

In []:

```
# Importing the necessary modules.

import pandas as pd
import tensorflow as tf
tf.__version__
from tensorflow.keras.layers import Embedding
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.preprocessing.text import one_hot
from tensorflow.keras.layers import LSTM
from tensorflow.keras.layers import Dense
import nltk
import re
from nltk.corpus import stopwords
nltk.download('stopwords')
from nltk.stem.porter import PorterStemmer
import numpy as np
from sklearn.model_selection import train_test_split
from tensorflow.keras.layers import Dropout
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score

# Importing the labeled dataset.
Labelled_data=pd.read_csv('data_to_label.csv')

Labelled_data.head()
Labelled_data=Labelled_data.dropna()

# Get the Independent Features
X=Labelled_data.drop('Skill',axis=1)

# Get the Dependent features
y=Labelled_data['Skill']
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

In []:

```
# Setting the Vocabulary size
voc_size=100000
PotentialWords=X.copy()
PotentialWords.reset_index(inplace=True)
```

In []:

```
# Building the corpus after removing stop words in each sentence in the job description.

ps = PorterStemmer()
corpus = []
for i in range(len(PotentialWords)):
    PotentialWord = re.sub('[^a-zA-Z]', ' ', PotentialWords['Word'][i])
    PotentialWord = PotentialWord.lower()
    PotentialWord = PotentialWord.split()

    PotentialWord = [ps.stem(word) for word in PotentialWord if not word in stopwords.words('english')]
    PotentialWord = ' '.join(PotentialWord)
    corpus.append(PotentialWord)

print(len(corpus))
print(corpus)
```

```
1674
['look', 'selenium', 'engin', '', 'must', 'solid', 'java', 'code', 'skill', 'sever', 'ope
n', 'month', 'hire', '', 'must', 'abl', 'go', 'perm', 'someon', '', 'year', 'experi', '',
'import', 'qualiti', 'eager', 'aptitud', '', 'posit', 'purpos', '', 'want', 'forefront',
```

'cut', 'edg', 'technolog', '', 'introduc', 'solut', 'problem', '', 'exist', '', 'abil', 'see', 'result', 'success', '', 'client', 'assur', 'growth', 'collabor', 'develop', 'team', 'creat', 'tool', 'aid', 'engin', 'build', '', 'test', '', 'debug', '', 'releas', 'softwar', '', 'touch', 'million', 'user', 'increas', 'rate', 'develop', 'ensur', 'product', 'method', 'test', '', 'expert', 'softwar', 'health', '', 'testabl', '', 'sustain', '', 'softwar', 'engin', 'test', 'client', '', 'expect', 'build', 'flexibl', 'scalabl', 'solut', 'work', 'complex', 'challeng', 'larg scale', 'comput', 'util', 'skill', 'data', 'structur', 'object', 'orient', 'program', '', 'major', 'task', '', 'respons', 'key', 'account', '', 'lead', 'contribut', 'engin', 'effort', 'plan', 'execut', 'deliveri', 'solv', 'complex', 'engin', 'problem', 'tool', 'test', '', 'design', 'build', 'advanc', 'autom', 'test', 'framework', '', 'build', 'tool', 'help', 'develop', 'measur', 'increas', 'veloc', '', 'adopt', 'best', 'practic', 'softwar', 'health', '', 'test', '', 'sustain', '', 'analyz', 'break', 'complex', 'softwar', 'system', 'collabor', 'develop', 'team', 'improv', 'overall', 'design', '', 'natur', 'scope', '', 'typic', 'report', 'manag', 'test', 'engin', '', 'associ', 'report', 'person', 'role', 'perman', 'basi', '', 'analyz', 'resolv', 'complex', 'qualiti', 'issu', 'critic', 'solut', 'deliveri', '', 'fast pace', 'environ', 'freedom', 'respons', 'prefer', 'qualif', '', 'hand', 'java', 'develop', 'experi', '', 'profici', 'ab', 'initio', 'etl', 'batch', 'test', '', 'hand', 'knowledg', 'soap', 'ui', 'itko', 'lisa', '', 'selenium', '', 'hand', 'experi', 'unix', '', 'ksh', 'bourn', 'shell', 'script', '', 'strong', 'script', 'knowledg', 'java', '', 'groovi', 'jmeter', 'strong', 'develop', 'experi', 'work', 'web', 'servic', 'rest', 'soap', 'experi', 'build', 'test', 'autom', 'framework', 'web', 'servic', 'use', 'soap', 'ui', 'groovi', 'lisa', 'profici', 'sql', 'work', 'knowledg', 'svn', 'jenkin', 'build', 'develop', 'process', 'ecommerc retail', 'qa', 'experi', 'excel', 'written', 'verbal', 'commun', 'skill', 'self motiv', 'strong', 'work', 'ethic', 'posit', 'attitud', 'desir', 'work', 'fast pace', 'result', 'orient', 'team', 'experi', 'script', 'languag', 'perl', 'passion', 'understand', 'learn', 'dissect', 'improv', 'new', 'technolog', 'profici', 'one', 'follow', 'retail', 'applic', 'window', 'platform', 'unix', 'platform', 'java', 'relat', 'databas', 'webspher', 'applic', 'server', 'webspher', 'commerc', 'webspher', 'mq', 'comprehens', 'understand', 'system', 'applic', 'network', 'knowledg', 'skill', 'abil', 'compet', 'profici', 'one', 'follow', 'retail', 'applic', 'window', 'platform', 'unix', 'platform', 'mv', 'cobol', 'java', 'j', 'ee', 'relat', 'databas', 'webspher', 'applic', 'server', 'webspher', 'commerc', 'webspher', 'mq', 'network', 'voic', 'lan', 'wan', 'sap', 'siebel', 'peoplesoft', 'comprehens', 'understand', 'system', 'applic', 'network', 'strong', 'problem', 'solv', 'analyt', 'skill', 'strong', 'oral', 'written', 'commun', 'skill', 'abil', 'coordin', 'collabor', 'across', 'cross function', 'team', 'univers', 'chicago', 'rapidli', 'grow', 'secur', 'program', 'mani', 'opportun', 'motiv', 'person', 'work', 'varieti', 'role', 'mani', 'technic', 'area', 'posit', 'respons', 'provid', 'incid', 'respons', 'inform', 'secur', 'assess', 'monitor', 'system', 'administr', 'gener', 'direct', 'inform', 'secur', 'offic', 'lead', 'inform', 'secur', 'engin', 'essenti', 'function', 'execut', 'secur', 'network', 'monitor', 'incid', 'respons', 'procedur', 'lead', 'particip', 'secur', 'project', 'administ', 'secur', 'system', 'includ', 'updat', 'backup', 'upgrad', 'consult', 'advis', 'provid', 'secur', 'assess', 'servic', 'applic', 'develop', 'acquisit', 'project', 'assess', 'secur', 'requir', 'control', 'ensur', 'secur', 'control', 'implement', 'plan', 'research', 'secur', 'issu', 'contribut', 'secur', 'commun', 'web', 'page', 'blog', 'awar', 'materi', 'document', 'intern', 'process', 'write', 'secur', 'standard', 'guidelin', 'duti', 'assign', 'qualif', 'bachelor', 'degre', 'minimum', 'four', 'year', 'experi', 'inform', 'technolog', 'field', 'equival', 'combin', 'educ', 'experi', 'requir', 'minimum', 'two', 'year', 'profession', 'experi', 'technic', 'inform', 'secur', 'role', 'requir', 'experi', 'administ', 'linux unix', 'system', 'requir', 'experi', 'administ', 'window', 'system', 'prefer', 'experi', 'program', 'script', 'languag', 'perl', 'python', 'strongli', 'prefer', 'univers', 'chicago', 'affirm', 'action equal', 'opportun disabl veteran', 'employ', 'discrimin', 'basi', 'race', 'color', 'religion', 'sex', 'sexual', 'orient', 'gender', 'ident', 'nation', 'ethnic', 'origin', 'age', 'statu', 'individu', 'disabl', 'protect', 'veteran', 'statu', 'genet', 'inform', 'protect', 'class', 'law', 'addit', 'inform', 'pleas', 'see', 'univers', 'notic', 'nondiscrimin staff', 'job', 'seeker', 'need', 'reason', 'accommod', 'complet', 'applic', 'process', 'call', 'email', 'talentacquisit', 'uchicago edu', 'request', 'galax solutionseveri', 'day', 'solut', 'affect', 'peopl', 'throughout', 'world', 'fortun', 'compani', 'start up', 'galax', 'develop', 'implement', 'strateg', 'project', 'critic', 'success', 'custom', 'busi', 'live', 'ten', 'million', 'peopl', 'twenty five', 'year', 'grown', 'evolv', 'multi nation', 'firm', 'employ', 'team', 'member', 'worldwid', 'done', 'evolv', 'took', 'collabor', 'innov', 'get', 'take', 'collabor', 'innov', 'get', 'custom', 'want', 'tomorrow', 'mean', 'employe', 'secur', 'establish', 'compani', 'benefit', 'work', 'compani', 'great', 'mind', 'hard', 'work', 'leadership', 'innov', 'highli', 'regard', 'reward', 'thoma', 'edison', 'said', 'way', 'better', 'find', 'want', 'employe', 'find', 'look', 'creativ', 'peopl', 'entrepreneuri', 'spirit', 'look', 'work', 'awesom', 'project', 'sound'

, 'like', '', 'come', 'work', 'us', '', 'find', 'mean', 'part', 'galax', 'team', '', '', 'alway', 'easi', '', 'import', 'work', 'never', '', '', 'wearegalax', 'develop', 'practic', 'roadmap', 'enterpris wide', 'bi', 'report', 'analyt', 'platform', 'consider', 'busi', 'object', '', 'technolog', 'consider', 'budget', '', 'defin', 'overall', 'bi', 'data', 'ar chitectur', 'could', 'includ', '', 'etl', 'process', '', 'od', '', 'edw', '', 'data', 'ma rt', 'data', 'lake provid', 'bi', 'technic', 'architectur', 'guidanc', 'project', 'projec t', 'team', 'ensur', 'new', 'initi', 'enabl', 'effect', 'data', 'report develop', 'new', 'report', 'solut', 'requiredidenti', '', 'evalu', '', 'provid', 'propos', 'solut', 'ide ntifi', 'process', 'data', 'qualiti', 'issuesdocu', 'best', 'practic', 'methodolog', 'dev elop', 'standard', '', 'data', 'govern', '', 'version', 'control', '', 'deploy', 'chang', 'manag perform', 'qualiti', 'assur', 'test', 'report', 'dashboard', '', 'perform', 'user', 'accept', 'test', '', 'uat', '', '', 'technic', 'experi', '', 'core', 'understand', 'bu si', 'intellig', '', 'data', 'wareh', 'technolog', '', 'methodolog', '', 'techniqu', '', 'data', 'discoveri', '', 'data', 'sourc', '', 'present', 'layer', '', 'etc', '', 'proven', 'experi', 'deliv', 'success', 'solutionsexcel', 'knowledg', 'data', 'warehous', 'princi pl', 'report experi', 'design', 'build', 'enterpris', 'data', 'warehous extens', 'experi', 'bi', 'report', '', 'web', 'mobil', 'dashboard', 'develop', '', 'experi', 'multipl', 'd atabas', 'platformsexperi', 'data', 'analysi', '', 'assess', 'data', 'qualiti', 'develop', 'correct', 'actionsthorough', 'understand', 'complex', 'larg', 'scale', 'data', 'models understand', 'common', 'sdlc', 'practic methodologiesexperi', 'work', 'end', 'user', 'gat her', 'requirementsexperi', 'use', 'data', 'discoveri', 'toolsunderstand', 'rdbm', 'java', 'developerful time direct hirebolingbrook', '', 'il', 'client', 'lead', 'edg', 'softwar', 'develop', 'organ', 'uniqu', 'interest', 'industri', '', 'provid', 'train', 'help', 'b oard', '', 'excel', 'cultur', 'energet', '', 'challeng', '', 'team', 'orient', '', 'casua l', '', 'jean', 'ok', '', '', 'work life', 'balanc', '', 'outstand', 'benefit', 'packag', '', 'encourag', 'commun', 'atmospher', 'creativ', 'bright', 'mind', 'reward', '', 'look', 'java', 'develop', 'join', 'compani', '', 'posit', 'give', 'someon', 'abil', 'input', 'sy stem', 'design', 'grow', 'compani', '', 'person', 'work', 'team', 'focus', 'develop', 'ma inten', 'variou', 'applic', 'support', 'client', '', 'posit', 'build', 'applic', 'scratch', 'involv', 'implement', 'process', '', 'abil', 'commun', 'work', 'directli', 'client', 'essenti', '', 'success', 'candid', 'self direct', 'problem', 'solver', 'experi', 'applic', 'design', '', 'java jdbc', 'develop', '', 'multi thread', '', 'interest', '', 'pleas', 'email', 'resum', 'contact', 'kgalecki', '', 'transtechit com', 'overview', '', 'primari', 'applic', 'person', 'initi', 'work', 'serv', 'middl', 'tier', '', 'back', 'relat', 'dat abas', '', 'oracl', '', 'sql', 'server', '', 'mysql', '', 'via', 'jdbc', '', 'front', 'co mpani', '', 'transact base', 'law', 'enforc', 'messag', 'switch', 'via', 'proprietary', 'tcp ip', 'protocol', '', 'primari', 'respons', '', '', 'develop', 'robust', '', 'high', 'perform', '', 'flexibl', 'applic', 'analyz', 'diagnos', 'applic', 'error', 'write', 'appl ic', 'document', 'clearli', 'commun', 'statu', 'manag', 'requir', 'skill', '', '', 'core', 'java', 'experi', 'creat', 'jdbc', 'interfac', 'relat', 'databas', 'multithread', '', 'basic', 'experi', '', '', 'abl', 'develop', 'work', 'linux', '', 'aix', '', 'unix', 'deri v', 'os', '', 'sql', 'socket', 'level', 'messag', '', 'basic', 'experi', '', '', 'underst and', 'data', 'structur', 'self direct', 'problem', 'solv', 'abl', 'clearli', 'commun', 'verbal', 'write', 'bachelor', 'degre', 'comput', 'scienc', 'prefer', 'skill', '', '', 'c', '', 'c', '', 'knowledg', 'one', 'script', 'languag', 'experi', 'type', 'version', 'cont rol', 'tool', '', 'prefer', 'dvc', 'socket', 'level', 'messag', 'non technic', 'requir', '', '', 'must', 'abl', 'pass', 'state', 'feder', 'background', 'check', 'state', 'union', 'willing', 'travel', 'throughout', 'unit', 'state', 'requir', '', 'less', '', '', 'mi nimum', 'year', 'relev', 'experi', 'midtown', 'base', 'high', 'tech', 'firm', 'immedi', 'need', 'innov', 'devop', 'engin', 'help', 'defin', 'dev', '', 'op', 'relat', 'process', 'build', 'autom', '', 'posit', '', 'dev', '', 'op', 'engin', 'need', 'engin', 'technic', 'solut', 'well', 'perform', 'process', 'engin', '', 'ideal', 'candid', 'strong', 'scri pt', 'skill', 'autom', 'well', 'experi', 'configur', 'manag', '', 'server', 'deploy', '', 'continu', 'integr', 'experi', 'tool', 'chef', '', 'puppet', '', 'etc', '', 'must', 'expe ri', 'virtual', 'use', 'vmware', 'linux', 'experi', 'prefer', '', 'experi', 'go', '', 'do cker', '', 'ansibl', '', 'microservic', '', 'cloud', 'highli', 'desir', '', 'perman', 'po sit', 'innov', 'technolog', 'firm', 'midtown', 'offer', 'excel', 'benefit', 'includ', 'bl eed', 'edg', 'technolog', '', 'gener', 'pto', '', 'match', '', 'k', '', '', 'high tech', 'offic', 'environ', '', 'casual', 'dress', 'collabor', 'cultur', '', 'need', 'immedi', '', 'categori', '', 'develop', '', 'architect', 'name', 'look', 'senior', 'sap', 'fico', 'architect', 'join', 'us', 'fulltim', 'support', 'ongo', 'project', 'custom', '', 'must', 'excel', 'pre sale', 'experi consult', 'readi', '', 'travel', '', 'interest', '', 'ple as', 'send', 'us', 'updat', 'resum', '', 'best', 'time', 'contact', 'number', 'girish', 'dot', '', 'bandaru', '', 'yash', '', 'dot', '', 'com', 'discuss', 'opportun', '', 'thank', '', 'regard', 'girish', 'bandaruyash', 'technolog', '', 'girish bandaru', 'yash com', 'network', 'engin', 'job', 'descript', 'network', 'engin', 'respons', 'anal yz', '', 'design', '', 'instal', '', 'configur', '', 'maintain', 'repair', 'network', 'in frastructur', 'applic', 'compon', '', 'network', 'engin', 'perform', 'wide', 'varieti', 'evalu', '', 'mainten', '', 'instal', 'train', 'task', 'ensur', 'comput', 'network', 'perf orm', 'meet', 'compani', 'user', 'satisfact', '', 'network', 'engin', 'also', 'provid', 'direct', '', 'inform', '', 'recommend', 'regard', 'network', 'configur', 'instal', '', 'p rimari', 'duti', 'respons', '', 'maintain', 'thorough', 'understand', 'basic', 'behin

```
d', 'internet', 'work', '', 'dn', 'secur', '', 'ip', 'rout', '', 'http', '', 'vpn', '', 'email', 'rout', '', 'spam', '', 'etc', '', '', 'configur', 'setup', 'cisco', 'firewal', '', 'vpn', 'concentr', 'secur', 'applianc', 'access', 'vital', 'busi', 'applic', '', 'design', '', 'setup', 'configur', 'complex', 'switch', 'environ', '', 'design', '', 'setu', 'p', 'configur', 'complex', 'wireless', 'network', 'support', 'open', 'secur', 'access', 'abil', 'support', 'voic', 'video', 'applic', '', 'maintain', 'thorough', 'understand', 'l', 'ocal', 'area', 'network', '', 'assist', 'design', 'multi server', 'environ', 'includ', 'i', 'p', 'address', 'scheme', '', 'dn', '', 'win', '', 'etherchannel', '', 'bond', '', '', 'et', 'c', '', '', 'configur', 'instal', 'client', 'server', 'network', 'softwar', 'upgrad', 'ma', 'intain', 'network', 'telecommun', 'system', '', '', 'maintain', 'multi site', 'network', 'oper', 'softwar', 'applic', '', 'oper', 'system', 'regular', 'mainten', 'privat', 'publi', 'c', 'facil', '', 'manag', 'assign', 'project', 'program', 'compon', 'deliv', 'servic', 'a', 'ccord', 'establish', 'object', '', '', 'respond', 'inquiri', 'staff', '', 'administr', '', 'servic', 'provid', '', 'site', 'personnel', 'outsid', 'vendor', 'etc', '', 'provid', 'technic', 'assist', 'support', '', 'supervis', 'administr', 'system', 'server', 'relat', 'network', 'ensur', 'avail', 'servic', 'author', 'user', '', 'date', 'post', '', '', '']
```

In []:

```
# Generating one-hot representations for each word in the corpus.
```

```
onehot_repr=[one_hot(words,voc_size)for words in corpus]
print(len(onehot_repr))
```

1674

In []:

```
# Creating embedding docs with a vector of 5 dimensions to be passed to the LSTM.
```

```
words_length=5
embedded_docs=pad_sequences(onehot_repr,padding='pre',maxlen=words_length)
print(len(embedded_docs))
```

1674

In []:

```
# Creating model
embedding_vector_features=10
model=Sequential()
model.add(Embedding(voc_size,embedding_vector_features,input_length=words_length))
model.add(Dropout(0.3))
model.add(LSTM(100))
model.add(Dropout(0.3))
model.add(Dense(1,activation='sigmoid'))
model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])

print(model.summary())
print(len(embedded_docs),y.shape)
```

Model: "sequential_3"

Layer (type)	Output Shape	Param #
embedding_3 (Embedding)	(None, 5, 10)	1000000
lstm_3 (LSTM)	(None, 100)	44400
dense_3 (Dense)	(None, 1)	101

Total params: 1,044,501
Trainable params: 1,044,501
Non-trainable params: 0

None
1674 (1674,)

In []:

```
# Creating training and testing data.
```

```

X_total=np.array(embedded_docs)
y_total=np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X_total, y_total, test_size=0.20, random_state=40)

# Training the model
model.fit(X_train,y_train,validation_data=(X_test,y_test),epochs=50,batch_size=64)

# Making predictions.
y_pred=model.predict_classes(X_test)

confusion_matrix(y_test,y_pred)
accuracy_score(y_test,y_pred)

```

```

Epoch 1/50
21/21 [=====] - 3s 48ms/step - loss: 0.6614 - accuracy: 0.8381 - val_loss: 0.5207 - val_accuracy: 0.8179
Epoch 2/50
21/21 [=====] - 0s 22ms/step - loss: 0.4625 - accuracy: 0.8369 - val_loss: 0.4718 - val_accuracy: 0.8179
Epoch 3/50
21/21 [=====] - 0s 22ms/step - loss: 0.4113 - accuracy: 0.8566 - val_loss: 0.4682 - val_accuracy: 0.8179
Epoch 4/50
21/21 [=====] - 0s 23ms/step - loss: 0.4243 - accuracy: 0.8439 - val_loss: 0.4654 - val_accuracy: 0.8179
Epoch 5/50
21/21 [=====] - 0s 22ms/step - loss: 0.4207 - accuracy: 0.8405 - val_loss: 0.4602 - val_accuracy: 0.8179
Epoch 6/50
21/21 [=====] - 0s 23ms/step - loss: 0.4338 - accuracy: 0.8273 - val_loss: 0.4524 - val_accuracy: 0.8179
Epoch 7/50
21/21 [=====] - 0s 22ms/step - loss: 0.3800 - accuracy: 0.8486 - val_loss: 0.4326 - val_accuracy: 0.8179
Epoch 8/50
21/21 [=====] - 0s 23ms/step - loss: 0.3756 - accuracy: 0.8392 - val_loss: 0.4151 - val_accuracy: 0.8179
Epoch 9/50
21/21 [=====] - 0s 22ms/step - loss: 0.3385 - accuracy: 0.8455 - val_loss: 0.3947 - val_accuracy: 0.8179
Epoch 10/50
21/21 [=====] - 0s 23ms/step - loss: 0.2806 - accuracy: 0.8598 - val_loss: 0.3619 - val_accuracy: 0.8478
Epoch 11/50
21/21 [=====] - 0s 22ms/step - loss: 0.2388 - accuracy: 0.8755 - val_loss: 0.3226 - val_accuracy: 0.8597
Epoch 12/50
21/21 [=====] - 0s 23ms/step - loss: 0.1788 - accuracy: 0.9222 - val_loss: 0.3468 - val_accuracy: 0.8716
Epoch 13/50
21/21 [=====] - 0s 22ms/step - loss: 0.1233 - accuracy: 0.9541 - val_loss: 0.3752 - val_accuracy: 0.8776
Epoch 14/50
21/21 [=====] - 1s 24ms/step - loss: 0.1510 - accuracy: 0.9321 - val_loss: 0.3462 - val_accuracy: 0.8806
Epoch 15/50
21/21 [=====] - 0s 22ms/step - loss: 0.1055 - accuracy: 0.9552 - val_loss: 0.3631 - val_accuracy: 0.8836
Epoch 16/50
21/21 [=====] - 0s 22ms/step - loss: 0.1271 - accuracy: 0.9465 - val_loss: 0.3533 - val_accuracy: 0.8746
Epoch 17/50
21/21 [=====] - 0s 23ms/step - loss: 0.1065 - accuracy: 0.9563 - val_loss: 0.3577 - val_accuracy: 0.8746
Epoch 18/50
21/21 [=====] - 1s 25ms/step - loss: 0.0988 - accuracy: 0.9556 - val_loss: 0.3644 - val_accuracy: 0.8806
Epoch 19/50
21/21 [=====] - 0s 22ms/step - loss: 0.1021 - accuracy: 0.9573 - val_loss: 0.3591 - val_accuracy: 0.8746

```

Epoch 20/50
21/21 [=====] - 0s 23ms/step - loss: 0.0960 - accuracy: 0.9623 -
val_loss: 0.3627 - val_accuracy: 0.8776
Epoch 21/50
21/21 [=====] - 0s 23ms/step - loss: 0.1022 - accuracy: 0.9537 -
val_loss: 0.3703 - val_accuracy: 0.8836
Epoch 22/50
21/21 [=====] - 0s 23ms/step - loss: 0.1052 - accuracy: 0.9536 -
val_loss: 0.3602 - val_accuracy: 0.8776
Epoch 23/50
21/21 [=====] - 0s 22ms/step - loss: 0.1024 - accuracy: 0.9509 -
val_loss: 0.3653 - val_accuracy: 0.8806
Epoch 24/50
21/21 [=====] - 0s 22ms/step - loss: 0.0947 - accuracy: 0.9581 -
val_loss: 0.3695 - val_accuracy: 0.8776
Epoch 25/50
21/21 [=====] - 1s 25ms/step - loss: 0.1125 - accuracy: 0.9434 -
val_loss: 0.3713 - val_accuracy: 0.8806
Epoch 26/50
21/21 [=====] - 0s 22ms/step - loss: 0.0917 - accuracy: 0.9627 -
val_loss: 0.3742 - val_accuracy: 0.8806
Epoch 27/50
21/21 [=====] - 0s 22ms/step - loss: 0.0943 - accuracy: 0.9558 -
val_loss: 0.3825 - val_accuracy: 0.8806
Epoch 28/50
21/21 [=====] - 0s 22ms/step - loss: 0.1014 - accuracy: 0.9501 -
val_loss: 0.3814 - val_accuracy: 0.8716
Epoch 29/50
21/21 [=====] - 0s 23ms/step - loss: 0.1032 - accuracy: 0.9505 -
val_loss: 0.3849 - val_accuracy: 0.8806
Epoch 30/50
21/21 [=====] - 0s 23ms/step - loss: 0.1047 - accuracy: 0.9507 -
val_loss: 0.3830 - val_accuracy: 0.8806
Epoch 31/50
21/21 [=====] - 0s 23ms/step - loss: 0.1055 - accuracy: 0.9466 -
val_loss: 0.4097 - val_accuracy: 0.8836
Epoch 32/50
21/21 [=====] - 0s 23ms/step - loss: 0.0999 - accuracy: 0.9548 -
val_loss: 0.3807 - val_accuracy: 0.8776
Epoch 33/50
21/21 [=====] - 0s 23ms/step - loss: 0.1036 - accuracy: 0.9498 -
val_loss: 0.3884 - val_accuracy: 0.8836
Epoch 34/50
21/21 [=====] - 0s 22ms/step - loss: 0.0894 - accuracy: 0.9608 -
val_loss: 0.3765 - val_accuracy: 0.8806
Epoch 35/50
21/21 [=====] - 0s 21ms/step - loss: 0.0879 - accuracy: 0.9582 -
val_loss: 0.3789 - val_accuracy: 0.8806
Epoch 36/50
21/21 [=====] - 0s 22ms/step - loss: 0.0985 - accuracy: 0.9561 -
val_loss: 0.3817 - val_accuracy: 0.8776
Epoch 37/50
21/21 [=====] - 0s 23ms/step - loss: 0.1012 - accuracy: 0.9516 -
val_loss: 0.3831 - val_accuracy: 0.8806
Epoch 38/50
21/21 [=====] - 0s 23ms/step - loss: 0.0988 - accuracy: 0.9517 -
val_loss: 0.4069 - val_accuracy: 0.8836
Epoch 39/50
21/21 [=====] - 0s 23ms/step - loss: 0.0978 - accuracy: 0.9579 -
val_loss: 0.3781 - val_accuracy: 0.8776
Epoch 40/50
21/21 [=====] - 0s 22ms/step - loss: 0.0995 - accuracy: 0.9502 -
val_loss: 0.3853 - val_accuracy: 0.8896
Epoch 41/50
21/21 [=====] - 0s 23ms/step - loss: 0.1090 - accuracy: 0.9413 -
val_loss: 0.3905 - val_accuracy: 0.8776
Epoch 42/50
21/21 [=====] - 1s 24ms/step - loss: 0.0918 - accuracy: 0.9573 -
val_loss: 0.3902 - val_accuracy: 0.8806
Epoch 43/50
21/21 [=====] - 0s 22ms/step - loss: 0.0918 - accuracy: 0.9561 -
val_loss: 0.3927 - val_accuracy: 0.8806

```

Epoch 44/50
21/21 [=====] - 0s 23ms/step - loss: 0.0972 - accuracy: 0.9517 -
val_loss: 0.3908 - val_accuracy: 0.8806
Epoch 45/50
21/21 [=====] - 0s 22ms/step - loss: 0.0853 - accuracy: 0.9598 -
val_loss: 0.3923 - val_accuracy: 0.8746
Epoch 46/50
21/21 [=====] - 0s 23ms/step - loss: 0.0880 - accuracy: 0.9572 -
val_loss: 0.3966 - val_accuracy: 0.8806
Epoch 47/50
21/21 [=====] - 0s 22ms/step - loss: 0.0942 - accuracy: 0.9567 -
val_loss: 0.3990 - val_accuracy: 0.8806
Epoch 48/50
21/21 [=====] - 0s 22ms/step - loss: 0.0977 - accuracy: 0.9513 -
val_loss: 0.4126 - val_accuracy: 0.8836
Epoch 49/50
21/21 [=====] - 0s 22ms/step - loss: 0.0995 - accuracy: 0.9484 -
val_loss: 0.3966 - val_accuracy: 0.8776
Epoch 50/50
21/21 [=====] - 0s 23ms/step - loss: 0.1057 - accuracy: 0.9469 -
val_loss: 0.3817 - val_accuracy: 0.8776

```

```

/usr/local/lib/python3.7/dist-packages/tensorflow/python/keras/engine/sequential.py:450:
UserWarning: `model.predict_classes()` is deprecated and will be removed after 2021-01-01
. Please use instead: * `np.argmax(model.predict(x), axis=-1)`, if your model does multi
-class classification (e.g. if it uses a `softmax` last-layer activation). * `(model.pre
dict(x) > 0.5).astype("int32")`, if your model does binary classification (e.g. if it
uses a `sigmoid` last-layer activation).
  warnings.warn("`model.predict_classes()` is deprecated and '

```

```
Out[ ]:
```

```
0.8208955223880597
```

```
In [77]:
```

```

# Importing necessary modules.

import pandas as pd
import nltk
import re
nltk.download('stopwords')
nltk.download('punkt')
from nltk.corpus import stopwords

# Shortlisting the columns to be used for building the skill set.

columns = ['LanguageWorkedWith', 'LanguageDesireNextYear', 'DatabaseWorkedWith', 'DatabaseDe
sireNextYear',
           'PlatformWorkedWith', 'PlatformDesireNextYear', 'FrameworkWorkedWith', 'Framework
kDesireNextYear',
           'IDE', 'OperatingSystem', 'VersionControl']
Skills = pd.read_csv(filepath_or_buffer='survey_results_public_skills.csv', usecols=column
s)
Skills = Skills.dropna()
Skills['Combined_Skills'] = Skills[Skills.columns[1:]].apply(lambda x: ';'.join(x.dropna
()).astype(str), axis=1)
Skills['Skill_List'] = Skills['Combined_Skills'].str.split(";")
Unique_Skills = set()
print(Skills)

# Adding all unique skills in a set.

for i in Skills["Skill_List"]:
    for j in i:
        Unique_Skills.add(j.lower())

print(len(Unique_Skills))
print(Unique_Skills)

# Filtering the rows with no skills.

```

```

Jobs = pd.read_csv(filepath_or_buffer='jobs.csv')
print(Jobs)

JobDescriptions = pd.read_csv(filepath_or_buffer='jobs.csv')
JobDescriptions = JobDescriptions.loc[JobDescriptions['skills'].str.contains('See', na=False)]
JobDescriptions = JobDescriptions.loc[JobDescriptions['skills'].str.count(',') == 0]
JobDescriptions = JobDescriptions.reset_index(drop=True)
print(JobDescriptions)

# Predicting the skills with the help of jobdescription column by using the model trained above.

for i in range(len(JobDescriptions)):
    JobDescriptions.loc[i, "skills"] = " "

    # Preprocessing the data
    JobDescriptions.loc[i, "jobdescription"] = re.sub(r'\[[0-9]*\]', ' ', str(JobDescriptions.loc[i, "jobdescription"]))
    JobDescriptions.loc[i, "jobdescription"] = re.sub(r'\s+', ' ', JobDescriptions.loc[i, "jobdescription"])
    JobDescriptions.loc[i, "jobdescription"] = JobDescriptions.loc[i, "jobdescription"].lower()
    JobDescriptions.loc[i, "jobdescription"] = re.sub(r'\d', ' ', JobDescriptions.loc[i, "jobdescription"])
    JobDescriptions.loc[i, "jobdescription"] = re.sub(r'\s+', ' ', JobDescriptions.loc[i, "jobdescription"])

    # Breaking down job description into sentences.
    sentences = nltk.sent_tokenize(JobDescriptions.loc[i, "jobdescription"])

    sentences = [nltk.word_tokenize(sentence) for sentence in sentences]

    for j in range(len(sentences)):
        sentences[j] = [word for word in sentences[j] if word not in stopwords.words('english')]

        for k in sentences[j]:
            if k in Unique_Skills and k not in JobDescriptions.loc[i, "skills"] :
                JobDescriptions.loc[i, "skills"] += k
                JobDescriptions.loc[i, "skills"] += ","
            elif k not in Unique_Skills:
                onehot_repr = one_hot(k, voc_size)
                embedded_docs = pad_sequences([onehot_repr], padding='pre', maxlen=words_length)
                if model.predict_classes(embedded_docs)[0][0] == 1:
                    JobDescriptions.loc[i, "skills"] += k
                    JobDescriptions.loc[i, "skills"] += ","

JobDescriptions["skills"] = JobDescriptions["skills"].str.rstrip(',')
print(JobDescriptions)

Jobs = pd.concat([Jobs, JobDescriptions]).reset_index(drop=True)
Jobs = Jobs.drop_duplicates(subset=['uniq_id'], keep='last')
Jobs = Jobs.reset_index(drop=True)
print(Jobs)

```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
```

```
[nltk_data] Package stopwords is already up-to-date!
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
```

```
[nltk_data] Package punkt is already up-to-date!
```

```
LanguageWorkedWith ...
```

```
Skill_List
```

```
0 JavaScript;Python;HTML;CSS ... [JavaScript, Python, HTML, CSS, Redis, SQL Ser...
```

```
1 JavaScript;Python;Bash/Shell ... [Go, Python, Redis, PostgreSQL, Memcached, Pos...
```

```
5 Java;JavaScript;Python;TypeScript;HTML;CSS ... [C#, Go, Java, JavaScript, Python, SQL, TypeSc...
```

```
13 Java ... [Java, Python, MongoDB, MySQL, Oracle, MariaDB...
```

```
17 C#;SQL;HTML;CSS;Bash/Shell ... [C#, F#, Haskell, SQL, Ocaml, SQ
```



```

L Server, Redi...
...
...
92430          PHP;Python;SQL;HTML;CSS ... [JavaScript, PHP, Python, SQL, Ko
tlin, HTML, C...
92447          PHP ... [PHP, MySQL, MySQL, WordPress, W
ordPress, Node...
92455          PHP;SQL ... [PHP, MySQL, MySQL, Oracle, Word
Press, Android...
92471          Python;SQL;HTML;Bash/Shell ... [Python, SQL, TypeScript, CSS, Ba
sh/Shell, SQL...
92507          TypeScript;HTML;CSS ... [TypeScript, HTML, CSS, PostgreS
QL, PostgreSQL...

```

```

[30607 rows x 13 columns]
130

```

```

{'rstudio', 'zip file back-ups', 'google cloud storage', 'ios', 'amazon rds/aurora', 'sha
repoint', 'cordova', 'java', 'apple watch or apple tv', 'windows phone', 'swift', 'linux'
, 'c++', 'google cloud platform/app engine', 'firebase', 'git', 'android studio', 'micros
oft azure (tables, cosmosdb, sql, etc)', 'android', 'rubymine', 'perl', 'zend', 'sql', '.
net core', 'wordpress', 'drupal', 'esp8266', 'angular', 'python', 'linux-based', 'scala',
'netbeans', 'hack', 'matlab', 'r', 'ruby', 'mercurial', 'intellij', 'apache hbase', 'pych
arm', 'f#', 'copying and pasting files to network shares', 'css', 'go', 'sublime text', '
amazon redshift', 'vim', 'spring', 'cobol', 'visual basic 6', 'django', 'apache hive', 'i
bm db2', 'clojure', 'postgresql', 'atom', 'heroku', 'objective-c', 'notepad++', 'subversi
on', 'typescript', 'salesforce', 'elasticsearch', 'eclipse', 'spark', 'visual studio code
', 'arduino', 'torch/pytorch', 'emacs', 'macos', 'sql server', 'kotlin', 'html', 'gaming
console', 'mongodb', 'phpstorm', 'tensorflow', 'c#', 'ipython / jupyter', 'sqlite', 'cass
andra', 'memcached', 'google home', 'amazon echo', 'mariadb', 'raspberry pi', 'bsd/unix',
'oracle', 'amazon dynamodb', 'vb.net', 'team foundation version control', 'light table',
'aws', 'rust', 'google bigquery', 'textmate', 'react', 'erlang', 'coda', 'bash/shell', 'h
adoop', 'ocaml', 'mac os', 'node.js', 'coffeescript', 'vba', 'delphi/object pascal', 'mys
ql', 'lua', 'neo4j', 'javascript', 'ibm cloud or watson', 'xamarin', 'windows', 'komodo',
'xcode', 'assembly', 'i don't use version control', 'mainframe', 'predix', 'visual studio
', 'c', 'azure', 'groovy', 'php', 'julia', 'haskell', 'serverless', 'windows desktop or s
erver', 'redis'}

```

```

advertiserurl ...
uniq_id
0      https://www.dice.com/jobs/detail/AUTOMATION-TE... ... 418ff92580b270ef4e7c14f0dd
fc36b4
1      https://www.dice.com/jobs/detail/Information-S... ... 8aec88cba08d53da65ab99cf20
f6f9d9
2      https://www.dice.com/jobs/detail/Business-Solu... ... 46baa1f69ac07779274bcd90b8
5d9a72
3      https://www.dice.com/jobs/detail/Java-Develope... ... 3941b2f206ae0f900c4fba4ac0
b18719
4      https://www.dice.com/jobs/detail/DevOps-Engine... ... 45efaf1f6bc65acc32bbbbb953a1
ed13b7
...
...
21995  https://www.dice.com/jobs/detail/Web-Designer-... ... 86e27ce6b7e631e55d69d142c7
d43df2
21996  https://www.dice.com/jobs/detail/Senior-Front-... ... 4287c7ee3317ccf1edd76e238c
f8e584
21997  https://www.dice.com/jobs/detail/QA-Analyst-Sa... ... d7512f0181d69f83f96db38cd7
7a4d08
21998  https://www.dice.com/jobs/detail/Tech-Lead%252... ... ec375268b494b3bcbed1635d64
226112
21999  https://www.dice.com/jobs/detail/C%2526%252347... ... 9a4e8c27f74af4c0d2f6efbd42
0a8a91

```

```

[22000 rows x 12 columns]

```

```

advertiserurl ... u
niq_id
0      https://www.dice.com/jobs/detail/AUTOMATION-TE... ... 418ff92580b270ef4e7c14f0ddfc
36b4
1      https://www.dice.com/jobs/detail/Java-Develope... ... 3941b2f206ae0f900c4fba4ac0b1
8719
2      https://www.dice.com/jobs/detail/Application-S... ... 95c9127e2770172f454f3b83981e
aa88
3      https://www.dice.com/jobs/detail/Windows-Syste... ... 9e5704d08bc07ddb6df9ef98b223
b036

```

```

4      https://www.dice.com/jobs/detail/Java-Architec...    ...    e4f57bc5366124a0a47cac27f557
f9ec
...
...
188    https://www.dice.com/jobs/detail/JR-Project-ma...    ...    72597cc210e6c518f05dc2cfdd6f
d17f
189    https://www.dice.com/jobs/detail/PC-Support-As...    ...    38e97bc0955ee08d7fff35bf9d2c
7b6b
190    https://www.dice.com/jobs/detail/Technical-Lea...    ...    ca3b2c097deffef3bef71e8e8dd7
b567
191    https://www.dice.com/jobs/detail/Applications-...    ...    4fa8a123b307765c21beef51693d
f0b5
192    https://www.dice.com/jobs/detail/Tech-Project-...    ...    28adfb1f1cfef4419a0b85db460d
7288

```

[193 rows x 12 columns]

```

/usr/local/lib/python3.7/dist-packages/tensorflow/python/keras/engine/sequential.py:450:
UserWarning: `model.predict_classes()` is deprecated and will be removed after 2021-01-01
. Please use instead: * `np.argmax(model.predict(x), axis=-1)`, if your model does multi
-class classification (e.g. if it uses a `softmax` last-layer activation). * `(model.pre
dict(x) > 0.5).astype("int32")`, if your model does binary classification (e.g. if it
uses a `sigmoid` last-layer activation).
  warnings.warn("`model.predict_classes()` is deprecated and '

```

```

                                advertiserurl    ...    u
uniq_id
0      https://www.dice.com/jobs/detail/AUTOMATION-TE...    ...    418ff92580b270ef4e7c14f0ddfc
36b4
1      https://www.dice.com/jobs/detail/Java-Develope...    ...    3941b2f206ae0f900c4fba4ac0b1
8719
2      https://www.dice.com/jobs/detail/Application-S...    ...    95c9127e2770172f454f3b83981e
aa88
3      https://www.dice.com/jobs/detail/Windows-Syste...    ...    9e5704d08bc07ddb6df9ef98b223
b036
4      https://www.dice.com/jobs/detail/Java-Architec...    ...    e4f57bc5366124a0a47cac27f557
f9ec
...
...
188    https://www.dice.com/jobs/detail/JR-Project-ma...    ...    72597cc210e6c518f05dc2cfdd6f
d17f
189    https://www.dice.com/jobs/detail/PC-Support-As...    ...    38e97bc0955ee08d7fff35bf9d2c
7b6b
190    https://www.dice.com/jobs/detail/Technical-Lea...    ...    ca3b2c097deffef3bef71e8e8dd7
b567
191    https://www.dice.com/jobs/detail/Applications-...    ...    4fa8a123b307765c21beef51693d
f0b5
192    https://www.dice.com/jobs/detail/Tech-Project-...    ...    28adfb1f1cfef4419a0b85db460d
7288

```

[193 rows x 12 columns]

```

                                advertiserurl    ...
uniq_id
0      https://www.dice.com/jobs/detail/Information-S...    ...    8aec88cba08d53da65ab99cf20
f6f9d9
1      https://www.dice.com/jobs/detail/Business-Solu...    ...    46baa1f69ac07779274bcd90b8
5d9a72
2      https://www.dice.com/jobs/detail/DevOps-Engine...    ...    45efa1f6bc65acc32bbbb953a1
ed13b7
3      https://www.dice.com/jobs/detail/SAP-FICO-Arch...    ...    e0ac9d926dda5e95162ef05ade
a7318c
4      https://www.dice.com/jobs/detail/Network-Engin...    ...    e7e326053c586bd94e59f1fd74
de4a1b
...
...
21995  https://www.dice.com/jobs/detail/JR-Project-ma...    ...    72597cc210e6c518f05dc2cfdd
6fd17f
21996  https://www.dice.com/jobs/detail/PC-Support-As...    ...    38e97bc0955ee08d7fff35bf9d
2c7b6b
21997  https://www.dice.com/jobs/detail/Technical-Lea...    ...    ca3b2c097deffef3bef71e8e8d
d7b567
21998  https://www.dice.com/jobs/detail/Applications-...    ...    4fa8a123b307765c21beef5169
3df0b5

```

5d1000

21999 https://www.dice.com/jobs/detail/Tech-Project-... 28adfb1f1cfef4419a0b85db46
0d7288

[22000 rows x 12 columns]

In [79]:

```
Jobs.to_csv('jobs_modified.csv')
```