

# Assignment7\_B

April 22, 2020

```
[1]: # Roll no: 33140 (Sahil Naphade)
# Batch: L9
# Assignment 7(B): WordCloud on Mark Zuckerberg text file
import pandas as pd
from wordcloud import wordcloud
import nltk
import numpy as np
from nltk import word_tokenize
import matplotlib.pyplot as plt
from nltk.corpus import stopwords
# nltk.download('stopwords')

[2]: data = open("G:/College/Sl-VI DataSets/TextMining/MarkZuckerberg.txt", "r").read()
data = data.lower()
print(data)
```

mark elliot zuckerberg (/ˈzʔkʔrbʔʔrg/; born may 14, 1984) is an american computer programmer and internet entrepreneur. he is a co-founder of facebook, and is currently its chairman and chief executive officer.[4][5] his net worth is estimated to be us\$62.2 billion as of march 25, 2018.[6]

zuckerberg launched facebook from his harvard university dormitory room on february 4, 2004 with college roommates and fellow harvard students eduardo saverin, andrew mccollum, dustin moskovitz, and chris hughes.[7] the group then introduced facebook to other college campuses. facebook expanded rapidly, reaching one billion users by 2012. during this time, zuckerberg became involved in various legal disputes brought by his friends and cofounders, who claimed they were due a share of the company based upon their involvement during its development phase.[8]

since 2010, time magazine has named zuckerberg among the 100 wealthiest and most influential people in the world as a part of its person of the year award.[3][9][10] in december 2016, zuckerberg was ranked 10th on forbes list of the world's most powerful

```
[3]: stop = stopwords.words('english')
```

```
[4]: # Function to remove numbers
def remove_num(text):
    text = ''.join([i for i in text if not i.isdigit()])
    return text

# Function to remove special characters and punctuations
def remove_u(text):
    text = text.replace('_', '')
    text = text.replace('?', '')
    text = text.replace('.', '')
    text = text.replace("@", '')
    text = text.replace(' ', '')
    text = text.replace("'", '')
    text = text.replace('"', '')
    text = text.replace('zkrbrg', '')
    text = text.replace('[', '')
    text = text.replace(']', '')
    text = text.replace(", ", "")
    return text
```

```
[5]: #executing functions to clean data
data = np.vectorize(remove_u)(data)
print(data)
```

mark elliot zuckerberg (//; born may 14 1984) is an american computer programmer and internet entrepreneur. he is a co-founder of facebook and is currently its chairman and chief executive officer.<sup>45</sup> his net worth is estimated to be us\$62.2 billion as of march 25 2018.<sup>6</sup>

zuckerberg launched facebook from his harvard university dormitory room on february 4 2004 with college roommates and fellow harvard students eduardo saverin andrew mccollum dustin moskovitz and chris hughes.<sup>7</sup> the group then introduced facebook to other college campuses. facebook expanded rapidly reaching one billion users by 2012. during this time zuckerberg became involved in various legal disputes brought by his friends and cofounders who claimed they were due a share of the company based upon their involvement during its development phase.<sup>8</sup>

since 2010 time magazine has named zuckerberg among the 100 wealthiest and most influential people in the world as a part of its person of the year award.<sup>39</sup> in december 2016 zuckerberg was ranked 10th on forbes list of the worlds most powerful

```
[6]: data = np.vectorize(remove_num)(data)
print(data)
```

mark elliot zuckerberg (//; born may ) is an american computer programmer and internet entrepreneur. he is a co-founder of facebook and is currently its chairman and chief executive officer. his net worth is estimated to be us\$.

billion as of march .

zuckerberg launched facebook from his harvard university dormitory room on february with college roommates and fellow harvard students eduardo saverin andrew mccollum dustin moskovitz and chris hughes. the group then introduced facebook to other college campuses. facebook expanded rapidly reaching one billion users by . during this time zuckerberg became involved in various legal disputes brought by his friends and cofounders who claimed they were due a share of the company based upon their involvement during its development phase.

since time magazine has named zuckerberg among the wealthiest and most influential people in the world as a part of its person of the year award. in december zuckerberg was ranked th on forbes list of the worlds most powerful

```
[7]: #removing extra spaces
def remove_extra_space(text):
    word_list = text.split()
    text = ' '.join(word_list)
    return text

#executing function
data = np.vectorize(remove_extra_space)(data)

# Convert data to list
data = data.tolist()
```

```
[8]: def create_word_cloud(string):
    cloud = wordcloud.WordCloud(width=800,height=800,background_color = "
↳"white", stopwords = stop, max_words = 20).generate(string)
    pplt.figure(figsize = (10, 10), facecolor = None)
    pplt.axis('off')
    pplt.imshow(cloud)
```

```
[9]: create_word_cloud(data)
```

entrepreneur programmer  
zuckerberg  
college founder  
currently may computer co  
facebook born  
world american  
time billion  
mark chairman  
harvard  
internet elliot

[ ]: