Week 10 Deliverables

Group Details

- Name: William Harvey
- Email: will.harvey@att.net
- Country: United States of America
- College: University of California, Berkeley
- Specialization: NLP Analyst

Problem Description

- Advanced NLP working with detecting hate speech.

EDA

- I created wordframes to show the most popular words in hate comments versus non-hate comments

Week 10: Deliverables | Untitled.ipynb (auto-G) - Jupyt

data100.datahub.berkeley.edu/user/wsharvey/lab/workspaces/auto-G/tree/DataGlacier/Final%20Proj/Untitled.ipynb

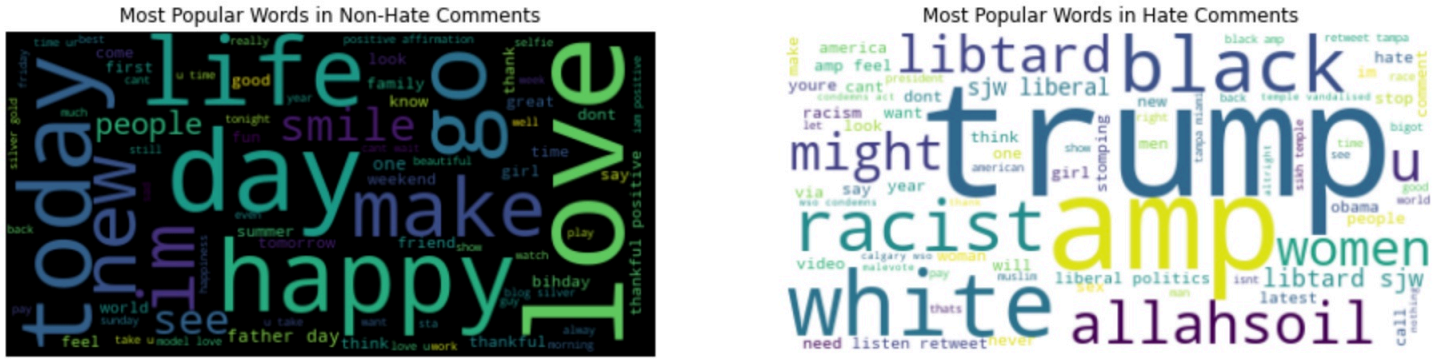
File Edit View Run Kernel Tabs Settings Help Share

Untitled.ipynb | train.csv | test_tweets_anuFYb8.csv

Code | Interface | Python 3 (ipykernel)

[78]:
```
from wordcloud import WordCloud
import matplotlib.pyplot as plt



no_hate = [token for tweet in train_tweets[train_tweets['label'] == 0]['Tweet'] for token in tweet]
hate = [token for tweet in train_tweets[train_tweets['label'] == 1]['Tweet'] for token in tweet]

# Create and generate a word cloud image:
wordcloud_hate = WordCloud(max_words=75, background_color="white").generate(' '.join(hate))
wordcloud_no_hate = WordCloud(max_words=75, background_color="black").generate(' '.join(no_hate))

# Display the generated images:
fig, axs = plt.subplots(1,2 , figsize=(16,8))
axs[0].imshow(wordcloud_no_hate, interpolation='bilinear')
axs[0].axis('off')
axs[0].set_title('Most Popular Words in Non-Hate Comments')
axs[1].imshow(wordcloud, interpolation='bilinear')
axs[1].axis('off')
axs[1].set_title('Most Popular Words in Hate Comments')
```

[78]: Text(0.5, 1.0, 'Most Popular Words in Hate Comments')



[ ]:

Simple | 0 | 1 | No Kernel | Idle | Mem: 636.95 / 2048.00 MB | Mode: Edit | Ln 12, Col 1 | Untitled.ipynb

Repo Link

https://github.com/wsharvey/DG-Harvey/blob/main/DG_Proj%20(1).ipynb