In [1]:
```python
pip install docx2txt
```

Requirement already satisfied: docx2txt in c:\users\acer\anaconda3\lib\site-packages (0.8)
Note: you may need to restart the kernel to use updated packages.

In [2]:
```python
import docx2txt
import warnings
warnings.filterwarnings('ignore')
```

In [3]:
```python
# Load the data
resume = docx2txt.process('python-sample-resume.docx')
job = docx2txt.process('python-job-description.docx')
```

In [4]:
```python
1  # printing the resume
2  resume
```

Out[4]: 'Loren Brekke \n\n9873 Ignacio Corners Boston \n\nMA \n\nPhone +1 (555) 570 0033 \n\n\nExperience \n\nHouston, TX \n\nPython Applications Developer \n\n10/2015 – present \n\n\tHelp set development standards in areas of best practices, code quality, test coverage, monitoring, logging, exception management, and others\n\n\tBecome proficient in Python, and the Athena platform's NoSql OO database, graph-oriented development model and developer toolset\n\n\tPerform unit testing and system integration testing of the newly developed functionality\n\n\tCreate technical design recommendations for developing and integrating programs per written specifications\n\n\tConduct code reviews on components and applications to ensure adherence to the development standards and best practices\n\n\tImprove SDLC\n\n\tIntegrate various systems to improve efficiency\n\nLos Angeles, CA \n\nPython Application Developer \n\n02/2013 – 08/2015 \n\n\tExecute change management activities supporting production deployment to Developers, Quality Control Analysts, and Environment Management personnel\n\n\tKnowledge of software development life cycle and agile development practices\n\n\tStrong knowledge of Spring frameworks and leveraging distributed data caching frameworks\n\n\tProvide guidance and oversight to development teams across Capital technology\n\n\tWorking with production support teams to ensure smooth transition of software developed\n\n\tDevelop code for intermediate to complex modules, following documentation and development standards\n\n\tWork with the development team to facilitate the Build effort and to ensure successful end-to-End integration / testing\n\nChicago, IL \n\nPython \n\n12/2005 – 09/2012 \n\n\tWork with production support teams to ensure smooth transition of software developed\n\n\tWork in an environment which encourages and promotes personal development\n\n\tProvide design recommendations, developing and integrating programs per high level specifications\n\n\tLiaise with other teams and providers\n\n\tBe an active part of a larger internal open source community of developers at Hearst\n\n\tDeveloping a world-class, machine learning-driven recommendation and personalization system to match our users with jobs\n\n\tUnderstand the project scope, identify activities/ tasks, task level estimates, schedule, dependencies, risks and provide inputs to Module Lead for review\n\nEducation \n\nBachelor's Degree in Computer Science \n\nKent State University \n\nSkills \n\n\tFront-office/risk management business knowledge as well as Knowledge of financial instrument features and trade processing would be a definite \n\n\tStrong analytical ability \n\n\tAble to react quickly and perform well under stress and high demand \n\n\tWorking Excel knowledge including VBA, pivot tables, formulas \n\n\tProven strong experience with databases (columnar, nosql and RDMS) and good SQL skills \n\n\tWillingness to take ownership of tasks and the ability to work on their own initiative while being a strong team player \n\n\tExcellent verbal and written communication skills, able to communicate accurately, concisely and with tact and diplomacy when appropriate \n\n\tDevelop a good understanding of End to End business processes and the associated technical blocks \n\n\tEquities business knowledge is an advantage but other relevant experience in other business areas is welcomed \n\n\tThe ability to work under pressure within agreed timelines and to support multiple tasks in parallel'

In [5]:
```python
#print the job description
job

```

Out[5]: 'Python Developer responsibilities include:\n\n\tWriting effective, scalable code \n\n\tDeveloping back-end components to improve responsiveness and over all performance \n\n\tIntegrating user-facing elements into applications \n\nJob brief\n\nWe are looking for a Python Developer to join our engineering team and help us develop and maintain various software products.\n\nPython D eveloper responsibilities include writing and testing code, debugging progra ms and integrating applications with third-party web services. To be success ful in this role, you should have experience using server-side logic and wor k well in a team.\n\nUltimately, you'll build highly responsive web applicat ions that align with our business needs.\n\nResponsibilities\n\n\tWrite effe ctive, scalable code \n\n\tDevelop back-end components to improve responsive ness and overall performance \n\n\tIntegrate user-facing elements into appli cations \n\n\tTest and debug programs \n\n\tImprove functionality of existin g systems \n\n\tImplement security and data protection solutions \n\n\tAsses s and prioritize feature requests \n\n\tCoordinate with internal teams to un derstand user requirements and provide technical solutions \n\nRequirements \n\n\tWork experience as a Python Developer \n\n\tExpertise in at least one popular Python framework (like Django, Flask or Pyramid) \n\n\tKnowledge of object-relational mapping (ORM) \n\n\tFamiliarity with front-end technologie s (like JavaScript and HTML5) \n\n\tTeam spirit \n\n\tGood problem-solving s kills \n\n\tBSc in Computer Science, Engineering or relevant field'

In [6]:
```python
# a list of text

text = [resume,job]

```

In [7]:
```python
from sklearn.feature_extraction.text import CountVectorizer

cv = CountVectorizer()
count_matrix = cv.fit_transform(text)
```

In [8]:
```python
# import cosine similarity

from sklearn.metrics.pairwise import cosine_similarity

#print similarity score

print('Similarity score : ',cosine_similarity(count_matrix))
```

```
Similarity score :  [[1.          0.63676947]
 [0.63676947 1.          ]]
```

In [9]:
```python
# get the percentage
matchpercentage = cosine_similarity(count_matrix)[0][1]
matchpercentage = round(matchpercentage*100,2)
print('Your Resume {} % match to the job description !'.format(matchperce
```

```
Your Resume 63.68 % match to the job description !
```

```
In [10]:    1  from sklearn.feature_extraction.text import CountVectorizer
            2  from sklearn.metrics.pairwise import cosine_similarity
            3  from sklearn import __version__ as sklearn_version
```

```
In [11]:    1  pip install joblib
            2
```

Requirement already satisfied: joblib in c:\users\acer\anaconda3\lib\site-pa
ckages (1.1.0)
Note: you may need to restart the kernel to use updated packages.

```
In [*]:     1  print('\nSimilarity score: {}%'.format(matchpercentage))
            2  print('Your Resume is {}% match to the job description!'.format(matchperc
            3
            4  # Save the model using joblib (No need to import joblib separately)
            5  if sklearn_version >= '0.24.0':
            6      from sklearn import set_config
            7      set_config(working_memory=False)  # This may be needed for compatibil
            8
            9      # Save the model
           10      cv_filename = 'model.pkl'
           11      joblib.dump(cv, cv_filename)
           12      print(f'Model saved as {cv_filename}')
           13  else:
           14      print('Please upgrade scikit-learn to version 0.24.0 or newer to use
```

```
In [ ]:     1
```