

Unraveling the Web of Fake News: An Analysis of Twitter's Information Landscape

Muhammad Hassan
23100199@lums.edu.pk

Aleem Malik
23100179@lums.edu.pk

Jawad Azhar
23100321@lums.edu.pk

Muneeb Munir
23110327@lums.edu.pk

Index Terms—news, tweet

I. METHODOLOGY

The methodology of the research paper involved several steps for preprocessing and analysis of the dataset including Frequent Pattern Mining and Clustering.

Firstly, a vocabulary was generated from the text data, and a bag of words representation was created using this vocabulary. The text data was tokenized to split it into individual words for further processing. Stop words, which are common words with little significance in text analysis, were removed to reduce noise in the data.

Next, frequent pattern mining was performed using the FP-Growth algorithm, a popular method for mining frequent itemsets. Three variations of FP-Growth were used with different support thresholds of 0.01, 0.03, 0.05, and 0.07.

Association rules were then observed from these three variations, allowing for insights into the relationships between frequent patterns and their implications.

We then applied the Apriori algorithm. Similar to the previous approach, we generated a vocabulary and created a bag of words representation from the text data. Stop words were removed, and the text data was tokenized for further processing.

Next, we performed frequent pattern mining using the Apriori algorithm with various support thresholds. We repeated the tasks of generating association rules and observing patterns for different threshold values to compare the results. However, we encountered similar challenges in terms of time and memory consumption with the Apriori algorithm, especially for lower support thresholds.

As a next step, we explored clustering techniques. We started with the K-means algorithm. We used the elbow method to determine the optimal number of clusters. We ran the K-means algorithm on the number of clusters suggested by the elbow graph, specifically for the variables "Followers" and "Following". We also applied this approach to another variation of the variables "Followers" and "Status Count".

In addition to the above-mentioned approaches, we also explored the use of DBSCAN. The data in our dataset was found to be highly clustered so we used DBSCAN with specific hyperparameter values, including an epsilon value of 5000 and a minimum number of samples set to 100. These values were determined after thorough experimentation and analysis of the dataset.

The DBSCAN approach allowed us to identify dense regions of data points, which could potentially represent meaningful patterns or groups within the data. By carefully tuning the hyperparameter values, we aimed to capture clusters that exhibited significant density, while disregarding noisy or sparse regions.

II. ISSUES

During the experimentation process, we encountered an issue with the 0.01 threshold variation of the FP-Growth algorithm. The execution of the algorithm at this threshold took an extensive amount of time and required a significant amount of RAM, resulting in the inability to complete the analysis. This issue may be attributed to the large size of the dataset or the complexity of the patterns being mined. As a result, it was not feasible to obtain meaningful results from this particular threshold variation.

III. RESULTS

A. Graphs

- 1) *Frequent Pattern Mining:*
- 2) *Clustering:*

support		itemsets
0	0.07	(wedding)
1	0.05	(royal)
2	0.03	(x)
3	0.12	(weinstein)
4	0.12	(harvey)
...
5440	0.05	(assault, allegati, sexual, charges, looking, ...
5441	0.05	(assault, allegati, face, sexual, looking, eve...
5442	0.05	(assault, allegati, face, sexual, charges, loo...
5443	0.05	(assault, face, sexual, charges, looking, even...
5444	0.05	(assault, allegati, face, sexual, charges, loo...

Fig. 1. FP growth with 0.03 support (Fake News)

	antecedents	consequents	confidence
0	(wedding)	(royal)	0.571429
1	(royal)	(wedding)	0.800000
2	(brad)	(weinstein)	0.384615
3	(weinstein)	(brad)	0.416667
4	(pitt)	(weinstein)	0.384615
...
386711	(even) (assault, allegati, face, sexual, charges, loo...		0.833333
386712	(could) (assault, allegati, face, sexual, charges, loo...		0.714286
386713	(harvey) (assault, allegati, face, sexual, charges, loo...		0.416667
386714	(investigations) (assault, allegati, face, sexual, charges, loo...		0.833333
386715	(weinstein) (assault, allegati, face, sexual, charges, loo...		0.416667

Fig. 2. FP growth Association rules with threshold 0.25 (Fake News)

	support	itemsets
0	0.07	(wedding)
1	0.05	(royal)
2	0.12	(weinstein)
3	0.12	(harvey)
4	0.05	(think)
...
2565	0.05	(assault, allegati, face, sexual, looking, eve...
2566	0.05	(assault, allegati, sexual, charges, looking, ...
2567	0.05	(assault, allegati, face, sexual, charges, loo...
2568	0.05	(assault, face, sexual, charges, looking, even...
2569	0.05	(assault, allegati, face, sexual, charges, loo...

Fig. 3. FP growth with 0.05 support (Fake News)

	antecedents	consequents	confidence
0	(brad)	(weinstein)	0.384615
1	(weinstein)	(brad)	0.416667
2	(pitt)	(weinstein)	0.384615
3	(weinstein)	(pitt)	0.416667
4	(brad, pitt)	(weinstein)	0.384615
...
185161	(even) (assault, allegati, face, sexual, charges, loo...		0.833333
185162	(could) (assault, allegati, face, sexual, charges, loo...		0.714286
185163	(harvey) (assault, allegati, face, sexual, charges, loo...		0.416667
185164	(investigations) (assault, allegati, face, sexual, charges, loo...		0.833333
185165	(weinstein) (assault, allegati, face, sexual, charges, loo...		0.416667

Fig. 4. FP growth Association rules with threshold 0.20 (Fake News)

	support	itemsets
0	0.07	(wedding)
1	0.12	(harvey)
2	0.12	(weinstein)
3	0.07	(amp)
4	0.13	(brad)
5	0.13	(pitt)
6	0.07	(youtube)
7	0.11	(angelina)
8	0.11	(jolie)
9	0.09	(kids)
10	0.08	(fighting)
11	0.07	(harry)
12	0.08	(sexual)
13	0.07	(could)
14	0.12	(harvey, weinstein)
15	0.13	(brad, pitt)
16	0.08	(pitt, angelina)
17	0.08	(brad, angelina)
18	0.08	(brad, pitt, angelina)
19	0.11	(angelina, jolie)
20	0.08	(pitt, jolie)
21	0.08	(brad, jolie)
22	0.08	(angelina, pitt, jolie)
23	0.08	(brad, angelina, jolie)

Fig. 5. FP growth with 0.07 support (Fake News)

	antecedents	consequents	confidence
0	(harvey)	(weinstein)	1.000000
1	(weinstein)	(harvey)	1.000000
2	(brad)	(pitt)	1.000000
3	(pitt)	(brad)	1.000000
4	(pitt)	(angelina)	0.615385
...
197	(weinstein, sexual)	(harvey)	1.000000
198	(harvey, weinstein)	(sexual)	0.666667
199	(sexual)	(harvey, weinstein)	1.000000
200	(harvey)	(weinstein, sexual)	0.666667
201	(weinstein)	(harvey, sexual)	0.666667

Fig. 6. FP growth Association rules with threshold 0.15 (Fake News)

	support	itemsets
0	0.03	(slams)
1	0.05	(husband)
2	0.03	(matthew)
3	0.07	(harry)
4	0.03	(day)
...
5440	0.05	(assault, allegati, face, sexual, charges, eve...
5441	0.05	(assault, face, sexual, charges, looking, even...
5442	0.05	(assault, allegati, sexual, charges, looking, ...
5443	0.03	(slams, wants, cop, simpsons, social, husband,...
5444	0.05	(assault, allegati, face, sexual, charges, loo...

Fig. 7. AP growth with 0.03 support (Fake News)

	antecedents	consequents	confidence
0	(slams)	(husband)	1.000000
1	(husband)	(slams)	0.600000
2	(slams)	(following)	1.000000
3	(following)	(slams)	1.000000
4	(slams)	(simpsons)	1.000000
...
386711	(even) (assault, allegati, face, sexual, charges, loo...		0.833333
386712	(could) (assault, allegati, face, sexual, charges, loo...		0.714286
386713	(harvey) (assault, allegati, face, sexual, charges, loo...		0.416667
386714	(investigations) (assault, allegati, face, sexual, charges, loo...		0.833333
386715	(weinstein) (assault, allegati, face, sexual, charges, loo...		0.416667

Fig. 8. AP growth Association rules with threshold 0.25 (Fake News)

	support	itemsets
0	0.05	(husband)
1	0.07	(harry)
2	0.06	(face)
3	0.11	(jolie)
4	0.09	(kids)
...
2565	0.05	(assault, allegati, face, sexual, looking, eve...
2566	0.05	(assault, allegati, face, sexual, charges, eve...
2567	0.05	(assault, face, sexual, charges, looking, even...
2568	0.05	(assault, allegati, sexual, charges, looking, ...
2569	0.05	(assault, allegati, face, sexual, charges, loo...

Fig. 9. AP growth with 0.05 support (Fake News)

	antecedents	consequents	confidence
0	(harry)	(prince)	0.714286
1	(prince)	(harry)	1.000000
2	(harry)	(wedding)	0.714286
3	(wedding)	(harry)	0.714286
4	(harry)	(meghan)	0.714286
...
185161	(even) (assault, allegati, face, sexual, charges, loo...		0.833333
185162	(could) (assault, allegati, face, sexual, charges, loo...		0.714286
185163	(harvey) (assault, allegati, face, sexual, charges, loo...		0.416667
185164	(investigations) (assault, allegati, face, sexual, charges, loo...		0.833333
185165	(weinstein) (assault, allegati, face, sexual, charges, loo...		0.416667

Fig. 10. AP growth Association rules with threshold 0.20 (Fake News)

	support	itemsets
0	0.07	(harry)
1	0.11	(jolie)
2	0.09	(kids)
3	0.13	(pitt)
4	0.12	(weinstein)
5	0.08	(sexual)
6	0.11	(angelina)
7	0.07	(youtube)
8	0.13	(brad)
9	0.12	(harvey)
10	0.07	(could)
11	0.08	(fighting)
12	0.07	(wedding)
13	0.07	(amp)
14	0.09	(kids, jolie)
15	0.08	(pitt, jolie)
16	0.11	(angelina, jolie)
17	0.08	(brad, jolie)
18	0.08	(fighting, jolie)
19	0.09	(kids, angelina)
20	0.08	(pitt, angelina)
21	0.13	(brad, pitt)
22	0.08	(fighting, pitt)
23	0.08	(sexual, weinstein)
24	0.12	(harvey, weinstein)
25	0.08	(harvey, sexual)
26	0.08	(brad, angelina)
27	0.08	(fighting, angelina)
28	0.08	(brad, fighting)

Fig. 11. AP growth with 0.07 support (Fake News)

	antecedents	consequents	confidence
0	(kids)	(jolie)	1.000000
1	(jolie)	(kids)	0.818182
2	(pitt)	(jolie)	0.615385
3	(jolie)	(pitt)	0.727273
4	(angelina)	(jolie)	1.000000
...
197	(brad) (fighting, jolie, pitt, angelina)		0.615385
198	(angelina) (brad, fighting, pitt, jolie)		0.727273
199	(fighting) (brad, jolie, pitt, angelina)		1.000000
200	(pitt) (brad, fighting, jolie, angelina)		0.615385
201	(jolie) (brad, fighting, pitt, angelina)		0.727273

Fig. 12. AP growth Association rules with threshold 0.15 (Fake News)

support		itemsets
0	0.10	(scares)
1	0.10	(little)
2	0.10	(help)
3	0.10	(bts)
4	0.10	(ellen)
...
10342	0.08	(one, video, best, kemper, ellie, making, kali...
10343	0.08	(one, video, best, mindy, kemper, ellie, makin...
10344	0.08	(one, video, best, mindy, kemper, ellie, makin...
10345	0.08	(one, video, best, mindy, kemper, ellie, kalin...
10346	0.08	(one, video, best, mindy, kemper, ellie, makin...

Fig. 13. FP growth with 0.03 support (Real News)

support		itemsets
0	0.10	(bts)
1	0.10	(help)
2	0.10	(little)
3	0.10	(fangirl)
4	0.10	(ellen)
...
4881	0.08	(one, video, best, mindy, kemper, ellie, makin...
4882	0.08	(one, video, best, mindy, kemper, ellie, makin...
4883	0.08	(one, video, best, mindy, kemper, ellie, kalin...
4884	0.08	(one, video, best, mindy, ellie, making, kalin...
4885	0.08	(one, video, best, mindy, kemper, ellie, makin...

Fig. 17. FP growth with 0.07 support (real News)

antecedents	consequents	confidence
0 (slams)	(husband)	1.000000
1 (husband)	(slams)	0.600000
2 (slams)	(following)	1.000000
3 (following)	(slams)	1.000000
4 (slams)	(simpsons)	1.000000
...
386711 (even)	(assault, allegati, face, sexual, charges, loo...	0.833333
386712 (could)	(assault, allegati, face, sexual, charges, loo...	0.714286
386713 (harvey)	(assault, allegati, face, sexual, charges, loo...	0.416667
386714 (investigations)	(assault, allegati, face, sexual, charges, loo...	0.833333
386715 (weinstein)	(assault, allegati, face, sexual, charges, loo...	0.416667

Fig. 14. FP growth Association rules with threshold 0.25 (Real News)

antecedents	consequents	confidence
0 (brad)	(weinstein)	0.384615
1 (weinstein)	(brad)	0.416667
2 (pitt)	(weinstein)	0.384615
3 (weinstein)	(pitt)	0.416667
4 (brad, pitt)	(weinstein)	0.384615
...
185161 (even)	(assault, allegati, face, sexual, charges, loo...	0.833333
185162 (could)	(assault, allegati, face, sexual, charges, loo...	0.714286
185163 (harvey)	(assault, allegati, face, sexual, charges, loo...	0.416667
185164 (investigations)	(assault, allegati, face, sexual, charges, loo...	0.833333
185165 (weinstein)	(assault, allegati, face, sexual, charges, loo...	0.416667

Fig. 18. FP growth Association rules with threshold 0.15 (Real News)

support		itemsets
0	0.10	(help)
1	0.10	(bts)
2	0.10	(degeneres)
3	0.10	(fangirl)
4	0.10	(ellen)
...
9220	0.08	(one, video, best, mindy, kemper, ellie, makin...
9221	0.08	(one, video, best, mindy, kemper, ellie, kalin...
9222	0.08	(one, video, best, mindy, ellie, making, kalin...
9223	0.08	(one, best, mindy, video, kemper, ellie, makin...
9224	0.08	(one, video, best, mindy, kemper, ellie, makin...

Fig. 15. FP growth with 0.05 support (real News)

support		itemsets
0	0.07	(trying)
1	0.19	(abuse)
2	0.03	(queen)
3	0.06	(royal)
4	0.13	(day)
...
10342	0.08	(one, best, mindy, video, kemper, making, kali...
10343	0.08	(one, video, best, mindy, kemper, ellie, makin...
10344	0.08	(one, best, mindy, video, kemper, ellie, makin...
10345	0.05	(something, trying, nikki, bella, excited, tot...
10346	0.08	(one, video, best, mindy, kemper, ellie, makin...

Fig. 19. AP growth with 0.03 support (Real News)

antecedents	consequents	confidence
0 (brad)	(weinstein)	0.384615
1 (weinstein)	(brad)	0.416667
2 (pitt)	(weinstein)	0.384615
3 (weinstein)	(pitt)	0.416667
4 (brad, pitt)	(weinstein)	0.384615
...
185161 (even)	(assault, allegati, face, sexual, charges, loo...	0.833333
185162 (could)	(assault, allegati, face, sexual, charges, loo...	0.714286
185163 (harvey)	(assault, allegati, face, sexual, charges, loo...	0.416667
185164 (investigations)	(assault, allegati, face, sexual, charges, loo...	0.833333
185165 (weinstein)	(assault, allegati, face, sexual, charges, loo...	0.416667

Fig. 16. FP growth Association rules with threshold 0.20 (Real News)

antecedents	consequents	confidence
0 (trying)	(thats)	1.000000
1 (thats)	(trying)	1.000000
2 (trying)	(total)	1.000000
3 (total)	(trying)	1.000000
4 (trying)	(nikki)	1.000000
...
1113587 (kaling)	(one, video, best, mindy, kemper, ellie, makin...	0.888889
1113588 (life)	(one, video, best, mindy, kemper, ellie, makin...	1.000000
1113589 (said)	(one, video, best, mindy, kemper, ellie, makin...	0.888889
1113590 (office)	(one, video, best, mindy, kemper, ellie, makin...	0.888889
1113591 (days)	(one, video, best, mindy, kemper, ellie, makin...	0.888889

Fig. 20. AP growth Association rules with threshold 0.25 (Real News)

	support	itemsets
0	0.07	(trying)
1	0.19	(abuse)
2	0.06	(royal)
3	0.13	(day)
4	0.06	(like)
...
9220	0.08	(one, best, mindy, video, kemper, making, kali...
9221	0.08	(one, video, best, mindy, kemper, ellie, makin...
9222	0.08	(one, best, mindy, video, kemper, ellie, makin...
9223	0.05	(something, trying, nikki, bella, excited, tot...
9224	0.08	(one, video, best, mindy, kemper, ellie, makin...

Fig. 21. AP growth with 0.05 support (Real News)

	antecedents	consequents	confidence
0	(trying)	(thats)	1.000000
1	(thats)	(trying)	1.000000
2	(trying)	(total)	1.000000
3	(total)	(trying)	1.000000
4	(trying)	(nikki)	1.000000
...
1075119	(kaling) (one, video, best, mindy, kemper, ellie, makin...		0.888889
1075120	(life) (one, video, best, mindy, kemper, ellie, makin...		1.000000
1075121	(said) (one, video, best, mindy, kemper, ellie, makin...		0.888889
1075122	(office) (one, video, best, mindy, kemper, ellie, makin...		0.888889
1075123	(days) (one, video, best, mindy, kemper, ellie, makin...		0.888889

Fig. 22. AP growth Association rules with threshold 0.20 (Real News)

	support	itemsets
0	0.07	(trying)
1	0.19	(abuse)
2	0.13	(day)
3	0.07	(thats)
4	0.09	(kaling)
...
6917	0.09	(one, video, best, mindy, kemper, ellie, makin...
6918	0.08	(one, best, mindy, video, kemper, making, kali...
6919	0.08	(one, video, best, mindy, kemper, ellie, makin...
6920	0.08	(one, best, mindy, video, kemper, ellie, makin...
6921	0.08	(one, video, best, mindy, kemper, ellie, makin...

Fig. 23. AP growth with 0.07 support (Real News)

	antecedents	consequents	confidence
0	(trying)	(thats)	1.000000
1	(thats)	(trying)	1.000000
2	(trying)	(total)	1.000000
3	(total)	(trying)	1.000000
4	(trying)	(nikki)	1.000000
...
718871	(kaling) (one, video, best, mindy, kemper, ellie, makin...		0.888889
718872	(life) (one, video, best, mindy, kemper, ellie, makin...		1.000000
718873	(said) (one, video, best, mindy, kemper, ellie, makin...		0.888889
718874	(office) (one, video, best, mindy, kemper, ellie, makin...		0.888889
718875	(days) (one, video, best, mindy, kemper, ellie, makin...		0.888889

Fig. 24. AP growth Association rules with threshold 0.15 (Real News)

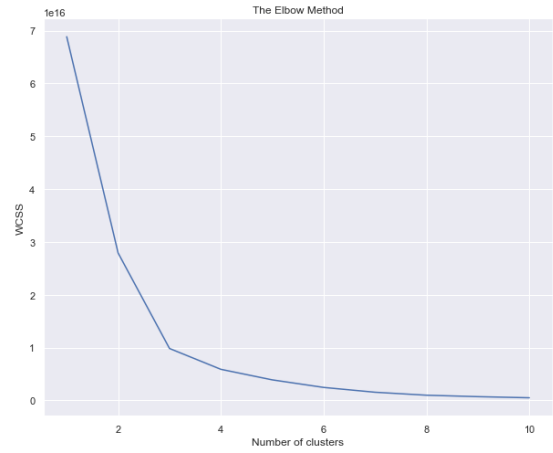


Fig. 25. (Elbow Method - Following (Fake News))

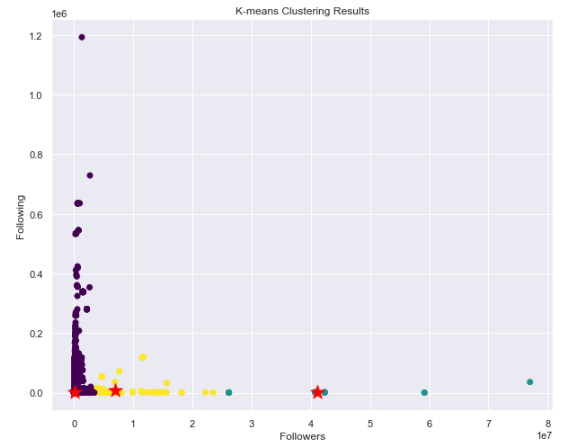


Fig. 26. (Kmeans - Following (Fake News))

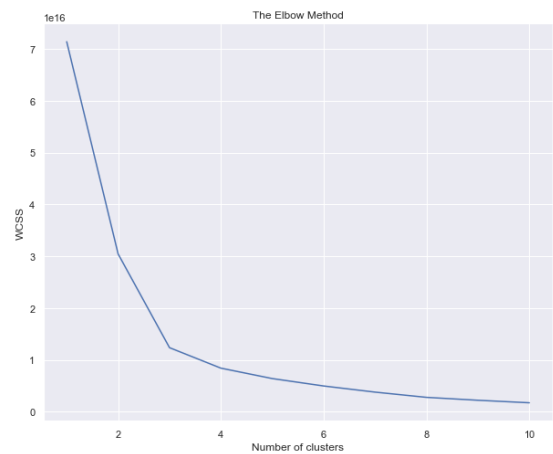


Fig. 27. (Elbow Method - Status (Fake News))

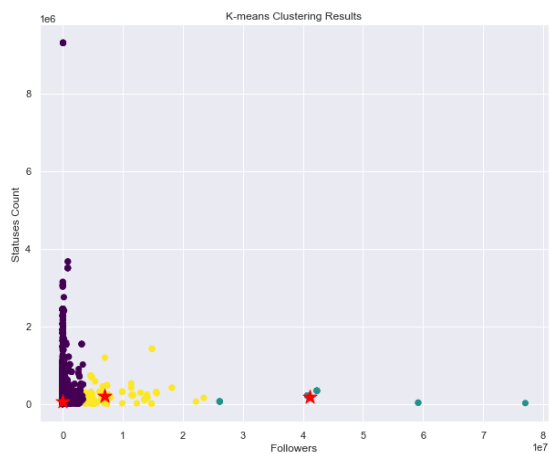


Fig. 28. (Kmeans - Status (Fake News))

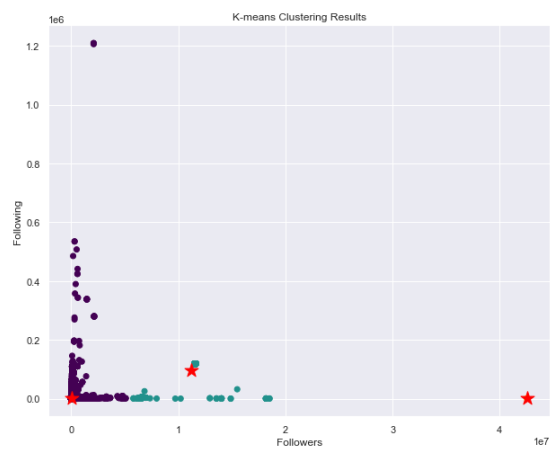


Fig. 31. (Kmeans - Following (Real News))

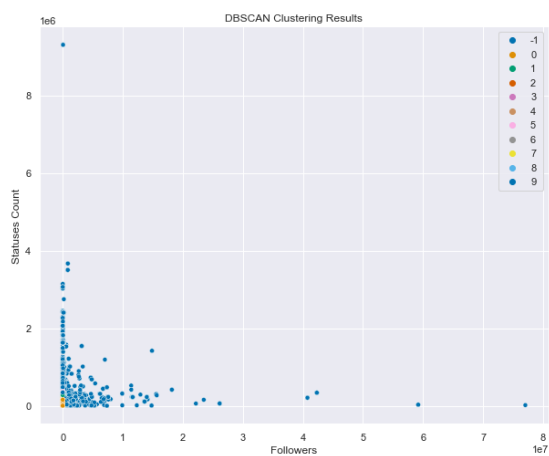


Fig. 29. (DB Scan - Status (Fake News))

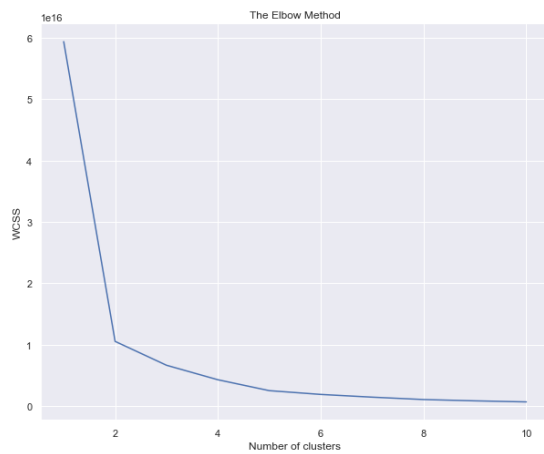


Fig. 32. (Elbow Method - Status (Real News))

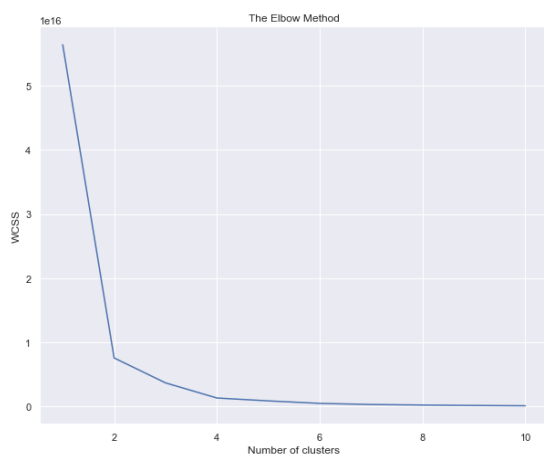


Fig. 30. (Elbow Method - Following (Real News))

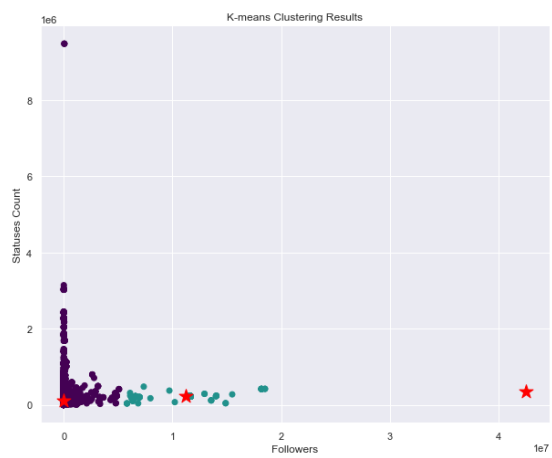


Fig. 33. (Kmeans - Status (Real News))

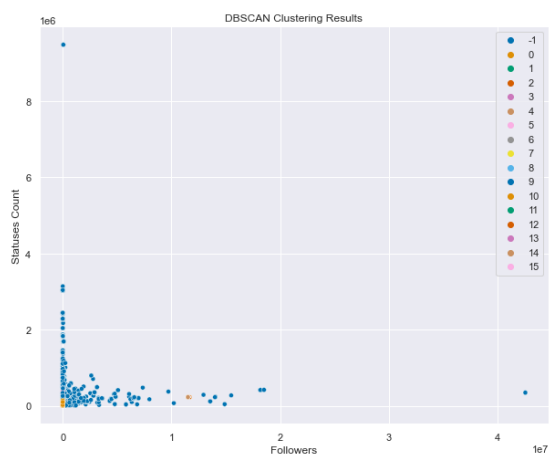


Fig. 34. (DB Scan - Status (Real News))