

MDS6212 Fintech Theory and Practice

Assignment 4

Yihang Li
220041006

1 Question 1-3

For the following news article,

"Tesla's third-quarter sales jumped 44% as global demand for its electric vehicles outpaced that of most other automakers. The company reported Friday that it had delivered 139,000 SUVs and sedans from July through September, compared with 97,000 deliveries during the same period a year ago. The sales topped even some of the most optimistic projections coming from Wall Street. Analysts polled by data provider FactSet expected the company to sell closer to 137,000. Tesla has been rewriting the script throughout the year amidst a pandemic that has closed factories and scrambled supply lines. This puts Musk & Co. in prime position to hit the area code of 500k units for the year which six months ago was not even on the map for the bulls," Daniel Ives of Wedbush wrote Friday. China was likely a major source of strength in the quarter, Ives said. Tesla could post its fifth consecutive quarter of profits later this month."

1.1 Present word cloud

By using Python, we present the word cloud as Figure 1. Here we add a picture of Tesla Model as the background. Also, we use the STOPWORD to clean the text, that is, to prevent some meaningless text.



Figure 1: Word Cloud for Given Text

1.2 Calculate the news sentiment variable using Loughran and McDonald Sentiment Word Lists

By using Python and the provided 'LoughranMcDonald_SentimentWordLists_2018.xlsx' file. We counted the number of negative words is 1 while the number of positive words is 2. Then also by using Python, **News Sentiment** = (# of positive words - # of negative words) / # of words = **0.00641025641025641**.

1.3 Calculate the Fog index

Recall that the fog index is the quality of being easy or enjoyable to read. By using textstat with `textstat.gunning_fog(text)` in Python, we calculate the Fog index of the given text: **Fog** = 0.4 (words / sentences + 100 * complex words / words) = **12.45**.

2 Question4

2.1 Report the summary statistics of sentiment, novelty, and impact

By using Python, the summary statistics of sentiment, novelty and impact are as in Table 1.

Table 1: Summary Statistics

	Sentiment	Novelty	Impact
count	1292.000000	1292.000000	1292.000000
mean	53.845975	28.877709	45.277864
std	14.824257	38.537946	9.996097
min	2.000000	0.000000	13.000000
25%	40.000000	0.000000	39.000000
50%	50.000000	3.000000	45.000000
75%	64.000000	56.000000	52.000000
max	100.000000	100.000000	77.000000

2.2 Present the correlation coefficient among sentiment, novelty, and impact

Table 2 is the correlation coefficient among sentiment, novelty, and impact.

Table 2: Correlation Coefficient

	<i>Sentiment</i>	<i>Novelty</i>	<i>Impact</i>
<i>Sentiment</i>	1.000000	0.165557	-0.130972
<i>Novelty</i>	0.165557	1.000000	-0.064844
<i>Impact</i>	-0.130972	-0.064844	1.000000

Table 3: Top 10 Categories

<i>Category</i>	<i>Frequency</i>	<i>Fraction</i>
<i>stock-loss</i>	409	0.316563
<i>stock-gain</i>	232	0.179567
<i>product-release</i>	111	0.085913
<i>business-contract</i>	63	0.048762
<i>capital-increase</i>	55	0.042570
<i>legal-verdict-favored</i>	42	0.032508
<i>price-target-upgrade</i>	40	0.030960
<i>acquisition-interest-acquirer</i>	33	0.025542
<i>fundraising</i>	33	0.025542
<i>product-price-cut</i>	30	0.023220

2.3 Show the frequency and fraction of top 10 news categories

Table 3 is the the frequency and fraction of top 10 news categories.