

Table 1: Veracity results for “sports” keyword

Content	Unverified	Verified
Number of unique users	582	17
Distribution of tweets by users	504	18
Number of unique locations	505	17
Total tweets	599	22
Diffusion index	0.9716193656093489	0.7727272727272727
Geographic spread index	0.8430717863105175	0.7727272727272727
Spam index	0.9632721202003339	0.75

From the above table we can assume that veracity of information is more in verified dataset

For further understanding of data, I have congregated the tweets using keywords “golf,” “cricket” and “nba” and calculated the veracity for each of the category. I have got the following results

Table 2: Veracity results for “golf” keyword

Content	Unverified	Verified
Number of unique users	2	1
Distribution of tweets by users	2	1
Number of unique locations	2	1
Total tweets	3	1
Diffusion index	0.6666666666666666	1.0
Geographic spread index	0.6666666666666666	1.0
Spam index	0.6666666666666666	1.0

From the above table we can assume that veracity of information is more in unverified dataset.

Table 3: Veracity results for “nba” keyword

Content	Unverified	Verified
Number of unique users	17	0
Distribution of tweets by users	12	0
Number of unique locations	17	0
Total tweets	17	0
Diffusion index	1.0	NA
Geographic spread index	1.0	NA
Spam index	1.0	NA

The veracity of tweets cannot be compared as there are no tweets for verified data. But we can assume that unverified tweets may contain false information as the 3 factors are equal to one.

Table 4: Veracity results for “cricket” keyword

Content	Unverified	Verified
Number of unique users	3	1
Distribution of tweets by users	4	1
Number of unique locations	3	1
Total tweets	4	2
Diffusion index	0.75	0.5
Geographic spread index	0.75	0.5
Spam index	0.625	0.5

From the above table we can assume that veracity of information is more in unverified dataset