferred Qualifications

Experience training machine learning models in frameworks like PyTorch. Experience applying machine learning and artificial intelligence to natural language specific applications. Additional domain application experience is preferred - geospatial intelligence, computer vision, few-shot learning, adversarial machine learning, social computing, etc.

Demonstrated ability to contribute that knowledge to the academic and research strength of PNNL and have experience writing scientific publications demonstrating their insight and discovery (e.g., publication in conferences such as ACL, ICWSM, EMNLP, NAACL, EACL, and COLING)

Hazardous Working Conditions/Environment

Not applicable.

Additional Information

This position requires the ability to obtain and maintain a federal security clearance.

行1, 列1 100% Windows (CRLF) UTF-8





PROCESS

- Prepossessing
- Word Frequencies
- Keyword Extraction
- Text Summarization
- Similarity Comparison
- Q&A samples





PREPROCESSING

PREPOCESSING

- Read data
- Import libraries
- Render all words in lowercase
- Remove punctuation
- Filter stopwords
- Filter job names

PREPOCESSING

- Read data
- Import libraries
- Render all words in lowercase
- Remove punctuation
- Filter numbers
- Filter stopwords
- Filter job names

```
DOC1 = OPEN("DS-JD1.TXT")
DOC2 = OPEN("DS-JD2.TXT")
DOC3 = OPEN("DS-JD3.TXT")
DOC4 = OPEN("DS-JD4.TXT")
DOC5 = OPEN("DS-JD5.TXT")
DOC6 = OPEN("DS-JD6.TXT")
DOC7 = OPEN("DS-JD7.TXT")
DOC8 = OPEN("DS-JD8.TXT")
DOC9= OPEN("DS-JD9.TXT")
DOC10= OPEN("DS-JD10.TXT")
DOC11 = OPEN("DS-JD11.TXT")
DOC12 = OPEN("DS-JD12.TXT")
DOC13 = OPEN("DS-JD13.TXT")
DOC14 = OPEN("DS-JD14.TXT")
DOC15 = OPEN("DS-JD15.TXT")
DOC16 = OPEN("DS-JD16.TXT")
DOC17 = OPEN("DS-JD17.TXT")
DOC18 = OPEN("DS-JD18.TXT")
DOC19= OPEN("DS-JD19.TXT")
DOC20= OPEN("DS-JD20.TXT")
DOC=
DOC1.READ() + DOC2.READ() + DOC3.READ() + DOC4.READ() + DOC5.READ() + DOC6.READ() + DOC7.READ() + DOC8.READ() + DOC9.READ() + DOC10.READ() + DOC11
.READ() + DOC12.READ() + DOC13.READ() + DOC14.READ() + DOC15.READ() + DOC16.READ() + DOC17.READ() + DOC18.READ() + DOC19.READ() + DOC20.READ()
```

PREPOCESSING

- Read data
- Import libraries
- Render all words in lowercase
- Remove punctuation
- Filter numbers
- Filter stopwords
- Filter job names

```
IMPORT MATPLOTLIB.PY PLOT AS PLT
IMPORT PANDAS AS PD
IMPORT NUMPY AS NP
IMPORT NLTK
NLTK.DOWNLOAD('PUNKT')
NLTK.DOWNLOAD('STOPWORDS')
FROM KEYBERT IMPORT KEYBERT
FROM COLLECTIONS IMPORT COUNTER
FROM GENSIM.SUMMARIZATION IMPORT SUMMARIZE
FROM PYPARSING.HELPERS IMPORT WORDSTART
FROM TRANSFORMERS IMPORT AUTOTOKENIZER, AUTOMODEL
FROM SKLEARN.METRICS.PAIRWISE IMPORT COSINE_SIMILARITY
FROM NLTK.CORPUS IMPORT STOPWORDS
FROM NLTK.TOKENIZE IMPORT WORD_TOKENIZE, SENT_TOKENIZE
```

PREPOCESSING

- Read data
- Import libraries
- **Render all words in lowercase**
- Remove punctuation
- Filter numbers
- Filter stopwords
- Filter job names

```
#TOKENIZE WORDS  
DOCX = WORD_TOKENIZE(DOC)
```

```
# CONVERT TO LOWER CASE  
DOCX = [W.LOWER() FOR W IN DOCX]
```

PREPOCESSING

- Read data
- Import libraries
- Render all words in lowercase
- **Remove punctuation**
- Filter numbers
- Filter stopwords
- Filter job names

```
# REMOVE PUNCTUATION FROM EACH WORD
IMPORT STRING
TABLE = STR.MAKETRANS("", "", STRING.PUNCTUATION)
STRIPPED = [W.TRANSLATE(TABLE) FOR W IN DOCX]
```

PREPOCESSING

- Read data
 - Import libraries
 - Render all words in lowercase
 - Remove punctuation
 - **Filter numbers**
 - Filter stopwords
 - Filter job names
- # remove remaining tokens that are not alphabetic*
docx = [word for word in stripped if word.isalpha()]

PREPOCESSING

- Read data
- Import libraries
- Render all words in lowercase
- Remove punctuation
- Filter numbers
- **Filter stopwords**
- Filter job names

```
# filter out stop words  
stop_words = set(stopwords.words('english'))  
docx = [w for w in docx if not w in stop_words]
```

PREPOCESSING

- Read data
- Import libraries
- Render all words in lowercase
- Remove punctuation
- Filter numbers
- Filter stopwords
- **Filter job names** $docx = [w \text{ for } w \text{ in } docx \text{ if not } w \text{ in } job_names]$

PREPOCESSING RESULTS



Tokenized Data

```
['responsibilities', 'science', 'pnnl', 'addresses', 'critical', 'national',  
'global', 'issues', 'applying', 'scientific', 'mathematical',  
'engineering', 'techniques', 'missionfocused', 'data', 'challenges',  
'position',.....]
```



PROCESSION & RESULTS



VALIDATION OF BERT



SIMILARITIES COMPARISON

SENTENCE SIMILARITIES

*Are sentences nearby or random sentence pairs have more similarities?
-Logically answer: sentences nearby.*

Bert can help us validate the answer

```
model_name = "bert-base-uncased" # Plain Bert Model
# model_name = "GanjinZero/UMLS Bert_ENG" # CODER UMLS Medical Bert Model
model = AutoModel.from_pretrained(model_name)
tokenizer = AutoTokenizer.from_pretrained(model_name)
query = sent_tokenize(doc)[2] #3rd sentence in JD
sent_1 = sent_tokenize(doc)[3] #4th sentence in JD
sent_2 = sent_tokenize(doc)[100] #100th sentence in JD
# query = "the MRI of the abdomen is normal and without evidence of malignancy."
# sent_1 = "no significant abnormalities involving the abdomen is observed."
# sent_2 = "deformity of the ventral thecal sac is observed."
# Tokenize Sentences
inputs_0 = tokenizer(query, return_tensors='pt') # Get tokens as tensors
inputs_1 = tokenizer(sent_1, return_tensors='pt') # Get tokens as tensors
inputs_2 = tokenizer(sent_2, return_tensors='pt') # Get tokens as tensors
# Get the Embeddings
sent_0_embed = np.mean(model(**inputs_0).last_hidden_state[0].detach().numpy(), axis=0, keepdims=True)
sent_1_embed = np.mean(model(**inputs_1).last_hidden_state[0].detach().numpy(), axis=0, keepdims=True)
sent_2_embed = np.mean(model(**inputs_2).last_hidden_state[0].detach().numpy(), axis=0, keepdims=True)
# Calculate Cosine Symilarity between the 2 texts
similarities_q1 = cosine_similarity(sent_0_embed, sent_1_embed) # Find Cosine Symilarity between the 2 texts
similarities_q2 = cosine_similarity(sent_0_embed, sent_2_embed) # Find Cosine Symilarity between the 2 texts
print("Query Symilarity with Sentence 1 ---", similarities_q1[0][0])
print("Query Symilarity with Sentence 2 ---", similarities_q2[0][0])
print("====")
print("End")
#sentences nearby have logical connections, thus has more similarities than random sentences pairs.
```

S3- third sentence in JD

S4- fourth sentence in JD

S100- third sentence in JD

Similarity(S3,S4) 0.67355597

Similarity(S3,S100) 0.54893076

The 3rd and 4th sentence has more similarity than 3rd and 100th sentences

VALIDATED!



Q&A SAMPLES

SENTENCE SIMILARITIES

Question: *What is a data scientist?*

Bert can help us validate the answer too.

```
#Q&A#
text = summary_Gensim #load summary text
tokenizer = AutoTokenizer.from_pretrained("deepset/bert-base-cased-squad2") #tokenize models
model = AutoModelForQuestionAnswering.from_pretrained("deepset/bert-base-cased-squad2")
inputs = tokenizer.encode_plus(question, text, return_tensors="pt") # Take the question & text as text
and returns tokens (model-ingestible format).
answer_start_scores, answer_end_scores = model(**inputs, return_dict=False) #Calculate Model Score
using AutoModelForQuestionAnswering
answer_start = torch.argmax(answer_start_scores) # get the most likely beginning of answer with the
argmax of the score
answer_end = torch.argmax(answer_end_scores) + 1 # get the most likely end of answer with the argmax
of the score
the_answer = tokenizer.convert_tokens_to_string(tokenizer.convert_ids_to_tokens(inputs["input_ids"])[0]
[answer_start:answer_end])) #get the answers
print('=====')
print('Question: ' + question)
print('Answer: ' + the_answer)
print('=====')
print('End') #print the answers
#Data sciencst are engineers, that's true!
```



Question: *What is data scientist?*
Answer: *engineers*

**Data scientists are engineers,
that's true!**

VALIDATED!



NEXT SENTENCE PREDICTION

NEXT SENTENCE PREDICTION

Can Bert help us to decide next sentence?

Bert can help us validate the answer

```
##NEXT SENTENCE PREDICTION##  
#Validation of the Bert algorithm's ability to recognise sentences  
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')  
model = BertForNextSentencePrediction.from_pretrained('bert-base-uncased')  
first_sentence = sent_tokenize(doc)[2] #34d sentence in jd1  
next_sentence1 = sent_tokenize(doc)[3] #4th sentence in jd  
next_sentence2 = sent_tokenize(doc)[103] #104th sentence in jd  
print("~~~~~")  
print(first_sentence)  
print("~~~~~")  
for next_sentence in [next_sentence1, next_sentence2]:  
    encoding = tokenizer.encode_plus(first_sentence, next_sentence, return_tensors='pt')  
    outputs = model(**encoding)[0]  
    softmax = F.softmax(outputs, dim=1)  
  
    if softmax[0][0] > softmax[0][1]:  
        print(next_sentence)  
        print("This sentence is the better following sentence")  
        print(softmax)  
        print("==== ===== === ")  
    else:  
        print(next_sentence)  
        print("This sentence is not the better following sentence")  
        print(softmax)  
        print("==== ===== === ")  
    print("End")  
#bert successfully chose the following sentences.
```

3rd "This rolebudget."

4th "This individual expertise. "

This sentence is the **better** following sentence

104th "Sure it's.....sentence."

This sentence is **not the better** following sentence

Bert successfully chose the correct sentence

VALIDATED!

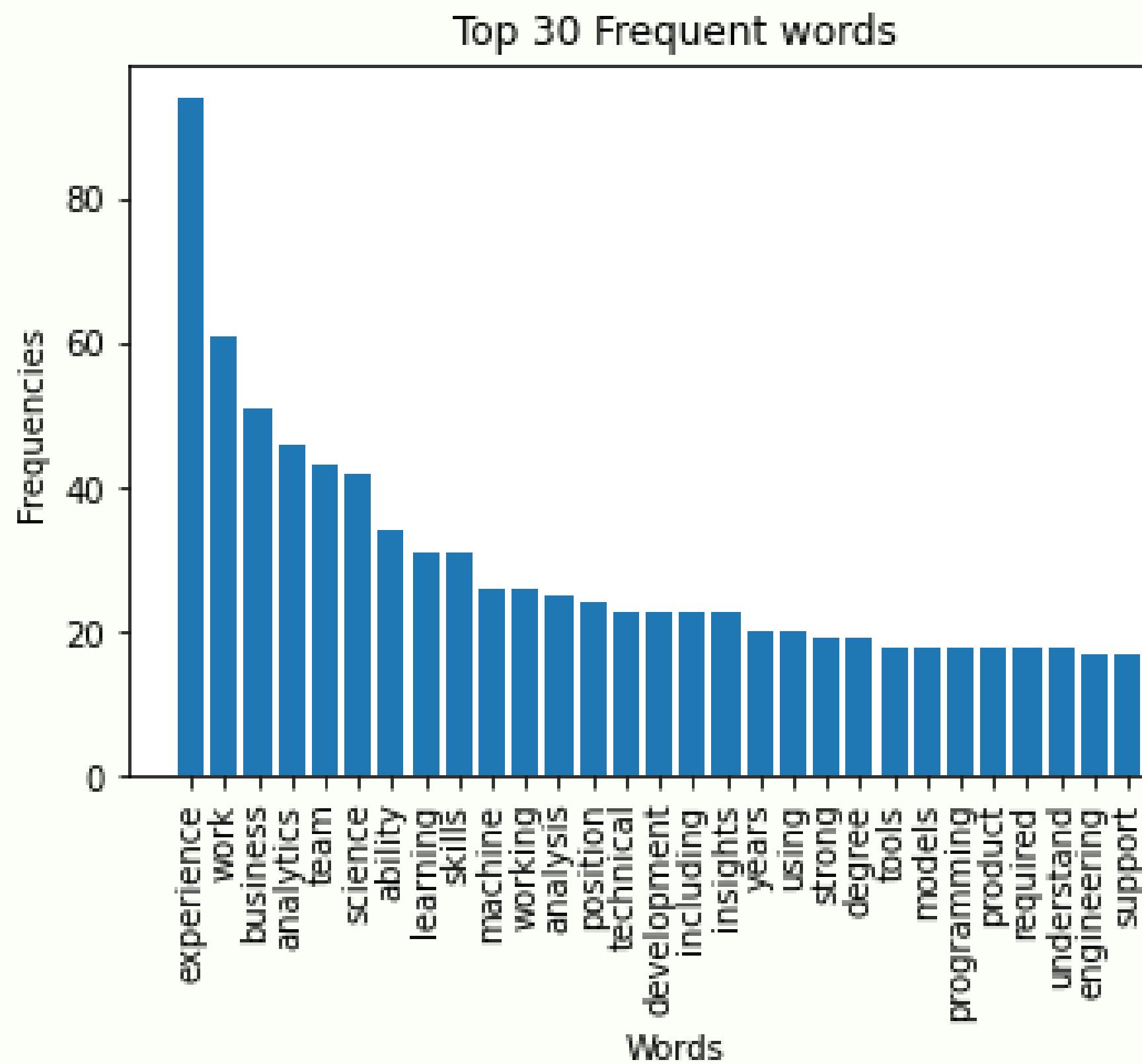


FINDING INSIGHTS

A large, stylized silhouette of a person's head and shoulders, facing right. The interior of the head is filled with a dense, glowing blue network of lines, dots, and small squares, resembling a circuit board or neural connections.

WORD FREQUENCIES

WORD FREQUENCIES-TOP 30 FREQUENT WORDS



```
Counter = Counter(docx) #input doc to counter
freq = Counter.most_common(75) #show most frequent 75 words
#freq = list(freq)
print(freq) #print frequency in list
#output barchart
list=[] #void list
i=0
c=range(0,30)
for i in c:
    if i <30:
        list += freq[i]
        i=i+1
    else:
        mostword = list[0:60:2] #get words
        frequency= list[1:60:2] #get frequencies
#plot frequencies in barplot
plt.bar(mostword, frequency)
plt.xticks(rotation=90)
plt.xlabel('Words')
plt.ylabel('Frequencies')
plt.title('Top 30 Frequent words')
plt.show()
```



KEYWORD EXTRACTION

KEYWORD EXTRACTION



Single word extraction

- [('scientists', 0.4431), ('teamwork', 0.4246), ('scientist', 0.4088), ('academic', 0.3588), ('engineering', 0.3569)]

Max Sum Similarity

- [('scientists develop highly', 0.5893), ('interdisciplinary engineering research', 0.5921), ('scientists engineers clinicians', 0.5948), ('strategy data scientist', 0.5974), ('science deep learning', 0.6082)]

Maximal Marginal Relevance

- [('scientists comprehensive analytical', 0.6565), ('new potential sponsorship', 0.4201), ('economics mathematics statistics', 0.3215), ('requests disability religious', 0.2311), ('marijuana 12 months', 0.1207)]

through KeyBERT with model

- [('scientists develop highly', 0.5893), ('interdisciplinary engineering research', 0.5921), ('scientists engineers clinicians', 0.5948), ('strategy data scientist', 0.5974), ('science deep learning', 0.6082)]



TEXT SUMMARIZATION

TEXT SUMMARIZATION-USING GENSIM

```
summary_Gensim = summarize(doc, ratio=0.01)  
print(summary_Gensim) #print Gensim
```

SUMMARIZATION OF JOB DESCRIPTION

*This position requires interactions with government, military, and industry officials nationwide for a variety of programs, projects, and tasks, including technical and programmatic concept development, planning, coordination, integration, and execution that can be supported by data science and deep learning techniques. Our team supports all phases of analytic work product development, from the identification of key business questions through data collection and **ETL**, and from performing analyses and using a wide range of statistical, **machine learning**, and applied mathematical techniques to delivery insights to decision-makers. Experience working with advanced analytics tools, and software such as Tableau, **Power BI** and **Python** as well as data languages and programming Advanced degree in Statistics, Data Mining, Machine Learning, Analytics, **Applied Math**, Computer Science, Electrical Engineering, Physics, or related fields. Strong interpersonal skills; the ability to understand business requirements and naturally explain complex technical topics to everyone — from data scientists to engineers to product marketing partners to executives. 3+ years of experience with data scientist, dashboard development, data analytics, statistics, or machine learning engineering*



CONCLUSION

CONCLUSION

- Text mining is an efficient way to generate insights from big data.
- The Bert tool is very efficient in NLP as validated in my JD case
- After data mining, I found companies are finding data scientists w/
 - Programming Language: Python, SQL, R
 - Skills: teamwork, analysis, experience, insights
 - Background: engineering, math

WHAT WILL I DO NEXT



Yuandi (Tyler) Tang
tyleryd@bu.edu | Boston, MA | 857-284-3248 | linkedin.com/in/yuanditang

Education
Boston University, College of Communication
MS in Marketing Communication Research with Data Analytics Graduate Certificate
Guangdong University of Foreign Studies, Faculty of English Language and Culture
BA in English Literature

Boston, MA
Expected May 2023
Guangzhou, China
Sep 2010 – Jun 2014

Professional Projects
CM724, Boston University College Students' Preference for soda purchasing (Dr Pepper)
Boston, MA
Jan 2022 – May 2022
•Conducted the qualitative and quantitative analysis through SPSS, Excel, and R by implementing model selections
•Analyzed the previous literature on soda marketing to figure out the correct sampling models and predictors to use
•Cleaned and classified data from self-reporting measurements by using red flag approaches to correct models
•Designed the research question and surveyed 300 questionnaires for sodas purchasing preferences
•Produced a 130-page final report with several findings including age is not a factor in soda consumption

Working Experience
Technical Consulting and Research Inc.
Marketing Communications Intern
Boston, MA
Jun 2022 – Aug 2022
•Collaborated with the team in various marketing campaigns, enhanced the user experience website traffic by 13%
•Conducted research and analysis on best practices of online marketing strategies
•Provided constructive suggestions to further TCR online presence using social media tools
•Utilized hashtags, photos, videos, articles, and contests to strengthen the program's presence

Hummingbird Marketing Agency Inc.
Marketing and Research Intern
Boston, MA
Feb 2022 – Aug 2022
•Organized media promotion kits for tech brands like Tron and Uoni, and improve the media contacts poll by 10%
•Increased a media pool by 40% through better qualitative labeling methods and qualitative analysis of coverages
•Delivered PowerPoint presentations with real-time data of ELLE and Grazia media kit on revenue of \$100,000
•Shared clients and executive board with their ROI performance with real-time dashboards in plain language

Talping Life Insurance Co., Ltd.
Marketing Manager
Hefei, China
Nov 2016 – May 2021
•Planned and organized events with trip.com for 1000+ people sales team and improved sales by \$1 million per year
•Wrote press releases and feature stories for 11 sales teams that reached an audience of 10,000+ people
•Analyzed sales data for 11 teams daily in real-time in Excel and Tableau for a visualized and interactive report
•Awarded consecutive excellency awards provincially for conducting excellent research projects in the whole province
•Designed data structure for 25 salesperson' performances by using data visualization and improved revenue by 25%
•Managed marketing research project by modeling sales and marketing data with CRM system

Shanghai U-tour International Tourism Agency Co., Ltd.
Marketing Specialist
Shanghai, China
May 2015 – Sep 2016
•Researched European seasonal travel routes through tourists surveys
•Improved satisfaction by 15% and decreased travel operation costs by 12%
•Established partnerships with major hospitality service suppliers including MIKI, GTA, and EH TRAVEL
•Organized regular meetings and festival events to maintain a cooperation renewal rate of 95%

Extracurricular Activities
Boston University International Graduate Student Guide
Initiator and Co-author
Boston, MA
Jun 2022 – Sep 2022
•Navigated challenges for International Graduate students to determine the editorial style and outline of the guide
•Recruited and led editorial team, wrote content and appendices, and organized review and proofreading
•Contacted all departments related to graduate students and international students at BU for promotion
•BU news report on the guide with integrated promotions including newsletter, facebook groups and Instagram
•Posted on all relevant departments' websites and social media, accumulated 10000+ exposures and 1000+ likes

Skills/Certificates
•Data Analytics Certificates: Google Data Analytics Certificate (May 2022)
•Research Technical Skills: SPSS, R, Python, SQL, Microsoft PowerBI, Tableau
•Marketing Technical Skills: Microsoft Office Suite; Adobe Photoshop, Illustrator, InDesign
•Marketing Management Certificates: ALMI issued by LOMA, the international insurance association
•Language Skills: Mandarin (Fluent), Cantonese (Intermediate)

- Mention the keywords found in the project more often in my resume
- Improve my programming skills according the high frequency of R/Python/SQL

Keep working on the right track

SUCCESS DREAM JOBS



Job Market

THANK YOU

Yuandi TANG U65674688

APPENDIX-DATA SOURCE

- <https://www.linkedin.com/jobs/search/?currentJobId=3359606365&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3363442868&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3367156013&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3342970208&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3360789017&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3359606365&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3356690656&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3365941184&keywords=data%20scientist&refresh=true>



APPENDIX-DATA SOURCE

- <https://www.linkedin.com/jobs/search/?currentJobId=3365941184&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3369274349&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3360750294&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3364565457&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3360763342&keywords=data%20scientist&refresh=true>
- <https://www.linkedin.com/jobs/search/?currentJobId=3358572872&keywords=data%20scientist&refresh=true&start=25>
- <https://www.linkedin.com/jobs/search/?currentJobId=3369157562&keywords=data%20scientist&refresh=true&start=25>
- <https://www.linkedin.com/jobs/search/?currentJobId=3360782629&keywords=data%20scientist&refresh=true&start=25>
- <https://www.linkedin.com/jobs/search/?currentJobId=3357035379&keywords=data%20scientist&refresh=true&start=25>



APPENDIX-CODE REFERENCES

- Course Materials-TP Start code
- <https://railsware.com/blog/python-for-machine-learning-indexing-and-slicing-for-lists-tuples-strings-and-other-sequential-types/>
- <https://www.mathworks.com/matlabcentral/answers/751809-removing-specific-numbers-from-matrix>

