News Risk Alerting System (NRAS): A Data-Driven LLM Approach to Proactive Credit Risk Monitoring

Adil Nygaard[*], Ashish Upadhyay[*], Lauren Hinkle[*], Xenia Skotti[*],
Joe Halliwell, Ian Brown, Glen Noronha
JPMorganChase
{adil.nygaard, ashish.x2.upadhyay, xenia.skotti}@jpmchase.com

Abstract

Credit risk monitoring is an essential process for financial institutions to evaluate the creditworthiness of borrowing entities and minimize potential losses. Traditionally, this involves the periodic assessment of news regarding client companies to identify events which can impact their financial standing. This process can prove arduous and delay a timely response to credit impacting events. The News Risk Alerting System (NRAS) proactively identifies credit-relevant news related to clients and alerts a corresponding Credit Officer (CO). This production system has been deployed for nearly three years and has alerted COs to over 2700 credit-relevant events with an estimated precision of 77%.

1 Introduction

Credit risk management is a systemic process to evaluate and monitor the solvency of borrowing entities, allowing financial institutions to understand and detect emerging risks. Previous algorithmic approaches for credit risk have focused on financial assessment, generally using a company's financial statements (Clements et al., 2020, Golbayani et al., 2020). However, these financial reports are produced relatively infrequently and often lack wider commercial context, so monitoring credit-impacting news events is a necessary part of effective credit risk management. As financial institutions often have credit portfolios that contain a multitude of clients, this manual analysis can be labor intensive. Leveraging natural language processing (NLP) and machine learning (ML) can help expedite this process.

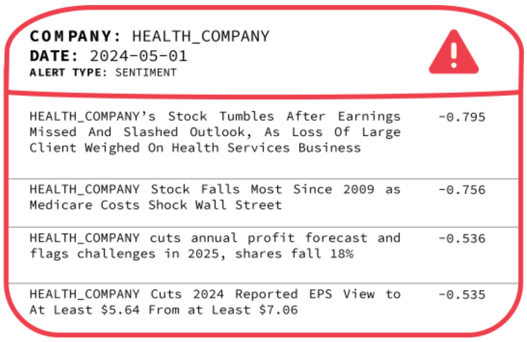

Figure 1: An anonymized NRAS alert

This paper presents the News Risk Alerting System (NRAS) which proactively alerts Credit Officers (COs) to credit-impacting news events about client entities in their portfolio. NRAS identifies these events through large language model (LLM) (Rogers and Luccioni, 2024) enabled high-precision content filtering and the volumetric analysis of news. Alert generation utilizes a dynamic volumetric threshold to account for the variability in news coverage of companies and to prevent spurious or duplicative alerts. NRAS consists of two alerting subsystems: negative sentiment and mergers & acquisitions, with additional components, such as filtering and deduplication of headlines applied to further enhance the effectiveness of generated alerts.

These different components are holistically integrated and deployed within a real-world scalable system. NRAS is designed to process over 20,000 news articles a day and generate event-driven, timely, and actionable alerts for COs. This allows for a more proactive and comprehensive credit risk review process. COs can promptly identify relevant events in personalized alerts generated from a diversified set of news sources (see Figure 1 for an example alert).

[*] Equal contributions; first name alphabetical order

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track, pages 429–439
November 12-16, 2024 ©2024 Association for Computational Linguistics

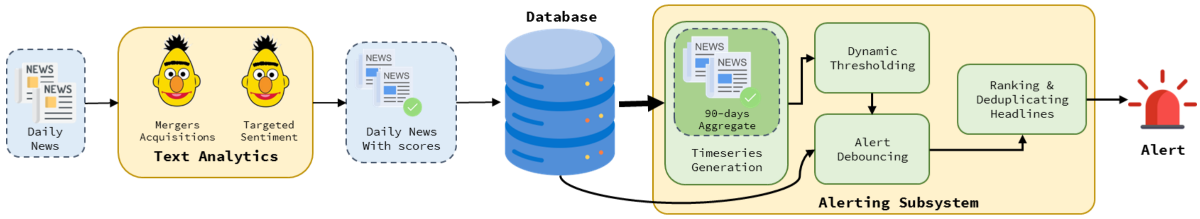

Figure 2: Overview of the News Risk Alerting System (NRAS)

This paper details the empirical approach taken in the development of NRAS and the real-world evaluation of news alerts. Section 2 of this paper describes the system holistically, including the LLM-based text analysis and the volumetric analysis used in alert generation. Section 3 explains the development of NRAS and the associated experimentation for model and parameter selection. Section 4 discusses the real-world system evaluation by end-users that is incorporated into NRAS. Related work is discussed in Section 5, including other NLP approaches for the credit risk monitoring of news data, before the paper concludes in Section 6.

2 System Overview

NRAS is comprised of two main components: text analysis and alert generation. This section explains the design decisions and structure of these components. An overview of the system architecture of NRAS can be seen in Figure 2.

2.1 NRAS Architecture

NRAS proactively identifies news events which may impact a company's credit risk. NRAS raises an alert if there is an anomalous increase in the volume of credit-impacting news about a company. The alerts are either about articles with negative sentiment about the company or a company related mergers & acquisitions event.

NRAS processes over 20,000 news articles daily to generate approximately 10 credit-relevant alerts per day. Each input article is accompanied by structured information, including identifiers for companies mentioned within the article and their corresponding spans within the text. The metadata also identifies which companies are focal, where the article is primarily about that company, and which are merely incidental, where the company is tangentially related to the news event.

Each news article is then processed by two text analytics to identify the credit-relevance of the underlying news event. The two analytics are: **Targeted Sentiment**, which assigns a sentiment score for each focal company mentioned within the article and **Mergers and Acquisitions** (M&A), which determines the probability that each news article is about M&A activity.

Finally, NRAS generates alerts through volumetric analysis and anomaly detection on daily news counts. This process starts by producing a timeseries of news volume over the previous 90-day period, which is used to dynamically calculate a threshold for alerting. Recent alerting activity can raise the minimum threshold required. If the news volume exceeds the calculated threshold an alert is generated and sent to COs for review.

2.2 Text Analytics

NRAS includes two text analysis models which evaluate the credit-relevance of each article through targeted sentiment analysis and relation to M&A activity.

2.2.1 Targeted Sentiment

The sentiment model identifies the scope of the positive or negative impact a news event may have on the financial standing of a company. Each focal entity within a news article is assigned a sentiment score which can range between -1 to +1, for negative and positive news respectively. The sentiment model consists of a fine-tuned BERT (Vaswani et al., 2017) model, which utilizes target-dependent sentiment (Gao et al., 2019) with a custom regression head. This targeted approach was taken as a news event may impact each of the companies mentioned in the news article to different degrees.

The sentiment model was fine-tuned on a dataset of 5348 news headlines sampled between 2019 and 2021, which consisted of 5194 headlines in a combined training and validation set and 154 headlines in a held-out test set (See Appendix 7.2 for descriptive statistics of data sets). Each annotated headline was assigned a sentiment score

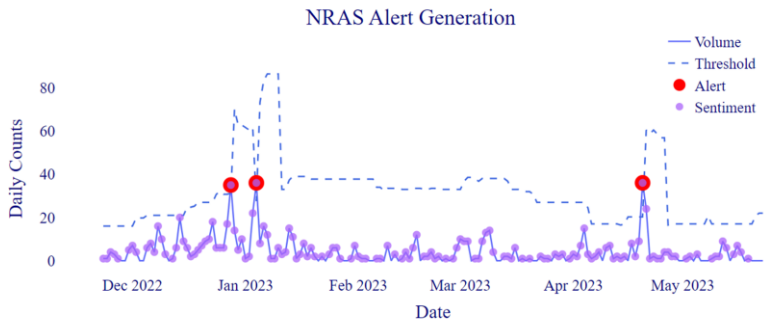

Figure 3: Alert generation for a company. Given the negative daily counts (volume), we compute the dynamic threshold and raise an alert if the volume is sufficiently high. When alert is generated, the threshold is doubled (alert debouncing)

for each company mentioned within the headline and annotations were performed by multiple annotators. Each company's annotated targeted sentiment score represents the averaged annotated score. Inter-annotator agreement was approximately 80% with 0.69 as an estimated lower bound for Krippendorff's alpha. The headlines were annotated with sentiment scores for each focal company to denote the relative positive or negative effect. The articles were sampled from the same top financial news sources which NRAS uses for daily alerting.

The references to these focal companies were masked in each article headline before training to ensure that biases did not arise. For hyper-parameter tuning, 5-fold cross validation was used.

Targeted sentiment using BERT was found to be the most efficacious approach when compared to traditional regression models such as XGBoost Regression and Support Vector Regression, with a mean squared error (MSE) of 0.0312 on the test set (see Appendix 7.1 for comprehensive model comparison).

2.2.2 Mergers & Acquisitions (M&A)

The M&A model is a binary classifier that predicts the probability of a news article being about an M&A event. It is a DeBERTa-based (He et al., 2023) classifier that takes as input a text created by concatenating the title and body of an article truncated at 128 tokens. All focal company references in the concatenated text are masked with generic tokens to prevent associative bias to any particular company. Each focal company with a title span in the news article is then assigned the same probability as its M&A score. This score ranges between 0 & 1 where a score greater than or equal to 0.5 is considered M&A credit-worthy and classified as an M&A article.

The M&A dataset consists of 1606 news articles divided into training, validation, and test sets using a 60:20:20 split. The timeseries nature of the data was taken into account when splitting the dataset, using their real-world publication date (see Appendix 7.2 for the distribution and timeframe for each set). Each of these news articles were labelled as either MA or NOT_MA. Each article was annotated by a Subject Matter Expert (SME) who considered the title and the first 500 words of body text.

The dataset was sampled to ensure the equal representation of M&A news. Each article sampled was published in 2022, and sampling was performed both from a pool of news articles identified as M&A by previous experimentation, and from a uniform sample. The articles were sampled from the same top financial news sources which NRAS uses for daily alerting.

Multiple models ranging from Random Forest, SVM with TF-IDF vectors, BERT, RoBERTa and DeBERTa were evaluated. DeBERTa was the best performing model with the highest Macro-F1 score of 92.1 on the test set (for more details, see Appendix 7.1).

2.3 Alerting Subsystem

NRAS utilizes timeseries analysis to determine when an anomalously high volume of relevant news is occurring for a client entity and raises an alert accordingly. The threshold of news volume required to generate an alert is determined dynamically based on recent news coverage for each company. The dynamic threshold allows each company to be assessed based on their distinct news volume history, which means that smaller client entities with lower average news coverage are as likely to raise an alert as a larger client. Figure 3 visualizes the different steps of alert generation for a single company.

2.3.1 Timeseries Generation

A timeseries of the daily counts of relevant news articles about a company is generated. Articles must contain company title spans, but are otherwise selected differently for each alerting subsystem. For sentiment, we consider relevant news to be articles with a score lower than –0.4 for the focal company. While for M&A, all articles classified as M&A are included in the daily counts.

2.3.2 Dynamic Thresholding

For an alert to be generated, the count of relevant articles for a company, **c**, on a given day must exceed a dynamic volume threshold $\boldsymbol{t_c}$, as defined in [Equation 1]. This threshold is defined as either a minimum article threshold, $\boldsymbol{MAT}$, or a robust scaled average based on the volume of relevant news articles over the prior **d** days, which is calculated using the mean, $\boldsymbol{\mu_{c,d}}$ interquartile range (IQR), $\boldsymbol{IQR_{c,d}}$ and a corresponding multiplier $\boldsymbol{m}$.

This variant of robust scaling is used to determine whether the current news volume is significantly elevated in comparison to the recent historical volume of relevant news. The threshold represents the minimum number of articles a company needs to have on any given day to trigger an alert. This requires companies with higher average volumes of historic news coverage to achieve a higher number of articles in a day to produce an alert, and for companies with low historic news coverage to require fewer. The use of a minimum volume for the threshold ensures that the alerting system is not overly responsive to noise.

This rolling, dynamic threshold represents a different minimum daily count for each company on each day and is defined as follows:

$$t_c = \max\left(MAT,\ \mu_{c,d} + IQR_{c,d} \times m\right) \quad (1)$$

As the baseline is clamped to a minimum value and is always greater than zero, an “alert level” can be calculated as the ratio $\boldsymbol{r_c}$, of the daily volume of news articles to the daily dynamic threshold. When $\boldsymbol{r_c \geq 1.0}$ the volume of news articles for a company on a particular day is above the daily minimum threshold and an alert is generated.

2.3.3 Alert Debouncing

As news stories develop, it is often the case that news articles concerning the underlying event will be published over the course of multiple days. It is undesired behavior to raise multiple alerts for the same on-going news event unless circumstances have significantly changed or worsened. Thus, in order to avoid producing alerts for the same news event, when an alert is raised, NRAS requires the dynamic threshold for the following seven days to be at least double the alert level, r_c. This means that any alert raised within seven days of the most recent previous alert requires more than double the news volume to be generated. This ensures that multiple alerts will only be raised within a 7-day period if the news coverage surrounding a company significantly increases. This alert threshold doubling can be seen in Figure 3.

2.3.4 Deduplicating & Ranking Headlines

An alert for a company displays the most informative headlines of the day. The headlines are ranked by sentiment (ascending order) or M&A score (descending order), and penalized if they are published by low-quality sources.

Headlines are de-duplicated using a Locality Sensitive MinHash (LSH) clustering algorithm, with each headline being assigned to a cluster, and only the top headline per cluster being included within the alert. Template generated articles are identified and filtered via regular expressions. The top four ranked headlines are then shown to end-users.

3 System Development

NRAS has two alerting subsystems, one for sentiment and another for M&A. These subsystems utilize the same underlying architecture, with minimal changes to hyperparameters. This architectural configuration was initially used for the sentiment alerting stream, but proved flexible enough to add a new M&A alerting stream with minimal modifications. This section discusses the experiments performed for the parameter selection of each subsystem. Parameters were selected to prioritize precision, but with a secondary consideration for the number of alerts which was used as a proxy for real-world recall.

PC	MAT	MST	Alerts	P	R
S1	5	-0.3	1514	96.7	48
S2	5	-0.35	1313	96.9	48
S3	5	-0.4	1113	97.9	48
S4	4	-0.35	1798	96.7	52
S5	**4**	**-0.4**	**1551**	**97.3**	**52**

Table 1: Sentiment alerting parameter selection.

3.1 Sentiment Alerting

News data from January 2018 to July 2021 for 800 companies was used to generate sentiment alerts using different parameters. These parameters were the minimum article threshold (MAT) and maximum sentiment score threshold (MST) for an article to be considered negative. Alerts were generated from April 2018 to July 2021, using the first 90 days of the dataset to backfill volume counts for timeseries generation. Each alert was then manually evaluated by an SME and was marked as relevant or irrelevant accordingly.

Table 1 shows the parameter configurations (PC) considered and the precision (P), recall (R), and the number of alerts generated (Alerts) over the 39-month test period. Precision is calculated as the percentage of relevant alerts over total alerts generated. Recall is calculated as the percentage of alert generated by the specific PC out of a superset of all alerts generated by all the PC variations.

Of the experiments above, S5 was selected as our production PC with MAT 4 and MST -0.4 because it has the second highest precision, 97.3, and the highest recall, 52. The IQR multiplier was fixed to 5 based on previous experiments with a generic sentiment model that was initially considered (for more details, see Appendix 7.3).

3.2 M&A Alerting

The estimation dataset for this experiment was composed of news articles from October 2021 to December 2022 for over 1100 companies. Alerts were generated from January 2022 to December 2022, using the first 90 days of the dataset to backfill volume counts for timeseries generation. Table 2 summarizes the different versions of the M&A subsystems tested with precision calculated similarly to sentiment.

The M&A alerting subsystem only considers articles which have the company mentioned in the title. In addition, a keywords-based rules overlay is applied as the post-alert relevancy filter to verify that at least one of the articles contains keywords such as buy, sell, deal, merge, or acquire in the title. The M1 version of the M&A Alerting Subsystem with MAT 4 and IQR 5 was selected. It achieved the second highest precision of 93.53, but with a higher number of alerts generated overall which was used as a proxy for recall. Additional experiments without title span requirement are shown in Appendix 7.3.

PC	MAT	IQR	Alerts	Precision
M1	**4**	**5**	**572**	**93.53**
M2	4	6	561	93.76
M3	5	5	447	93.28
M4	5	6	438	93.60
M5	6	5	365	92.87
M6	6	6	364	92.85
M7	7	5	300	92.66
M8	7	6	299	92.64

Table 2: M&A alerting parameter selection

4 Real-World Evaluation

Each alert produced by NRAS is reviewed by COs, who assess the relevance of the alert to the credit rating and risk review process. The production system incorporates this ongoing performance monitoring and inbuilt end-user feedback to enable continuous system improvements. Alerts are categorized using the following five class typology:

- **New Information**: Alerted event represents new and relevant information to the client's credit profile.
- **Recently Considered**: Alerted event pertains to credit relevant news, but it was already evaluated by COs, either through previous alerts or manual news review.
- **Financial Factor**: Alerted event would have otherwise been relevant to the client's credit risk, but the impact was mitigated by the client's financial standing.
- **Other Factors**: Alerted event would have otherwise been relevant to the client's credit risk, but the impact was mitigated by other factors related to the client entity, such as collateral support.
- **Irrelevant Event**: Alerted event was irrelevant to the client's credit profile.

Further clarification of this typology and example alerts can be found in Appendix 7.4.

Action	Sentiment (Sep21-May24)		M&A (Aug23-May24)	
	Count	%	Count	%
New Information	509	21.31	301	27.87
Recently Considered	418	17.5	158	14.63
Financial Factor	892	37.35	265	24.54
Other Factor	127	5.32	57	5.28
Irrelevant	442	18.51	299	27.69
Total	2388	-	1080	-

Table 3: Distribution of overall Sentiment and M&A alerts by categories

4.1 Sentiment

Table 3 represents the 2000+ sentiment alerts generated by NRAS since its inception in September 2021, of which 21.31% were classed as New Information. This corresponds to over 500 news events about which COs were successfully alerted. To improve the efficacy of NRAS and optimize system performance, an effort was taken to reduce the number of Irrelevant Alerts generated. Analysis of the Irrelevant Alerts showed that many originated from the same news sources and were produced automatically via templates. These templates contained semantically charged diction and were thereby assigned negative sentiment. Steps were taken in early 2022 to mitigate the overrepresentation of these headlines among the sentiment alerts by filtering out these templates. This effectively halved the number of irrelevant alerts produced over the subsequent year and a half, as seen in Figure 4.

4.2 Mergers & Acquisitions

Table 3 shows the distribution of all M&A alerts generated by NRAS, of which almost 73% are about credit-worthy M&A events. 27% are categorized as New Information, meaning NRAS proactively alerted COs about credit-worthy M&A events over 300 times. 27% of alerts were classified as Irrelevant. About 25% of these Irrelevant Alerts are considered immaterial events by COs, which means that these are M&A activities, but have an insignificant impact on the client's credit rating. Similarly, 30% of these alerts are about business transactions between two entities that are not considered M&A activity. Figure 5 shows the percentage of alerts in each category distributed by month.

5 Related Work

M&A using NLP: Research has explored the utility of different ML and NLP techniques in predicting M&A events and their associated roles (Routledge et al., 2013, Katsafados et al., 2021, Moriarty et al., 2019). Traditional NLP approaches to predicting M&A activity utilize textual data from 10-K SEC filling reports of publicly traded US companies (Lohmeier and Stitz, 2023). A few notable approaches include the use of Logistic Regression models with n-gram features to predict the likelihood of the filing company being involved in an M&A deal within the next year. Some works have also found that using tabular financial indicator data combined with information from filling reports can substantially improve the performance of M&A prediction models (Sanchez-Blanco Gómez, 2022).

Sentiment Analysis for Credit Risk Monitoring: Sentiment analysis of financial news data has proven to be an effective and indicative method of monitoring credit risk (Duan and Yao, 2022). Identifying news articles which contain semantically charged diction can be used as a proxy to indicate the credit impact of the underlying news event (Tran-The, 2020). The degree of negative sentiment in news data can also adequately predict credit rating downgrades for corporate entities (Tsai et al., 2010). A high volume of semantically negative news data regarding an entity is correlated with an increase in the risk of credit default for that entity (Tsai et al., 2016).

Credit Risk Alerting for News Events: There are a few commercial products that produce alerts about credit-adverse news events for companies using sentiment analysis (Dow Jones, 2024, FitchRatings, 2024, Moody's, 2024, Zanders, 2024). Though the details about most of these systems are not publicly available, Ahbali et al. (2022) detail their approach. There are fundamental differences between their approach and NRAS: including the volumetric analysis, the use of fixed credit risk scores and the sentiment analysis models. NRAS uses continuous sentiment scores as opposed discrete categorizations. Additionally, instead of classifying the news event and then assigning a fixed severity score, NRAS implicitly encapsulates the event severity within the sentiment score itself. NRAS also offers different streams of risk alerting other than sentiment, such as M&A, without requiring any significant changes to the system architecture.

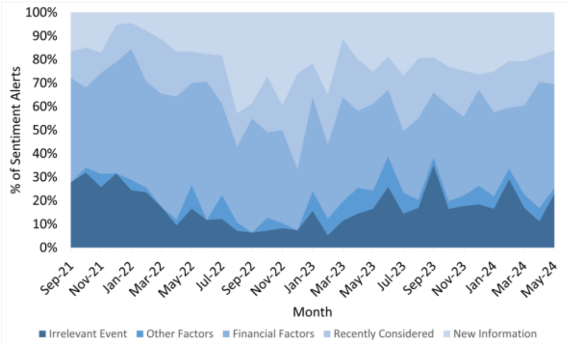

Figure 4: Sentiment alerts over time

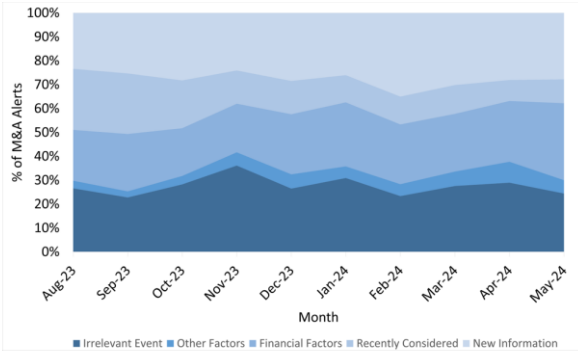

Figure 5: M&A alerts over time

6 Conclusion

This paper presents NRAS, a system to proactively identify and alert for credit adverse news events. It is designed to process thousands of news articles daily, and proactively generate real-time actionable alerts. In the three years that NRAS has been in production, it has generated 1946 and 781 credit-relevant alerts for sentiment and M&A respectively, of which 509 and 301 were marked as new information.

NRAS automates the news monitoring process, enhancing credit risk management by analyzing large news volumes from a diverse set of sources, including smaller publications. This enables the evaluation of smaller client entities who might not have as much media coverage in major financial publications. The custom dynamic threshold allows for each client entity to be assessed based on their distinct news volume history, which means that entities with lower average news coverage are just as likely to have alerts raised.

This paper presents a detailed account of the system's development and evaluation processes, addressing the general lack of public information on how commercial systems are built and assessed. NRAS is designed to scale to multiple alerting streams without requiring changes to the overall architecture, while its modular design allows for seamless integration of additional components such as filtering of templated headlines, deduplication, and clustering.

The individual components of NRAS, such as sentiment analysis and M&A classification, are based on established techniques, however, the novelty of our work lies in their holistic integration and deployment within a real-world, scalable system. This integration and the system's ability to dynamically adjust thresholds to prevent spurious or duplicative alerts are key innovations that enhance its practical applicability and effectiveness.

While NRAS has been developed for credit risk management, the underlying framework is versatile and adaptable to other domains requiring real-time news monitoring and alerting. This adaptability can be achieved by integrating models which recognize relevant news events in other domains. The system can also extend monitored entities to include individuals or countries, in addition to corporate entities. For instance, specifying the target entity to be a person or location allows NRAS to generate alerts based on their news volume while utilizing the same underlying mechanisms. This flexibility demonstrates the system's broader applicability beyond just company monitoring for credit risk.

Concept drift may be a limitation of NRAS due to an evolved understanding of the problem over time. For example, new types of events may become relevant to COs, and the underlying model should reflect that change. This can only be identified through discussions with SMEs as with the current real-world evaluation setup it would not be detected.

Possible future enhancements include: the continuous evaluation and retraining of the underlying models with expanded datasets; increasing the number of alerting streams; as well as integrating new components which perform detailed information extraction. For example, identifying the buyers, sellers, and deal size of an M&A transaction to measure its material impact on the parties involved. Additionally, the information presented to COs can be improved by leveraging recent summarization advancements with GenAI, which can provide more fine-grained information about the cause of an alert.

Acknowledgement

We would like to acknowledge the contributions and assistance that the Research and Engineering team provided throughout the development and implementation process. We would also like to thank our partners within the Credit Risk division who have helped us with their expertise in the domain and with their continuing support.

References

Noujoud Ahbali, Xinyuan Liu, Albert Nanda, Jamie Stark, Ashit Talukder, and Rupinder Paul Khandpur. 2022. Identifying Corporate Credit Risk Sentiments from Financial News. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Track*, pages 362–370, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Jillian M Clements, Di Xu, Nooshin Yousefi, and Dmitry Efimov. 2020. Sequential deep learning for credit risk monitoring with tabular financial data. *arXiv preprint arXiv:2012.15330*.

Jin-Chuan Duan and Xuan Yao. 2022. Media sentiments for enhanced credit risk assessment. *NUS Credit Research Initiative*.

FitchRatings, 2024. Fitch Solutions News sentiment. https://www.fitchsolutions.com/credit/bank-sovereign-news (date accessed: 17/7/2024)

Zhengjie Gao, Ao Feng, Xinyu Song, and Xi Wu. 2019. Target-dependent sentiment classification with BERT. *IEEE Access, 7:154290–154299.*

Parisa Golbayani, Dan Wang, and Ionut Florescu. 2020. Application of deep neural networks to assess corporate credit rating. arXiv preprint arXiv:2003.02334.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2023. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. The *Eleventh International Conference on Learning Representations.*

Dow Jones, 2024. Factiva sentiment signals. https://www.dowjones.com/professional/factiva/factiva-sentiment-signals/ (date accessed: 17/7/2024)

Apostolos G Katsafados, Ion Androutsopoulos, Ilias Chalkidis, Emmanouel Fergadiotis, George N Leledakis, and Emmanouil G Pyrgiotakis. 2021. Using textual analysis to identify merger participants: Evidence from the us banking industry. *Finance Research Letters,* 42:101949.

Nils Lohmeier and Lennart Stitz. 2023. Identifying m&a targets from textual disclosures: A transformer neural network approach. *Available at SSRN 4373306*.

Moody's, 2024. Credit Sentiment Score. https://www.moodysanalytics.com/-/media/products/MA-NewsEdge%20Credit%20CSS%20Overview%203.pdf (date accessed: 17/7/2024)

Ryan Moriarty, Howard Ly, Ellie Lan, and Suzanne K McIntosh. 2019. Deal or no deal: Predicting mergers and acquisitions at scale. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 5552-5558. IEEE.

Anna Rogers, and Alexandra Sasha Luccioni. 2024. "Position: Key Claims in LLM Research Have a Long Tail of Footnotes." arXiv preprint arXiv:2308.07120.

Bryan R Routledge, Stefano Sacchetto, and Noah A Smith. 2013. Predicting merger targets and acquirers from text. *Carnegie Mellon University working paper.*

Juan David Sánchez-Blanco Gómez. 2022. Identifying likelihood merger participants through natural language processing in the european banking sector. Repositorio de la Universidad Pontificia Comillas.

Tam Tran-The. 2020. Modeling institutional credit risk with financial news. *arXiv preprint arXiv:2004.08204.*

Feng-Tse Tsai, Hsin-Min Lu, and Mao-Wei Hung. 2010. The effects of news sentiment and coverage on credit rating analysis. In *Proc. of Pacific Asia Conference on Information Systems (PACIS)*.

Feng-Tse Tsai, Hsin-Min Lu, and Mao-Wei Hung. 2016. The impact of news articles and corporate disclosure on credit risk valuation. *Journal of Banking & Finance*, 68:100–116.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Anna Rogers, and Sasha Luccioni. 2024. Position: Key Claims in LLM Research Have a Long Tail of Footnotes. *Forty-first International Conference on Machine Learning*.

Zanders, 2024. News Sentiment Analysis and Credit Risk. https://magazine.zandersgroup.com/zm2021-2-eng/news-sentiment-analysis-and-credit-risk (date accessed: 17/7/2024)

7 Appendices

7.1 Text Analysis Model Selection

Table 4 demonstrates the performance of different regression approaches for sentiment and their Mean Absolute Error (MAE) and Mean Squared Error (MSE). While Table 5 shows the different models considered for M&A classifier.

As demonstrated in the tables, the highest performing model for Targeted Sentiment was BERT with 0.136 MAE and 0.0312 MSE and therefore was our model of choice. Similarly, for M&A classification DeBERTa had the highest mean F1 score of 92.9, with a standard deviation of 2.47 over 10 runs.

Model	MAE	MSE
BERT	**0.1360**	**0.0312**
XGBoost Regressor	0.1696	0.0525
Support Vector Regression (SVR)	0.1682	0.0571
Gradient Boosting Regressor	0.2009	0.0683
Linear Regression	0.6397	0.6610

Table 4: Targeted Sentiment model selection experiments

Model	Precision	Recall	F1
BERT	92.54±1.21	91.63±1.88	91.78±1.81
RoBERTa	92.97±2.28	92.1±2.47	92.25±2.49
DeBERTa	**93.29±1.87**	**92.83±2.54**	**92.9±2.47**
SVM	85.9	85.9	85.9
Random Forest	84.85±1.26	84.89±1.26	84.85±1.25
xGBoost	82.43	82.53	82.41

Table 5: M&A model selection experiments

7.2 Text Analysis Estimation Datasets

The label distributions for the estimation datasets used for training and evaluating the two LLMs in the text analysis module of NRAS, are demonstrated in the following subsections.

The distribution of the annotated sentiment scores for the training/validation and test sets for the Targeted Sentiment model can be seen in Figure 6 and Figure 7. The label distribution for the training, validation, and test sets for the M&A model are shown in Table 6 and the timeframe for each set is shown in Figure 8.

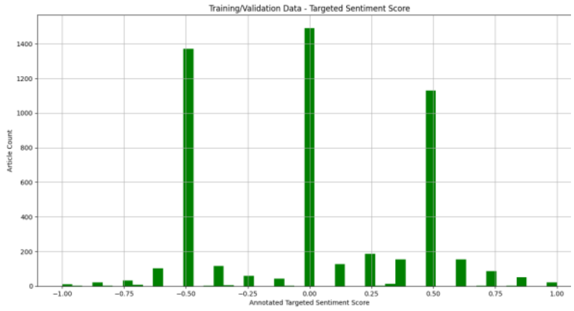

Figure 6: Histogram of Annotated Sentiment Scores – Training/Validation Set

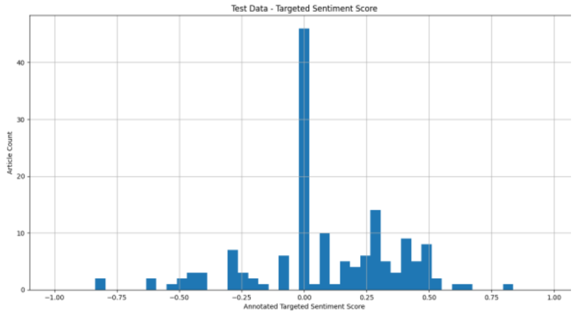

Figure 7: Histogram of Annotated Sentiment Scores - Test Set

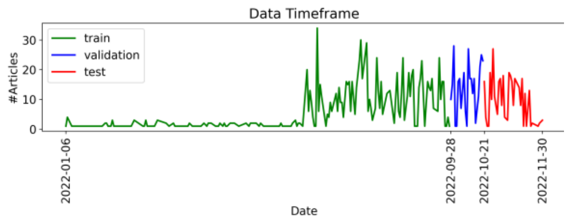

Figure 8: M&A classification model's estimation dataset – timeframe of set split

Set	MA	NOT_MA	Total
Training	476	511	987
Validation	174	132	306
Testing	165	148	313
All	815	791	1606

Table 6: Label distribution of M&A classification model's estimation dataset

7.3 Alerting Subsystem Model Selection

Our current sentiment model is targeted; however, we initially also considered a generic (G) sentiment model. A generic model considers the sentiment at a headline level instead of predicting a sentiment score with respect to a particular

PC	IQR	Alerts	Precision	Recall
G1	5	1320	91.2	52
G2	4	1372	91.3	52
G3	3	1475	90.6	52
G4	2	1615	90.0	52

Table 7: Sentiment alerting subsystem parameter selection

company. This generic sentiment model was classification based and instead of considering the maximum sentiment score for selecting relevant articles, we considered relevant articles to be any that were classified as either NEGATIVE and VERY_NEGATIVE.

At this stage we experimented only with the IQR multiplier. Table 7 shows the results, where G2 achieves the highest precision. On the next phase of experiments with a targeted model, we fixed the IQR multiplier to 5 based on discussions with SMEs and experimented with the remaining parameters.

For the M&A alerting subsystem, we also experimented with a version where news articles considered for alerting were not required to have the company mentioned in the title. The results from this experiment are shown in Table 8.

Model	MAT	IQR	Alerts	Precision
M9	4	5	791	92
M10	4	6	798	92
M11	5	5	646	91.8
M12	5	6	623	91.7
M13	6	5	531	92.1
M14	6	6	523	92
M15	7	5	448	92.6
M16	7	6	441	92.5

Table 8: M&A alerting subsystem parameter selection

7.4 Alert Evaluation

Each evaluation categorization for an NRAS alert encapsulates the novelty and utility of the underlying event to credit risk analysis. The determination of whether an alert belongs in each category is made by credit risk analysts. Examples of a redacted alert belonging to each category in the typology is shown in Table 9.

Evaluation Category	Example Alert	Comments/Explanation
New Information	COMPANY Shares Plunge After Warning of Losses from Metal Theft. COMPANY's Copper Theft and Uncertainty Prompt Ratings Downgrade. Massive Metals Theft Reported at one of Europe's Largest Copper Producers; COMPANY shares dropped 15% after the company said it could face losses of hundreds of millions of euros.	Represents information which is relevant to credit risk analysis and new actionable information for credit risk officers
Recently Considered	COMPANY Shares Fall After FDA Advisers Weigh In On Heart-disease Drug COMPANY's Ratings Tumble After FDA Advisors Dash Its Hopes of Releasing Heart-disease Drug	Presents new actionable information (as above), but was previously identified by credit risk officers before alert
Financial Factor	COMPANY Quarterly Profit Drops With Rise in Provision for Credit Losses. COMPANY reports $1.34B Q3 profit, down from $1.76B a year ago. COMPANY profits down on higher loan loss provisions after revised economic outlook	Represents an impactful news event, but the financial impacts are mitigated by the financial strength and standing of the company
Other Factor	COMPANY's legal loss could cost £113m in sales and higher prices for consumers. COMPANY's loses court battle over new regulations that will cost firm millions. COMPANY's loses legal challenge over new food promotion rules.	Represents an impactful news event, but the legal and reputational impacts are mitigated by the company's reputational strength and credit history
Irrelevant	COMPANY to Close Stores in New York, San Francisco Citing Safety, Theft Concerns COMPANY to Close Stores in San Francisco, Other Cities, Citing Theft; Nine stores, including in Portland, Ore., New York City and Seattle, are also on the list. COMPANY to shut 9 stores across 4 US states amid rising retail crime	Represents news that is irrelevant to credit risk analysis or otherwise unimpactful to company's credit standing.

Table 9: Examples of alerts from different categories