INDIVIDUAL PROJECT – APPLIED DATA SCIENCE

HURMET NOKA | 18TH OF MAY 2021

RELIGION OF TWITTER – TOPIC MODELING, SENTIMENT ANALYSIS AND HATE SPEECH DETECTION.

INTRODUCTION

Many are those claiming that the light of scientific and technological revolution will diminish the presence of religion in human life. Yet, it is 2021, and one thing seems inevitable, religion is still here and is here to stay! That being said, it is worth investigating the public opinion towards religion in modern times.

This article presents a comprehensive guide of using Python to extract, pre-process, and analyze tweets about religion. The analysis is about implementing *Topic Modeling* (LDA), *Sentiment Analysis* (Gensim), and *Hate Speech Detection* (HateSonar) models. The step-by-step tutorial is presented below alongside the code and results. The **complete code and data** can be found on my GitHub profile *here.* I also published an article on Towards Data Science which can be found *here*.

DATA

Figure 1: Castle of Liechtenstein

To get the tweets, we use a public python script, which enables capturing old tweets, thus bypassing the limitation of the 7-days period of Twitter API. The script is free and can be found here on GitHub. All you need to do is adjust the searching filters and run the program. For our study, we extract tweets containing the phrase "religion is." To reduce the bias of certain isolated events affecting the feeling towards religion (i.e., Charlie Hebdo attacks), we extend the timeframe to about five years. We extracted 1000 tweets/month starting from January 2015 until October 2019. This results in about **57'351 tweets** that are then loaded into a dataframe, ready for pre-processing.

Below is the whole analysis process in a schematic format

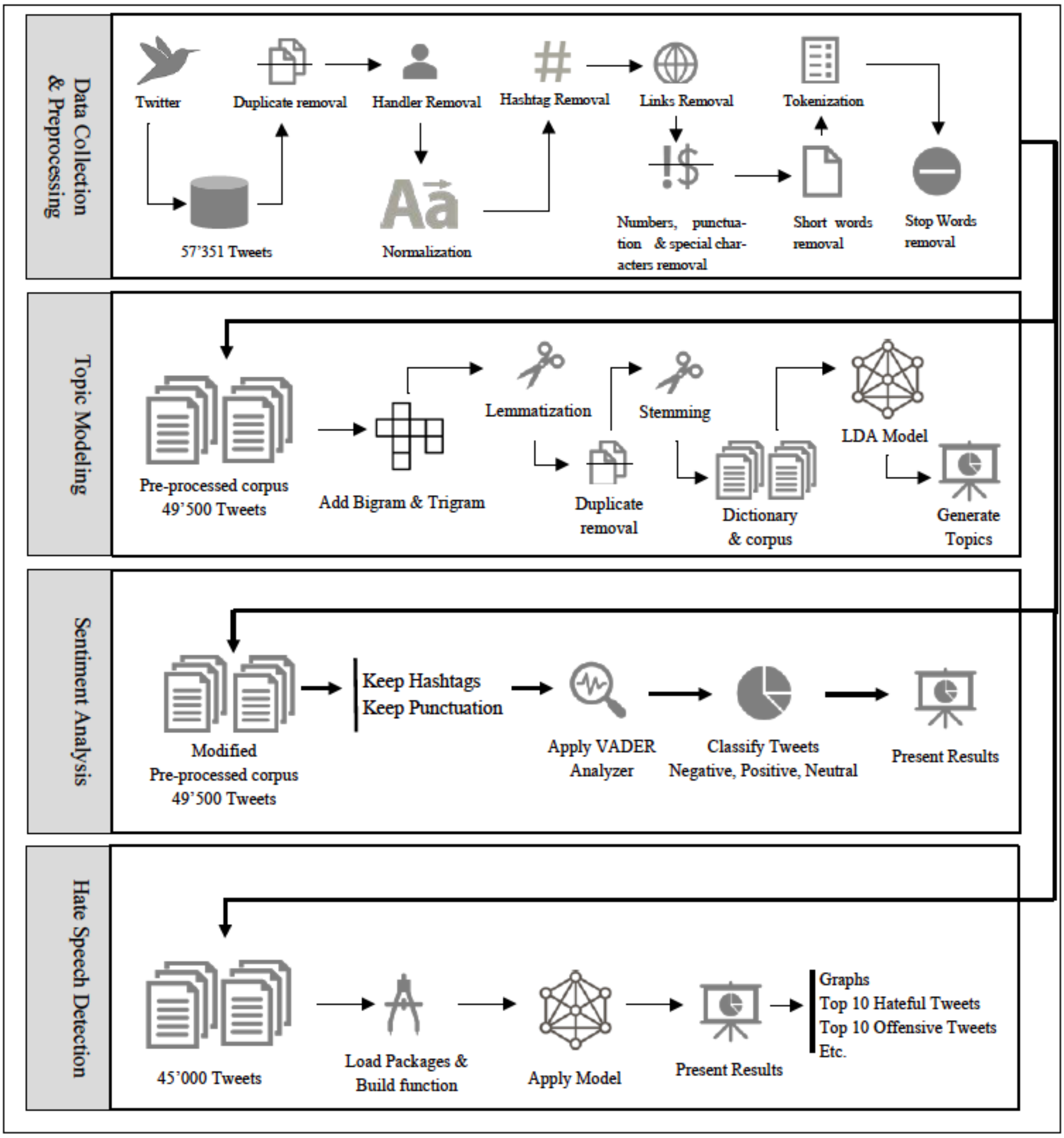


Figure 2: Analysis Steps

PRE-PROCESSING

Most of the Twitter data, as most of the text data, are highly unstructured and contain noise. There could be typos, slang usage, and grammar mistakes. Then cleansing steps are applied to the documents to generate structured data, which then are subjected to the model (Yang & Zhang, 2018, p. 527). The steps undertaken for the preprocessing phase are different according to the model intended to be applied. For the topic modeling, more preprocessing steps are required; these steps are listed in table below and will be sequentially described in more detail.

No.	Step
1	Remove Duplicates
2	Remove (@) Users
3	Normalization
4	Remove Hashtags
5	Remove Links
6	Remove Collection Words
7	Remove Punctuation
8	Remove Numbers
9	Remove Special Characters
10	Remove Short Words
11	Tokenization
12	Remove Stop Words
13	Remove Tweets less than 3 tokens

Table 1: Pre-processing Steps

These are the necessary packages to perform pre-processing phase.

```
import pandas as pd
import numpy as np
import re

# Plotting
import seaborn as sns
import matplotlib.pyplot as plt

# Gensim
import gensim
from gensim.utils import simple_preprocess

# NLTK
import nltk
from nltk.corpus import stopwords

from collections import Counter
from wordcloud import WordCloud

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

%matplotlib inline
```

```
df = pd.read_csv('full_data.csv', index_col=[0])
df.head()
```

	tweet_text
0	Islam is a religion full of blessings and good...
1	@ wagner_claire Religion and pity are antonyms...
2	if the only reason people are on twitter is to...
3	@15MeterClassYas @NotoriousDachi is there such...
4	Religion is needed to provide a moral compass ...

Data Overview

Before any further step, we remove the duplicates as it is common in twitter that people copy paste different quotes and retweet something. After duplicate removal, we get a total of 53'939 unique tweets.

```
# DROP DUPLICATES
df.drop_duplicates(subset=['tweet_text'], keep='first',inplace=True)
df.shape
Out: (53939, 1)
```

Next we generate some general descriptive statistics:

Count total number of characters and mean length of a tweet

```
count = df['tweet_text'].str.split().str.len()
count.index = count.index.astype(str) + ' words:'
count.sort_index(inplace=True)
```

```
print("Total number of words:", count.sum(), "words")
```

```
Total number of words: 1350392 words
```

```
print("Mean number of words per tweet:", round(count.mean(),2), "words")
```

```
Mean number of words per tweet: 25.04 words
```

```
df["tweet_length"] = df["tweet_text"].str.len()
print("Total length of the dataset is:", df.tweet_length.sum(), "characters")
```

```
Total length of the dataset is: 8278911 characters
```

```
print("Mean Length of a tweet is:", round(df.tweet_length.mean(),0), "characters")
df = df.drop(['tweet_length'], axis=1)
```

```
Mean Length of a tweet is: 153.0 characters
```

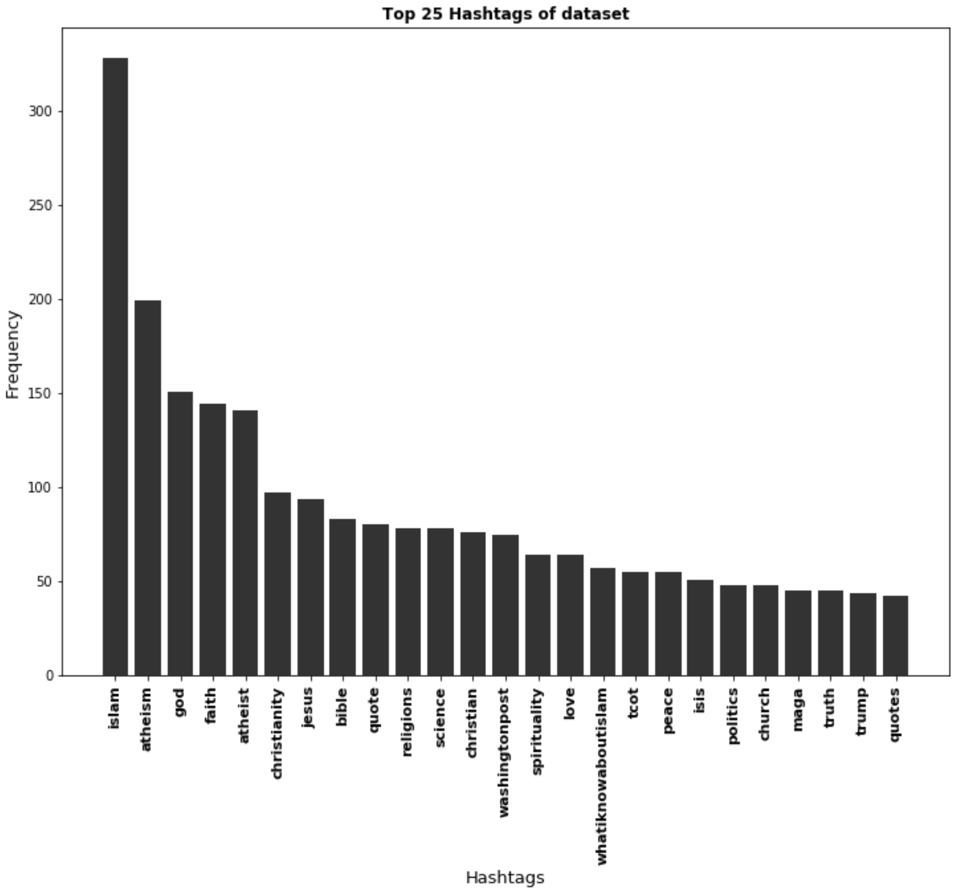
Twitter enables including usernames within tweets through the symbol "@." These do not possess any value for our analysis; hence they are removed from the dataset using a function.

```
# REMOVE '@USER'
def remove_users(tweet, pattern1, pattern2):
    r = re.findall(pattern1, tweet)
    for i in r:
        tweet = re.sub(i, '', tweet)

    r = re.findall(pattern2, tweet)
    for i in r:
        tweet = re.sub(i, '', tweet)
    return tweet

df['tidy_tweet'] = np.vectorize(remove_users)(df['tweet_text'],
"@ [\w]*", "@[\w]*")
```

Normalization, a method which converts all tweets into lowercase, so that “Token” and “token” are not considered two different words.

```
# NORMALIZATION
df['tidy_tweet'] = df['tidy_tweet'].str.lower()
```

Same as with usernames, hashtags also are considered not of significant value for topic modeling analysis, in particular, therefore, are removed.



Above are the top 25 hashtags of the dataset, we note that #islam is the most used hashtag, indicating it is the most discussed religion in particular. Hashtags are removed using the function:

```
# REMOVE HASHTAGS
def remove_hashtags(tweet, pattern1, pattern2):
    r = re.findall(pattern1, tweet)
    for i in r:
        tweet = re.sub(i, '', tweet)

    r = re.findall(pattern2, tweet)
    for i in r:
        tweet = re.sub(i, '', tweet)
    return tweet


df['tidy_tweet'] = np.vectorize(remove_hashtags)(df['tidy_tweet'],
"# [\w]*", "#[\w]*")
```

Next in line to be removed are URL's:

```
# REMOVE LINKS
def remove_links(tweet):
    tweet_no_link = re.sub(r"http\S+", "", tweet)
    return tweet_no_link


df['tidy_tweet'] = np.vectorize(remove_links)(df['tidy_tweet'])
```

Using the same function as with links, we also remove collection words, those words used for filtering the tweets in the first place, in this case: 'religion,' 'religious.'

Afterward, it is necessary to remove numbers, punctuation (only for topic modeling), and special characters (@,&,#,%, etc.).

```
# REMOVE Punctuations, Numbers, and Special Characters
df['tidy_tweet'] = df['tidy_tweet'].str.replace("[^a-zA-Z#]", " ")
```

Then words with less than three characters (short words) are removed from the dataset, simplifying feature extraction for the analysis.

```
# REMOVE SHORT WORDS
df['tidy_tweet'] = df['tidy_tweet'].apply(lambda x:' '.join([w for w
in x.split() if len(w)>3]))
```

An essential step of pre-processing is known as Tokenization. It is the process where the text is split according to whitespaces, and every word and punctuation is saved as a separate token. We perform this step using the simple_preprocess method from Gensim.

```
# TOKENIZATION
def tokenize(tweet):
    for word in tweet:
        yield(gensim.utils.simple_preprocess(str(word), deacc=True))


df['tidy_tweet_tokens'] = list(tokenize(df['tidy_tweet']))
```

Next, we remove stopwords that have no analytic value, usually articles, prepositions, or pronouns, for instance, 'a,' 'and,' 'the,' etc. The default list can be adjusted and extended as desired. We added some new words to the predefined list of Natural Language Toolkit (NLTK), which contains 179 words.

```
# Prepare Stop Words
stop_words = stopwords.words('english')
stop_words.extend(['from', 'https', 'twitter', 'religions',
'pic','twitt',])

# REMOVE STOPWORDS
def remove_stopwords(tweets):
    return [[word for word in simple_preprocess(str(tweet)) if word
not in stop_words] for tweet in tweets]

df['tokens_no_stop'] = remove_stopwords(df['tidy_tweet_tokens'])
```

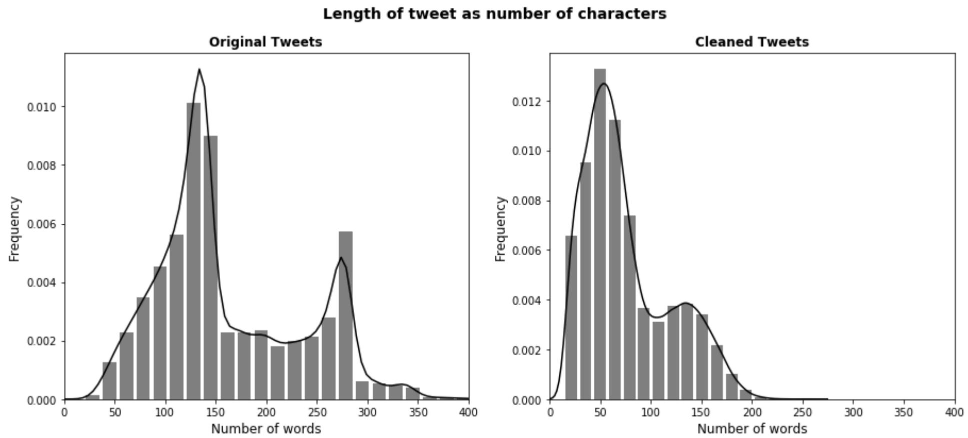
In the last step of pre-processing, we remove tweets with less than three tokens; this results in a total of 49'458 tweets to be considered further on the analysis phase. For topic modeling and also sentiment analysis, documents with less than three tokens are not suitable to generate enough information.

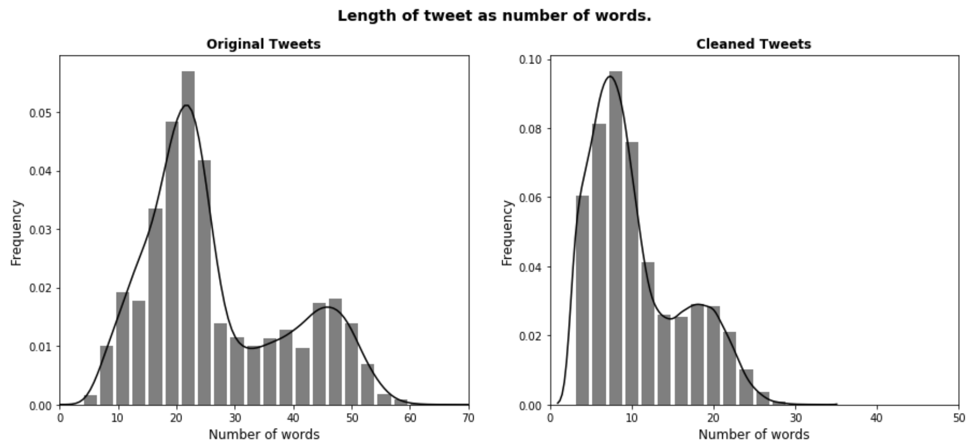
```
# REMOVE TWEETS LESS THAN 3 TOKENS
df['length'] = df['tokens_no_stop'].apply(len)
df = df.drop(df[df['length']<3].index)
df = df.drop(['length'], axis=1)
df.shape

df.reset_index(drop=True, inplace=True)
```

Now for an illustrative purpose to identify and visualize the most used words of the dataset, we generate a WordCloud of 200 most frequent words.

Figure 3: Word-Cloud of the All tweets

The preprocessing phase effects are illustrated below, where the length of cleaned tweets is reduced considerably. As shown, most tweets have less than ten tokens after preprocessing, unlike in original tweets where most tweets have about 20 words. On the second set of graphs, the effects are more visible as the tweet length went from about 150 for most of the tweets to about 50 characters after the cleaning stage. This phase is critical as it reduces the dimensionality and results in significantly valuable tokens for the models, which is explained in the next sections.





Lastly we save the pre-processed dataframe as a pickle, which is then used for Topic Modeling phase. That concludes the pre-processing phase of the analysis.

```
df.to_pickle('pre-processed.pkl')
```

TOPIC MODELING – LDA

These are the required packages to implement LDA (Latent Dirichlet Allocation) algorithm.

```
# IMPORTS
import pandas as pd
import numpy as np
import networkx as nx
import itertools
import collections
import spacy
from pprint import pprint

# Plotting
import matplotlib.pyplot as plt
import seaborn as sns
import pyLDAvis
import pyLDAvis.gensim

# Gensim
import gensim
import gensim.corpora as corpora
from gensim.utils import simple_preprocess
from gensim.models import CoherenceModel
from gensim.models.wrappers import LdaMallet

# NLTK
from nltk import bigrams
from nltk.stem import PorterStemmer

sns.set(font_scale=1.5)
sns.set_style("whitegrid")

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

We are not going to dive into an explanation of how LDA works; details can be found in the original paper by [1]. The core idea of the LDA algorithm can be understood as a generative process where documents are defined by a probability distribution over a set of topics T, and a probability distribution of discrete words, in turn, establishes each topic. Having said that, we load the pre-processed data:

	tweet_text	tidy_tweet	tidy_tweet_tokens	tokens_no_stop	no_stop_joined
0	Islam is a religion full of blessings and good...	islam full blessings good deeds which also res...	[islam, full, blessings, good, deeds, which, a...	[islam, full, blessings, good, deeds, also, re...	islam full blessings good deeds also respected...
1	@ wagner_claire Religion and pity are antonyms...	pity antonyms disaster other humanity live wit...	[pity, antonyms, disaster, other, humanity, li...	[pity, antonyms, disaster, humanity, live, wit...	pity antonyms disaster humanity live without pity
2	if the only reason people are on twitter is to...	only reason people twitter argue about hitler ...	[only, reason, people, twitter, argue, about, ...	[reason, people, argue, hitler, count]	reason people argue hitler count
3	Religion is needed to provide a moral compass ...	needed provide moral compass those lack empathy	[needed, provide, moral, compass, those, lack,...	[needed, provide, moral, compass, lack, empathy]	needed provide moral compass lack empathy
4	plus my religion has never been 'holier' than ...	plus never been holier than regular life regul...	[plus, never, been, holier, than, regular, lif...	[plus, never, holier, regular, life, regular, ...	plus never holier regular life regular life in...

After data is loaded, we proceed by adding Bigrams and Trigrams. A sequence of words which often occurs together when expressing a particular meaning. A sequence of N words is known as N-Grams, as theoretically N can be of any length; the most common are pairs of words (Bi-grams) and a sequence of three words (Tri-grams). First, we need to tokenize the no_stop_joined column and convert it to a list holding the tokens for each tweet; we name this list data_words as shown below:

```
# TOKENIZE
def sent_to_words(sentences):
    for sentence in sentences:
        yield(gensim.utils.simple_preprocess(str(sentence),
deacc=True))  # deacc=True removes punctuations

data_words = list(sent_to_words(data))
```

Now we are ready to add bigrams and trigrams to our corpus.

```
# Build the bigram and trigram model
bigram = gensim.models.Phrases(data_words, min_count=10,
threshold=100)
trigram = gensim.models.Phrases(bigram[data_words], threshold=100)

# Faster way to get a sentence clubbed as a bigram
bigram_mod = gensim.models.phrases.Phraser(bigram)
trigram_mod = gensim.models.phrases.Phraser(trigram)

def make_bigrams(texts):
    return [bigram_mod[doc] for doc in texts]

def make_trigrams(texts):
    return [trigram_mod[bigram_mod[doc]] for doc in texts]

# Form Bigrams
data_words_bigrams = make_bigrams(data_words)
```

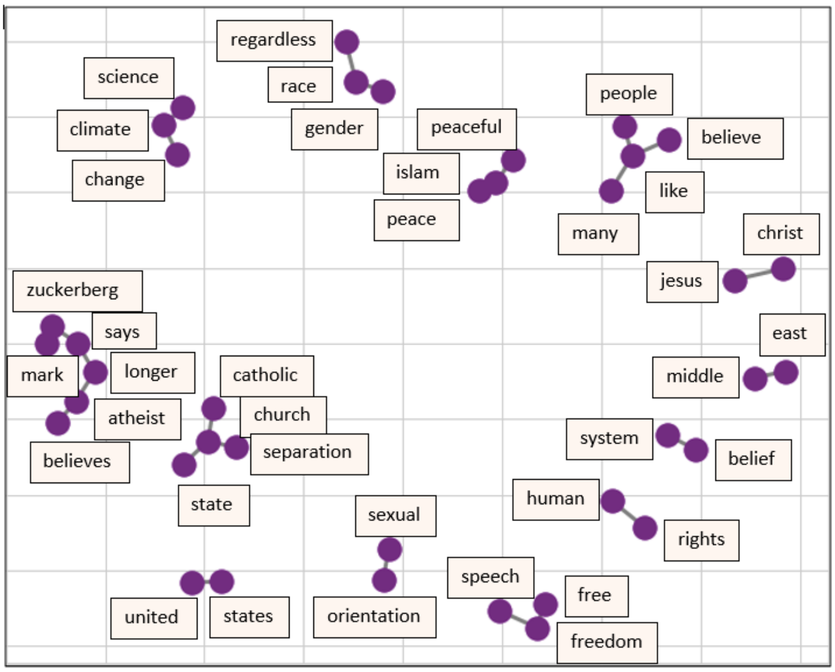


Figure 4: Bigrams

The next step is lemmatization, an essential step for many applications of text mining. Lemmatization takes into consideration the context and converts the word to its base form; for instance, the term "hugging" is converted to "hug" and "best" to "good." For the lemmatization task, the package used is spaCy, an open-source library with many pre-built models for natural language processing.

```
# LEMMATIZATION
def lemmatization(tweets, allowed_postags=['NOUN', 'ADJ', 'VERB',
'ADV']):
    """https://spacy.io/api/annotation"""
    tweets_out = []
    for sent in tweets:
        doc = nlp(" ".join(sent))
        tweets_out.append([token.lemma_ for token in doc if
token.pos_ in allowed_postags])
    return tweets_out


# Initialize spacy 'en' model, keeping only tagger component
# python3 -m spacy download en
nlp = spacy.load('en', disable=['parser', 'ner'])


# Lemmatization keeping only noun, adj, vb, adv
df['lemmatized'] = pd.Series(lemmatization(data_words_bigrams,
allowed_postags=['NOUN', 'ADJ', 'VERB', 'ADV']))
```

After this step, we dropped duplicates once more, as short tweets composed of very few tokens when lemmatized may lead to duplicate rows. After this step, we have 48'013 unique rows.

Another useful technique is word stemming, which is the process of transforming a word into its root form. Unlike lemmatization, which was mentioned earlier, stemming is a more aggressive approach as suffix cuttings often result in non-meaningful English words. For instance, the word "animals" would be lemmatized as "animal," but the Porter Stemmer yields "anim." We decided to implement both to help in dimensionality reduction.

```
# STEMMING
stemmer = PorterStemmer()
df['stemmed'] = df['lemmatized'].apply(lambda x : [stemmer.stem(y)
for y in x])
```

Before building the LDA model, we have to create two main inputs: the dictionary and the corpus, which are created using functions from the Gensim package.

```
# Create Dictionary
id2word_stemmed = corpora.Dictionary(df['stemmed'])
IN: print(id2word_stemmed)
OUT: Dictionary(26748 unique tokens: ['also', 'bless', 'blood',
'deed', 'fact']...)\

# Create Corpus
tweets_stemmed = df['stemmed']
IN: df['stemmed'][1]
OUT: ['piti', 'antonym', 'disast', 'human', 'live', 'piti']
```

Here is how corpus looks like, it has a length of 48'013:

```
tweets_stemmed
```

```
0         [islam, full, bless, good, deed, also, respect...
1                [piti, antonym, disast, human, live, piti]
2                      [reason, peopl, argu, hitler, count]
3              [need, provid, moral_compass, lack, empathi]
4         [never, holi, regular, life, regular, life, in...
5         [impot, human, mind, deal, occurr, not, unders...
6         [other, would, do, univers, matter, cultur, be...
7         [principl, follow, ahimsa, satyagraha, expel, ...
8         [whole, peac, thing, start, make, trend, musli...
9             [jorg, lui, borg, fall, love, creat, fallibl]
10             [christian, call, piti, friedrich, nietzsch]
```

Gensim assigns a unique Id to each word, and then the corpus is represented as a tuple (word_id, word_frequency).

```
# Term Document Frequency
corpus_stemmed = [id2word_stemmed.doc2bow(tweet) for tweet in
tweets_stemmed]
```

For instance, the following unprocessed tweet:

'@ wagner_claire Religion and pity are antonyms. One is disaster, other is humanity. We can live without religion but not pity. Yet we hug …'

After pre-processing in the corpus will be presented as a list of tuples:

[(11, 1), (12, 1), (13, 1), (14, 1), (15, 2)]

BUILD LDA MODEL

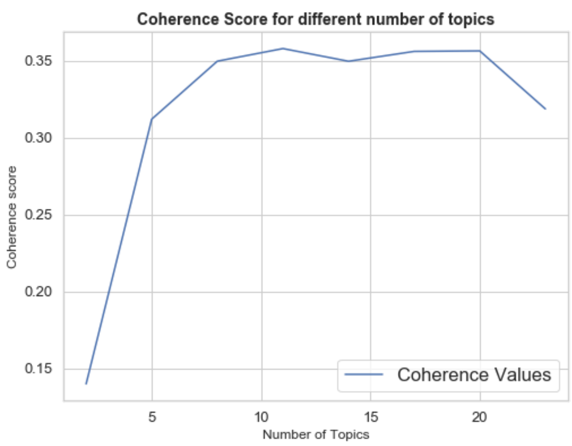
Now it is the moment that we initialize with number of topics k=10, which then will be subject of adjustments.

```
# Build LDA model
lda_model_stemmed =
gensim.models.ldamodel.LdaModel(corpus=corpus_stemmed,
                                           id2word=id2word_stemmed,
                                           num_topics=10,
                                           random_state=100,
                                           update_every=1,
                                           chunksize=100,
                                           passes=15,
                                           alpha='auto',
                                           per_word_topics=True)
```

After this step, we may directly generate the topics or search for the optimal model, using coherence score as a measure for each model having a different number of topics.

```
# OPTIMAL MODEL
def compute_coherence_values(dictionary, corpus, texts, limit,
start=2, step=3):

    coherence_values = []
    model_list = []
    for num_topics in range(start, limit, step):
        model = gensim.models.wrappers.LdaMallet(mallet_path,
corpus=corpus_stemmed, num_topics=num_topics,
id2word=id2word_stemmed)
        model_list.append(model)
        coherencemodel = CoherenceModel(model=model, texts=texts,
dictionary=dictionary, coherence='c_v')
        coherence_values.append(coherencemodel.get_coherence())

return model_list, coherence_values

model_list, coherence_values =
compute_coherence_values(dictionary=id2word_stemmed,
corpus=corpus_stemmed, texts=df['stemmed'], start=2, limit=26,
step=3)
```

Below is illustrated the coherence score for each model generated, the optimal model we choose has k=8 topics.



After building the model and running the optimal model we yeild the following topics:

Topic		Keywords
0	Religion & Politics	polit, freedom, govern, nation, forc, support, trump, america, control
1	Christianity	christian, true, call, follow, church, mani, teach, atheist, fuck, cathol,
2	Religion & Science	believ, human, faith, belief, culture, scienc, part, fact, find, bibl
3	Religious Doctrine	love, good, noth, someon, anyon, truth, child, anyth, evil, cult
4	Religion & Society	peopl, race, hate, person, differ, matter, everyon, base, stop, realli
5	Islam	islam, muslim, world, peac, countri, women, kill, read, book, leav
6	Personal Belief	peopl, thing, wrong, problem, reason, understand, respect, make, agre, practic
7	Diverse Opinions	make, life, time, give, live, talk, real, feel, question, year

Table 12: Topics addressed in online discussion

Interpretation:

Topic **[0]** Religion & Politics is self-evidently represented by terms like politics, control, government, nation, trump, etc., which illustrate how religion is a sensitive aspect of its role related to politics.

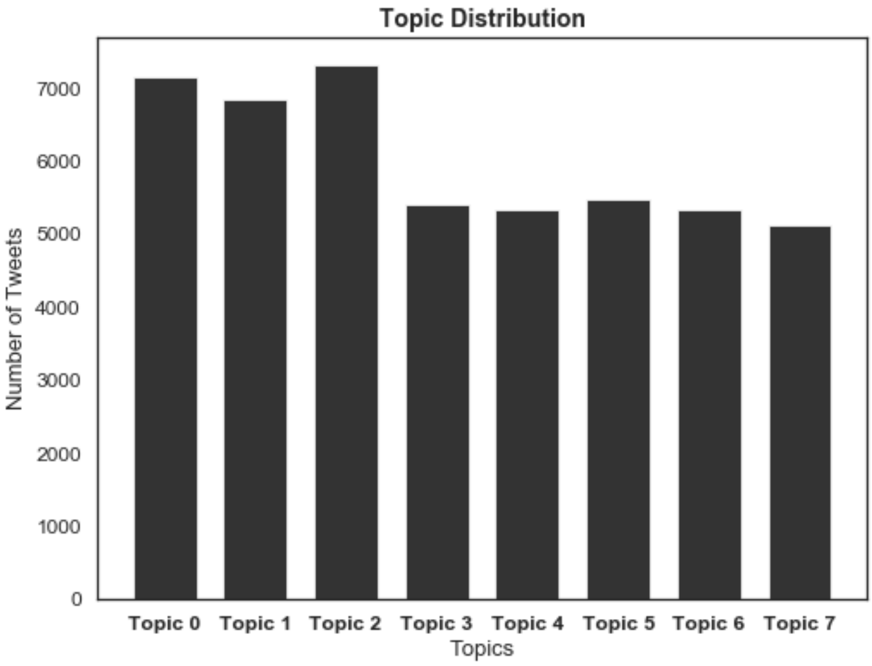
Topic **[1]** Christianity is constituted with tweets which contain Christianity as the subject of the discussions and its representative key-terms such as church, catholic, christian etc., such topic is expected for an English speaking audience mainly belonging to western world, having Christianity as the leading religion.

Topic **[2]** Religion & Science is concerned with the never ending debate between religion and science, a debate intensified in the modern world which is led by a technological revolution.

Topic **[3]** Religious Doctrine , **[6]** Personal Belief and **[7]** Diverse Opinions seem to be more closer to each other with slight differences, based on the keywords.

Topic **[5]** Islam is a collection of tweets with Islam as the main subject of discussion. As seen by the number of hashtags and the word-frequency statistics, Islam is highly discussed on social media for its different aspects, which trigger diverse reactions among people.

Below is given the distribution of tweets among topics, we can notice that the first three topics are more dominant:



SENTIMENT ANALYSIS (VADER)

For this part the required packages are as follows:

```
# IMPORTS
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import spacy
import re
from pprint import pprint

import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.corpus import stopwords

from collections import Counter
from wordcloud import WordCloud

import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

%matplotlib inline
```

VADER stands for Valence Aware Dictionary for sEntiment Reasoning and was developed by [2] as a rule-based model for sentiment analysis. Detailed information how this model was developed and its features can be found on original paper (see References). It takes into consideration punctuation, capitalization, degree modifiers, conjunctions, Tri-gram preceding when assigning sentiment values: negative, positive, neutral. These features make VADER sentiment analyzer achieve remarkable results when classifying social media texts like tweets and make it a suitable tool to conduct our analysis.

The pre-processing phase is slightly different from the one employed for Topic Modeling because of the features that VADER sentiment analyzer emodies, the steps taken are listed below, whereas the code is the same as in the pre-processing phase.

- Data Loading
    Total Tweets: 49421.
- Pre-Processing (Different from Topic Modelling)
    - Links Removal
    - Handles (@User) Removal
    - Hashtag Symbol Removal (We keep hashtags for that they contain sentiment value)
    - Collection Words Removal
    - We Keep Punctuation as VADER takes them into consideration
    - Remove Duplicates [New Dataset size: 48580]

```
df.head()
```

	tweet_text	tokens_no_stop
0	Islam is a religion full of blessings and good...	[islam, full, blessings, good, deeds, also, re...
1	@ wagner_claire Religion and pity are antonyms...	[pity, antonyms, disaster, humanity, live, wit...
2	if the only reason people are on twitter is to...	[reason, people, argue, hitler, count]
3	Religion is needed to provide a moral compass ...	[needed, provide, moral, compass, lack, empathy]
4	plus my religion has never been 'holier' than ...	[plus, never, holier, regular, life, regular, ...

Data Overview after dropping unnecessary columns

VADER MODEL

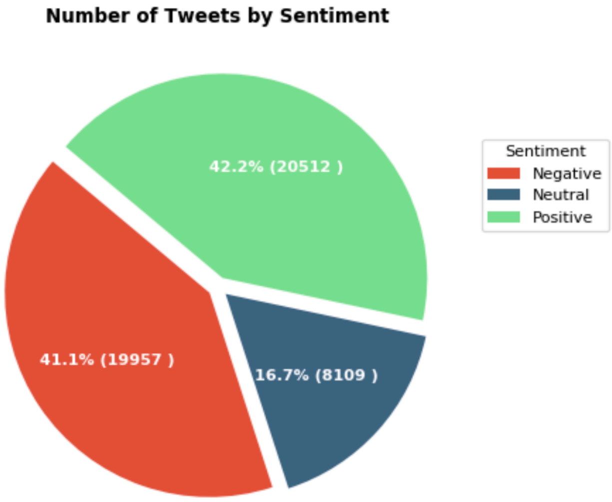
```
# Create an object of Vader Sentiment Analyzer
vader_analyzer = SentimentIntensityAnalyzer()


negative = []
neutral = []
positive = []
compound = []


def sentiment_scores(df, negative, neutral, positive, compound):
    for i in df['tweet_text_p']:
        sentiment_dict = vader_analyzer.polarity_scores(i)
        negative.append(sentiment_dict['neg'])
        neutral.append(sentiment_dict['neu'])
        positive.append(sentiment_dict['pos'])
        compound.append(sentiment_dict['compound'])


# Function calling
sentiment_scores(df, negative, neutral, positive, compound)


# Prepare columns to add the scores later
df["negative"] = negative
df["neutral"] = neutral
df["positive"] = positive
df["compound"] = compound


# Fill the overall sentiment with encoding:
# (-1)Negative, (0)Neutral, (1)Positive
sentiment = []
for i in df['compound']:
    if i >= 0.05 :
        sentiment.append(1)

    elif i <= - 0.05 :
        sentiment.append(-1)

    else :
        sentiment.append(0)
df['sentiment'] = sentiment
```

After applying the model to our data we have the following sentiment scores assigned to each tweet. The last column sentiment indicates the final classification (1 - positive, 0 - neutral, -1- negative).

```
df.head()
```

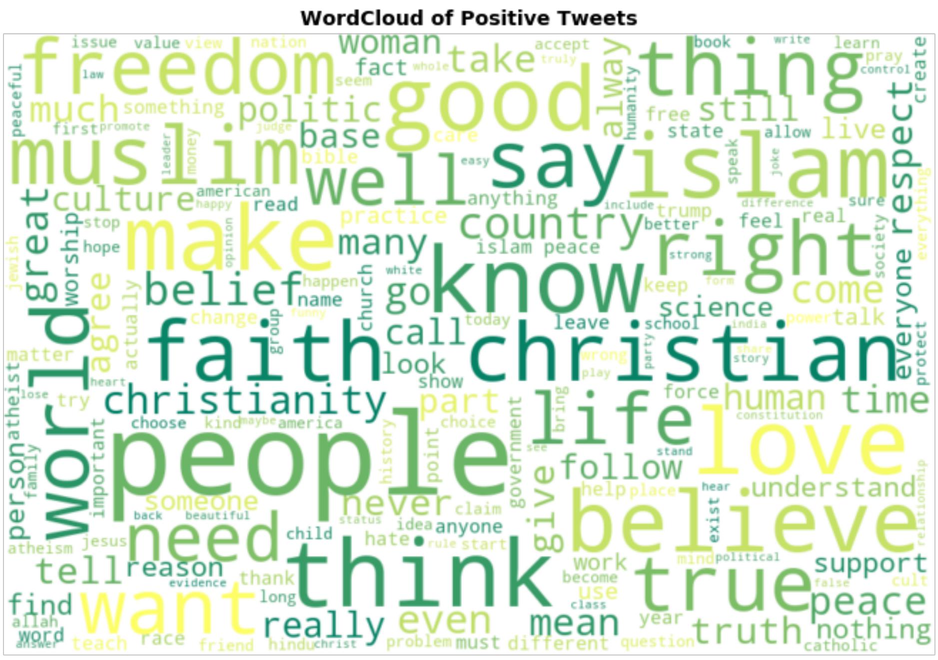
	tweet_text	tokens_no_stop	tweet_text_p	negative	neutral	positive	compound	sentiment
0	Islam is a religion full of blessings and good...	[islam, full, blessings, good, deeds, also, re...	Islam is a full of blessings and good deeds, ...	0.000	0.564	0.436	0.8910	1
1	@ wagner_claire Religion and pity are antonyms...	[pity, antonyms, disaster, humanity, live, wit...	and pity are antonyms. One is disaster, othe...	0.247	0.605	0.148	-0.2023	-1
2	if the only reason people are on twitter is to...	[reason, people, argue, hitler, count]	if the only reason people are on twitter is to...	0.130	0.870	0.000	-0.3400	-1
3	Religion is needed to provide a moral compass ...	[needed, provide, moral, compass, lack, empathy]	is needed to provide a moral compass to those...	0.173	0.827	0.000	-0.3182	-1
4	plus my religion has never been 'holier' than ...	[plus, never, holier, regular, life, regular, ...	plus my has never been 'holier' than regular ...	0.000	0.811	0.189	0.4939	1

Figure below, gives the overall classification result of tweets to the sentiment class, where 42.2% (20'512) of tweets are classified as positive. 41.1% (19'957) were classified as negative, and the remaining 16.7% (8'109) of tweets are considered neutral (neither positive or negative). The public opinion seems to be balanced in terms of the sentiments on discussions about religion on social media.



Classification of Tweets by Sentiment

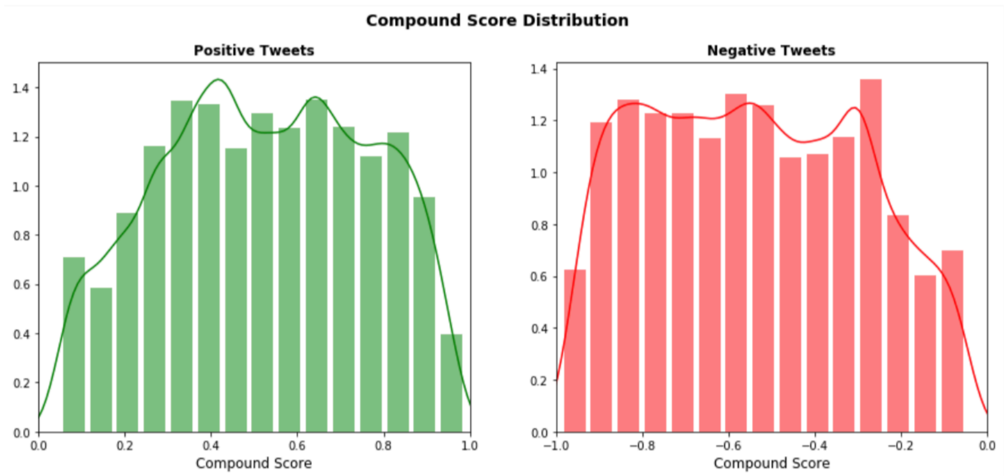
Moreover we can have a look at the negative and positive tweets using word-cloud to illustrate the dominant words for each category.



Most frequent words are as expected domain-related words such as *Islam, Muslim, Christian, Christianity*, whereas some of the sentiment-related words are: love, peace, support, great, respect, good, etc., which naturally imply a positive sentiment score of the tweet.

In case of negative tweets, beside domain-related words, some of the words carrying negative sentiment are: *kill, hate, attack, stupid, violence, problem, evil, fuck, shit, etc.*.

As the opinions are shown to be almost equally split regarding positive and negative sentiment, the positive tweets beside the slight numerical advantage also have a small negative difference in terms of the compound score, as shown in the figure below. Positive tweets have a mean of 0.525122, whereas the negative tweets a mean of -0.541433.



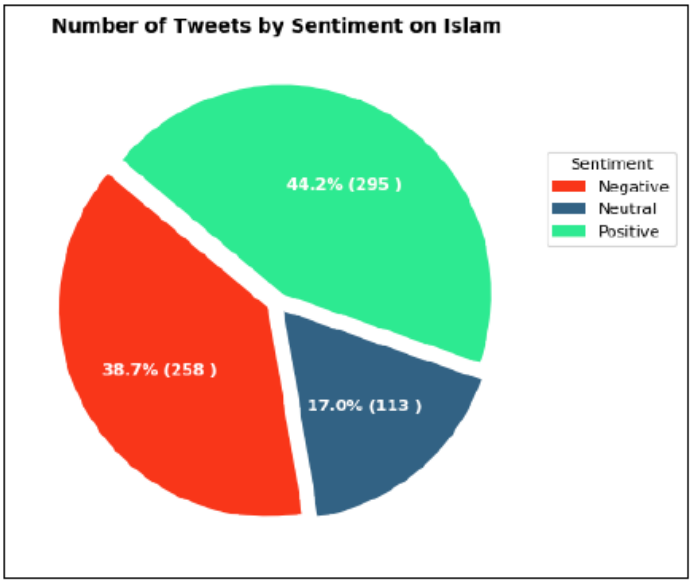
Compound Score Distribution

When it comes to specific religions, most of the discussions are oriented towards Islam with the highest number of tweets, 679 tweets after pre-processing steps, followed by Christianity with 127 tweets and then the other popular religions like Hinduism, Buddhism, and Judaism.

No.	Religion	No. Tweets
1	Islam	679
2	Christianity	127
3	Atheism	52
4	Judaism	35
5	Hinduism	27

Most Discussed Religions

Islam is the most discussed religion on social media. Hence, we had a closer look into the sentiments towards Islam particularly. A sentiment analysis on tweets which talk about Islam yielded the following results:



Tweets were classified using the same method (VADER), and analysis resulted in 44.2% (295) of tweets to be positive against 38.7% (258) negative tweets, while 17.0% (113) are considered neutral.

Table below illustrates the most frequent words on positive and negative tweets about Islam accordingly.

Positive Tweets		Negative Tweets	
Word	Frequency	Word	Frequency
Peace	105	Political	30
Love	31	Kill	23
Peaceful	17	Hate	20
Beautiful	15	Violence	19
Fastest	14	Death	18
Justice	12	Terrorist	14

Word Frequency based on Sentiment about Islam

Now we present top 10 most positive tweets:

Nr.	Tweet Text	Negative	Neutral	Positive	Compound
1	Religion tells you to clean yourself up, then God will love you, but that's not real love. That's conditional love. Real love is not dependent on how much you love Him; it's about how much He loves you.	0.000	0.458	0.542	**0.989**
2	Prayer is the most genuine and sincere thing people of religion know. sometimes it's everything they have to offer spiritually, in their eyes. it's not always understood but its best to appreciate it and see it as a very sincere "I want the best for you" sympathetic attitude :)	0.000	0.554	0.446	**0.986**
3	Be kind, polite, humble, respectful, caring and grateful no matter what, love everyone despite race, religion, wealth, life-style, we all are above things and we are worthy of all the love in the world. God is with us all and He loves each one of Us!!!	0.030	0.499	0.471	**0.986**
4	My name is peace. My religion is peace. I pray for peace. My heart is peace. I live in peace. I wish you #peace. #peacefeelgood #love you	0.000	0.317	0.683	**0.983**

5	The only socialism needed is charity, generosity stemming from the fruits of capitalism worked hard for by the base of the pyramid, Judge. That's why we need religion to encourage the GIVINGNESS OF GOD to one another as show of gratefulness to HIS LOVE, GRACE, BLESSINGS.	0.021	0.549	0.439	**0.981**
6	Do I have a religion? Yes. Do I tell people about it unsolicited? Definitely not. Everyone has choice. Everyone is beautiful. Everyone is loved and I love everyone. And yeah, I will hold EVERYONES HAND BECAUSE THEY DESERVE LOVE NO MATTER FROM WHO OR WHERE. THAT'S NOT MY BUSINESS.	0.044	0.496	0.460	**0.981**
7	It is a great part of our perfection to bear with one another in our imperfections. – St Francis de Sales #Catholic #Christian #religion #faith #faithful #faithfulness #God #morality #virtue #spirituality #charity #kindness #fellowship #sympathy #empathy # patience.	0.000	0.458	0.542	**0.981**
8	My true religion, my simple faith is in LOVE and COMPASSION. There is no need for a complicated philosophy, doctrine or dogma. Our own HEART, our own MIND, is the temple. The doctrine is COMPASSION. LOVE for others and respect for their rights and dignity.	0.033	0.531	0.436	**0.981**
9	This is who @USER truly is! And I'm proud of him! Thank you USER not only for being a great role model & a devout Christian but also for inviting our fellow brothers & sisters to get to know our great religion, which is Christianity! Nothing wrong with it, it's great!	0.000	0.584	0.416	**0.980**
10	#Islam is a beautiful # religion. It teaches #patience, #charity, #tolerance, #peace, love, the importance of seeking knowledge. Islam teaches us how to be a good neighbor & that our neighbors have rights over us regardless of their faith so we should live a good life.	0.000	0.458	0.542	**0.981**

The above-shown tweets are highly rated as positive, where the 10th tweet has a compound score of 0.9810, and the 1st tweet is rated with a compound score of 0.989. The classifier has produced correct labels, as the tweets are clearly positive and do not hide underlying patters of irony or sarcasm that undermine the meaning of the sentences. It is interesting to see how out of 10, only 4 of these tweets mention the word God, whereas the word Love is mentioned 16 times on 7 out of 10 most positive tweets. One tweet is about Islam, and two are about Christianity. The rest are general opinions on religion focusing on the relationship with others, namely loving each other as the core religious principle one should follow. Most of the tweets (6/10) have a negative score of 0.000; other tweets have a very minimal above zero negative score because of the words: dogma, race, unsolicited, etc. The tweet with the highest individual positive score is the 4th with a score of 0.683 because it mentions the word peace seven times, which increases the positivity of the sentence while it does not contain negation words.

Same, here are the top 10 most negative tweets and their respective scores:

Nr.	Tweet Text	Negative	Neutral	Positive	Compound
1	Religion caused the division of India and the formation of Pakistan. Hinduism versus Buddhism. Ban religion. That's the cause of so much death, killing, torture, abuse, rape, pain and suffering. There is no religion in India or Pakistan. Period. Take all books, etc.	0.522	0.478	0.000	**-0.988**
2	That's what they live for. The desalination of all others but Muslims. The religion of hate, murder, rape, war, paedophilia. Islam is pure evil.	0.649	0.351	0.000	**-0.987**
3	Threatening murder against a poor woman is not protecting Islam. If anything you are harming your religion by engaging in such barbaric language. Threatening violence does not lead to peace. It just leads to more violence. This is why peeps don't like Islam. Because it's violence.	0.649	0.351	0.000	**-0.987**
4	Agreed, for all but Islam. It's a cult of war and death the calls itself a religion to confuse Christians. Public beheadings rape and murder all part of their religion. To know them is to fear them.	0.525	0.448	0.027	**-0.987**
5	Christ is dead too you know, but I hear y'all, the last thing we need is another bullshit religion pushing even worse laws than the previous. Islam is the worst of the worst I believe in no god but would stand besides Christians to fight this threat against our freedoms.	0.452	0.478	0.070	**-0.986**

The above-shown tweets are very highly (in a negative sense) rated as negative where the 10th tweet has a compound score of -0.982, and the 1st most negative tweet is rated with a compound score of -0.988. It is evident that the classifier has produced correct labels, as the tweets are negative, expressing refusal, dislike, and negative feelings towards religion or a religious group i.e., Muslims. Some of the dominant words and their respective frequency on this table are death (8), Islam (6), rape (4), etc. The most addressed religion is Islam, where eight tweets out of 10 focus on Islam when expressing their negative emotions, while one was about Christianity and another for the Buddhism/Hinduism in India.

6	Its horrendous and pathetic. I don't think Saudi owns twitter though. It owns most of the political leaders but on here, I just think twitter tries not to be seen as racist or fascist but is failing appallingly. Any religion that is violent, oppresses women should be spoken against.	0.446	0.484	0.069	**-0.987**
7	That is not what China is saying. China is saying you can have your other religion anywhere else but here, we hate all religion. Enemy of My Enemy can kiss my ass too. China and Islam can go to hell. But I will not clap for China's fascist imperialism.	0.417	0.532	0.050	**-0.984**
8	You live in existential fear of your own death. Only humans can fear death when it is not immanent. But this fear of death is what gave rise to all religion. When the religion is attacked, the fear comes out. The fear of Nothing.	0.482	0.518	0.000	**-0.984**
9	I'm not interested in your sentiments. Religion is open for critics. Grow up! Get used to it. Mohamed was a terrorist, paedophile and a rapist. He enslaved and killed many people. He who considers this criminal a 'prophet' is either uneducated or a criminal himself.	0.476	0.524	0.000	**-0.983**
10	A religion gives a person the right to leave it. Sharia law says Criticizing or denying any part of the Quran is punishable by death, Criticizing Muhammad or denying that he is a prophet is punishable by death, A Muslim who becomes a non-Muslim is punishable by death.	0.503	0.497	0.000	**-0.982**

Most Negative tweets

HATE SPEECH DETECTION (SONAR)

The model used to classify tweets which discuss about religion is a pre-trained model based on the work of [3]. This model is considered useful for our task, to investigate the presence of hate speech on tweets about religion, and to measure it. It will contribute to a general understanding of the religious landscape in an online environment.

The required packages for this task are:

```
# IMPORTS
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


from hatesonar import Sonar # This is the hate speech detection
library


import warnings
warnings.filterwarnings('ignore')
```

The process starts by loading the data in a dataframe format on the notebook. The data consist of about 48’500 tweets that are not pre-processed as the model is trained to process the input to extract the necessary features. Some pre-processing steps used in the model are: remove hashtags, remove links, remove mentions, tokenization, and stemming using Porter Stemmer.

	tweet_text	negative	neutral	positive	compound	sentiment
0	Islam is a religion full of blessings and good...	0.000	0.564	0.436	0.8910	1
1	@ wagner_claire Religion and pity are antonyms...	0.247	0.605	0.148	-0.2023	-1
2	if the only reason people are on twitter is to...	0.130	0.870	0.000	-0.3400	-1
3	Religion is needed to provide a moral compass ...	0.173	0.827	0.000	-0.3182	-1
4	plus my religion has never been 'holier' than ...	0.000	0.811	0.189	0.4939	1

SONAR Model

Now we apply this classification model to our data:

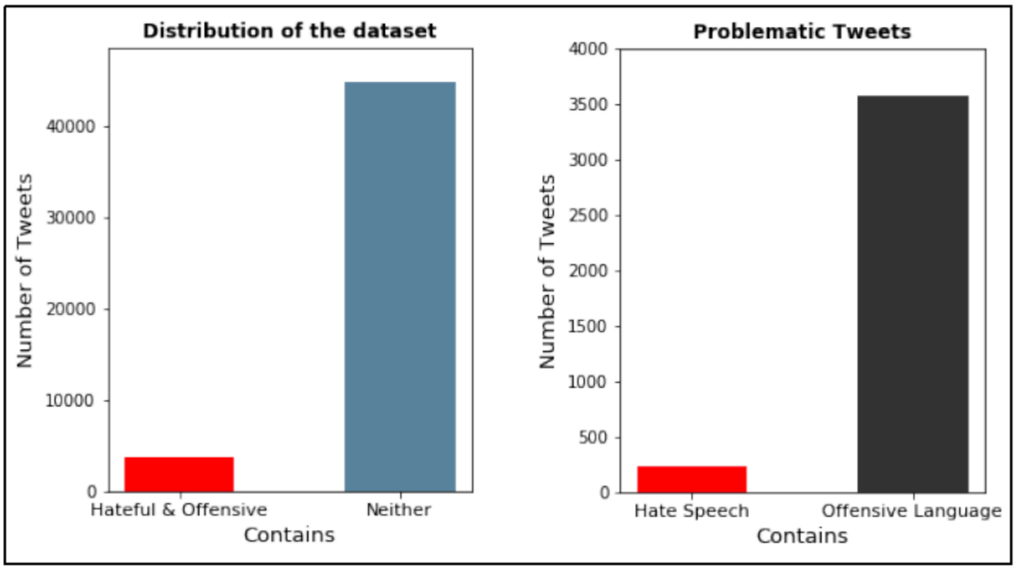
```
# Create an object of Sonar Hate Speech Detection
sonar = Sonar()

Class = []
hate = []
offensive = []
neither = []

def hate_speech_classifier(df, Class, hate, offensive, neither):
 for i in df['tweet_text']:
        sonar_dict = sonar.ping(text=i)
        Class.append(list(sonar_dict.values())[1])
 hate.append(list(list(sonar_dict.values())[2][0].values())[1])
 offensive.append(list(list(sonar_dict.values())[2][1].values())[1])
 neither.append(list(list(sonar_dict.values())[2][2].values())[1])

# Function calling
hate_speech_classifier(df, Class, hate, offensive, neither)

# Prepare columns to add the scores later
df["Class"] = Class
df["hate"] = hate
df["offensive"] = offensive
df["neither"] = neither
```

The resulted dataframe looks like this:

	tweet_text	negative	neutral	positive	compound	sentiment	Class	hate	offensive	neither
0	Islam is a religion full of blessings and good...	0.000	0.564	0.436	0.8910	1	neither	0.042266	0.310822	0.646912
1	@ wagner_claire Religion and pity are antonyms...	0.247	0.605	0.148	-0.2023	-1	neither	0.052652	0.314402	0.632946
2	if the only reason people are on twitter is to...	0.130	0.870	0.000	-0.3400	-1	neither	0.050406	0.405948	0.543646
3	Religion is needed to provide a moral compass ...	0.173	0.827	0.000	-0.3182	-1	neither	0.069441	0.317860	0.612700
4	plus my religion has never been 'holier' than ...	0.000	0.811	0.189	0.4939	1	neither	0.028691	0.459400	0.511909

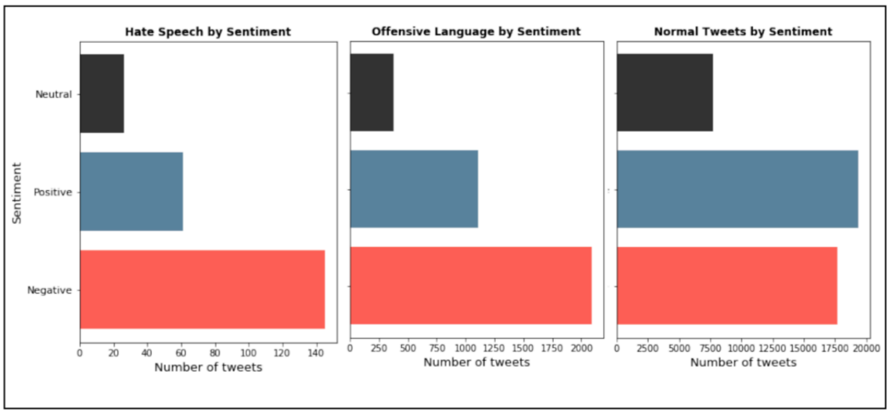
Result Overview

After the model is applied, the dataset consisting of 48’528 tweets is split into three categories Hate Speech, Offensive Language, and Neither (indicating neither hateful nor offensive). The first chart in figure below gives the overall results of tweets distribution across these categories. Hateful and offensive tweets combined resulted in a total of 3’802 tweets or 7.83% of the dataset. Thus, leaving the non-problematic tweets to a total of 44’726 or 92.16% of the dataset. Whereas in the second chart is shown the distribution among the so-called problematic tweets, Hate speech results in a total of 232 Tweets (6.10% of problematic tweets) as Offensive Language dominates this category with a total of 3’570 (93.89%).



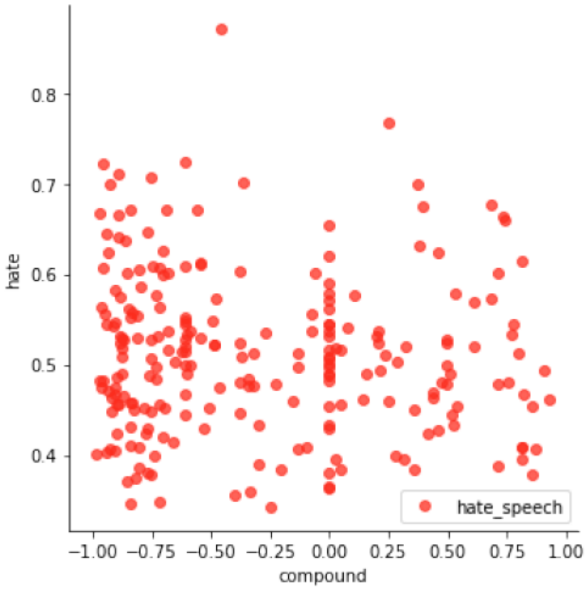
Hate Speech Detection Results

It is interesting to see how the sentiment feature correlates with the hate speech classification results. Figure below shows the distribution of tweets per category of hate speech, offensive language, and neither, based on the sentiment, being negative, positive, or neutral. As expected hate speech is occurred more in negative tweets, numerically speaking 60% (145 out of 232) tweets that contain hate speech are negative. In the case of offensive tweets, the proportions are similar to hate speech, as customarily expected the majority of tweets: 58.8% (2089 out of 3570 tweets) contain negative sentiment. From the rest, 1106 or 30.98% are positive, and 375 or 10.50% of tweets classified as offensive are neutral by sentiment.



Hate Speech Detection By Sentiment

Figure below illustrates the distribution of hateful tweets based on their compound sentiment score. The left side corresponding to negative tweets is more populated (60% of tweets), which makes sense that a tweet using hate speech is negative in sentiment.



Now we present top 10 most hateful tweets:

Nr.	Tweet Text	Hate	Offensive	Neither
1	<USER> "your religion is misogynistic" a white feminist abusing me on twitter.	**0.870**	0.069	0.059
2	<USER> White is a race, anyone can be a Muslim. It isn't racist to disagree with a dated, sexist stupid religion.	**0.767**	0.061	0.171
3	I don't know where White Racist Americans get this myth that there's a White Genocide in South Africa. This is how religion started.	**0.724**	0.212	0.062
4	Your baby is ugly, your life is ugly, your shapeless floral dress that suppresses you into accordance with your cult religion is ugly, and your heart is ugly so fuck you.	**0.721**	0.181	0.097
5	Hate is the new religion of some white folks. Their god of hate, is course, the devil.'	**0.709**	0.160	0.129
6	Niggas whip is going to be shittin on your niggas 200 dollar true religions faggot <LINK>'	**0.706**	0.292	0.001
7	How awake are you? Level 4: White Nationalism is good. Jews are evil. Christianity is a White religion.	**0.701**	0.261	0.037

8	The majority of American Jews are of white European descent. Judaism is a religion, not a race you dipshit.	**0.699**	0.105	0.196
9	<USER>: "A blanket ban based on race or religion is racist and bigoted!" Also <USER>: "BAN ALL WHITE MEN!""	**0.699**	0.096	0.205
10	Welcome to the white boys club Where your religion is allowed But only if it is a geek reference Or something to please our fetish. Welcome to the white boys club Well, I'm not a white boy I don\'t belong here "But you a different" They say. Just to justify their racist jokes.	**0.676**	0.258	0.064

Top 10 Hateful Tweets

Out of 232 tweets classified as hate speech, the above-presented tweets have the highest hate score, the 10th tweet has a hate score of 0.676, and the 1st tweet is rated with a hate score of 0.870. From a general manual analysis, we see that these tweets are dominated by a presence of offensive words like tweet number [4]. The model is trained to overcome the thin border between offensive language and hate speech by focusing on a vocabulary of hateful words. The most dominant aspect of hate is race. Eight out of 10 most hateful tweets contain a reference to white people, using words like white feminist, white racist Americans, white folks, white Nationalism, white European descent, white Men, and white boys club. Tweet number [2] expresses hate against Muslims calling their religion dated, sexist, stupid, while tweets number [7] and [8] contain references to Jews. Tweet number [6] is hate speech with a racist motive with reference to black people using the slang word niggas.

Similarly we present the most offensive tweets.

Offensive language scores dominate the ratings of the list as it varies from 96.9% to 98.9% for the first tweet. Only tweet number [10] has references to Islam, and tweet number [5] uses the word Hebrew but with no clear evidence of any reference to Judaism. Offensive tweets are populated vastly by offensive words of sexual reference such as bitch, which appears in 8 out of 10 tweets of the list. Tweets [3, 5, 6] contain strong sex-related words, while only tweets [1, 2] use the phrase “is my religion” in a figurative way.

Nr.	Tweet Text	Hate	Offensive	Neither
1	<USER> speaking "son of a bitch" is my religion.	0.010	**0.989**	0.000
2	<USER> saying "vegan bitches!" in her last Instagram story is my religion.	0.009	**0.990**	0.000
3	No more drugs for me, pussy n religion is all I need.	0.011	**0.987**	0.000
4	Whatever (if any) religion is right I'm fighting the head bitch of it for making me live my life with this fucking face and brain.	0.016	**0.983**	0.000
5	#Hebrew Religions sex teen pussy porn of is internet may: Of a teen pussy porn not to Are by involves almost: <LINK>	0.016	**0.982**	0.001
6	I don't know how chicks think pussy is the most powerful thing out pussy never made a nigga change religions	0.020	**0.979**	0.000
7	I need a religion to follow i swear to god that I got more problems than there is bitches in Carlo	0.019	**0.977**	0.000
8	Got bitches sending me titty pictures, she said no religion is the new religion	0.018	**0.973**	0.000
9	<USER> There is no religion that has "Paedophiles" ...Greedy Sloppy ass bitches can't close a room properly, things Morph <LINK>	0.027	**0.972**	0.000
10	Like idk if that's islamophobia jumping out or she really thinks her religion is superior and doesn't trust other shit out there. i cant share nice things with my family for ONCE? I can't be interested in another religion and experience it for once???? closed minded ass bitch.	0.028	**0.969**	0.001

10 Most Offensive Tweets

CONCLUSION

After cleaning and pre-processing steps, the data are fetched to the LDA algorithm to perform topic modeling. This phase aimed to investigate the topics related to religion that the public is concerned with, thus investigating how the multidimensionality of religion is materialized on online discussions. Then VADER sentiment analyzer tool was used for sentiment analysis of these tweets to uncover the sentiment of the public towards religion, whether it is negative, positive, or neutral. Lastly, SONAR, a hate speech detection tool, was used to investigate and measure the presence of hate speech and offensive language on discussions about religion.

The results empirically showed that the most relevant topics of public discussion related to religion are religion & politics, religion & science, Christianity, and Islam. Sentiment analysis showed that the public is divided when it comes to their sentiment towards religion, 42.2% (20'512) of tweets are classified as positive. 41.1% (19'957) were classified as negative with the rest classified as neutral. These results can be used to argue on the level of secularization and religion affiliation that the modern society has embraced. Hate speech detection tool presented evidence that indeed there is presence of hate speech and offensive language expressed on social media with religion as a motivation. A total of 3'802 tweets or 7.83% of the dataset resulted in containing either hate speech (232 tweets) or offensive language (3'570 tweets). This shows that religion is a sphere where the public opinion is divided and may motivate hate and offensive language in between their discussions. Thus, social media platforms should make more effort to prevent such abusive behavior of their users in their online discussions.

THANK YOU

References

[1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003, 01). Latent Dirichlet Allocation. (J. Lafferty, Ed.) *Journal of Machine Learning Research*, 993–1022.

[2] Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth international AAAI conference on weblogs and social media.* Michigan: AAAI Press.

[3] Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. *Eleventh International AAAI Conference on Web and Social Media* (pp. 512–515). Montreal: AAAI Press.