

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import spacy
import pickle
import random
```

```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
→ /kaggle/input/resume-data-with-annotations/resume v6.pdf
/kaggle/input/resume-data-with-annotations/Alice Clark CV.docx
/kaggle/input/resume-data-with-annotations/Alice Clark CV.txt
/kaggle/input/resume-data-with-annotations/AshlyLauResume.pdf
/kaggle/input/resume-data-with-annotations/Smith Resume.pdf
/kaggle/input/resume-data-with-annotations/train_data.txt
/kaggle/input/resume-data-with-annotations/Smith Resume.docx
/kaggle/input/resume-data-with-annotations/Jinesh_Dhruv_poster.pdf
/kaggle/input/resume-data-with-annotations/Alice Clark CV.pdf
/kaggle/input/resume-data-with-annotations/train_data.pkl
/kaggle/input/resume-data-with-annotations/Sample_Resume.pdf
/kaggle/input/resume-data-with-annotations/Xinni_Chng.pdf
/kaggle/input/resume-data-with-annotations/sample_input.pdf
```

```
train_data = pickle.load(open('/kaggle/input/resume-data-with-annotations/train_
train_data[0]
```

```
→ ('Govardhana_K_Senior_Software_Engineer_Bengaluru,Karnataka,Karnataka -
```

Email me on Indeed: [indeed.com/r/Govardhana-K/ b2de315d95905b68](https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN) Total IT experience 5 Years 6 Months Cloud Lending Solutions INC 4 Month • Salesforce Developer Oracle 5 Years 2 Month • Core Java Developer Languages Core Java, Go Lang Oracle PL-SQL programming, Sales Force Developer with APEX. Designations & Promotions Willing to relocate: Anywhere WORK EXPERIENCE Senior Software Engineer Cloud Lending Solutions – Bangalore, Karnataka – January 2018 to Present Present Senior Consultant Oracle – Bangalore, Karnataka – November 2016 to December 2017 Staff Consultant Oracle – Bangalore, Karnataka – January 2014 to October 2016 Associate Consultant Oracle – Bangalore, Karnataka – November 2012 to December 2013 EDUCATION B.E in Computer Science Engineering Adithya Institute of Technology – Tamil Nadu September 2008 to June 2012 <https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN> SKILLS APEX. (Less than 1 year), Data Structures (3 years), FLEXCUBE (5 years), Oracle (5 years), Algorithms (3 years) LINKS <https://www.linkedin.com/in/govardhana-k-61024944/> ADDITIONAL INFORMATION Technical Proficiency: Languages: Core Java, Go Lang, Data Structures & Algorithms, Oracle PL-SQL programming, Sales Force with APEX. Tools: RADTool, Jdeveloper, NetBeans, Eclipse, SQL developer, PL/SQL Developer, WinSCP, Putty Web Technologies: JavaScript, XML, HTML, Webservice Operating Systems: Linux, Windows Version control system SVN & Git-Hub Databases: Oracle Middleware: Web logic, OC4J Product FLEXCUBE: Oracle FLEXCUBE Versions 10.x, 11.x and 12.x <https://www.linkedin.com/in/govardhana-k-61024944/>,

```
{'entities': [(1749, 1755, 'Companies worked at'),
(1696, 1702, 'Companies worked at'),
(1417, 1423, 'Companies worked at'),
(1356, 1793, 'Skills'),
(1209, 1215, 'Companies worked at'),
(1136, 1248, 'Skills'),
(928, 932, 'Graduation Year'),
(858, 889, 'College Name'),
(821, 856, 'Degree'),
(787, 791, 'Graduation Year'),
(744, 750, 'Companies worked at'),
(722, 742, 'Designation'),
(658, 664, 'Companies worked at'),
(640, 656, 'Designation'),
(574, 580, 'Companies worked at'),
(555, 573, 'Designation'),
(470, 493, 'Companies worked at'),
(444, 469, 'Designation'),
(308, 314, 'Companies worked at'),
(234, 240, 'Companies worked at'),
(175, 198, 'Companies worked at'),
(93, 137, 'Email Address'),
(39, 48, 'Location'),
(13, 38, 'Designation'),
(0, 12, 'Name')]]})
```

```
# Load Blank Model
nlp = spacy.blank('en')
```

```

def train_model(train_data):
    # Remove all pipelines and add NER pipeline from the model
    if 'ner' not in nlp.pipe_names:
        ner = nlp.create_pipe('ner')
        # adding NER pipeline to nlp model
        nlp.add_pipe(ner, last=True)

    #Add labels in the NLP pipeline
    for _, annotation in train_data:
        for ent in annotation.get('entities'):
            ner.add_label(ent[2])

    #Remove other pipelines if they are there
    other_pipes = [pipe for pipe in nlp.pipe_names if pipe != 'ner']
    with nlp.disable_pipes(*other_pipes): # only train NER
        optimizer = nlp.begin_training()
        for itn in range(10): # train for 10 iterations
            print("Starting iteration " + str(itn))
            random.shuffle(train_data)
            losses = {}
            index = 0
            for text, annotations in train_data:
                try:
                    nlp.update(
                        [text], # batch of texts
                        [annotations], # batch of annotations
                        drop=0.2, # dropout - make it harder to memorise data
                        sgd=optimizer, # callable to update weights
                        losses=losses)
                except Exception as e:
                    pass

            print(losses)

# Start Training model
train_model(train_data)

```

```

gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)

```

```
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
/opt/conda/lib/python3.7/site-packages/spacy/language.py:482: UserWarning:
gold = GoldParse(doc, **gold)
{'ner': 15730.53110635804}
Starting iteration 1
{'ner': 9380.157330286129}
Starting iteration 2
{'ner': 10739.070459000464}
Starting iteration 3
{'ner': 6388.888260545632}
Starting iteration 4
{'ner': 6133.455884867987}
Starting iteration 5
{'ner': 5401.049414239988}
Starting iteration 6
{'ner': 6698.737881477294}
Starting iteration 7
{'ner': 4742.5542419806925}
Starting iteration 8
{'ner': 5257.23237508259}
Starting iteration 9
{'ner': 4542.069310597443}
```

```
# Saving the model
nlp.to_disk('nlp_ner_model')
```

```
#Loading Model
nlp_model = spacy.load('nlp_ner_model')
```

```
nlp_model = nlp_model.from_instances(train_data_loader.iter_instances(),
```

```
# trying and seeing the prediction of the model
doc = nlp_model(train_data[0][0])
for ent in doc.ents:
    print(f"{ent.label_.upper():{30}}-{ent.text}")
```

```
⇒ NAME -Nitin Tr
DESIGNATION -PeopleSoft Consultant
EMAIL ADDRESS -indeed.com/r/Nitin-Tr/e7e3a2f5b4c1e24e
DESIGNATION -PeopleSoft consultant
COMPANIES WORKED AT -Oracle
COMPANIES WORKED AT -Oracle India
DEGREE -Btech information science in BCET
COLLEGE NAME -Vtu
LOCATION -Bengaluru
```

```
!pip install PyMuPDF
```

```
⇒ Collecting PyMuPDF
  Downloading PyMuPDF-1.18.4-cp37-cp37m-manylinux2010_x86_64.whl (6.3 MB)
    |████████████████████████████████████████| 6.3 MB 630 kB/s
Installing collected packages: PyMuPDF
Successfully installed PyMuPDF-1.18.4
```

```
import sys, fitz
fname = '/kaggle/input/resume-data-with-annotations/Alice Clark CV.pdf'
doc = fitz.open(fname)
text = ""
for page in doc:
    text = text + str(page.getText())
tx = " ".join(text.split('\n'))
print(tx)
```

```
⇒ Alice Clark AI / Machine Learning Delhi, India Email me on Indeed • 2
```

```
# Applying the model
doc = nlp_model(tx)
for ent in doc.ents:
    print(f'{ent.label_.upper():{30}}- {ent.text}')
```

```
⇒ NAME - Alice Clark
DESIGNATION - AI / Machine Learning
```

LOCATION	- Delhi
DESIGNATION	- Software Engineer
COMPANIES WORKED AT	- Microsoft
LOCATION	- Bangalore
GRADUATION YEAR	- 2000
COMPANIES WORKED AT	- Microsoft
COMPANIES WORKED AT	- Microsoft
COMPANIES WORKED AT	- Microsoft
COLLEGE NAME	- Indian Institute of Technology
SKILLS	- Machine Learning, Natural Language Processi

Start coding or [generate](#) with AI.