Wang, Han, Wen Wang, Feng Li, Yanfei Kang, and Han Li. 2025. "Catastrophe Duration and Loss Prediction via Natural Language Processing." *Variance* 18 (April). https://doi.org/10.66573/001c.133589.

Data Management and Information

Catastrophe Duration and Loss Prediction via Natural Language Processing

Han Wang[a], Wen Wang[b], Feng Li[c], Yanfei Kang[d], Han Li[e]

Keywords: Catastrophe duration, Catastrophe loss, Predictive analysis, Wildfire, Hurricane, Natural language processing

https://doi.org/10.66573/001c.133589

Variance

Vol. 18, 2025

Textual information from online news is more timely than insurance claim data during catastrophes, and there is value in using this information to achieve earlier damage estimates. This research used text-based information to predict the duration and severity of catastrophes. We constructed text vectors using Word2Vec and BERT models, then used Random Forest, LightGBM, and XGBoost as learners, all of which showed more satisfactory prediction results. This new approach provides timely warnings of catastrophe severity, which can aid decision making and support appropriate responses.

Address for Correspondence: han.li@unimelb.edu.au

1. INTRODUCTION

Climate change poses one of the gravest risks to humanity in the 21st century (Reichstein et al. 2013). For insurance industries, the impact of natural catastrophes is increasingly important for managing and mitigating adverse claim experiences, and thus for solvency considerations. According to the Swiss Re Group, natural catastrophes caused total economic losses of US$190 billion in 2020, leading to US$89 billion in insurance payouts (Swiss Re Institute 2021). In 2022, the total economic losses from natural catastrophes rose to US$275 billion, resulting in US$125 billion in insurance payouts (Swiss Re Institute 2023), which is the fourth highest total for a single year.

Accurately predicting the severity of catastrophes remains a challenge that requires further scholarly attention and research. The catastrophe modeling literature includes classical research, such as Grossi (2005), Panjer and Willmot (1992), and Kozlowski and Mathewson (1995), as well as more recent publications, such as Reed et al. (2020), Vijayan et al. (2021), and Zanardo and Salinas (2022). Our research explored the potential of leveraging text data from news articles for predicting the duration of and losses from catastrophes. Natural language processing (NLP), an emerging field in artificial intelligence, provides the ability to automatically read, understand, and derive meaning from text data. A typical NLP process requires data scraping, sentimental analysis, text embedding, feature extraction, and topic modeling. While text-based analytics are a relatively recent development in insurance, they have proved to be useful in the social science, economics, and finance fields, as evidenced by the work of Cong, Liang, and Zhang (2019) and Hanley and Hoberg (2019), along with comprehensive survey papers by Gentzkow, Kelly, and Taddy (2019) and Grimmer and Stewart (2013).

Our research proposes a natural language learning-based approach to catastrophe duration and loss prediction, and is among the first of its kind in actuarial research. Our primary objective was to gain new insights on how information from online news articles can be used to help achieve a timely estimation of catastrophe duration and insurable economic loss. Applying the proposed approach, we investigated the following questions:

- Can online news provide an early warning of extreme catastrophe events?
- What words (or topics) are the most significant indicators of catastrophe severity?

Answers to these questions are of particular value to insurance and reinsurance companies with large exposure in catastrophe risk. The paper is organized as follows: Section 2 introduces the proposed NLP-based approach. Section 3 describes the catastrophe data used in this research. Sec-

a Han Wang has a master's degree from the School of Statistics and Mathematics, Central University of Finance and Economics, China.

b Wen Wang has a master's degree from the School of Statistics and Mathematics, Central University of Finance and Economics, China.

c Feng Li is an Associate Professor in Guanghua School of Management, Peking University, China.

d Yanfei Kang is an Associate Professor in School of Economics and Management, Beihang University, China.

e Han Li is an Associate Professor in Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia.

tion 4 presents empirical results based on wildfire and hurricane case studies. Section 5 concludes the paper.

The news text information used in our model is not only readily obtainable but also allows swift identification of disaster risks and early-stage loss predictions during catastrophic events. The results indicate that machine learning tools exhibit high efficiency and accuracy in predicting disaster losses, which will facilitate early warnings and timely decision-making. These findings hold substantial implications for the insurance industry.

2. METHODOLOGY

2.1. A QUICK INTRODUCTION TO NLP

Natural language is a crucial component of human society and represents the most fundamental aspect of modern data sources. With advances in machine learning algorithms and computer languages, it is now possible to analyze large collections of unstructured machine-readable texts (corpora).

First, statistical features can be extracted using n-grams (Cavnar and Trenkle 1994) and skip-grams (Cheng, Greaves, and Warren 2006). These features are usually organized in a *term frequency, inverse document frequency* (TF-IDF) matrix. The information is then represented using frequency measures. Second, NLP can efficiently extract news semantics from a given corpus. Semantic features can be extracted via Word2Vec (Mikolov et al. 2013) and BERT (Devlin et al. 2018) methods, based on a neural network model that learns word associations from a large corpus of text. This process allows us to represent text information in terms of both statistical features and semantic features.

2.2. OVERVIEW OF THE APPROACH

Stage 1: Daily news source for catastrophe events. We identified and collected relevant news from news aggregator websites such as Google by searching with a combination of location and catastrophe keywords, restricting the search time to the calendar month in which the catastrophe occurred. News aggregator sites generally aggregate headlines from trusted news sources worldwide, group similar catastrophe events together, and display them according to certain patterns.

Stage 2: Full text scraping. For each news headline displayed in Google, we built a web crawler to extract its news contents using the Python programming language, allowing us to retrieve the news title, news source, published time, and full contents. The crawler was deployed onto a server so that it executed the scraping task periodically. This ensured that we obtained the news in a consecutive manner. News was cleaned so that only catastrophe event-related contents were retained. See Section 3 for more details about text data scraping.

Stage 3: NLP steps. We employed standard NLP algorithms to extract features from scraped news, then calculated and computed two types of numerical features, statistical and semantic. Statistical features, including word frequency and catastrophe loss information (such as event location, time of occurrence, type of loss, and amount of loss) were computed by the TF-IDF method. Semantic features were associated with a lexical item from the catastrophe news, and extracted via Word2Vec and BERT methods.

Stage 4: Catastrophe duration and loss prediction. To detect early warnings of severe catastrophe events, we employed anomaly detection tools for a given variable of interest from statistical features. We extracted disaster-related words from existing textual information to construct a disaster lexicon, and matched disaster duration and loss data with the content extracted from the textual information. By converting textual content into vectors, we could construct classification and regression models to predict future disaster severity.

Based on the computed news features and the corresponding catastrophe data, we employed the RandomForest (Breiman 2001), LightGBM[1] (Ke et al. 2017), and XGBoost (Chen and Guestrin 2016) algorithms to determine the relationship between news features and catastrophe duration and losses. Moreover, the relationship between the number of news items, number of news topics, and the size of loss could also be studied along with the news meta information. The proposed modeling framework is illustrated in Figure 1.

3. DATA

The project's empirical study analyzed U.S. wildfire and hurricane data from 2020, and the text data for analyzing catastrophe loss was crawled from Google. We converted the textual data into numerical variables to facilitate our modeling and prediction strategies. We first combined location and catastrophe keywords as search phrases for each catastrophe and restricted the search time to the calendar month in which the catastrophe occurred. We then used Python's scraping module `request` to construct a URL object that requested resources from the server and parsed the HTMLs using `BeautifulSoup`. The result was 62 pieces of data for the wildfires and 207 pieces of data for the hurricanes. The searched results were saved as machine-readable CSV files. News meta information, such as title, source, and time, were saved to identify the comprehensiveness of the news. For the data preprocessing step, we cleaned invalid information or abnormal text by assigning regular expressions and matched the crawled news content with the corresponding U.S. catastrophe duration/loss data.

1 Light GBM is a fast, distributed, high-performance gradient boosting framework based on a decision tree algorithm. It has been used for ranking, classification, and many other machine learning tasks. See details at https://www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/

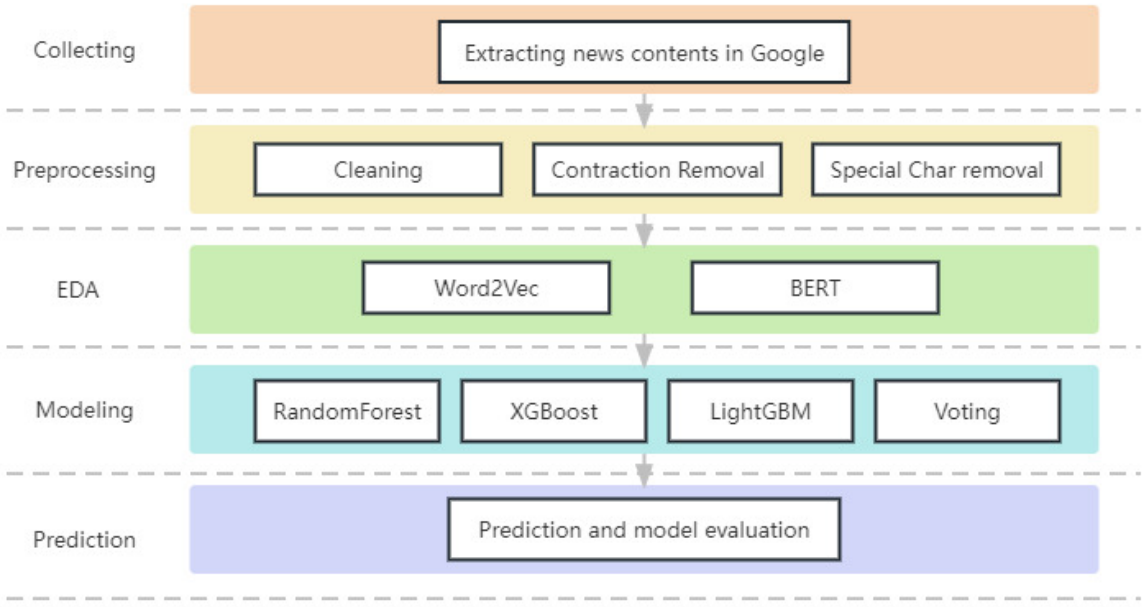

Figure 1. The framework for catastrophe duration and loss prediction.

Table 1. Catastrophe data description

Variable	Description
State Name County Name	Catastrophe location
Hazard	Catastrophe type
Year Month	Catastrophe time
CropDmg PropertyDmg	Catastrophe loss of crop and property
Duration_Days	Catastrophe duration
Text	News text information crawled from Google

Our primary data source for catastrophe duration and loss in the U.S. was the Spatial Hazard Events and Losses Database (SHELDUS). This database is maintained by Arizona State University and is commonly used by the actuarial community. It provides detailed catalogs of all major catastrophes in the U.S., such as droughts, earthquakes, floods, hurricanes, thunderstorms, tornadoes, and wildfires. Available information includes event location and duration, time of occurrence, type of loss, and amount of loss. We collected catastrophe duration and loss data at the county level for the same time period as the scraped news. See Table 1 for a detailed variable description.

4. EMPIRICAL STUDIES

We considered two types of natural catastrophes, wildfires and hurricanes, and predicted and evaluated the losses of each catastrophe-affected county using machine learning models such as Random Forest, XGBoost, and LightGBM. Each type of natural catastrophe has unique characteristics, influences, and indicators used to measure severity. For wildfires and hurricanes, we selected duration days and loss amounts, respectively, to measure event severity. We compared the performance of various machine learning models to identify the most suitable model for a specific type of natural catastrophe, and to develop a general model that can be applied across multiple types of catastrophes.

4.1. PREMODELING PREPARATION

For the wildfire and hurricane data, we used Word2Vec and BERT to construct the document vectors. We first cleaned the text using regular expressions, and then used Python's `NLTK` library for English word separation. We expanded the deactivated lexicon in the `NLTK` library to make the separation method more applicable to catastrophe data. The wildfire and hurricane datasets yielded 24,306 and 24,023 words, respectively. We selected 300 words as the keyword database, based on words that described the specifics and severity of the events and their frequency of occurrence, then we divided the words into categories according to their meaning. *Environment* describes the surrounding conditions, *Actions* describes the rescue activities, and *Wildfire* and *Hurricane* describe the event itself. The cloud maps for these words are shown in Figure 2 and Figure 3. Some of the keywords and their corresponding frequencies are shown in Table 2 and Table 3.

The word vector was obtained using a Word2Vec training model, where the frequency of each keyword was used as the weight. The weighted Word2Vec word vector was obtained by multiplying the word vector and the correspond-

Figure 2. Keywords cloud map for wildfire events.

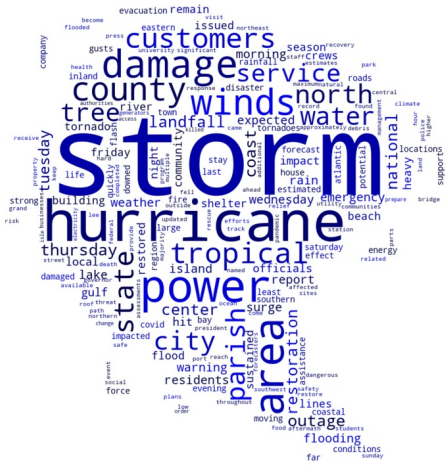

Figure 3. Keywords cloud map for hurricane events.

ing weight. The weighted word vector was accumulated to obtain the weighted text vectorized representation. Each text is represented by a 128-dimensional vector, as expressed in the Equation 4.1:

$$y_i = \sum_{j=1}^{j=300} w_{ij} x_j, \tag{4.1}$$

where y_i denotes the vectorized representation of each document, for $i = 1, 2, \ldots, 62$; x_j denotes the vectorized representation of each word for $j = 1, 2, \ldots, 300$; and w_{ij} denotes the frequency of each word occurring in each document.

The word vector trained by the Word2Vec model is static, which means that the same word represents the same expression in different contexts. A BERT model, which can reflect the expression of sentence semantics in different contexts, offers a better solution. We used the pretrained *BERT-base, multilingual cased* model published by Google,

Table 2. Category and word frequency of keywords for wildfire events

Environment		Actions		Wildfire	
1-2 Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
air	1737	evacuation	2545	fire	15018
forest	1590	reported	835	burned	1981
creek	1237	orders	808	wildfire	1896
winds	1101	help	798	burning	1638
valley	1081	service	666	smoke	1637
lake	1071	containment	650	emergency	843
weather	964	evacuate	624	flames	758
climate	631	closed	523	blaze	744
water	605	issued	461	heat	572
canyon	559	control	381	spread	567

Table 3. Category and word frequency of keywords for hurricane events

Environment		Actions		Hurricane	
1-2 Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
power	3620	restoration	1354	storm	6708
damage	2618	service	1309	hurricane	6285
water	1375	outages	1274	winds	2303
landfall	1371	emergency	1150	lake	1208
coast	1480	report	844	flooding	1180
island	989	warning	833	surge	1083
trees	987	moving	553	storms	1056
heavy	668	stay	541	rain	1025
beach	642	forecast	526	gulf	893
disaster	464	shelter	437	tornado	736

which has 12 coding layers and can output 768-dimensional word vectors. For the hurricane data, we only trained the model using BERT.

4.2. CATASTROPHE INDEX AND ACCURACY MEASUREMENT

Wildfires can cause significant damage, and their severity is often determined by the duration of their combustion. Therefore, we used *Duration Days* as the severity index for wildfires. Since *Duration Days* is a continuous positive variable, we employed a regression model. The duration of a hurricane event may not effectively reflect its severity, so we used the total amount of crop and property loss as the hurricane catastrophe index. We added *Crop Damage* and *Property Damage* to get *Total Damage*, the loss index for the hurricane data. However, there were large differences in total damage for different catastrophe events, such as 1 million dollars, 10 million dollars, or even 1 billion dollars of losses, spanning several orders of magnitude. For this analysis, we set two threshold values arbitrarily and divided *Total Damage* into three grades: less serious, serious, and very serious. We then used this as a new label loss index and converted the hurricane loss forecast into a classification model.

For evaluation indicators, we selected root mean square error (RMSE) and mean absolute percentage error (MAPE). It should be noted that RMSE measures the deviation between the predicted and true values and is more sensitive to outliers in the data. It is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)}, \tag{4.2}$$

where y_i is the true value, $\hat{y}_i$ is the forecasting value and n is the number of samples. Unlike RMSE, which has the same unit as the forecast variable, MAPE is *unit less* and can be used to compare different datasets using the following formula:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{4.3}$$

In addition, we used a triple classified confusion matrix, as shown in Table 4, where R0, R1, R2 represent the actual labels, 0, 1, 2, respectively, and P0, P1, P2 represent the predicted labels 0, 1, 2, respectively. The confusion matrix interpretation is as follows: N denotes the total number of samples in the test set, $T0$, $T1$, and $T2$ denote the num-

Table 4. Confusion matrix with three classifications

Actual	Predicted			Total
	P0	P1	P2	
R0	$T00$	$E01$	$E02$	$T0$
R1	$E10$	$T11$	$E12$	$T1$
R2	$E20$	$E21$	$T22$	$T2$
Total	$E0$	$E1$	$E2$	N

ber of samples for each of the true labels, and $E0$, $E1$, and $E2$ denote the number of samples for each of the predicted labels, respectively. Among them, the number of samples that will be correctly predicted to class i is Tii, and the number of samples of class i incorrectly predicted to class j is Eij. We selected Accuracy, Precision, Recall, and F1 score as evaluation indicators to compare the prediction effects of each model (see Equation 4.4–4.7).

$$Accuracy = \frac{T00 + T11 + T22}{N} \tag{4.4}$$

$$Precision = \frac{Tii}{Ei}(i = 0, 1, 2) \tag{4.5}$$

$$Recall = \frac{Tii}{Ti}(i = 0, 1, 2) \tag{4.6}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4.7}$$

It should be clarified that the main objective of the research was to demonstrate the potential of NLP as a more timely method than classic loss prediction models for catastrophe loss prediction, but it may not necessarily have higher predictive accuracy. We acknowledge that if we had perfect information on geographic data, building characteristics, climate data, socioeconomic data, and policyholder data, classic loss prediction models might work better than the NLP methods we proposed. However, perfect data are unlikely to be the case in practice. This highlights another advantage of the NLP method.

4.3. WILDFIRE DURATION PREDICTION

In Section 2, we described two ways to build variable sets: text processing methods based on Word2Vec and BERT. We compared the loss prediction effects of each model under two variable sets and used *optuna*, to optimize the super parameters of each model. In addition, we used the voting model to combine the prediction results of different models, where the weight of each model was the reciprocal of the RMSE of its prediction results. The voting model, using vectors embedded by BERT, only combines the prediction results of XGBoost and LightGBM. The model parameters used are shown in Table 5 and the prediction results are shown in Table 6, where numbers in boldface indicate the model with the best forecasting performance. We also show the prediction results from the voting model in Figure 4, where the x-axis represents the index of testing samples.

Overall, the prediction results of most models were good and fairly consistent, except for the Random Forest model, which performed poorly on the BERT variable set. The voting model demonstrated good prediction results for both variable sets. This suggests that XGBoost and LightGBM may have certain advantages for predicting wildfire duration.

In addition, the prediction effect of each model using BERT was generally better than that of Word2Vec. This shows that the method of manually selecting keywords and constructing variable sets based on Word2Vec is reasonable, and it achieves results similar to the existing mature BERT model. However, the BERT method extracts and contains more effective information than the Word2Vec method, which improves the subsequent model prediction effect.

4.4. HURRICANE LOSS PREDICTION

As mentioned in Section 4.2, according to the features of the variable *Total Damage*, we assigned two threshold values, 10^5 and 10^7, divided *Total Damage* into three levels, less serious, serious, and very serious, as a new label loss index. See Table 7 for more details, where *Dmg* represents *Total Damage*. We used a Box-Cox transformation in Figure 5 to unify the order of magnitude and improve the normality of *Total Damage*.

We constructed training and test sets through stratified sampling from multiple samples, on which we fit three classification models using the same method. We also used a voting model to combine the prediction results, where the weight of each model was determined by the accuracy of its prediction, as shown in Table 8.

According to the results shown in Table 9 and 10, the overall accuracy of the prediction results of each model was around 70%, which indicates that the text information reflects a basic and accurate understanding of the level of hurricane loss. For each level, taking the voting model as an example, the precision, recall, and F1 score indicators on various samples are also around 70%, showing a good classification effect. Although the model may occasionally misclassify some serious disasters as less serious, it still shows good recognition accuracy at the very serious level. This increases our confidence in using text information to determine whether a certain hurricane is very serious and enables us to provide relevant early warning and decision support for such catastrophic events in a timely manner.

Finally, as with the Section 4.3 wildfire data, we also attempted to directly perform regression prediction on hurricane catastrophe loss after a Box-Cox transformation. However, because of significant differences in losses caused by hurricane catastrophes of different severity levels, in addition to regressing the overall data, we also conducted re-

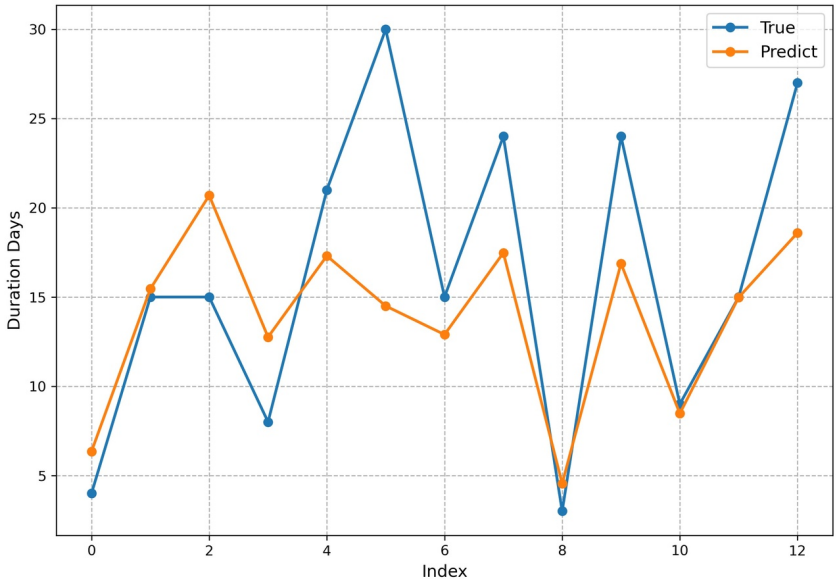

Figure 4. Voting model prediction effect for BERT.

Table 5. Model parameters for wildfire duration

Input	Output	Model	Params
Vector embedded by Word2Vec	Predicted *Duration Days*	RandomForest	'max_depth': 7, 'max_features': 25, 'ccp_alpha': 0.0421
		XGBoost	'max_depth': 3, 'learning_rate': 0.0984, 'reg_alpha': 0.6909, 'reg_lambda': 0.4471
		LightGBM	'max_depth': 4, 'learning_rate': 0.9990, 'reg_alpha': 0.0742, 'reg_lambda': 0.5885
		Voting	weight = [0.3378,0.3507,0.3115]
Vector embedded by BERT	Predicted *Duration Days*	RandomForest	'max_depth': 7, 'max_features': 15, 'ccp_alpha': 0.0393
		XGBoost	'max_depth': 3, 'learning_rate': 0.4362, 'reg_alpha': 0.7601, 'reg_lambda': 0.8517
		LightGBM	'max_depth': 3, 'learning_rate': 0.4170, 'reg_alpha': 0.0448, 'reg_lambda': 0.3310
		Voting	weight = [0.5300,0.4700]

gression on three severity levels, which we first classified and then regressed. Results are shown in Figure 6, Figure 7, and Table 11.

5. CONCLUSION

Accurately predicting catastrophe severity has been a challenge in the insurance industry. This research provides a unique and meaningful exploration of this issue based on natural language processing methods and machine learning

Table 6. Prediction accuracy for wildfire based on Word2Vec and BERT

Embedding	Word2Vec		BERT	
Model	RMSE	MAPE	RMSE	MAPE
RandomForest	7.632	0.6552	8.3679	0.6694
XGBoost	7.35	0.4119	6.0531	0.3966
LightGBM	8.275	0.5694	6.8242	0.3577
Voting	**7.3039**	**0.5117**	**6.1121**	**0.2984**

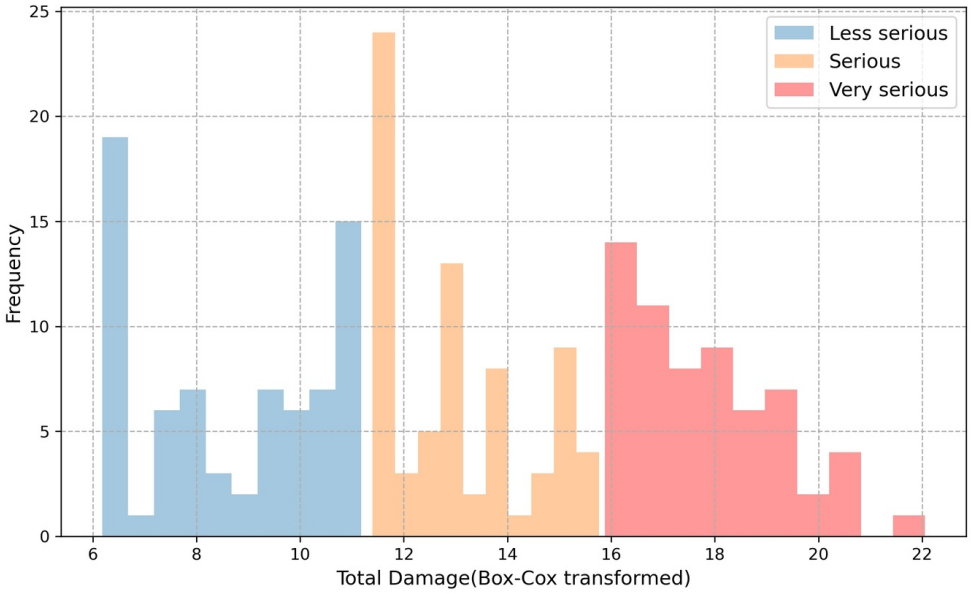

Figure 5. Label loss index for hurricane events.

Table 7. Label loss index for hurricane events

Label	Level	Range	Sample size
0	Less serious	$Dmg < 1 \times 10^5$	73
1	Serious	$1 \times 10^5 \leqslant Dmg < 1 \times 10^7$	72
2	Very serious	$Dmg \geqslant 1 \times 10^7$	62

models. Our results show that using news text information effectively improves the accuracy of predicting catastrophe severity. Both the Word2Vec and BERT methods effectively extract information from news text, while machine learning models, such as Random Forest, XGBoost, and LightGBM, accurately predict catastrophe duration and loss. The BERT method outperformed Word2Vec in text information extraction. Both XGBoost and LightGBM models performed well for predicting the severity of catastrophes when using BERT-generated text vectors. The voting model provides the best prediction performance on wildfire duration and showed nearly 70% accuracy for hurricane losses.

As a fast and open data source, news text information provides a large amount of timely and effective information in the early stages of a catastrophe, which is of great significance for the insurance industry. First, NLP models excel at real-time event detection by scanning news articles and social media posts for mentions of catastrophic events. This early awareness allows insurers to swiftly assess potential risks and inform the relevant departments. Integrating NLP with predictive modeling would potentially improve accuracy by incorporating textual data, such as news reports and expert analysis, into their risk assessment frameworks. Furthermore, during catastrophic events, NLP aids in disaster response and management by analyzing real-time reports and sensor data, providing insurers with the most up-to-date information to deploy resources effectively and adapt their risk assessment strategies in real time. Therefore, us-

Table 8. Model information for hurricane losses

Input	Output	Model	Params
Vector embedded by Word2Vec	Predicted *Label*	RandomForest	'max_depth': 5, 'max_features': 25, 'ccp_alpha': 0.0132
		XGBoost	'max_depth': 6, 'learning_rate': 0.4312, 'reg_alpha': 0.4747, 'reg_lambda': 0.4007
		LightGBM	'max_depth': 8, 'learning_rate': 0.1320, 'reg_alpha': 0.5265, 'reg_lambda': 0.6743
		Voting	weight = [0.6279, 0.6512, 0.6512]
Vector embedded by BERT	Predicted *Label*	RandomForest	'max_depth': 7, 'max_features': 27, 'ccp_alpha': 0.0163
		XGBoost	'max_depth': 3, 'learning_rate': 0.6081, 'reg_alpha': 0.6474, 'reg_lambda': 0.5731
		LightGBM	'max_depth': 5, 'learning_rate': 0.8688, 'reg_alpha': 0.0388, 'reg_lambda': 0.7788
		Voting	weight = [0.3496,0.3252,0.3252]

ing NLP in predictive modeling and disaster response allows insurers to be better equipped to predict, manage, and respond to catastrophic losses.

Finally, we acknowledge the limitation of NLP methods in terms of interpretability compared with classical models. Improving the interpretability of NLP methods is an important area for future research. Future research could also explore other catastrophe types beyond wildfires and hurricanes. In addition, while our research provides valuable insights into catastrophe loss modeling, it is important to acknowledge our dataset's limitations. The scope of our study is constrained by data availability, and our findings may not fully capture the broader trends and variations that could emerge over a more extended period.

The code for this article is based on Python. It is available at: https://github.com/feng-li/catastrophe-loss-prediction-with-NLP.

ACKNOWLEDGMENT

This research is sponsored by the Casualty Actuarial Society through the 2022 Individual Grant Competition project titled "Catastrophe loss prediction via natural language processing."

Submitted: March 29, 2023 EDT. Accepted: May 22, 2024 EDT.
Published: April 02, 2025 EDT.

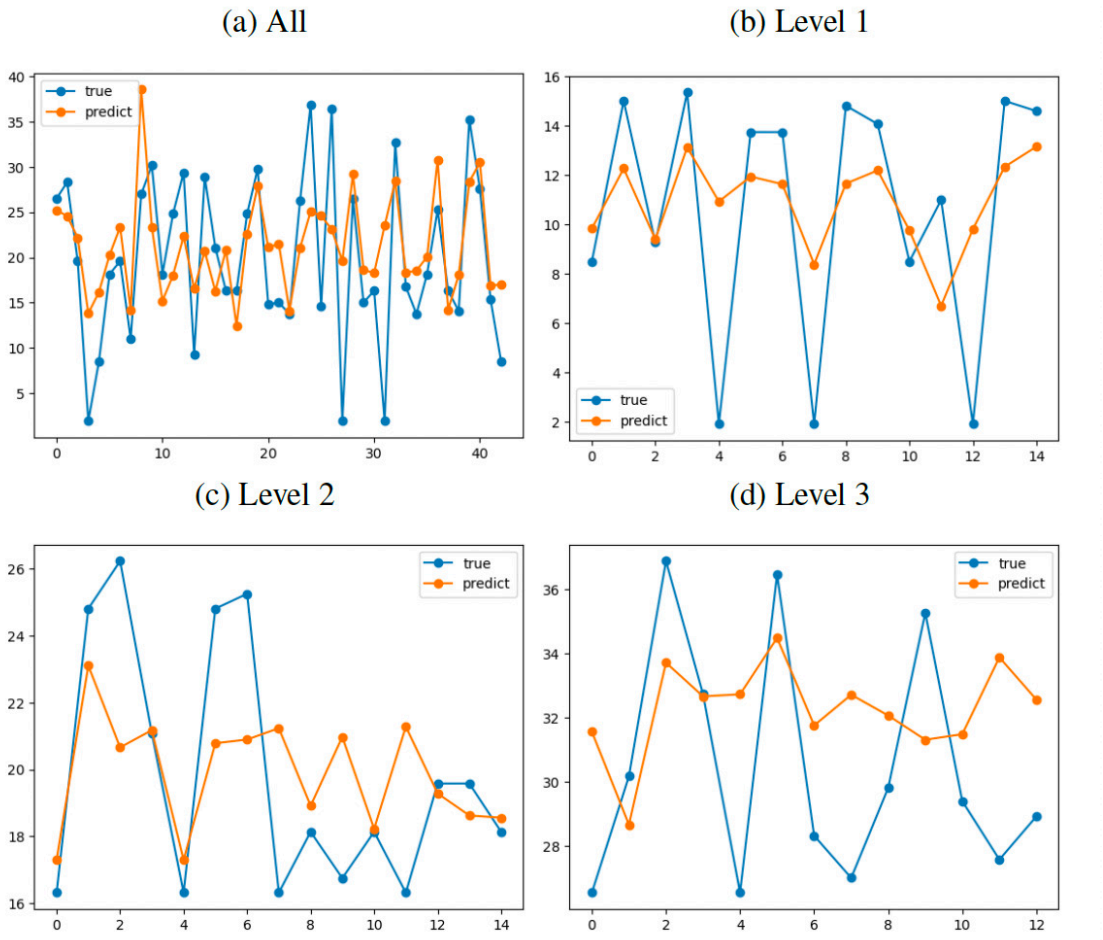

Figure 6. Voting regression model prediction effect for Word2Vec.

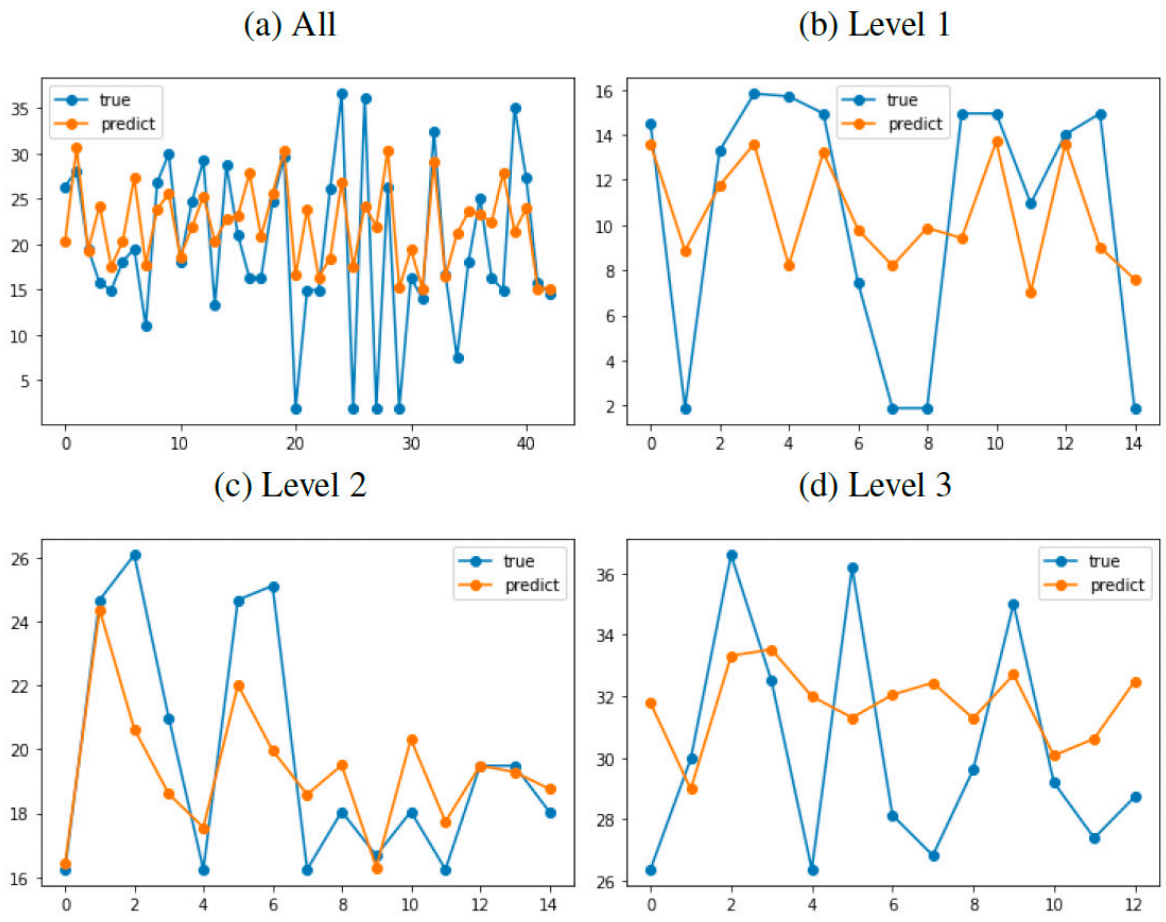

Figure 7. Voting regression model prediction effect for BERT.

Table 9. Prediction accuracy for hurricane based on Word2Vec

Model	Actual	Predicted			Precision	Recall	F1-score	Accuracy
		P0	P1	P2				
RandomForest	R0	9	3	3	0.64	0.6	0.62	0.6279
	R1	3	10	2	0.62	0.67	0.65	
	R2	2	3	8	0.62	0.62	0.62	
XGBoost	R0	8	3	4	0.57	0.53	0.55	0.6511
	R1	3	10	2	0.77	0.67	0.71	
	R2	3	0	10	0.62	0.77	0.69	
LightGBM	R0	8	4	3	0.62	0.53	0.57	0.6635
	R1	3	11	1	0.69	0.73	0.71	
	R2	2	1	10	0.71	0.77	0.74	
Voting	R0	8	4	3	0.62	0.53	0.57	0.6744
	R1	3	11	1	0.69	0.73	0.71	
	R2	2	1	10	0.71	0.77	0.74	

Table 10. Prediction accuracy for hurricane based on BERT

Model	Actual	Predicted			Precision	Recall	F1-score	Accuracy
		P0	P1	P2				
RandomForest	R0	12	1	2	0.71	0.80	0.75	0.6509
	R1	4	9	2	0.75	0.60	0.67	
	R2	1	2	10	0.71	0.77	0.74	
XGBoost	R0	10	3	2	0.62	0.67	0.65	0.6744
	R1	4	10	1	0.67	0.67	0.67	
	R2	2	2	9	0.75	0.69	0.72	
LightGBM	R0	10	5	0	0.71	0.67	0.69	0.6744
	R1	4	9	2	0.53	0.60	0.56	
	R2	0	3	10	0.83	0.77	0.80	
Voting	R0	11	2	2	0.65	0.73	0.69	0.6977
	R1	5	9	1	0.69	0.60	0.64	
	R2	1	2	10	0.77	0.77	0.77	

Table 11. Regression prediction results for hurricane based on Word2Vec and BERT

Embedding		Word2Vec		BERT	
label	Model	RMSE	MAPE	RMSE	MAPE
ALL	RandomForest	7.9349	0.9756	8.2727	1.0511
	XGBoost	7.5926	0.7171	7.8822	1.0513
	LightGBM	8.0563	0.9905	8.2993	1.0175
	Voting	7.2943	0.8818	7.7824	1.0343
Label 1	RandomForest	4.6399	1.0571	5.0508	1.2273
	XGBoost	4.1406	0.9686	4.9730	1.0423
	LightGBM	4.4032	0.9581	5.2482	1.1666
	Voting	4.0908	0.9575	4.7620	1.1180
Label 2	RandomForest	3.3100	0.1409	3.1321	0.1249
	XGBoost	3.0795	0.1132	2.4643	0.0790
	LightGBM	3.6140	0.1534	2.4870	0.0922
	Voting	3.0509	0.1156	2.3955	0.0834
Label 3	RandomForest	4.0585	0.1150	4.5422	0.1302
	XGBoost	4.3053	0.1264	3.7278	0.1125
	LightGBM	4.0132	0.1252	4.0362	0.1268
	Voting	3.9546	0.1204	3.7046	0.1123

REFERENCES

Breiman, L. 2001. "Random Forests." *Machine Learning* 45:5–32.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–94. KDD '16. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785.

Cheng, Winnie, Chris Greaves, and Martin Warren. 2006. "From N-Gram to Skipgram to Concgram." *International Journal of Corpus Linguistics* 11 (4): 411–33.

Cong, Lin William, Tengyuan Liang, and Xiao Zhang. 2019. "Textual Factors: A Scalable, Interpretable, and Data-Driven Approach to Analyzing Unstructured Information." *Interpretable, and Data-Driven Approach to Analyzing Unstructured Information (September 1, 2019)*.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *arXiv Preprint arXiv:1810.04805*.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57 (3): 535–74.

Grimmer, Justin, and Brandon M Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267–97.

Grossi, Patricia. 2005. *Catastrophe Modeling: A New Approach to Managing Risk*. Vol. 25. Springer Science & Business Media.

Hanley, Kathleen Weiss, and Gerard Hoberg. 2019. "Dynamic Interpretation of Emerging Risks in the Financial Sector." *The Review of Financial Studies* 32 (12): 4543–4603.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems*, 3146–54.

Kozlowski, Ronald T, and Stuart B Mathewson. 1995. "Measuring and Managing Catastrophe Risk."

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv Preprint arXiv:1301.3781*.

Panjer, Harry H, and Gordon E Willmot. 1992. *Insurance Risk Models*. Society of Acturaries.

Reed, Kevin A, AM Stansfield, MF Wehner, and CM Zarzycki. 2020. "Forecasted Attribution of the Human Influence on Hurricane Florence." *Science Advances* 6 (1): eaaw9253.

Reichstein, Markus, Michael Bahn, Philippe Ciais, Dorothea Frank, Miguel D Mahecha, Sonia I Seneviratne, Jakob Zscheischler, et al. 2013. "Climate Extremes and the Carbon Cycle." *Nature* 500 (7462): 287.

Swiss Re Institute. 2021. *Sigma 1/2021: Natural Catastrophes in 2020: Secondary Perils in the Spotlight, but Don't Forget Primary-Peril Risks*.

———. 2023. *Sigma 1/2023: Natural Catastrophes and Inflation in 2022: A Perfect Storm*.

Vijayan, Linoj, Wenrui Huang, Kai Yin, Eren Ozguven, Simone Burns, and Mahyar Ghorbanzadeh. 2021. "Evaluation of Parametric Wind Models for More Accurate Modeling of Storm Surge: A Case Study of Hurricane Michael." *Natural Hazards* 106:2003–24.

Zanardo, Stefano, and Jose Luis Salinas. 2022. "An Introduction to Flood Modeling for Catastrophe Risk Management." *Wiley Interdisciplinary Reviews: Water* 9 (1): e1568.