Lend-Rx Data Feed Specifications

Authors : Olivier Leherle
Eric Fokou
Philippe Very

1 Frequency of delivery

lend-Rx files (cf section 2) are delivered every Monday at 12:00:00 PM GMT on Lend-Rx ftp server.

2 Specifications for Feed dataset

The company file provides data at the drug level. Each line provides quantities computed on a weekly basis for a given company.

Column Name	Description	Format	Example
release_time	Timestamp for the data release	yyyy-MM-dd HH:mm:ss zzz	2019-09-23 12:00:00 GMT
computation_time	Timestamp for the data computation	yyyy-MM-dd HH:mm:ss zzz	2019-09-23 06:01:12 was that the GMT
ticker	Company Bloomberg equiy ticker	string	MRK US
count_drug	number of drugs developed by the company	long integer	10
volume_company_weekly	number of posts related to drugs developed by the company over the last week	long integer	50

volume_company_monthly	number of posts related to drugs developed by the company over the last month	long integer	250
volume_company_quarterly	number of posts related to drugs developed by the company over the last quarter	long integer	500
volume_company_positive_weekly	number of positive posts over the last week	long integer	10
volume_company_positive_monthly	number of positive posts over the last month	long integer	60
volume_company_positive_quarterly	number of positive posts over the last quarter	long integer	100
volume_company_neutral_weekly	number of neutral posts over the last week	long integer	20
volume_company_neutral_monthly	number of neutral posts over the last month	long integer	120
volume_company_neutral_quarterly	number of neutral posts over the last quarter	long integer	250
volume_company_negative_weekly	number of negative posts over the last week	long integer	10
volume_company_negative_monthly	number of negative posts over the last month	long integer	60
volume_company_negative_quarterly	number of negative posts over the last quarter	long integer	100
positive_daily_average	contribution of positive tweets to the sentiment factor (cf 3)	double	1.5
negative_daily_average	contribution of negative tweets to the sentiment factor (cf 3)	double	0.7

neutral_daily_average	daily average of the number of neutral twets	double	1.2
sentiment_factor	sentiment factors computed by aggregating twitter data at the company level	double	0.45
trading_signal	trading signal based on the sentiment factor (sell, buy or neutral)	double	sell

3 Remarks

We have added in the data-feed file contributions of positive and negative tweets to the sentiment factor. Indeed as explained in ([FV19]), the sentiment factor can be written as

$$S_T = \frac{P_T - N_T}{N_T + P_T} \tag{1}$$

where P_T (resp N_T) are exponential moving averages of the number of daily positive (resp negative) tweets computed with a half-life assumption of 25 days.

P_T (resp N_T) can therefore be found as **positive_daily_average** (resp **negative_daily_average**) (cf 2) in the data feed. To be completed , we have added the exponential moving averages of the number of daily neutral tweets using the same assumptions as **neutral_daily_average** (cf 2)

References

[FV19] Eric Fokou and Philippe Very. Cross-sectional analysis of twitter sig-nals in the pharmaceutical industry. 2019.