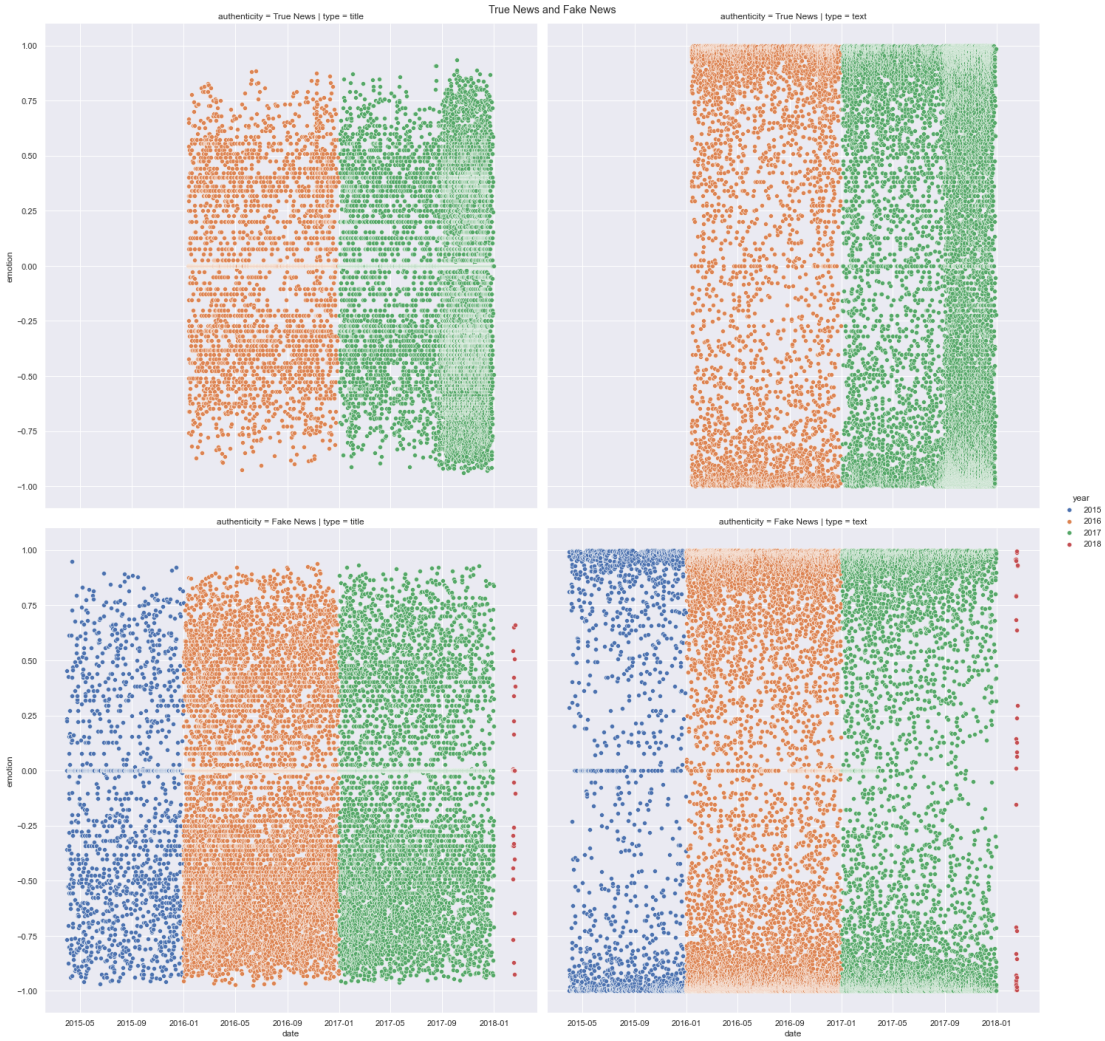
Project

In order to predict the true and false news later and have a deeper understanding of the data distribution, a comparative analysis is carried out on the two categories of true and false news. The analysis includes four parts: sentiment analysis of true and false news, named entity recognition of true and false news, subject analysis of true and false news and word cloud map.

1. Sentiment analysis of true and false news

Analyze the emotion of true and false news respectively, and select the scatter diagram to fully display the emotion analysis results of each news. A large picture includes four small pictures. The upper two pictures show the title and content emotion of real news respectively, and the lower two pictures show the title, content and emotion of fake news respectively.



By comparing the scatter diagram of true and false news, we can see that the emotional distribution of true news is relatively uniform, but the content distribution of false news is relatively extreme. The false news content is mostly gathered in the positions of 1 and -1, 1 represents positive and -1 represents negative, while the news gathered in 0 is less than the true news.

2. Named entity recognition of true and false news

data sources: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset

Project

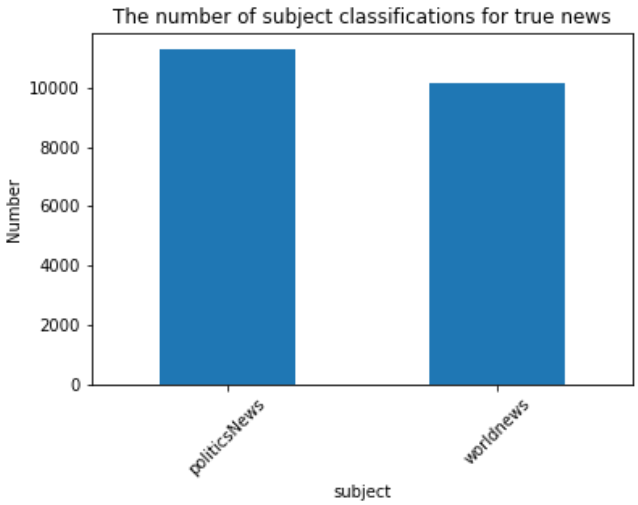
The following table shows the comparison of real and false news named entity recognition results. The first column is true news and the second column is false news. Although the number of true and false news is different, the difference can be clearly seen. The GPE of true news is much higher than that of false news.
GPE usually represents geo political items, such as city, state, country, state, etc. In addition to the above, LOCATION can also represent famous mountains and rivers. GPE and LOCATION are different. Real news should appear more GPE than LOCATION. Just get the desired result in the table, which may be helpful to the identification of true and false news.

ENTITY TYPE	TRUE NEWS	FALSE NEWS
FACILITY	745	680
GPE	16397	10109
GSP	480	289
LOCATION	419	193
ORGANIZATION	14237	13220
PERSON	15821	17043

3. Subject analysis of true and false news

On the basis of named entity recognition, topic classification analysis is carried out. Under the same classification standard, true and false news shows different topic classification results.

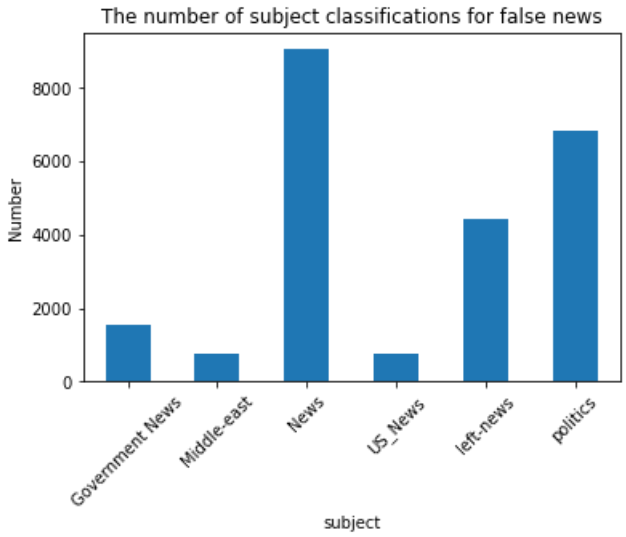
As shown in the figure below, the subjects of true news have two categories: political news and world news.



There are six themes of fake news, namely government news, Middle-East news, US news, left-news and politics. This may be helpful for the classification of true and false news, but it is worth noting that the number of false news and political news accounts for a large proportion, and there is little difference compared with true news.

data sources: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset

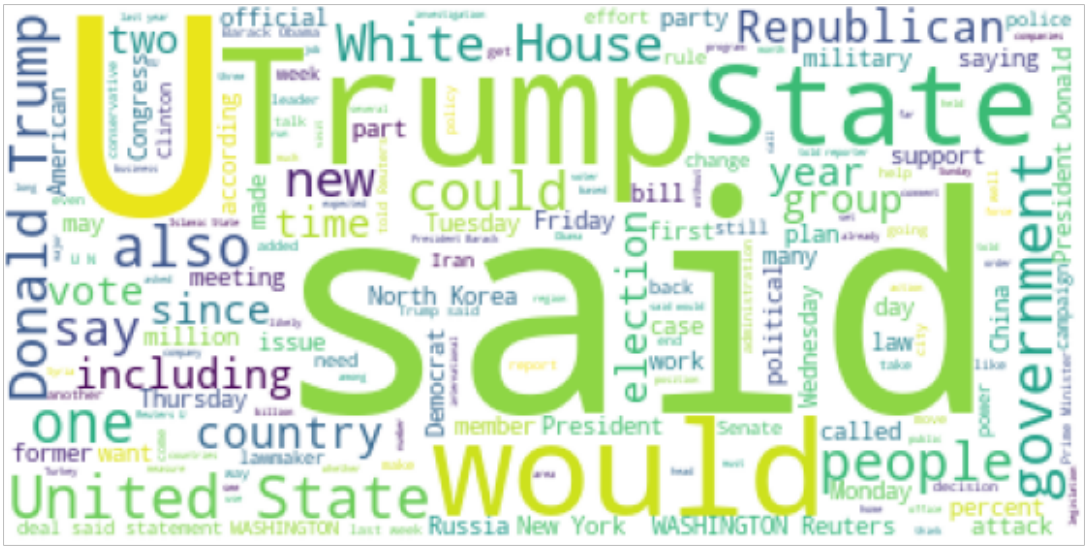
Project



4. Word cloud map of true and false news

Finally, it is the word cloud picture of true and false news. Observe the differences of true and false news in words through word frequency.

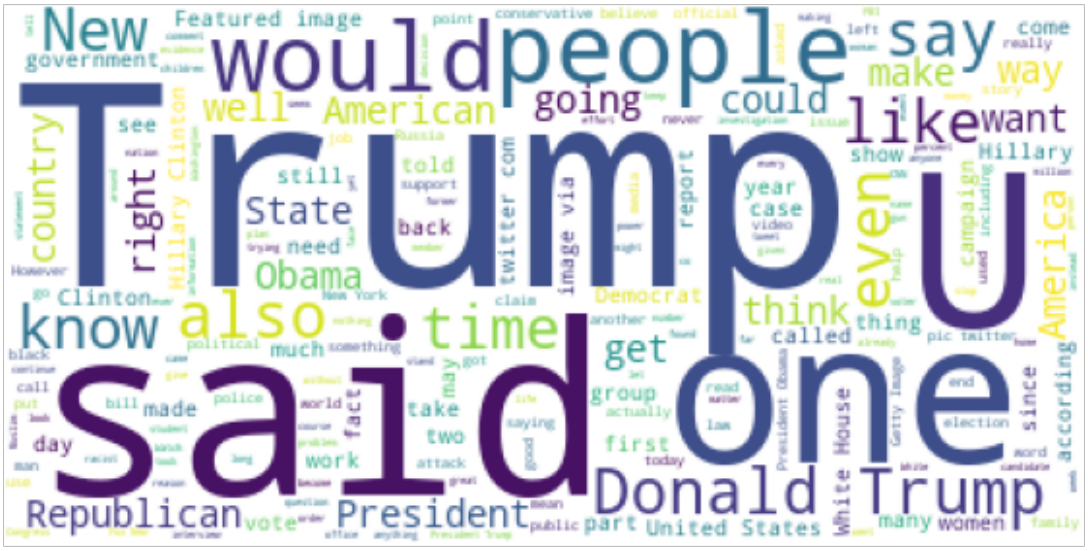
The following figure shows the word cloud of true news. You can see that "said" is the word with the highest word frequency. Followed by “Trump” and “would”. “said” has the highest word frequency, indicating that true news tends to directly introduce everyone's objective expression rather than subjective imagination. And "would" expresses a possibility, which can better reflect the accuracy and preciseness of the news.



The following is the cloud of the false news, which shows that “Trump” frequency is the highest, followed by “said”, “U” and “one”. It can be seen that the high-frequency words of false news and true news are somewhat different. The meaning of "U" is uncertain, which needs to be further determined by looking at the sentence. The word frequency of "one" is also very high. Maybe it represents an uncertain person in fake

data sources: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset

news, which shows that the accuracy and preciseness of fake news are not as high as real news.



In short, true and false news are different in many aspects.

data sources: https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset