

GG

**BUSINESS REPORT ON:-**

**FAKE NEWS DETECTION**

**Submitted By:-**

**ABHISHEK JHA**

**BUSINESS REPORT**

**ACKNOWLEDGMENT**

I express my deepest sense of regards and gratitude to my mentor **Mr.Mohd.Kasif( Flip Robo) & Datatrained Team,** for their valuable guidance and encouragement; the present project has been Completed. Their positive attitude, valuable suggestion, lively discussion, constructive Criticism gave me immense help in completing the project. This thesis would not have been possible without them.

I express my gratefulness to **Mr. Vijay Raj Singh ( Data Scientist ), Mr.Anshuman Pandey & Mr. Steffin John** for helping & guiding me to complete this project. References taken from [www.Amazon.com](http://www.Amazon.com),

**Datasets: True.csv , Fake.csv**

1. International journal of recent technology and engineering (IJRTE) ISSN: 2277-3878, volume-7, issue-6, march 2019

2. Building a fake news classifier using natural language processing BY NATHAN

[(https://towardsdatascience.com/building-a-fake-news-classifier-using-natural-language-processing-83d911b237e1)]((https:/towardsdatascience.com/building-a-fake-news-classifier-using-natural-language-processing-83d911b237e1)%20)

3. Fake news detector: NLP project by ishant juyal

[(https://levelup.gitconnected.com/fake-news-detector-nlp-project-9d67e0177075)]((https:/levelup.gitconnected.com/fake-news-detector-nlp-project-9d67e0177075)%20)

4. Shloka Gilda,“Evaluating Machine Learning Algorithms for Fake News Detection” ,2017 IEEE 15th Student Conference on Research and Development (SCOReD).

5. Mykhailo Granik, Volodymyr Mesyura, “Fake News Detection Using Naive Bayes Classifier”, 2017 IEEEFirst Ukraine Conference on Electrical and Computer Engineering (UKRCON).

6. Gravanis, G., et al., Behind the cues: A benchmarking study for fake news detection. Expert Systems with Applications, 2019. 128: p. 201- 213.

7. Zhang, C., et al., Detecting fake news for reducing misinformation risks using analytics approaches. European Journal of Operational Research, 2019.

**LIST OF CONTENTS**

|  |  |
| --- | --- |
| **Sr. No.** | *Table Name* |
| 1 | *List of Tables* |
| 2 | *Introduction & Problem Statement* |
| 3 | *Dataset Head ,Dataset Tail, Shape of the Data* |
| 4 | *Describe of the Data* |
| 5 | *Info of the Data* |
| 6 | *Find the Null Values* |
| 7 | *Remove Unnecessary Values* |
| 8 | *Categorisation of Words to Real & Fake*  *1.Visualize of Real Words*  *2.Visualize of Fake Words* |
| 9 | *Preprocessing the text* |
| 10 | *Calculation the division points & Dividing the Datasets* |
| 11 | *Building AwordVec Model & K Means Model & The Prediction* |
| 12 | *Comparision of both models & Conclusion* |

# 

**LIST OF TABLES**

|  |  |
| --- | --- |
| **Sr. No.** | **Table Name** |
| 1 | *Dataset Head* |
| 2 | *Dataset Tail* |
| 3 | *Shape of the Data* |
| 4 | *Describe of the Data* |
| 5 | *Info of the Data* |
| 6 | *Find the Null Values* |
| 7 | *Remove Unnecessary Values* |
| 8 | *Categorisation of Words to Real & Fake*  *1.Visualize of Real Words*  *2.Visualize of Fake Words* |
| 9 | *Preprocessing the text* |
| 10 | *Calculation the division points & Dividing the Datasets* |
| 11 | *Building AwordVec Model & K Means Model & The Prediction* |
| 12 | *Comparision of both models & Conclusion* |

**INTRODUCTION**

**Business Problem Framing**

Fake news has become one of the biggest problems of our age. It has serious impact on our online as well as offline discourse. One can even go as far as saying that, to date, fake news poses a clear and present danger to western democracy and stability of the society.

**-Content**

What's inside is more than just rows and columns. Make it easy for others to get started by describing how you acquired the data and what time period it represents, too.

**What is a Fake News?**

Fake news's simple meaning is to incorporate information that leads people to the wrong path. Nowadays fake news spreading like water and people share this information without verifying it. This is often done to further or impose certain ideas and is often achieved with political agendas.

For media outlets, the ability to attract viewers to their websites is necessary to generate online advertising revenue. So it is necessary to detect fake news.

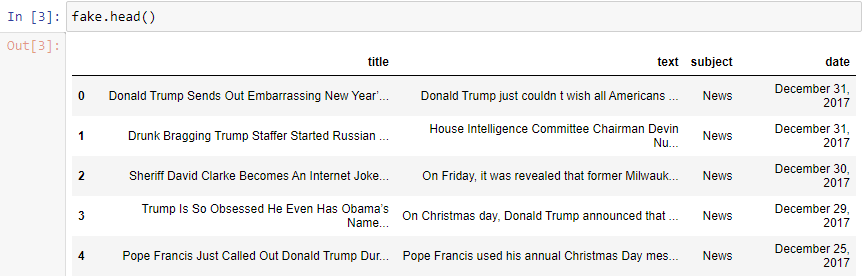
**Data Collection Phase**

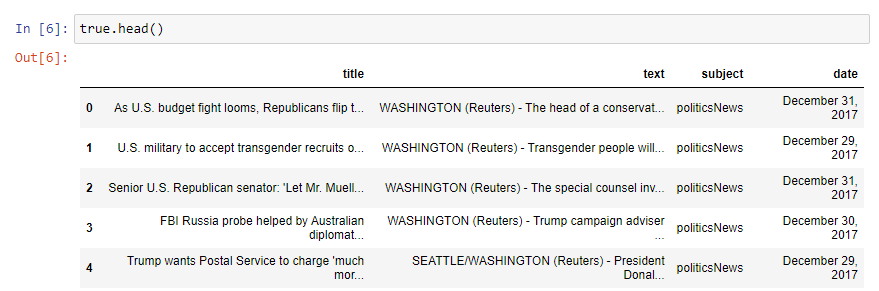
You can find many datasets for fake news detection on Kaggle or many other sites. I download these datasets from Kaggle. There are two datasets one for fake news and one for true news. In true news, there is 21417 news, and in fake news, there is 23481 news. You have to insert one label column zero for fake news and one for true news. We are combined both datasets using pandas built-in function.

**Data Ingestion:   
1. Read the dataset. Do the descriptive statistics and do the null value condition check. Write an inference on it.**

**Exploratory Data Analysis (EDA)**

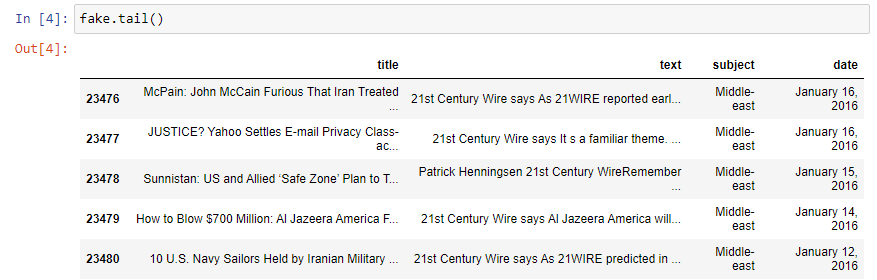
**Head of the Dataset**

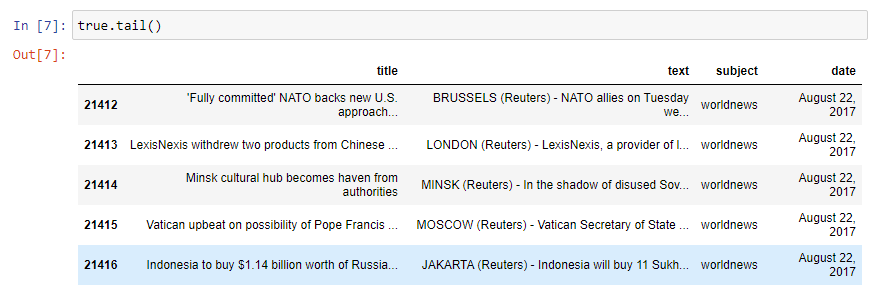
****

****

*Table 1:* *Dataset Head*

**Tail of the dataset**

****

****

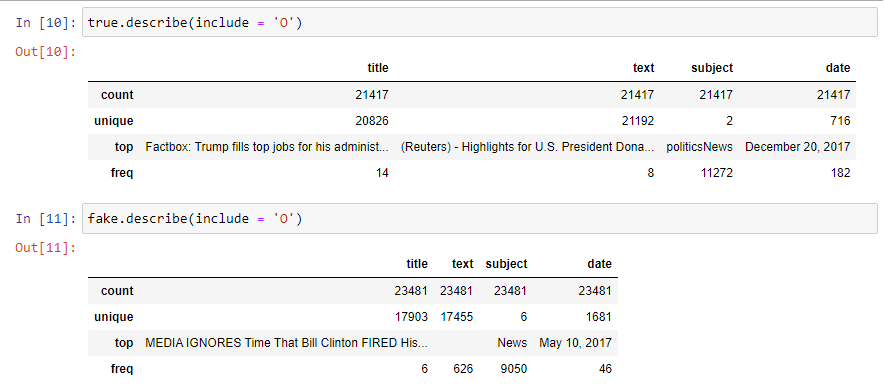
*Table 2:* *Dataset Tail*

**Shape of the Dataset**

Shape of fake is (23481, 4)

Shape of true is (21417, 4)

**Describe of the Dataset:**

****

*Table 3:* *Dataset Describe*

**Info of the Dataset**

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 23481 entries, 0 to 23480

Data columns (total 4 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 title 23481 non-null object

1 text 23481 non-null object

2 subject 23481 non-null object

3 date 23481 non-null object

dtypes: object(4)

memory usage: 733.9+ KB

None

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 21417 entries, 0 to 21416

Data columns (total 4 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 title 21417 non-null object

1 text 21417 non-null object

2 subject 21417 non-null object

3 date 21417 non-null object

dtypes: object(4)

memory usage: 669.4+ KB

None

**FIND THE NULL VALUES:**

title 0

text 0

subject 0

date 0

dtype: int64

\*\*\*\*\*\*\*\*\*\*\*\*

title 0

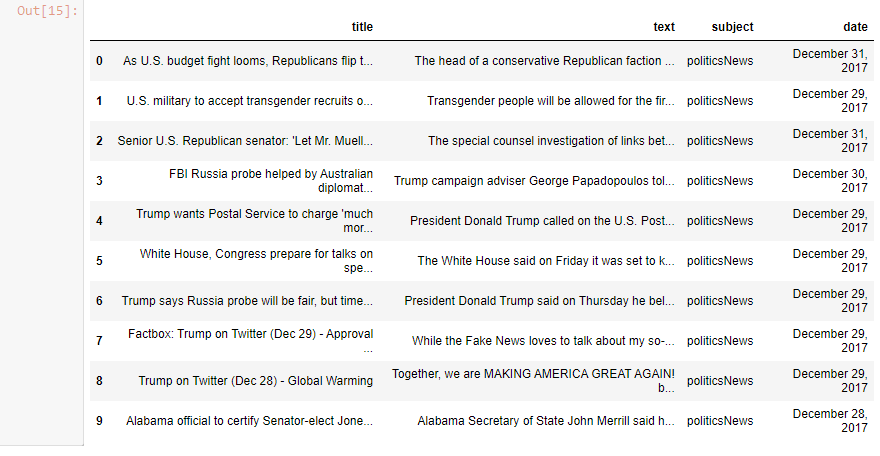
text 0

subject 0

date 0

dtype: int64

**REMOVE UNNECESSARY DATA:**

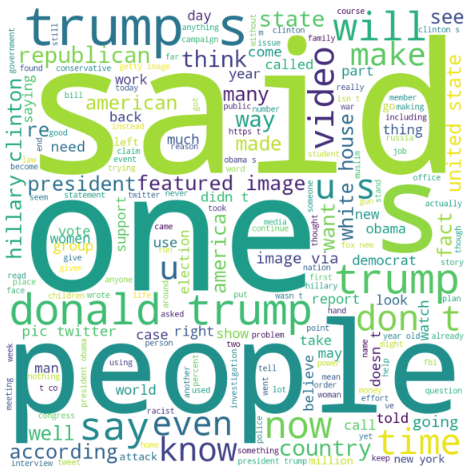


**CATEGORIZING WORDS TO REAL AND FAKE:**

**1.VISUALIZE REAL WORDS:**



**2.VISUALIZE FAKE WORDS:**

****

**PRE PROCESSING THE TEXT:**

44889

**CALCULATING THE DIVISION POINT:**

35912

8977

**DIVIDING THE DATASETS:**

8977

These 2 Orlando Hospitals Just Saved Pulse Nightclub Shooting Victims Over $5 Million In Costs The two Orlando-area hospitals that treated the shooting victims from the tragic Pulse nightclub attack are doing something amazing: They won t bill those patients for their out-of-pocket costs. Considering the fact that ER visits and hospital stays can run tens of thousands, if not hundreds of thousands of dollars, and our healthcare industry prioritizes getting their due dollars in any way possible over care, this is amazing news indeed.Instead, Orlando Regional Medical Center, which treated most of the victims, will seek additional payment from insurance and from a victims fund that city officials established in the wake of the shooting. Whatever insurance and the victims fund don t cover will be written off.Florida Hospital, which treated 12 victims, said it actually won t bill any of them or their insurance, nor will it bill them for follow-up treatment stemming from the attack.The two hospitals together will write off at least $5.5 million in treatment and care.David Strong, president and CEO of Orlando Health, said: The pulse [sic] shooting was a horrendous tragedy for the victims, their families and our entire community. During this very trying time, many organizations, individuals and charities have reached out to Orlando Health to show their support. This is simply our way of paying that kindness forward. The families of those who died won t find themselves charged, either.The victims who received care are relieved they were set to get fucked sideways with a rusty chainsaw by the abysmal payer structure that exists in our healthcare industry, which actually bankrupts entire families. Indeed, most of them (if not all of them) can t afford to pay for this level of hospital care.These are people who simply went out for a night of fun, and ended up badly hurt and in the hospital. They have to deal with the far more insidious mental and emotional trauma of the shooting in addition to healing from their physical injuries, and victims families are in a lot of pain, too. Nobody should ever have to deal with the added stress of hugantic bills for medical care they need.Now these victims can relax a bit about their treatment, and focus more of their energy on healing in all ways. If only more in our greedy, for-profit healthcare system cared as much as these hospitals.Featured image by Gerardo Mora/Getty Images

['koreans', 'japan', 'wary', 'tensions', 'north', 'worry', 'backlash', 'ethnic', 'koreans', 'living', 'japan', 'nervously', 'watching', 'growing', 'tensions', 'north', 'korea', 'wary', 'possible', 'backlash', 'community', 'pyongyang', 'ramps', 'sabre', 'rattling', 'public', 'antipathy', 'koreans', 'appear', 'escalated', 'reaction', 'north', 'latest', 'nuclear', 'test', 'missile', 'launches', 'community', 'target', 'abuse', 'japanese', 'nationalists', 'similar', 'incidents', 'past', 'north', 'conducted', 'powerful', 'nuclear', 'test', 'week', 'late', 'august', 'fired', 'ballistic', 'missile', 'hokkaido', 'northern', 'japan', 'new', 'force', 'western', 'japanese', 'city', 'osaka', 'home', 'country', 'largest', 'population', 'ethnic', 'koreans', 'willing', 'talk', 'publicly', 'north', 'korea', 'mixed', 'views', 'pyongyang', 'actions', 'kyon', 'owner', 'store', 'selling', 'korean', 'traditional', 'clothes', 'second', 'generation', 'korean', 'resident', 'japan', 'said', 'felt', 'north', 'pursuit', 'nuclear', 'weapons', 'natural', 'reaction', 'threats', 'united', 'states', 'loudly', 'secretly', 'think', 'north', 'korea', 'development', 'missile', 'nuclear', 'capabilities', 'said', 'north', 'korea', 'great', 'pressure', 'international', 'society', 'believe', 'end', 'watching', 'current', 'situation', 'great', 'hope', 'added', 'hand', 'chung', 'kap', 'second', 'generation', 'korean', 'town', 'said', 'hopes', 'north', 'korea', 'stops', 'provocations', 'pursues', 'peaceful', 'path', 'seeking', 'dialogue', 'south', 'korea', 'north', 'korea', 'defeat', 'united', 'states', 'hope', 'changes', 'way', 'south', 'koreans', 'north', 'koreans', 'lived', 'japan', 'year', 'according', 'government', 'data', 'forced', 'japan', 'occupation', 'korean', 'peninsula', 'world', 'war', 'japan', 'passed', 'anti', 'hate', 'law', 'year', 'discouraging', 'fresh', 'acts', 'abuse', 'community', 'issue', 'bears', 'watching', 'said', 'moon', 'gyeong', 'professor', 'ethnic', 'korean', 'studies', 'ritsumeikan', 'university', 'kyoto', 'north', 'korea', 'launches', 'missile', 'conducts', 'nuclear', 'tests', 'example', 'korean', 'schools', 'easy', 'target', 'bullies', 'accusations', 'japan', 'added']

**BUILDING A AWORD2VEC MODEL:**

**FINDING THE SENTENCE VECTOR:**

['koreans', 'japan', 'wary', 'tensions', 'north', 'worry', 'backlash', 'ethnic', 'koreans', 'living', 'japan', 'nervously', 'watching', 'growing', 'tensions', 'north', 'korea', 'wary', 'possible', 'backlash', 'community', 'pyongyang', 'ramps', 'sabre', 'rattling', 'public', 'antipathy', 'koreans', 'appear', 'escalated', 'reaction', 'north', 'latest', 'nuclear', 'test', 'missile', 'launches', 'community', 'target', 'abuse', 'japanese', 'nationalists', 'similar', 'incidents', 'past', 'north', 'conducted', 'powerful', 'nuclear', 'test', 'week', 'late', 'august', 'fired', 'ballistic', 'missile', 'hokkaido', 'northern', 'japan', 'new', 'force', 'western', 'japanese', 'city', 'osaka', 'home', 'country', 'largest', 'population', 'ethnic', 'koreans', 'willing', 'talk', 'publicly', 'north', 'korea', 'mixed', 'views', 'pyongyang', 'actions', 'kyon', 'owner', 'store', 'selling', 'korean', 'traditional', 'clothes', 'second', 'generation', 'korean', 'resident', 'japan', 'said', 'felt', 'north', 'pursuit', 'nuclear', 'weapons', 'natural', 'reaction', 'threats', 'united', 'states', 'loudly', 'secretly', 'think', 'north', 'korea', 'development', 'missile', 'nuclear', 'capabilities', 'said', 'north', 'korea', 'great', 'pressure', 'international', 'society', 'believe', 'end', 'watching', 'current', 'situation', 'great', 'hope', 'added', 'hand', 'chung', 'kap', 'second', 'generation', 'korean', 'town', 'said', 'hopes', 'north', 'korea', 'stops', 'provocations', 'pursues', 'peaceful', 'path', 'seeking', 'dialogue', 'south', 'korea', 'north', 'korea', 'defeat', 'united', 'states', 'hope', 'changes', 'way', 'south', 'koreans', 'north', 'koreans', 'lived', 'japan', 'year', 'according', 'government', 'data', 'forced', 'japan', 'occupation', 'korean', 'peninsula', 'world', 'war', 'japan', 'passed', 'anti', 'hate', 'law', 'year', 'discouraging', 'fresh', 'acts', 'abuse', 'community', 'issue', 'bears', 'watching', 'said', 'moon', 'gyeong', 'professor', 'ethnic', 'korean', 'studies', 'ritsumeikan', 'university', 'kyoto', 'north', 'korea', 'launches', 'missile', 'conducts', 'nuclear', 'tests', 'example', 'korean', 'schools', 'easy', 'target', 'bullies', 'accusations', 'japan', 'added']

Out[29]:

(8977, 100)

**K MEANS ON TEST DATASET:**

E:\ANACONDA\lib\site-packages\sklearn\cluster\\_kmeans.py:1077: ConvergenceWarning: Number of distinct clusters (1) found smaller than n\_clusters (2). Possibly due to duplicate points in X.

return self.fit(X, sample\_weight=sample\_weight).labels\_

**PREDICTING THE OUTPUT:**

# 

**COMPARING ORIGINAL TO PREDICTED OUTCOMES:**

Correctly clustered news: 100.0%

# CONCLUSION:-

# we are predicting whether an article is a real or fake article based on the relationship between the words . We have used the 2017 US president election datasets for creation of this system . We used Word2Vec model for building model and K -Means for the prediction and obtained an accuracy of 87%.