NATURAL LANGUAGE PROCESSING ENABLED NATURAL CALAMITY PROACTIVE CRISIS MANAGEMENT

Michael Angelo T. Galeste
Wilhelm Henri R. Alegrado

Operations Excellence
Western Digital, 109 Technology Ave., SEPZ, Laguna Technopark, Biñan, Laguna, Philippines 4024
michael.angelo.galeste@wdc.com, wilhelmhenri.alegrado@wdc.com

ABSTRACT

Calamities and untimely Natural Accidents happen in any country and company, and proper knowledge of these occurrences can help companies to ensure the continuity of its operation, safety of employees and its resiliency.

With such situation and environmental conditions, the company needs real-time information on natural calamities from government websites, to provide to the Emergency Response Team (ERT) and Crisis Management Team (CMT) to perform quick decisions minimizing the impacts of the natural calamities.

In this study, the goal is to increase the proactiveness of the company in responding to such Natural Calamities in alignment with its Business Continuity Plan through the means of 4th Industrial Revolution (4IR) technologies, in particular, Cognitive Computing. Using methods of Lemmatization, Tokenization and Normalization to feed the data through Unsupervised and Supervised Learning to categorize each News Article and feeding those to a Sentiment Analysis algorithm, identifying if it is a positive or negative impact to the company, targeting equal to or more than 70% accuracy. And with success of this paper, apply it to the other sites.

1. 0 INTRODUCTION

Philippines is situated inside the Pacific Ring of Fire, which has islands with volcanoes and active moving tectonic plates, posing dangers and untimely calamities to occur within the country. Being known as one of the most hazard prone country companies must devise a Business Continuity Plan that ensures the continuity of its business but more importantly the safety of its employees.

Quick decision making, minimizing risks, proper communication and business continuity are the most important aspect that are addressed by developing a solution that integrates real-time information on natural calamities from government websites through searching the latest RSS or Rich Site Summary news feeds.

In addition to information on the natural calamity, information about the plant's utilities, environmental conditions and other services are also required by the site's Emergency Response Team (ERT) and Crisis Management Team (CMT), giving them the informational edge for the immediate deployment of critical decisions and actions, minimizing the impacts of natural calamities.

There are plenty of websites that cater to News, Weather and Current Events that can support and provide the needed information for this quick decision making but with the altering environment situations, going through those websites manually takes time and eliminates or rather lessens the effectivity of the quick decision making, minimizing risk and employee safety, thus the need for such a solution to be modern and quick. Even with these available multiple sources of information, it also takes time to sift through what is related, relevant and coherent to the current situation, and often if not most of the time will cause confusion into making an ill-informed decision.

Doing this manually can and if not will lessen the effectiveness of the Business Continuity Plan set in place, exposing more risks to the company and overall, its employees.

Because of this manual process it takes longer to disseminate the information and come to a decision, instead of being proactive to such calamities the company retains being reactive once a calamity enters the region having less time to prepare and initiate the Business Continuity Plan.

But with the current advancement of technologies leading to the 4th industrial revolution, solutions to help alleviate such troublesome conditions can make the decision making, risk minimization, business continuity and employee safety more advanced and quicker.

1.1 The Fourth Industrial Revolution

The 4th Industrial Revolution or most known as 4IR conceptualizes rapid change to technology, industries, and

societal patterns and processes in the 21st century due to increasing interconnectivity and smart automation, this encompasses the following benefits:

- Autonomous actions
- Self-optimizing
- Agility & flexibility
- Prescriptive solutions
- End-2-End connectivity
- Optimized resources
- Analytics

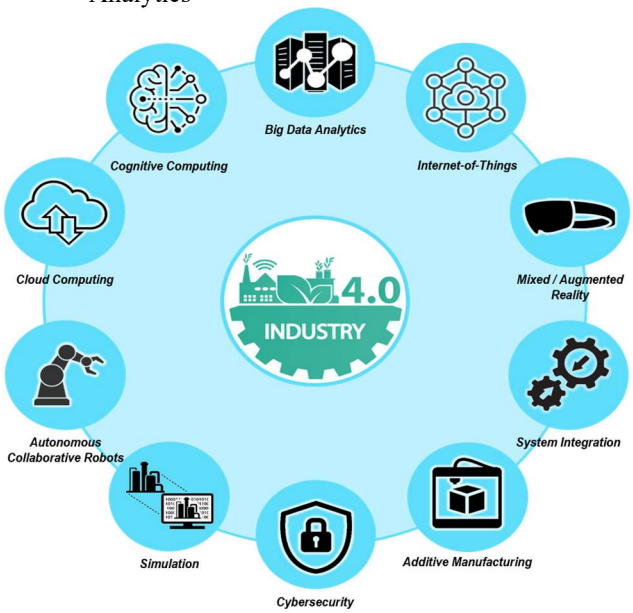

Fig. 1. List of 4IR technologies made known and readily accessible.

With the above ready solutions, the project utilized Industrial Internet-of-Things or most known as IIOT and Cognitive Computing. The combination of these 4IR technologies fully automates the information gathering process of reading through each news article and categorizing them down to providing the information to the ERT and CMT team, efficiently boosting the decision making of the team.

1.1.1 Industrial Internet-of-Things

Industrial Internet-of-Things or IIOT is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction made to cater, or tailor fit even to the company industrial needs.

Technologies under this 4IR branch utilize more on interactions be it human-to-human, human-to-computer or even computer-to-computer, and these interactions commonly transfer information requested across a network. This is one of the fundamental technologies to search through multiple RSS or Rich Site Summary news feeds and gather each of the news articles contents. The contents are then stored for later categorization and analysis of the second 4IR technology utilized.

1.1.2 Cognitive Computing

Cognitive computing is an attempt to have computers mimic the way the human brain works. To accomplish this, cognitive computing uses artificial intelligence or AI and other underlying technologies, including the following: Expert systems. Neural networks. Machine learning.

Cognitive computing is used to classify each news article whether it is about Calamity, Pandemic, or other type of news. After which it will then analyze and mimic a human's sentiment whether the news article is Positive, Negative or Neutral.

2. 0 REVIEW OF RELATED WORK

Not Applicable.

3.0 METHODOLOGY

With the multitude of News Articles to be categorized and analyzed, there will be paragraphs, sentences or even a single word that could change how the news article is viewed. With Cognitive Computing trying to mimic a Human's sentiment it is possible by first performing several processes to clean out the data.

3.1 Data Processing

Prior to any classification and identification of the News Article the text must first be processed, this will eliminate wrong classification and sentiment analysis. This is done by Text Normalization which includes Case Normalization, Punctuation Removal, Stop Word Removal, Tokenization and Lemmatization.

Case Normalization works by converting all the letters in each of the words to either all lower cases or all upper cases, this is used to eliminate the varying and mixed text cases in each of the text data being processed. Eliminating indication of proper nouns and emphasis that could sway the sentiment of the article.

Punctuation Removal is the process of removing special characters and punctuation marks from the text. This technique is useful when working with text data containing many punctuation marks, which can make the text harder to

process. It removes unnecessary characters, making the text cleaner and easier to process, it also reduces the dimensionality of the data, which can improve the performance of NLP algorithms.

Stop Word Removal is also commonly used to improve the performance of the NLP algorithm by removing stop words such as "the" and "a".

Tokenization is the process of breaking text into individual words or phrases, also known as "tokens". This technique is useful when working with text data that needs to be analyzed at the word or phrase level, such as in text classification or language translation tasks. It allows for analyzing and manipulating individual words or phrases in the text data, it also can improve the performance of NLP algorithms that rely on word or phrase-level analysis.

Lastly, Lemmatization is reducing words to their base form by considering the context in which they are used, such as "running" becoming "run". This technique is like stemming, but it is more accurate as it considers the context of the word. As it reduces the dimensionality of the data, which can improve the performance of NLP algorithms, also making it easier to identify the core meaning of a word while preserving context.

3.2 Unsupervised Learning

Once the text of the News article has been processed, normalized and cleaned each News article undergoes an unsupervised learning technique called Latent Dirichlet Allocation or LDA which is used to cluster or group each article into natural groups when we are not sure of what we are looking for or how we want to group the News Articles. The same News article the goes through Bidirectional Encoder Representations from Transformers to find similar texts that share the same Semantic or rather text that share the same meaning or context.

We then concatenate or combine those text results from the above classification in each of the News Article using Latent Space or Mathematically representing them where similar texts are items will be grouped using Autoencoder, a neural network consisting of an encoder that translates the input to an intermediate latent representation and a decoder that translates the latent representation into the output.

We then cluster the Autoencoder results via K-Means, a clustering method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster, into 7 groups:

- covid/pandemic,
- government/business/socioeconomics/environment
- floods/earthquake/volcano
- Inter-Agency Task Force (IATF)/lockdown/government/police/drugs
- hospitals/patients/energy/oil/travel,
- typhoon/rain/flooding
- campaign/elections/foreign relations.

Fig. 2. Customized sentiment analysis categorization using VADER.

After extensive clustering and identification, we then process those into Valence Aware Dictionary and Sentiment Reasoner or Vader, a sentiment analysis that not only tells if the statement is positive or negative along with the intensity of emotion. Figure 2 shows how we leveraged VADER to represent the weather intensity on a scale of 1 to -1 based on the article. A positive score indicates good weather while a negative score indicates the presence of a typhoon.

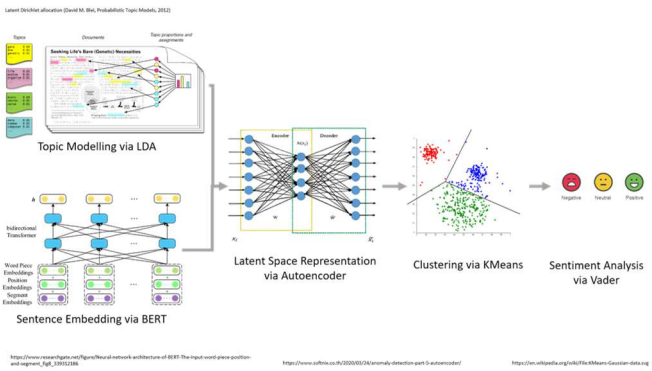

Fig. 3. Diagram version of the above processes explained in analyzing each news article.

Figure 3 shows the overall unsupervised learning architecture used for this project.

3.3 Supervised Learning

As a final step, a Recurrent Neural Network (RNN) is constructed and trained to determine whether the news article title is calamity-related or not.

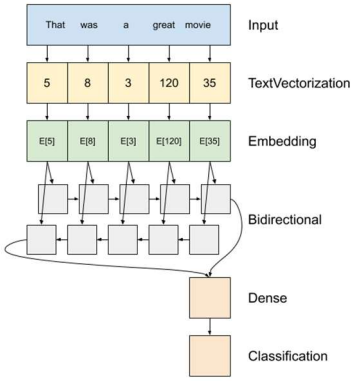

Fig. 4. Diagram of a Recurrent Neural Network (RNN) architecture in a Many-to-One configuration.

Recurrent Neural Networks (RNNs) are a family of artificial neural networks that are uniquely able to capture sequential dependencies in data, which sets it apart from standard feedforward networks and makes them ideal for processing and comprehending sequential information. RNNs are particularly good at evaluating the contextual links between words which helps them identify patterns and semantics that are essential for correctly classifying textual information.

We are using the Many-to-One configuration of the RNN shown in Figure 4 to take in a sequence of input or in this case, sentences within a given document and produces only one output which indicates whether the given document is calamity-related or not.

4.0 RESULTS AND DISCUSSION

Using the outlined pipeline, we were able to get around 70% to 84% accuracy in the classification of calamity-related articles.

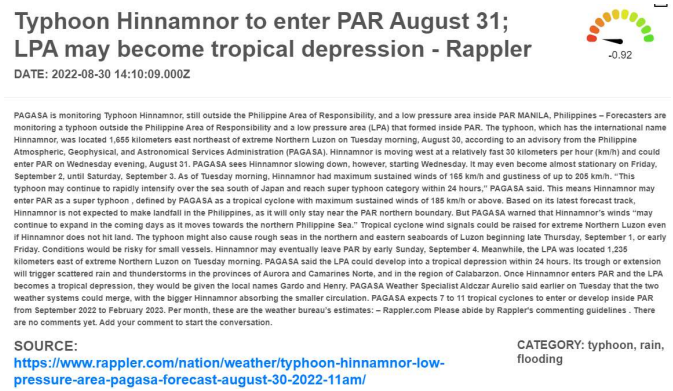

Fig. 5. Sample of a news Article classification using the NCPCM pipeline with the news article content shown at the bottom.

As shown in the sample NCPCM output in Figure 5, the article classification and VADER score shows that the news article content is pertaining to calamity-related news and the VADER score highlights the urgency of the content. Finally, the news article content is shows to supplement the NCPCM classification.

Table 1. Response time upon NCPCM Implementation

Calamity	Unit	Without NCPCM	With NCPCM
Typhoon	Days	2-3 Within Area of Responsibility	2-3 Outside Area of Responsibility
Volcano Eruption	Hours	20 - 24	5 - 12

With NCPCM, more preparation time is afforded to the Emergency Response Team (ERT) and Crisis Management Team (CMT). Table 1 shows the improvement of lead time reaction for the teams which allows them more effective actions being thought out in advance such as readiness in switching power sources from main power sources to alternative sources during calamity events and adjusting shuttle pickup times and routes to continue to bring employees onsite.

5.0 CONCLUSION

In conclusion, due to this advanced method of obtaining information of oncoming natural calamities, we can safely eliminate the need to manually look through multiple News websites. More importantly, this has also helped reduce the number of days or hours in lead time before the company has to make preparations, from knowing the information about the typhoon when it is within the Area of Responsibility to knowing the information about the typhoon before it even enters the Area of Responsibility.

6.0 RECOMMENDATIONS

With the success of this project other Western Digital manufacturing sites are considering learning from this project and applying it.

7.0 ACKNOWLEDGMENT

The team acknowledges the contributions of the members of Operations Excellence, specifically Dave Tablante, Brian Javelosa, Maria Venus Gambito and Janoah de los Santos as core members of the Scrum team. Edgard Reyes, Noel Divinagracia and Alexander Vega as the Product Owners and Stakeholders of the Project.

8.0 REFERENCES

1. https://spotintelligence.com/2023/01/25/text-normalization-techniques-nlp/#:~:text=Text%20normalization%20is%20a%20key,removal%2C%20stemming%2C%20and%20lemmatization.

2. https://medium.com/@shaikhrayyan123/a-comprehensive-guide-to-understanding-bert-from-beginners-to-advanced-2379699e2b51#:~:text=BERT%20(Bidirectional%20Encoder%20Representations%20from,context%20and%20nuances%20in%20language.

3. https://medium.com/@Mustafa77/semantic-textual-similarity-with-bert-e10355ed6afa#:~:text=Use%20BERT%20to%20measure%20the,textual%20similarity%20(STS)%20tasks.

4. https://en.wikipedia.org/wiki/File:KMeans-Gaussian-data.svg

5. https://www.softonix.co.th/2020/03/24/anomaly-detection-part-5-autoencoder/

6. https://www.reasearchgate.net/figure/Neural-network-architecture-of-BERT-The-input-word-piece-position-and-segment_fig8_339312186

7. https://en.wikipedia.org/wiki/Fourth_Industrial_Revolution#:~:text=%22Fourth%20Industrial%20Revolution%22%2C%20%22,advancement%20in%20the%2021st%20century.

8. https://www.rappler.com/nation/weather/typhoon-hinnamnor-low-pressure-area-pagasa-forecast-august-30-2022-11am

9.0 ABOUT THE AUTHORS

Michael Angelo T. Galeste is a Manager at Western Digital Storage Technologies for Operation Excellence, leading the Web Application Pillar within the Department, graduated Bachelor of Science in Software Engineering at Central Philippine University. He has been with the company for 4 years, bringing the past 14 years of experiences and knowledge from his previous companies under Web Applications and Manufacturing background.

Wilhelm Henri R. Alegrado is a Sr. Data Scientist at Western Digital Storage Technologies for Operation Excellence. He holds a master's degree in data science at Asian Institute of Management. He has been with the company for 12 years with experience from previous departments with an Engineering and Development background.

10.0 APPENDIX