



Presented By: Saheli Saha

Cleaning and normalizing the Data using and build corpus: -

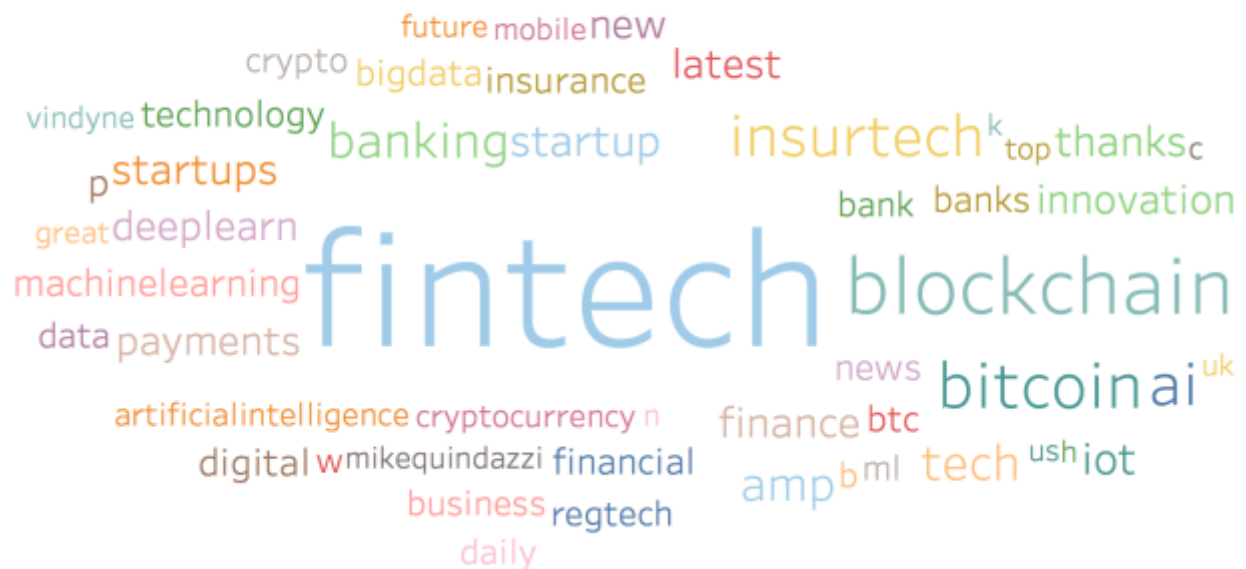
- Stop word
- Unicode to String
- Stemmer
- Tokenizing

corpus - List (5000 elements)				
idx	Type	Size	Value	
0	str	1	bitcoin price spikes etf decision nears llvtxzlgm btc fintech altcoin markets	
1	str	1	laura alozie exec team spacnation made strongg tobiadex online finance money ent ...	
2	str	1	china biggest social network useful business wechat wechatofficial fintech ddpcg ...	
3	str	1	friendsurance proud named shifting business model insurtech trend report talltve ...	
4	str	1	yostartups checkout activehours fintech startup watch lending usa yostartups che ...	
5	str	1	away fintech scoop rnq zlpaooy stories nicky whiting feliciatedx boarfinance bloc ...	
6	str	1	wealthtechbook thanks steven dryall sdryall abstract fintech wealthtech kqpotixi ...	
7	str	1	ubs wealth chief fintech beyond fighting business insider tmkw k k fintech	
8	str	1	rcinext tool everyone innovative champion systems cooperation innovation fintech ...	
9	str	1	bitcoin fintech primary litecoin pool upgrade segwit says ltc btc founder b wet ...	
10	str	1	theblockchain bitfinex nuked wa state customers verjrcjv bitcoin blockchain fint ...	
11	str	1	deeplearn ibm amp visa collaborate iot project machinelearning ml ai fintech tec ...	
12	str	1	assets seeing jump tweets fintech aapl blockchain eurUSD u od pw wowlduk aw	
13	str	1	malix fintech technology startups business journal ggrlaaz fintech lending	
14	str	1	biometrics borders banks u learn nations fintech biometrics w uuwo bwx f old	

china biggest social network useful business wechat wechatofficial fintech ddpcgp lh

In the cleaned corpus, I got sentence like above which is not giving any proper construction of sentence. Hence I checked the word frequency and visualized the top words with most word frequency as Word Cloud.

Top Words from Tweets



Further I implemented co-occurrence matrix to get better knowledge about Fintech. co-occurrence matrix gave me top two words that are coming together with their frequency. I visualized the top two words with most word frequency as Word Cloud.



- People discussed about Virtusa Polaris, Fintech Lab. So, they also discussed about the improvement of Fintech.
- People discussed about Australian Fintech world and its clash, collaboration with Banks.
- People are concerned about the latest news about Fintech, how integrate new things in it. Such as token's new payment ecosystem that uses end to end secure protocol with digital signature.
- We can see the word announcement has a high frequency, hence tweet users are mostly discussed about new partners, company updates and latest news of Fintech.

Sentiment Analysis: -

Performer sentiment analysis formed two list where each sentence got points for positive and negative sentence. As we can see below positive sentences got higher score compare to negative sentences. Few sentence even got 3 points whereas most of the sentence in negative list got '0', i.e. neutral. Some of them got 1. Hence, we can say people are mostly talking positive and neutral things about Fintech.

Points for Positive Sentence Points for Negative Sentence

Type	Size	
int	1	0
int	1	0
int	1	1
int	1	1
int	1	1
int	1	0
int	1	1
int	1	0
int	1	0
int	1	1
int	1	1
int	1	0
int	1	2
int	1	0
int	1	3
int	1	0
int	1	1

[illegible]

The chosen measure of “closeness” is Pointwise Mutual Information (PMI), calculated as follows (t_1 and t_2 are terms):

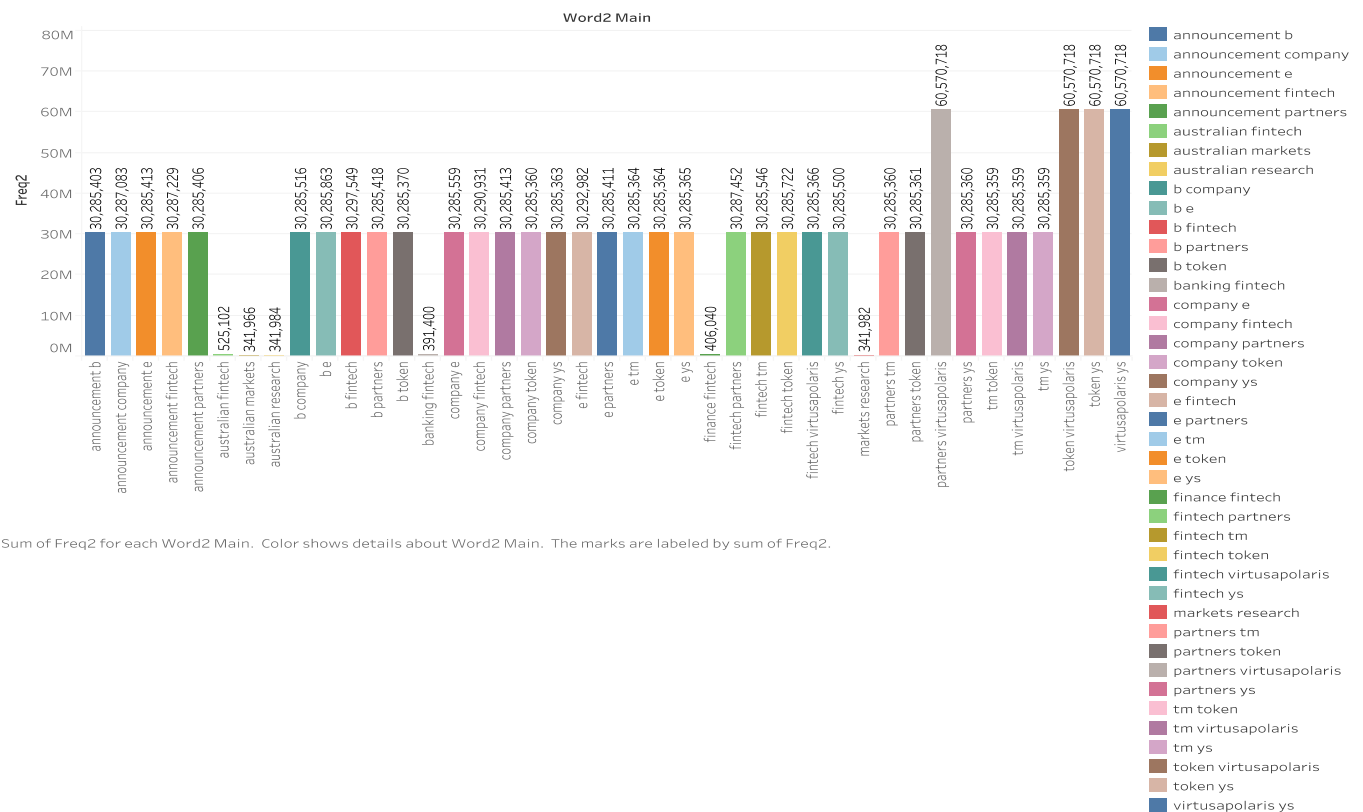
$$\text{PMI}(t_1, t_2) = \log\left(\frac{P(t_1 \wedge t_2)}{P(t_1) \cdot P(t_2)}\right)$$

Semantic Orientation Calculation: -

$$\text{SO}(t) = \sum_{t' \in V^+} \text{PMI}(t, t') - \sum_{t' \in V^-} \text{PMI}(t, t')$$

I plotted the semantic orientation in the bar graph as follows: -

Top Words from Tweets Bar Graph by Co Occurrence Matrix



Location Analysis: -

I performed location analysis to see people from which places are mainly talking about Fintech. Below are the locations that I got after cleaning the data. Further I plotted that in the world Map.

idx ▲	Type	Size	
19	str	1	south holland netherlands
20	str	1	n
21	str	1	red hill brisbane
22	str	1	hunting satoshi nakamoto
23	str	1	n
24	str	1	arex helsinki barcelona london
25	str	1	ottawa ontario
26	str	1	london england
27	str	1	florida usa
28	str	1	london england
29	str	1	new york city
30	str	1	london uk
31	str	1	smyrna valdosta
32	str	1	atlanta
33	str	1	singapore

Top Locations by Co Occurrence Matrix



Topic Analysis: -

Top 5 Topic using Non-Negative Matrix Factorization Algorithm: -

- People discussed positive, negative and neutral aspects of Fintech. Where people are hopeful for Fintech future, with application of Big Data, IOT, Data Analytics.
- People discussed about the role of Regetch in diminishing the cyber security issues.
- Fintech is growing in Singapore with the help of PWC. Innovation under Hood engines are praised and disruption for Fintech using traditional method.
- Indian community's contribution with latest technologies, where bank fails.
- On the contrary Russian Modulbank acquires three Fintech startups. Lastly people discussed about the famous SXSW conference and its impact on Fintech.

Top 5 Topic using Latent Dirichlet Allocation Algorithm: -

- People mostly discussed about the growth and development of Fintech. How far it came using Machine Learning, Artificial Intelligence, Deep Learning and how far it will go using these technologies.
- Latest news and updates like PWC Singapore contribution, IOT updates, positive and negative effects of Brexit.
- Other top topics are SXSW conference, and its updates on startups competition, panel discussions and mentoring workshops and payment modes other than bitcoin for Asian Fintech World.
- Regulation Technology is growing in Fintech and Machine Learning applications are applying more rigorously in finance.
- Banks are still slow compare to Fintech but they are progressing to adopt digital sales capabilities.