

# resume-jd matching

April 11, 2024

```
[1]: from gensim.models.doc2vec import Doc2Vec, TaggedDocument
from nltk.tokenize import word_tokenize
from numpy.linalg import norm
from termcolor import colored
import pandas as pd
import numpy as np
import requests
import PyPDF2
import re
import plotly.graph_objects as go
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\lenovo\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

[1]: True

[ ]:

```
[2]: df = pd.read_csv('nyc-jobs-1.csv')
```

```
[3]: df.head()
```

```
[3]:
```

	Job ID	Agency Posting Type	# Of Positions	\
0	87990 DEPARTMENT OF BUSINESS SERV.	Internal	1	
1	97899 DEPARTMENT OF BUSINESS SERV.	Internal	1	
2	102221 DEPT OF ENVIRONMENT PROTECTION	External	1	
3	102221 DEPT OF ENVIRONMENT PROTECTION	Internal	1	
4	114352 DEPT OF ENVIRONMENT PROTECTION	Internal	5	

	Business Title	Civil Service Title	\
0	Account Manager	CONTRACT REVIEWER (OFFICE OF L	
1	EXECUTIVE DIRECTOR, BUSINESS DEVELOPMENT	ADMINISTRATIVE BUSINESS PROMOT	
2	Project Specialist	ENVIRONMENTAL ENGINEERING INTE	
3	Project Specialist	ENVIRONMENTAL ENGINEERING INTE	
4	Deputy Plant Chief	SENIOR STATIONARY ENGINEER (EL	

	Title	Code No	Level	Job Category	Full-Time/Part-Time indicator	...	\
0		40563	1			...	
1		10009	M3		F	...	
2		20616	0		F	...	
3		20616	0		F	...	
4		91639	0		F	...	

	Additional Information		\
0	Salary range for this position is: \$42,405 - \$...		
1			
2	Appointments are subject to OMB approval		
3	Appointments are subject to OMB approval		
4	Appointments are subject to OMB approval Fo...		

	To Apply	Hours/Shift	\
0			
1	In addition to applying through this website, ...		
2	click the apply now button	35 hours per week/day	
3	click the apply now button	35 hours per week/day	
4	Click "Apply Now" button	40 per week / day	

	Work Location 1	Recruitment Contact	\
0		NaN	
1		NaN	
2		NaN	
3		NaN	
4	Various	NaN	

	Residency Requirement	Posting Date	\
0	New York City residency is generally required ...	2011-06-24T00:00:00	
1	New York City residency is generally required ...	2012-01-26T00:00:00	
2	New York City Residency is not required for th...	2012-06-21T00:00:00	
3	New York City Residency is not required for th...	2012-06-21T00:00:00	
4	New York City residency is generally required ...	2012-12-12T00:00:00	

	Post Until	Posting Updated	Process Date
0	NaN	2011-06-24T00:00:00	2018-07-17T00:00:00
1	NaN	2012-01-26T00:00:00	2018-07-17T00:00:00
2	NaN	2012-09-07T00:00:00	2018-07-17T00:00:00
3	NaN	2012-09-07T00:00:00	2018-07-17T00:00:00
4	NaN	2012-12-13T00:00:00	2018-07-17T00:00:00

[5 rows x 28 columns]

```
[12]: df.columns
```

```
[12]: Index(['Job ID', 'Agency', 'Posting Type', '# Of Positions',
        'Civil Service Title', 'Title Code No', 'Level', 'Job Category',
        'Full-Time/Part-Time indicator', 'Salary Range From', 'Salary Range To',
        'Salary Frequency', 'Work Location', 'Division/Work Unit',
        'Additional Information', 'To Apply', 'Hours/Shift', 'Work Location 1',
        'Recruitment Contact', 'Residency Requirement', 'Posting Date',
        'Post Until', 'Posting Updated', 'Process Date', 'data'],
        dtype='object')
```

```
[13]: df =pd.read_csv('nyc-jobs-1.csv')
```

```
[14]: df[['Business Title', 'Job Description', 'Minimum Qual Requirements',
        ↪'Preferred Skills']]
df=df.head(500)
df
```

```
[14]:
```

	Job ID	Agency	Posting Type	# Of Positions	\
0	87990	DEPARTMENT OF BUSINESS SERV.	Internal	1	
1	97899	DEPARTMENT OF BUSINESS SERV.	Internal	1	
2	102221	DEPT OF ENVIRONMENT PROTECTION	External	1	
3	102221	DEPT OF ENVIRONMENT PROTECTION	Internal	1	
4	114352	DEPT OF ENVIRONMENT PROTECTION	Internal	5	
..	...	...	...	...	
495	285409	DEPT OF ENVIRONMENT PROTECTION	External	1	
496	285451	HRA/DEPT OF SOCIAL SERVICES	Internal	1	
497	285451	HRA/DEPT OF SOCIAL SERVICES	External	1	
498	285463	POLICE DEPARTMENT	Internal	1	
499	285539	DEPARTMENT OF TRANSPORTATION	External	1	

	Business Title	\
0	Account Manager	
1	EXECUTIVE DIRECTOR, BUSINESS DEVELOPMENT	
2	Project Specialist	
3	Project Specialist	
4	Deputy Plant Chief	
..	...	
495	Civil Engineer 2	
496	ARCHITECT	
497	ARCHITECT	
498	Deputy Director of Support and Special Service...	
499	Civil Engineer 3	

	Civil Service Title	Title Code No	Level	\
0	CONTRACT REVIEWER (OFFICE OF L	40563	1	
1	ADMINISTRATIVE BUSINESS PROMOT	10009	M3	
2	ENVIRONMENTAL ENGINEERING INTE	20616	0	
3	ENVIRONMENTAL ENGINEERING INTE	20616	0	

4	SENIOR STATIONARY ENGINEER (EL	91639	0
..	...	...	...
495	CIVIL ENGINEER	20215	2
496	ARCHITECT	21215	1
497	ARCHITECT	21215	1
498	ASSOC SUPVR OF SCHL SEC (MGRL)	6082A	M2
499	CIVIL ENGINEER	20215	3

Job Category \

0	
1	
2	
3	
4	
..	...
495	Engineering, Architecture, & Planning
496	Engineering, Architecture, & Planning
497	Engineering, Architecture, & Planning
498	Administration & Human Resources Public Safety...
499	Engineering, Architecture, & Planning

Full-Time/Part-Time indicator ... \

0	...
1	F ...
2	F ...
3	F ...
4	F ...
..	... ..
495	F ...
496	F ...
497	F ...
498	F ...
499	F ...

Additional Information \

0	Salary range for this position is: \$42,405 - \$...
1	
2	Appointments are subject to OMB approval
3	Appointments are subject to OMB approval
4	Appointments are subject to OMB approval Fo...
..	...
495	Appointments are subject to OMB approval. For ...
496	
497	
498	
499	TO BE APPOINTED TO ANY CIVIL ENGINEERING POSIT...

To Apply \

0

1 In addition to applying through this website, ...

2 click the apply now button

3 click the apply now button

4 Click "Apply Now" button

..

495 Click "Apply Now" button.

496 Click "Apply Now" Button

497 Click "Apply Now" Button

498 Please submit your resume and cover letter. Pl...

499 Resumes may be submitted using the following m...

	Hours/Shift	Work Location 1 \
0		
1		
2	35 hours per week/day	
3	35 hours per week/day	
4	40 per week / day	Various
..	...	...
495	35 hours per week / day	96-05 Horace Harding Expway Corona, NY 11368
496		
497		
498		28-11 Queens Plaza No., L.I.C.
499	35hrs	55 Water Street, NY, NY

	Recruitment Contact	Residency Requirement \
0	NaN	New York City residency is generally required ...
1	NaN	New York City residency is generally required ...
2	NaN	New York City Residency is not required for th...
3	NaN	New York City Residency is not required for th...
4	NaN	New York City residency is generally required ...
..	...	...
495	NaN	New York City Residency is not required for th...
496	NaN	New York City Residency is not required for th...
497	NaN	New York City Residency is not required for th...
498	NaN	New York City residency is generally required ...
499	NaN	New York City Residency is not required for th...

	Posting Date	Post Until	Posting Updated	Process Date
0	2011-06-24T00:00:00	NaN	2011-06-24T00:00:00	2018-07-17T00:00:00
1	2012-01-26T00:00:00	NaN	2012-01-26T00:00:00	2018-07-17T00:00:00
2	2012-06-21T00:00:00	NaN	2012-09-07T00:00:00	2018-07-17T00:00:00
3	2012-06-21T00:00:00	NaN	2012-09-07T00:00:00	2018-07-17T00:00:00
4	2012-12-12T00:00:00	NaN	2012-12-13T00:00:00	2018-07-17T00:00:00
..	...	...	...	...
495	2017-03-24T00:00:00	NaN	2017-03-24T00:00:00	2018-07-17T00:00:00

496	2017-11-22T00:00:00	NaN	2018-07-10T00:00:00	2018-07-17T00:00:00
497	2017-11-22T00:00:00	NaN	2018-07-10T00:00:00	2018-07-17T00:00:00
498	2017-04-13T00:00:00	NaN	2017-05-31T00:00:00	2018-07-17T00:00:00
499	2017-04-01T00:00:00	NaN	2017-04-04T00:00:00	2018-07-17T00:00:00

[500 rows x 28 columns]

```
[15]: df['data'] = df[['Business Title', 'Job Description', 'Minimum Qualifications', 'Preferred Skills']].apply(lambda x: ' '.join(x.dropna().astype(str)), axis=1)

df.drop(['Business Title', 'Job Description', 'Minimum Qualifications', 'Preferred Skills'], axis=1, inplace=True)

print(df.head())
```

	Job ID	Agency Posting Type	# Of Positions	\
0	87990	DEPARTMENT OF BUSINESS SERV.	Internal	1
1	97899	DEPARTMENT OF BUSINESS SERV.	Internal	1
2	102221	DEPT OF ENVIRONMENT PROTECTION	External	1
3	102221	DEPT OF ENVIRONMENT PROTECTION	Internal	1
4	114352	DEPT OF ENVIRONMENT PROTECTION	Internal	5

	Civil Service Title	Title Code	No Level	Job Category	\
0	CONTRACT REVIEWER (OFFICE OF L	40563	1		
1	ADMINISTRATIVE BUSINESS PROMOT	10009	M3		
2	ENVIRONMENTAL ENGINEERING INTE	20616	0		
3	ENVIRONMENTAL ENGINEERING INTE	20616	0		
4	SENIOR STATIONARY ENGINEER (EL	91639	0		

	Full-Time/Part-Time indicator	Salary Range From	...	\
0		42405	...	
1	F	60740	...	
2	F	43349	...	
3	F	43349	...	
4	F	109850	...	

	To Apply	Hours/Shift	\
0			
1	In addition to applying through this website, ...		
2	click the apply now button	35 hours per week/day	
3	click the apply now button	35 hours per week/day	
4	Click "Apply Now" button	40 per week / day	

	Work Location 1 Recruitment Contact	\
0	NaN	
1	NaN	
2	NaN	

3		NaN
4	Various	NaN

	Residency Requirement	Posting Date \
0	New York City residency is generally required ...	2011-06-24T00:00:00
1	New York City residency is generally required ...	2012-01-26T00:00:00
2	New York City Residency is not required for th...	2012-06-21T00:00:00
3	New York City Residency is not required for th...	2012-06-21T00:00:00
4	New York City residency is generally required ...	2012-12-12T00:00:00

	Post Until	Posting Updated	Process Date \
0	NaN	2011-06-24T00:00:00	2018-07-17T00:00:00
1	NaN	2012-01-26T00:00:00	2018-07-17T00:00:00
2	NaN	2012-09-07T00:00:00	2018-07-17T00:00:00
3	NaN	2012-09-07T00:00:00	2018-07-17T00:00:00
4	NaN	2012-12-13T00:00:00	2018-07-17T00:00:00

data

0	Account Manager Division of Economic & Financi...
1	EXECUTIVE DIRECTOR, BUSINESS DEVELOPMENT The N...
2	Project Specialist Under direct supervision, p...
3	Project Specialist Under direct supervision, p...
4	Deputy Plant Chief Under general direction, is...

[5 rows x 25 columns]

```
[16]: data = list(df['data'])
tagged_data = [TaggedDocument(words = word_tokenize(_d.lower()), tags =_
↳[str(i)]) for i, _d in enumerate(data)]
```

```
[17]: model = Doc2Vec(vector_size = 50,
min_count = 5,
epochs = 50,
alpha = 0.001
)
```

```
[18]: model.build_vocab(tagged_data)
```

```
[19]: keys = model.wv.key_to_index.keys()
```

```
[20]: print(len(keys))
```

3201

```
[21]: for epoch in range(model.epochs):
print(f"Training epoch {epoch+1}/{model.epochs}")
model.train(tagged_data,
total_examples=model.corpus_count,
```

```
epochs=model.epochs)

model.save('cv_job_maching.model')
print("Model saved")
```

```
Training epoch 1/50
Training epoch 2/50
Training epoch 3/50
Training epoch 4/50
Training epoch 5/50
Training epoch 6/50
Training epoch 7/50
Training epoch 8/50
Training epoch 9/50
Training epoch 10/50
Training epoch 11/50
Training epoch 12/50
Training epoch 13/50
Training epoch 14/50
Training epoch 15/50
Training epoch 16/50
Training epoch 17/50
Training epoch 18/50
Training epoch 19/50
Training epoch 20/50
Training epoch 21/50
Training epoch 22/50
Training epoch 23/50
Training epoch 24/50
Training epoch 25/50
Training epoch 26/50
Training epoch 27/50
Training epoch 28/50
Training epoch 29/50
Training epoch 30/50
Training epoch 31/50
Training epoch 32/50
Training epoch 33/50
Training epoch 34/50
Training epoch 35/50
Training epoch 36/50
Training epoch 37/50
Training epoch 38/50
Training epoch 39/50
Training epoch 40/50
Training epoch 41/50
Training epoch 42/50
Training epoch 43/50
```



Training epoch 44/50  
Training epoch 45/50  
Training epoch 46/50  
Training epoch 47/50  
Training epoch 48/50  
Training epoch 49/50  
Training epoch 50/50  
Model saved

```
[24]: pdf = PyPDF2.PdfReader('Prakyath Peters.pdf')
resume = ""
for i in range(len(pdf.pages)):
    pageObj = pdf.pages[i]
    resume += pageObj.extract_text()
```

```
[26]: jd = """
Proactively engage with clients in the branch and deliver an outstanding_
    ↳service experience through completing transactions and by identifying_
    ↳opportunities for advice, solutions, digital enablement and partner_
    ↳introductions
Focus on education and demonstration, leverage technology to deliver a_
    ↳memorable client experience, drive solutions and retain business
Contribute to team results by listening and spotting opportunities to offer_
    ↳additional advice, introduce clients to the capability of RBC partners, or_
    ↳personally fulfil client solutions
Proactively take ownership of resolving and preventing client banking problems
Cultivate and maintain relationships with partners to work as one RBC team
Manage risks by adhering to compliance routines, processes, and controls to_
    ↳protect client and shareholder interests while completing transactions
What do you need to succeed?

Must-have

Goal-oriented individual with a demonstrated passion for putting clients first.
Drive and self-motivation, as well as excellent communication skills and_
    ↳emotional intelligence
Digital literacy across a broad range of devices (i.e., smartphones, tablets,_
    ↳laptops, etc.)
Personal flexibility to work flex hours
Eagerness to learn and determination to succeed
Confidence and ability to learn financial concepts and willingness to obtain_
    ↳the Investment Funds in Canada or the Canadian Securities Course
Nice-to-have

Track record in building rapport and maintaining client relationships within_
    ↳the financial, service or retail industry
Mutual Funds accreditation

```

Is this job right for you? Check out our video and decide for yourself!

What's in it for you?

We thrive on the challenge to be our best, progressive thinking to keep  
↳growing, and working together to deliver trusted advice to help our clients  
↳thrive and communities prosper. We care about each other, reaching our  
↳potential, making a difference to our communities, and achieving success  
↳that is mutual.

A comprehensive Total Rewards Program including bonuses and flexible benefits,  
↳competitive compensation, commissions, and stock where applicable

A world-class training program in financial services

Excellent career development and access to a variety of job opportunities  
↳across business and geographies

Leaders who support your development through coaching and managing opportunities

Work in a dynamic, collaborative, progressive, and high-performing team

We also strive to provide an accessible candidate experience for our  
↳prospective employees with different abilities. Please let us know if you  
↳need any accommodations during the recruitment process.

Join our Talent Community

Stay in-the-know about great career opportunities at RBC. Sign up and get  
↳customized info on our latest jobs, career tips and Recruitment events that  
↳matter to you.

Expand your limits and create a new future together at RBC. Find out how we use  
↳our passion and drive to enhance the well-being of our clients and  
↳communities at [rbc.com/careers](http://rbc.com/careers).

""

```
[27]: def preprocess_text(text):  
  
    text = text.lower()  
  
    text = re.sub('[^a-z]', ' ', text)  
  
    text = re.sub(r'\d+', '', text)  
  
    text = ' '.join(text.split())  
  
    return text
```

```
[28]: input_CV = preprocess_text(resume)  
input_JD = preprocess_text(jd)
```

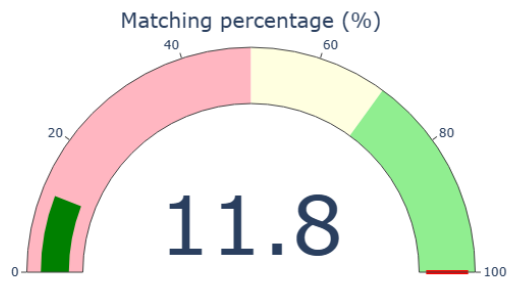
```
[29]: model = Doc2Vec.load('cv_job_maching.model')
v1 = model.infer_vector(input_CV.split())
v2 = model.infer_vector(input_JD.split())
similarity = 100*(np.dot(np.array(v1), np.array(v2))) / (norm(np.array(v1)) *
    ↪norm(np.array(v2)))
print(round(similarity, 2))
```

11.75

```
[30]: fig = go.Figure(go.Indicator(
    domain = {'x': [0, 1], 'y': [0, 1]},
    value = similarity,
    mode = "gauge+number",
    title = {'text': "Matching percentage (%)"},
    #delta = {'reference': 100},
    gauge = {
        'axis': {'range': [0, 100]},
        'steps' : [
            {'range': [0, 50], 'color': "#FFB6C1"},
            {'range': [50, 70], 'color': "#FFFFE0"},
            {'range': [70, 100], 'color': "#90EE90"}
        ],
        'threshold' : {'line': {'color': "red", 'width': 4}, 'thickness':
    ↪0.75, 'value': 100}}))

fig.update_layout(width=600, height=400) # Adjust the width and height as
    ↪desired
fig.show()

# Print notification
if similarity < 50:
    print(colored("Low chance, need to modify your CV!", "red", attrs=["bold"]))
elif similarity >= 50 and similarity < 70:
    print(colored("Good chance but you can improve further!", "yellow",
    ↪attrs=["bold"]))
else:
    print(colored("Excellent! You can submit your CV.", "green",
    ↪attrs=["bold"]))
```



Low chance, need to modify your CV!

[ ]: