Import Library

```
import requests
import pandas as pd
from bs4 import BeautifulSoup
import string
import spacy
import re
```

Text Analysis (1 url)

Scrap Data

```
url="""https://insights.blackcoffer.com/how-is-login-logout-time-tracking-for-employees-in-office-done-by-ai/"""

headers = {"User-Agent": "Mozilla/5.0 (X11; Linux x86_64; rv:60.0) Gecko/20100101 Firefox/60.0"}
page = requests.get(url, headers=headers)
soup = BeautifulSoup(page.content, 'html.parser')
```

We need to pass argument called Headers by passing "User-Agent" to the request to bypass the mod-security error.

```
soup=BeautifulSoup(page.content, 'html.parser')
```

Extract Title from articel

```
title=soup.find('h1',class_="entry-title")
title=title.text.replace('\n'," ")
title
```

```
'How is Login Logout Time Tracking for Employees in Office done by AI?'
```

Extract Content from articel

```
content=soup.findAll(attrs={'class':'td-post-content'})
content=content[0].text.replace('\n'," ")
content
```

```
' When people hear AI they often think about sentient robots and magic boxes. AI today is much more mundane and simple—but that doesn’t mean it’s not powerful. Another misconception is that high-profile research projects can be applied directly to any business situation. AI done right can create an extreme return on investments (ROIs)—for instance through automation or precise prediction. But it does take thought, time, and proper implementation. We have seen that success and value generated by AI projects are increased when there is a grounded understanding and expectation of what the technology can deliver from
```

the C-suite down. “Artificial Intelligence (AI) is a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense, learn, reason and take action.”3 Lately there has been a big rise in the day-to-day use of machines powered by AI. These machines are wired using cross-disciplinary approaches based on mathematics, computer science, statistics, psychology, and more.4 Virtual assistants are becoming more common, most of the web shops predict your purchases, many companies make use of chatbots in their customer service and many companies use algorithms to detect fraud. AI and Deep Learning technology employed in office entry systems will bring proper time tracking of each employee. As this system tries to learn each person with an image processing technology whose data is feed forwarded to a deep learning model where Deep learning isn’t an algorithm per se, but rather a family of algorithms that implements deep networks (many layers). These networks are so deep that new methods of computation, such as graphics processing units (GPUs), are required to train them, in addition to clusters of compute nodes. So using deep learning we can take detect the employee using face and person recognition scan and through which login, logout timing is recorded. Using an AI system we can even identify each employee’s entry time, their working hours, non-working hours by tracking the movement of an employee in the office so that system can predict and report HR for the salary for each employee based on their working hours. Our system can take feed from CCTV to track movements of employees and this system is capable of recognizing a person even he/she is being masked as in this pandemic situation by taking their iris scan. With this system installed inside the office, the following are some of the benefits: 1)Compliance/litigation needs For several countries, regulations insist that the employer must keep documents available that can demonstrate the working hours performed by each employee. In the event of control from the labor inspectorate or a dispute with an employee, the employer must be able to explain and justify the working hours for the company. This can be made easy as our system is tracking employee movements 2)Information security needs This is about monitoring user connection times to detect suspicious access times. In the event where compromised credentials are used to log on at 3 a.m. on a Saturday, a notification on this access could alert the IT team that an attack is possibly underway. 3)Employee login logout software To manage and react to employees’ attendance, overtime thresholds, productivity, and suspicious access times, our system records and stores detailed and interactive reporting on users’ connection times. These records allow you to better manage users’ connection times and provide accurate, detailed data required by management. 4)If you want to avoid paying overtime, make sure that your employees respect certain working time quotas or even avoid suspicious access. Our system will alert the HR officer about each employee’s office in and out time so that they can accordingly take action. 5)Last but not least it reduces human resource needs to keep

```
track of the records and sending the report to HR and HR officials has to check through the report so this system will reduce times and human resource needs With the use of AI and Deep Learning technologies, we can automate some routines stuff with more functionality which humans need more resources to keep track thereby reducing time spent on manual data entry works rather companies can think of making their position high in the competitive world. Blackcoffer Insights 33: Suriya E, Vellore Institute of Technology '
```

Remove punctuation from the content

```
#Punctuation
content = content.translate(str.maketrans('', '', string.punctuation))

content
```

```
' When people hear AI they often think about sentient robots and magic boxes AI today is much more mundane and simple—but that doesn’t mean it’s not powerful Another misconception is that highprofile research projects can be applied directly to any business situation AI done right can create an extreme return on investments ROIs—for instance through automation or precise prediction But it does take thought time and proper implementation We have seen that success and value generated by AI projects are increased when there is a grounded understanding and expectation of what the technology can deliver from the Csuite down “Artificial Intelligence AI is a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense learn reason and take action”3 Lately there has been a big rise in the daytoday use of machines powered by AI These machines are wired using crossdisciplinary approaches based on mathematics computer science statistics psychology and more4 Virtual assistants are becoming more common most of the web shops predict your purchases many companies make use of chatbots in their customer service and many companies use algorithms to detect fraud AI and Deep Learning technology employed in office entry systems will bring proper time tracking of each employee As this system tries to learn each person with an image processing technology whose data is feed forwarded to a deep learning model where Deep learning isn’t an algorithm per se but rather a family of algorithms that implements deep networks many layers These networks are so deep that new methods of computation such as graphics processing units GPUs are required to train them in addition to clusters of compute nodes So using deep learning we can take detect the employee using face and person recognition scan and through which login logout timing is recorded Using an AI system we can even identify each employee’s entry time their working hours nonworking hours by tracking the movement of an employee in the office so that system can predict and report HR for the salary for each employee based on their working hours Our system can take feed from CCTV to track movements of employees and this system is capable of
```

```
recognizing a person even heshe is being masked as in this pandemic situation by taking their iris scan With this system installed inside the office the following are some of the benefits 1Compliancelitigation needs For several countries regulations insist that the employer must keep documents available that can demonstrate the working hours performed by each employee In the event of control from the labor inspectorate or a dispute with an employee the employer must be able to explain and justify the working hours for the company This can be made easy as our system is tracking employee movements 2Information security needs This is about monitoring user connection times to detect suspicious access times In the event where compromised credentials are used to log on at 3 am on a Saturday a notification on this access could alert the IT team that an attack is possibly underway 3Employee login logout software To manage and react to employees' attendance overtime thresholds productivity and suspicious access times our system records and stores detailed and interactive reporting on users' connection times These records allow you to better manage users' connection times and provide accurate detailed data required by management 4If you want to avoid paying overtime make sure that your employees respect certain working time quotas or even avoid suspicious access Our system will alert the HR officer about each employee's office in and out time so that they can accordingly take action 5Last but not least it reduces human resource needs to keep track of the records and sending the report to HR and HR officials has to check through the report so this system will reduce times and human resource needs With the use of AI and Deep Learning technologies we can automate some routines stuff with more functionality which humans need more resources to keep track thereby reducing time spent on manual data entry works rather companies can think of making their position high in the competitive world Blackcoffer Insights 33 Suriya E Vellore Institute of Technology '
```

convert into Tokens

```
#Tokenization
from nltk.tokenize import word_tokenize
text_tokens = word_tokenize(content)
print(text_tokens[0:50])
```

```
['When', 'people', 'hear', 'AI', 'they', 'often', 'think', 'about',
'sentient', 'robots', 'and', 'magic', 'boxes', 'AI', 'today', 'is',
'much', 'more', 'mundane', 'and', 'simple—but', 'that', 'doesn', '’',
't', 'mean', 'it', '’', 's', 'not', 'powerful', 'Another',
'misconception', 'is', 'that', 'highprofile', 'research', 'projects',
'can', 'be', 'applied', 'directly', 'to', 'any', 'business',
'situation', 'AI', 'done', 'right', 'can']
```

lenghts of tokens before removing stopwords

```
len(text_tokens)
```

```
728
```

Remove stopwords from the tokens

```
#Remove stopwords
import nltk
from nltk.corpus import stopwords
nltk.download('punkt')
nltk.download('stopwords')

my_stop_words = stopwords.words('english')
my_stop_words.append('the')
no_stop_tokens = [word for word in text_tokens if not word in
my_stop_words]
print(no_stop_tokens[0:40])
```

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

['When', 'people', 'hear', 'AI', 'often', 'think', 'sentient',
'robots', 'magic', 'boxes', 'AI', 'today', 'much', 'mundane', 'simple—
but', '’', 'mean', '’', 'powerful', 'Another', 'misconception',
'highprofile', 'research', 'projects', 'applied', 'directly',
'business', 'situation', 'AI', 'done', 'right', 'create', 'extreme',
'return', 'investments', 'ROIs—for', 'instance', 'automation',
'precise', 'prediction']

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

lenghts of tokens after removing stopwords

```
len(no_stop_tokens)
```

```
456
```

Check for positive words

```
with open("G:/Data science/assignment/text minning(done)/positive-
words.txt","r") as pos:
    poswords = pos.read().split("\n")
    poswords = poswords[5:]
```

Download the positive words dictionary and store in local system to speed up the process

```
pos_count = " ".join ([w for w in no_stop_tokens if w in poswords])
pos_count=pos_count.split(" ")
```

Positive Score

```
Positive_score=len(pos_count)
print(Positive_score)
```

```
16
```

Check for negative words

```
with open("G:/Data science/assignment/text minning(done)/negative-
words.txt","r",encoding = "ISO-8859-1") as neg:
    negwords = neg.read().split("\n")

negwords = negwords[36:]

neg_count = " ".join ([w for w in no_stop_tokens if w in negwords])
neg_count=neg_count.split(" ")
```

Negative score

```
Negative_score=len(neg_count)
print(Negative_score)
```

```
9
```

```
filter_content = ' '.join(no_stop_tokens)

data=[[url,title,content,filter_content,Positive_score,Negative_score]
]

data=pd.DataFrame(data,columns=["url","title","content","filter_conten
t","Positive_Score","Negative_Score"])
```

calculate Polarity Score & Subjectivity Score

```
from textblob import TextBlob

# Get The Subjectivity
def sentiment_analysis(data):
    sentiment = TextBlob(data["content"]).sentiment
    return pd.Series([sentiment.polarity,sentiment.subjectivity ])

# Adding Subjectivity & Polarity
data[["polarity", "subjectivity"]] = data.apply(sentiment_analysis,
axis=1)

data
```

```
                                                 url  \
0  https://insights.blackcoffer.com/how-is-login-...

                                                title  \
0  How is Login Logout Time Tracking for Employee...
```

```
                                                content  \
0   When people hear AI they often think about se...

                                          filter_content
Positive_Score  \
0  When people hear AI often think sentient robot...               16


   Negative_Score  polarity  subjectivity
0               9   0.14304      0.478514
```

Average sentence length

```
#AVG SENTENCE LENGTH
AVG_SENTENCE_LENGTH = len(content.replace('
','',))/len(re.split(r'[?!.]', content))
print('Word average =', AVG_SENTENCE_LENGTH)
```

```
Word average = 3673.0
```

```
import textstat
```

Textstat is an easy to use library to calculate statistics from text. It helps determine readability, complexity, and grade level.

FOG INDEX

```
FOG_INDEX=(textstat.gunning_fog(content))
print(FOG_INDEX)
```

```
289.41
```

AVG NUMBER OF WORDS PER SENTENCE

```
AVG_NUMBER_OF_WORDS_PER_SENTENCE = [len(l.split()) for l in
re.split(r'[?!.]', content) if l.strip()]
AVG_NUMBER_OF_WORDS_PER_SENTENCE=print(sum(AVG_NUMBER_OF_WORDS_PER_SEN
TENCE)/len(AVG_NUMBER_OF_WORDS_PER_SENTENCE))
```

```
712.0
```

COMPLEX WORD COUNT

```
def syllable_count(word):
    count = 0
    vowels = "AEIOUYaeiouy"
    if word[0] in vowels:
        count += 1
    for index in range(1, len(word)):
        if word[index] in vowels and word[index - 1] not in vowels:
            count += 1
```

```
            if word.endswith("es"or "ed"):
                count -= 1
    if count == 0:
        count += 1
    return count



COMPLEX_WORDS=syllable_count(content)
print(COMPLEX_WORDS)
```

```
1279
```

Word Count

```
Word_Count=len(content)
print(Word_Count)
```

```
4386
```

Percentage of Complex words

```
pcw=(COMPLEX_WORDS/Word_Count)*100
print(pcw)
```

```
29.1609667122663
```

Personal Pronouns

```
def ProperNounExtractor(text):
    count = 0
    sentences = nltk.sent_tokenize(text)
    for sentence in sentences:
        words = nltk.word_tokenize(sentence)
        tagged = nltk.pos_tag(words)
        for (word, tag) in tagged:
            if tag == 'PRP': # If the word is a proper noun
                count = count + 1

    return(count)



# Calling the ProperNounExtractor function to extract all the proper
nouns from the given text.
Personal_Pronouns=ProperNounExtractor(content)
```

Average Word Length

```
Average_Word_Length=len(content.replace(' ','‘))/len(content.split())
print(Average_Word_Length)
```

```
5.158707865168539
```

SYLLABLE PER WORD

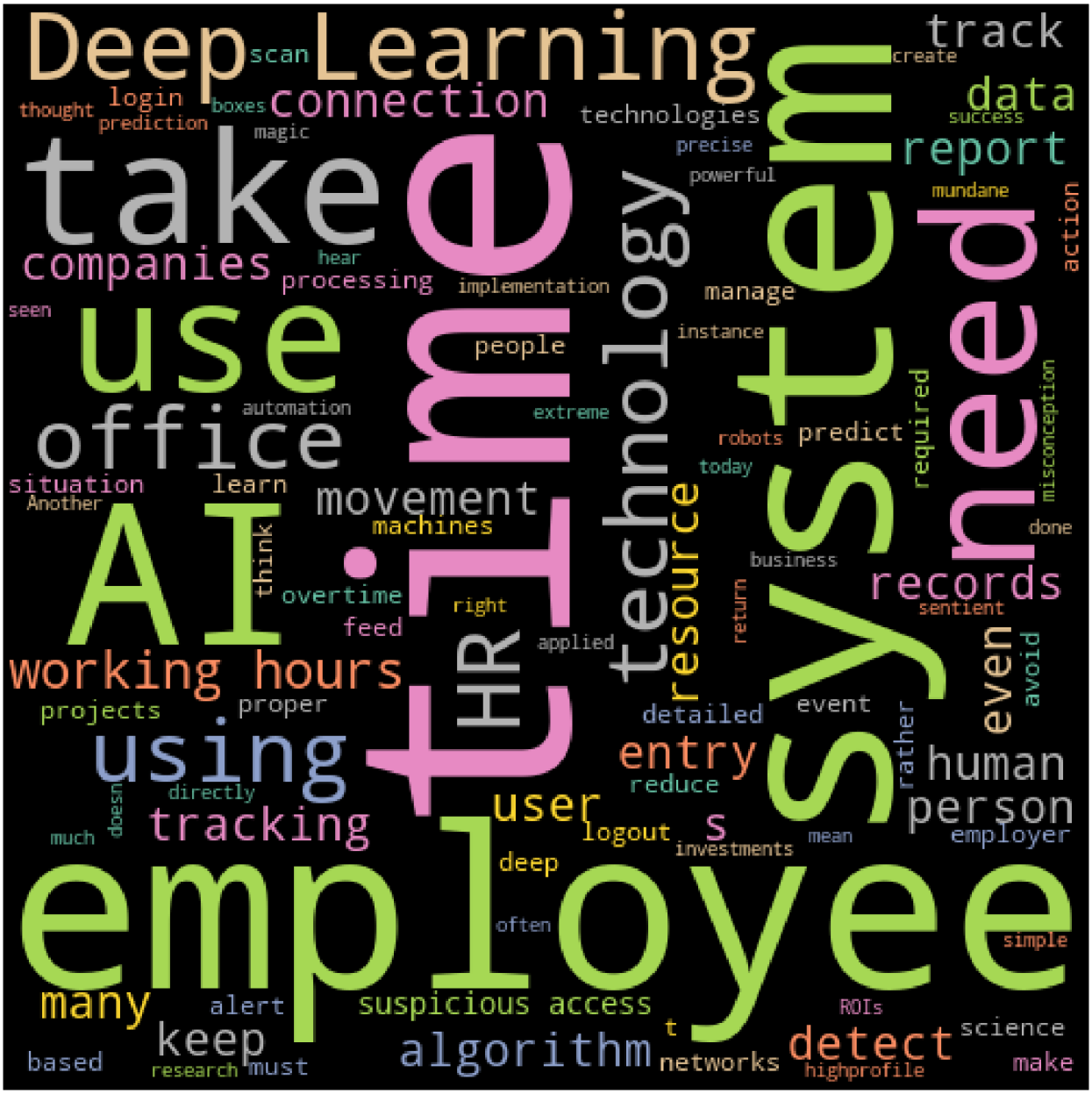
```
word=content.replace(' ','')
syllable_count=0
for w in word:
      if(w=='a' or w=='e' or w=='i' or w=='o' or w=='y' or w=='u' or
w=='A' or w=='E' or w=='I' or w=='O' or w=='U' or w=='Y'):
            syllable_count=syllable_count+1
print("The AVG number of syllables in the word is: ")
print(syllable_count/len(content.split()))
```

```
The AVG number of syllables in the word is:
2.109550561797753
```

For WordCloud

```
# Import packages
import matplotlib.pyplot as plt
%matplotlib inline
from wordcloud import WordCloud, STOPWORDS
# Define a function to plot word cloud
def plot_cloud(wordcloud):
    # Set figure size
    plt.figure(figsize=(40, 30))
    # Display image
    plt.imshow(wordcloud)
    # No axis details
    plt.axis("off");

# Generate wordcloud
stopwords = STOPWORDS
stopwords.add('will')
wordcloud = WordCloud(width = 500, height = 500,
background_color='black',
max_words=100,colormap='Set2',stopwords=stopwords).generate(content)
# Plot
plot_cloud(wordcloud)
```



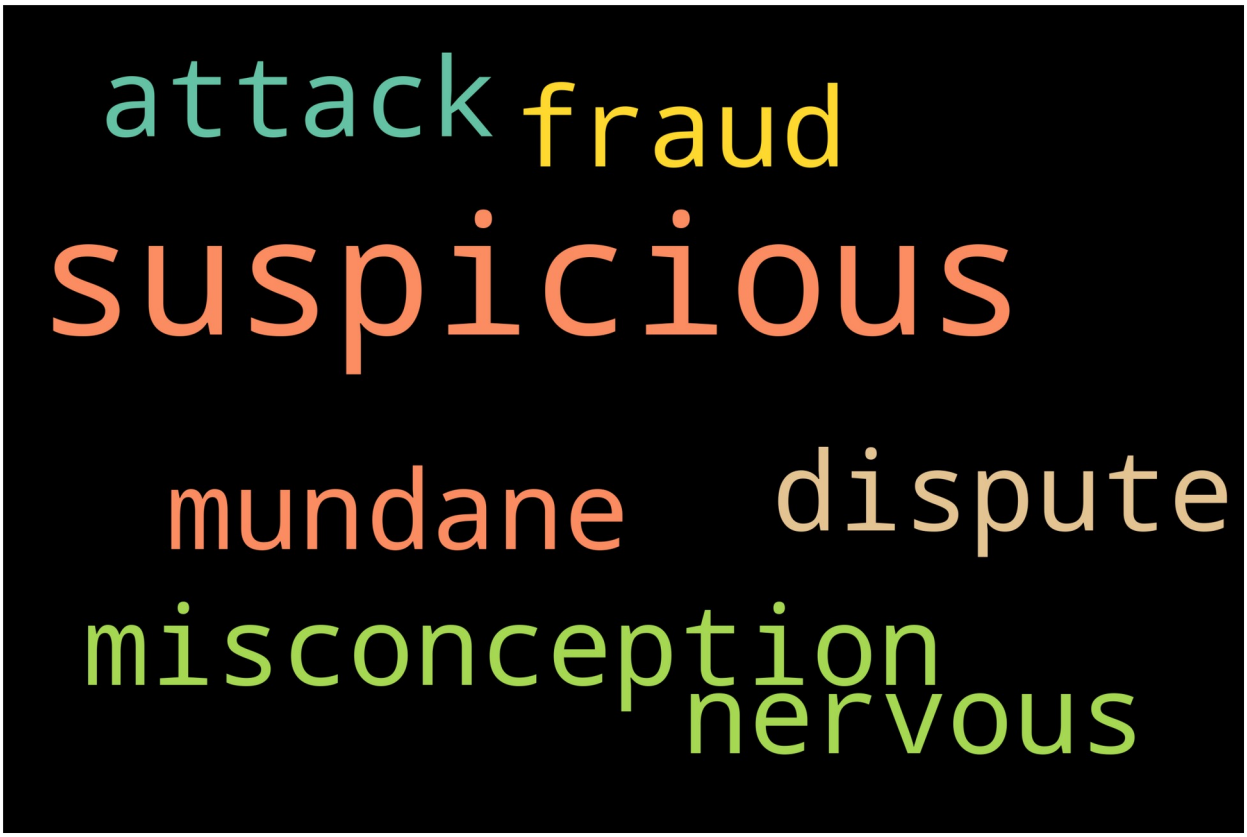
positive word cloud

```
# Choosing the only words which are present in posword
pos_review = " ".join ([w for w in pos_count if w in poswords])
wordcloud = WordCloud(width = 3000, height = 2000,
background_color='black',
max_words=100,colormap='Set2',stopwords=stopwords).generate(pos_review
)
#Plot
plot_cloud(wordcloud)
```

negative word cloud

```
# Choosing the only words which are present in negwords
neg_review = " ".join ([w for w in neg_count if w in negwords])

wordcloud = WordCloud(width = 3000, height = 2000,
background_color='black',
max_words=100,colormap='Set2',stopwords=stopwords).generate(neg_review
)
#Plot
plot_cloud(wordcloud)
```


attack
fraud
suspicious
mundane
dispute
misconception
nervous