

## **Project: Does media tone predict stock returns in India?**

### **Update 1 (Date:)**

- Summary submitted

### **Update 2 (Date:17/07/2020)**

1. Extracted news articles of 2018,2019 of all 50 nse firms from the sources: economic times,Livemint,the hindu Business line.
2. Retained the timestamp of each article too.
3. Performed tone analysis and collected:
  - a). Word count of all the firms for all the days.
  - b). Article count of all the firms for all the days.
  - c). Negative tone count of all the firms for all the days.
  - d). Negative tone percentage count of all the firms for all the days.

#### **Definition of output values generated in 'tone' folder:**

- 1.Article\_count: No. of articles per day(for all the 730 days) for each firm from all the three news sources.
- 2.ncount : no. of negative words per day( days for which article was available) from all articles, corresponding to each firm.
- 3.Word\_count : total no. of words per day(for all 730 days) from all articles, corresponding to each firm.
- 4.Pct : percentage of no. of negative words divided by total no. of words( $(\text{ncount}/\text{word\_count} * 100)$ ) for each day(days for which article was available) from all articles, corresponding to each firm.

#### **Definition of output value generated in 'summary\_statistics' folder:**

- 1.Count-count : Statistical measures such as mean,median,max of no. of articles per day (Article\_count) ,for all 730 days, corresponding to each firm.
- 2.Ncount-ncount: Statistical measures such as mean.median,variance,skew of no. of negative words per day(ncount) ,for days on which article was available, corresponding to each firm.
- 3.Pcts-pcts : Statistical measures such as mean.median,variance,skew of percentage of no. of negative words divided by total no. of words per day (pct), for days on which article was available,corresponding to each firm.
- 4.word-word : mean of total no. of words per day(word\_count) ,for all 730 days,corresponding to each firm.
5. No article: No. of days where no article was found from any of the three sources corresponding to each firm.

.