

Early Warning System using Global News for Creditors and Investors

Sneha Tiwari R19MBA08 REVA University, Bengaluru

Introduction and Business Understanding:



- One of the fundamentals for any Investor or Financial institution is Risk management.
- Risk management's goal is to ensure that the loan/credit fund is safe, profitable, and fluid.
- Considering the unexpected financial situation and credit risk quantitative research is necessary and a dire requirement to rethink the current methodologies and revise them



Present Day Scenario:

- Currently most of the Credit Risk Management teams determine the level of risk by using Credit Risk Modelling
 Techniques which calculates the risks considering variables such as Balance sheets, Cash Flow Statements, economic
 conditions, etc.
- These models Few teams depend on external rating agencies such as Moody's, Fitch Ratings, S & P Global Ratings, etc.
- Credit MIS reporting, another method where the borrower will be asked to submit pre-determined financial statements to the lender periodically.

Challenges and Problem Statement:



Challenges:

- With the ever-changing socio-economic landscape makes it extremely
 difficult to predict such emerging risks, hence companies looking
 forward for a solution which can be larger inclusive, faster, proactive
 and used to shield itself from external perils.
- Forget complex spreadsheet-based processes—to gain the most valuable insights, data and analysis must be presented in an intuitive, clean and clearly visualized way.



Problem Statement:

- Threats and Risks comes from various sources like Uncertainty in financial markets, Legal liabilities, Natural causes and disasters and many more.
- These kind of risks can not only affect a company's financial stability but can also damage its reputation in the market, due to which the companies have been investing in Risk Management related activities.
- Need for a mechanism which is capable of capturing, interpreting and focusing the underlined threats and send an alert to the Risk Management team.

Solution and Scope of Work:



Solution:

- Developing an early warning system with the help of Data Science techniques.
- The system will access the news media from both web and T.V. channels, from across the world and provides a dashboard with various quantifiers, components and graphs.
- It aids the risk management team to make decisions with better coverage amongst the huge data by considering intuitions of public data and important events across the globe.
- Enables the team to take all the necessary actions well in advance to avoid any kind of financial /reputational damage to the company.

Scope of Work:

- Select one industry to focus the final product/ solution on .
- Accessing news articles from web and all other possible sources .
- Trying different modelling approaches and interpreting the results.
- Creating the MVP.
- Running the MVP through industry/business experts to gather their feedback.

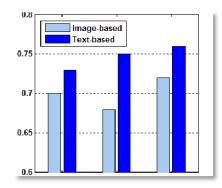
Solution Approach:



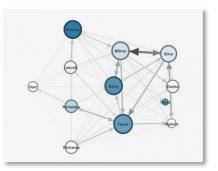
Phase 1

Dashboard Components: Data Preparation: DASHBOARD **GDELT** Select Company Name: XYZ Filter and List of top 10 recent Articles: Download all the news articles of Automobile List of top 10 positive Articles: Industry. List of top 10 negative Articles: Volume of News across the period Architecture Orator Tourism Ν Sentiments of the Articles Word Cloud Name Entity Data

Phase 2



Sentiments in Social Media



Network Graph

To provide ratings, by studying the correlation between the traditional methods and our results

(Graphs are only indicative)

processing

Recognition

Data Understanding:



- GDELT or Global Database of Events, Language, and Tone
- A forum that tracks the world's news media in print, radio, and online formats, in over 100 languages, across the globe, each moment of the day, extending back to January 1, 1979, to the present day which updates every 15 minutes.
- Today GDELT considers countless broadcasts, print, and online news sources like The Washington Post, United Press International, The New York Times, Foreign Broadcast Information Service, etc from all over the world in not less than a hundred languages, and its source list develops every day

Time window considered:	Industry under the scope:	Companies of Interest:		
November 2019 to August 2020	Automobile Industry	Volkswagen, Hertz, Ford, Tesla, Scania, Renault, Nissan, and Toyota		

DocumentIdentifier	Tone	Positive_S	Negative_	Polarity	Date_N	Month_Ye	Quotations
https://www.foxnews.com/tech/	0.411523	2.469136	2.057613	4.526749	20191001	201910	1404 205 There is a sense , though , that ma
https://www.nzherald.co.nz/fires	0	1.934524	1.934524	3.869048	20191001	201910	2696 25 It was lucky to save it .
https://indianexpress.com/article	-1.78926	1.192843	2.982107	4.17495	20191001	201910	2689 213 to approach various departments
https://www.houstoniamag.com	1.173709	4.225352	3.051643	7.276995	20191001	201910	5876 36 We're going on this grand adventur
https://governorswindenergycoa	-0.30534	3.053435	3.358779	6.412214	20191001	201910	665 34 may violate federal antitrust laws

(https://www.gdeltproject.org/, 2011).

Data preparation:

- a) Web Scraping and Named Entity Recognition:
- Web Scraping: All the content of a given URL is converted to a plain text using BeautifulSoup.
- Named Entity Recognition: Named Entity Recognition model is applied on the plain Text obtained after web scraping to identify all the organization the article is speaking about or mentioned.
- Further, label encoding is performed on the organizations' list for the given URL and is grouped under the eight specific companies.



						_) T 4 T
	0			ocumentidenti	fier		
	0 http	s://www.sfgat	e.com/news/b	ayarea/article/N	10		
	1	https://toronto	.citynews.ca/2	2019/09/30/off-d	ut		
	2 http	ps://www.jourr	nalgazette.net	/business/20190	09		
	3 http	s://www.sfgat	e.com/news/b	ayarea/article/N	10		
	4	https://toronto	.citynews.ca/2	2019/09/30/off-d	ut		
	5 htt	ps://www.jourr	nalgazette.net	/business/20190	9		
	6 http	s://www.foxne	ws.com/tech/	apples-steve-job	DS		
	7 http:	s://www.thesta	ar.com.my/nev	vs/nation/2019/1	10		
	8 htt	ps://www.auto	express.co.ul	/volkswagen/ca	lif		
	9	https://www.n	zherald.co.nz/	fires/news/articl	e		
1	0 http	ps://carbuzz.c	om/news/the-	atest-popemobi	le		
		Ĺ	<u> プ</u>				
Na					pa County S		
Eο					ards Bureau nizations,		
					pa County S		
					ards Bureau		
re					nizations,: Volkswagen		
					;Volkswage		
Vw	Grou				en,1168;Vol 331;Toyota		
Ni	ssan,				ssan,787;N		
		4	\Box				
For	d_Tag	Hertz_Tag	Tesla_Tag	Scania_Tag	Renault_Tag	Nissan_Tag	Toyota_Tag
	0	0	0	0	0	0	0
	0	0	0	0	0	1	
)							0
1	0	0	0	0	0	0	0
	0	0	0	0	0	0	
							0
	0	0	0	0	0	0	

Data preparation:



Quotations

992 210 | Autonomous vehicles will transform personal mobility x2026; xA0; reap the benefits of a new market which promises to ramp from 969 189 | All Tesla Model 3 Standard Range will not accept order after Oct 13th, this will be the last shipment of the Made in US Model 3, plz

5870 | 27 | | green industrial revolution

3119 | 75 | Is NHTSA going to successfully get Tesla to send out a new software update?

11730 | 134 | | We have a shot at achieving our first 100,000 vehicle delivery quarter, which is an incredibly exciting milestone for our company

5544 | 123 | | Subaru quandary is one of many reasons why California shouldn't be allowed to set de facto emission standards for the U.S.

5756 | 123 | | Subaru quandary is one of many reasons why California shouldn't be allowed to set de facto emission standards for the U.S.



b) WordCloud Generation:

Wordcloud is generated using the Text from the Quotation column for all companies to know what perspectives and emotions the participant is putting in.

- Missing rows are first eliminated since not all articles/ news published have excerpted statements
- Text is normalized by converting all the words to lower case and remove all unique character.
- NLTK is used to Stop word removal and perform stemming on the normalized Text to create a corpus.

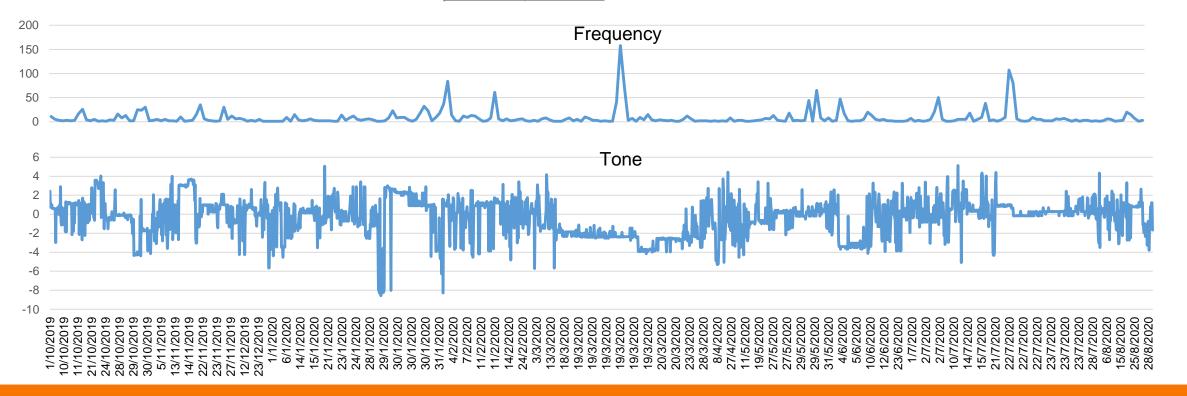
Transformed Datasets for Dashboard:



Date	Tone	Positive_S	Negative_	Polarity
1/10/2019	2.421308	4.72155	2.300242	7.021792
1/10/2019	0.728155	1.334951	0.606796	1.941748
1/10/2019	1.136364	2.272727	1.136364	3.409091
1/10/2019	0.980392	2.941176	1.960784	4.901961
1/10/2019	0.980392	2.941176	1.960784	4.901961
1/10/2019	0.60423	3.323263	2.719033	6.042296

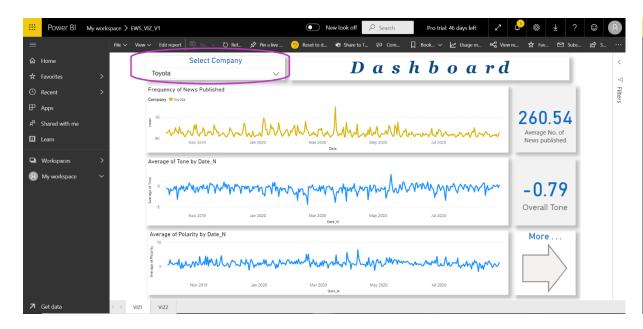
Date	count
20191001	11
20191002	5
20191003	3
20191004	2
20191005	3

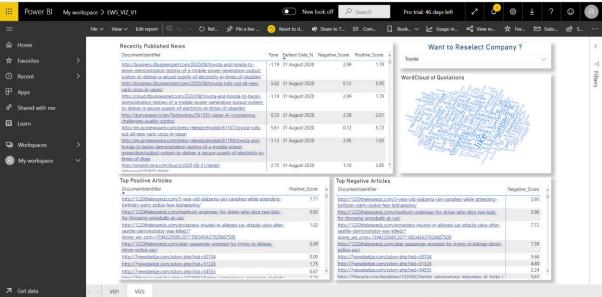
Date	DocumentIdentifier	Positive_S	Negative_So
31/8/2020	https://www.theglobeandmail.com/inves	2.216066	3.8781163
31/8/2020	https://www.torquenews.com/1083/toy	2.347418	1.1267606
31/8/2020	https://www.torquenews.com/comment	2.349624	1.1278195
29/8/2020	https://www.wealthdaily.com/articles/ba	2.51004	3.0120482
28/8/2020	https://www.moneycontrol.com/news/v	1.388889	2.4305556
28/8/2020	https://www.thesundaily.my/world/reute	2.496656	6.2862238



Dashboard on PowerBI:



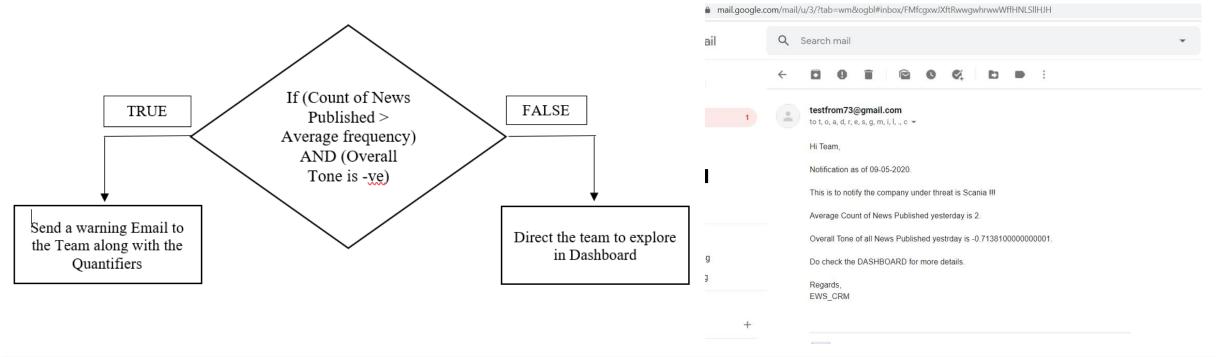




- The transformed data used to generate comprehensive components to create an interactive dashboard
- The analyst can select the company of interest from the provided dropdown, and all the components for the company will be displayed.
- One can even click on the URLs and navigate to the article he wishes from the given list.

Rule Based Email Generation:





- A simple condition-based model is created to automatically send out a warning mail to the team when the condition is satisfied.
- This model is designed to compare the News published for a particular company on the latest Date with the Average frequency of the News published till Date.
- When the overall tone of the News published for the company on the latest Date is negative, a warning mail is generated and sent to the team.

Conclusion and Future Work Recommendations :



- This Dashboard provides a list of recent articles and their respective sentiment scores about the company of interest can navigate to those articles and read .
- Along with a WordCloud that gives an essence of emotions in excerpted statements, a dropdown given to choose the company of interest and explore all the Dashboard mentioned above components.
- Also, few quantifiers are highlighted on the Dashboard, depending on which a rule-based model is created to send out automated generated warning Emails to the concerned team.
- Dynamic analysis and affinity analysis on the data extracted and discover any kind of relationships or insights on the activities, incidents, etc. recorded about the company of interest.
- Creating a network graph would help to focus other companies prone to be in the same bandwidth of risk as the company of interest.
- The next major step would be to bridge an association between the Dashboard's output and the traditional credit risk modelling methods by converting the Dashboard outputs to a generalized rating system given to the company of interest.

Code Repo and References



GitHub Link

https://github.com/tiwars10/Capstone_1/blob/main/EWS_CRM1.ipynb

References

- -, Z. Y., & -, S. Z. (2013). Bank Credit Risk Management Early Warning and Decision-making based on BP Neural Networks.

 INTERNATIONAL JOURNAL ON Advances in Information Sciences and Service Sciences. https://doi.org/10.4156/aiss.vol5.issue9.51
- Davis, E. P., & Karim, D. (2008). Comparing early warning systems for bankingDavis, E. P., & Karim, D. (2008). Comparing early warning systems for banking crises. Journal of Financial Stability. https://doi.org/10.1016/j.jfs.2007.12.004 crises. *Journal of Financial Stability*. https://doi.org/10.1016/j.jfs.2007.12.004
- Fuertes, A. M., & Kalotychou, E. (2007). Optimal design of early warning systems for sovereign debt crises. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2006.07.001
- https://www.gdeltproject.org/. (2011). *The GDELT Project*.
- McKinsey&Company. (2012). Building credit monitoring for competitive advantage. *McKinsey Working Papers on Risk*. https://www.mckinsey.com/~/media/mckinsey/business functions/risk/our insights/a better way for banks to monitor



Thank You!



Q&A