sustainability

MDPI

Article

Analyzing a Decade of Wind Turbine Accident News with Topic Modeling

Gürdal Ertek * and Lakshmi Kailas

College of Business and Economics, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates; lakshmi.kailas@gmail.com
* Correspondence: gurdal@uaeu.ac.ae

Abstract: Despite the significance and growth of wind energy as a major source of renewable energy, research on the risks of wind turbines in the form of accidents and failures has attracted limited attention. Research that applies data analytics methodologically in this context is scarce. The research presented here, upon construction of a text corpus of 721 selected wind turbine accident and failure news reports, develops and applies a custom-developed data analytics framework that integrates tabular analysis, visualization, text mining, and machine learning. Topic modeling was applied for the first time to identify and classify recurring themes in wind turbine accident news, and association mining was applied to identify contextual terms associated with death and injury. The tabular and visual analyses relate accidents to location (offshore vs. onshore), wind turbine life cycle phases (transportation, construction, operation, and maintenance), and the incidence of death and injury. As one of the insights, more incidents were found to occur during operation and transportation. Through topic modeling, topics associated most with deaths and injuries were revealed. The results could benefit wind turbine manufacturers, service providers, energy companies, insurance companies, government bodies, non-profit organizations, researchers, and other stakeholders in the wind energy sector.

Keywords: wind energy; wind turbine accidents; accident analysis; topic modeling; association mining; data analytics

check for updates

Citation: Ertek, G.; Kailas, L. Analyzing a Decade of Wind Turbine Accident News with Topic Modeling. *Sustainability* **2021**, *13*, 12757. https://doi.org/10.3390/su132212757

Academic Editor: Tomonobu Senjyu

Received: 7 October 2021
Accepted: 12 November 2021
Published: 18 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The inevitable demand for energy across the globe is projected to expand over the years and decades ahead. The World Energy Outlook 2020 report by the International Energy Agency (IEA) predicts a growth of 4% to 9% in global energy demand from 2019 to 2030, despite the global COVID-19 pandemic that began in late 2019 [1]. Renewable energy sources, including wind, solar, biomass, geothermal, and hydropower, can help reduce the consumption of fossil fuels and the release of greenhouse gases [2]. According to the 2019 outlook by the IEA, at least 66% to 80% of the newly added global energy capacity is forecasted to come from renewables. Wind, along with solar power, is anticipated to constitute the majority of additional power generation by 2040 [3]. For example, in the USA, wind power has grown more than three times from 2009 (35,130 MW) to 2019 (105,591 MW) and was the largest contributor to renewable power in the USA as of 2019 [4].

Wind turbines are mechatronic equipment that generate electricity from wind by first transforming wind power into mechanical power and then to electric power [5] (Figure 1). Wind turbines can be constructed onshore (on land) or offshore (in the sea). The power output can range from a few hundred watts (W) in residential models to hundreds of kilowatts (kW) and even megawatts (MW) in commercial models [6]. The efficiency of a wind turbine is determined by how the individual parts are configured and designed, including the blades (which need to generate the greatest amount of torque), yaw control system (which helps in maintaining the orientation of the nacelle with respect to the wind

direction), rotors (which rotate the shafts), shafts (which propel the generator), gears (which control the shaft speeds), brakes (which stop or decrease the speed of the machines), generator (which generates electricity), and anemometer (which is used to measure wind speeds) [5,7].

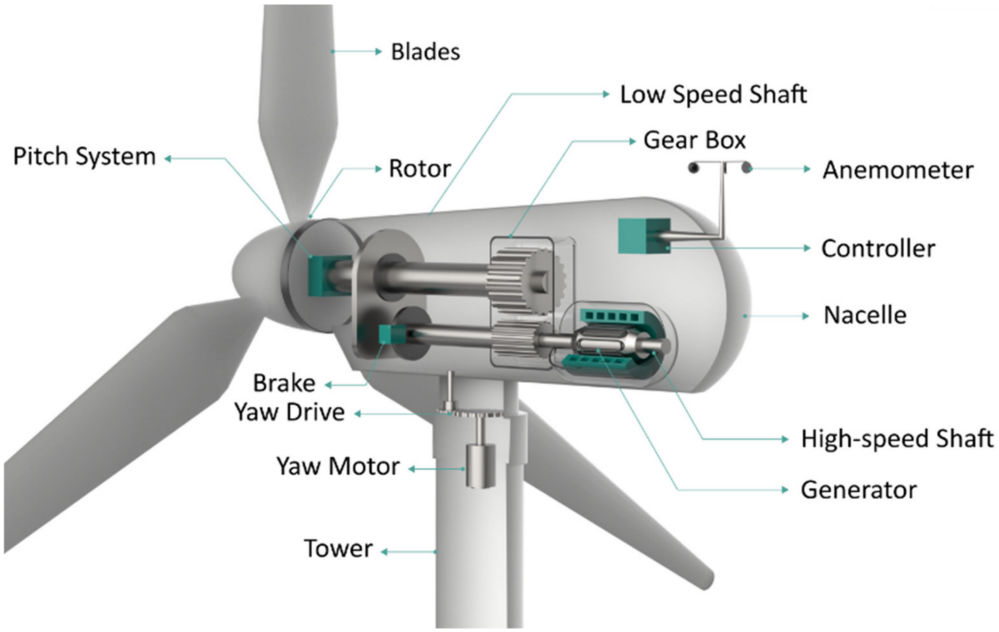

Figure 1. Components of a wind turbine. An animated video created by the research team is available online: https://bit.ly/3zphcUc/ (accessed on 7 October 2021).

While the growth in wind energy is praised owing to its renewable nature, as the literature review (Section 2) reveals, there is only limited research on one of the most important risks of wind turbines, namely, wind turbine accidents, failures, and breakdowns. Furthermore, earlier work is mostly based on summary statistics and cross-tabular analysis (Section 2.2), and only two research studies [5,8] were identified (Section 2.4), where advanced data analytics techniques were applied for systematic analysis of wind turbine accident news.

Here, by the terms "accident," "failure," or "breakdown," we refer to a wide spectrum of unforeseen and undesired events, which hereon will be referred to under the umbrella term "accidents". Wind turbine accidents may be caused by mechatronic failures, natural events, or human interventions. They may result in damage to wind turbines, wind farms, and associated properties, such as roads. Most importantly, wind turbine accidents may result in death or injury in humans. The underlying causes for the shortage of research on wind turbine accidents may be to maintain confidentiality, as well as the industry's motivation to protect its mostly positive public image [8].

Collections of accident news can be found in many online portals, especially at the Caithness Windfarm Information Forum (CWIF http://www.caithnesswindfarms.co.uk/, accessed on 7 October 2021), WindAction Group (http://www.windaction.org/, accessed on 7 October 2021), National Wind Watch (http://www.windaction.org, accessed on 7 October 2021), Friends Against Wind (http://en.friends-against-wind.org, accessed on 7 October 2021), and Gegenwind-Vogelsberg (http://www.gegenwind-vogelsberg.de/ueber-uns/, accessed on 7 October 2021). Other organizations such as "Rettet den Odenwald" ("Save the Odenwald" https://rettet-den-odenwald.de/, accessed on 7 October 2021), "Stop These Things" (https://stopthesethings.com/about/, accessed on 7 October 2021), "saveourskylineohio" (https://saveourskylineohio.com/, accessed on 7 October 2021), "WindWahn-Vogelseberg" (http://ww-vb.de/pages/home.php, accessed on 7 October 2021), "Save the Huron Mountains" (https://savethehuronmountains.org/, accessed on 7

October 2021), and "Landesverband Vernunftkraft-Hessen eV" (http://www.vernunftkraft-hessen.de/wordpress/ueber-uns/, accessed on 7 October 2021) have also published articles highlighting wind turbine failures. While several papers (as detailed in the next section) communicate important statistics, the systematic application of data mining and scientific methods for the analysis of accident news data can be found only in [5,8].

The lack of extensive research on the application of data analytics (data mining and data science) methods to accident news, despite the existence of thousands of accidents reported in the news and multiple summary statistics, highlights a significant gap in knowledge, particularly with respect to obtaining the most insights from accident data. This may partially be due to issues of reliability and verifiability of publicly available data on wind turbine accidents, in particular CWIF. As observed earlier in [5,8], within the CWIF database, a significant portion of the links refer to the WindAction Group and National Wind Watch, which typically contain the original content. Still, a notable portion of the remaining links are broken, raising concerns about the originality of the reported accident and possibly inhibiting the application of data analytics.

The motivation and objective of this study was to add to the stream of research [5,8] on the advanced analytical study of wind turbine accident news. The present research analyzes extensive recent and verified data and applies topic modeling and association analysis for the first time in the field. Thus, in this paper, we address the shortage of research regarding advanced studies of wind turbine accidents and develop a new understanding to be added to the body of knowledge on the topic. The research was aimed at providing practical and actionable insights to wind turbine manufacturers, service providers, energy companies, insurance companies, government bodies, non-profit organizations, researchers, and other stakeholders in the wind energy industry.

The unique contributions of the presented research are as follows:

First, the research study assembles the most extensive verifiable dataset in the world on wind turbine accidents, with full source text and evidence of originality. While extensive news collections are readily created by the organizations mentioned earlier, many of the links provided for the news are not active, making it impossible to verify the originality of the news. In the present research, both public data from these sources and Internet search results for "wind turbine accidents" and "wind turbine failures" were combined. Every piece of news regarding an accident was verified by accessing them directly from their sources. Eventually, only the accidents whose texts were verifiably available with sufficient information to make deductions were included in the analysis. Those that could only be found in news aggregator websites such as WindAction Group and National Wind Watch were not included. To serve as an immutable verified proof of originality, screenshots of the news were archived. In effect, a database with 721 accident news records was assembled, including the text and images, links to the original source (Weblink), screenshots of the source, translations to English where applicable, and a database of manually created metadata. The fields included in the metadata database include year, month, day, location, country, language, IsThereInjury, IsThereDeath, IsThereInjuryOrDeath, OffshoreOnshore, PhaseOfLifeCycle, and ModeOfTransportation. The constructed dataset is the largest data collection and dataset for wind turbine accident news, with verified proof of originality (while there are other datasets with more accidents reported, such as the Caithness Windfarms (CWIF) dataset, many of the links are not active and proof of accidents do not exist). In addition to the text collection and a database of the new metadata, for the first time in the literature, a comprehensive list of other interesting cases and incidents is presented in the Supplement to the paper.

Second, based on this extensively verified dataset on wind turbine accidents from around the world, statistics are derived across countries, locations (onshore vs. offshore), and phase of the turbine's life cycle. Analysis was conducted considering these categories, as well as the text of the news, and how different attributes relate to the two main outcomes, namely, human deaths and injuries. The effects of different attributes on the occurrence of accidents were also analyzed.

Third, the text of accident news has been analyzed using a multitude of data analytics techniques, some of which are applied for the first time in the domain. While [5], as the closest work to this paper, also applies text analytics, the research presented in this paper applies topic modeling through LDA (Latent Dirichlet allocation) for the first time to analyze wind turbine accidents, yielding novel insights. Another unique application is association mining to find relevant terms in the text of the news and the contextual relationship between these terms, and death and injury.

The paper is organized as follows: In this section (Section 1), the research topic, motivation, and contributions are presented. Section 2 reviews important literature on wind turbine accidents. Section 3 describes the data mining methodologies applied. Section 4 details the analysis framework and its steps. Section 5 provides details on the collected data, including the data dimensions, descriptions of the attributes, and the steps in data collection and cleaning. Section 6 presents the analysis results and provides insights. Section 7 concludes the paper with a summary of the work conducted, discussions, and future research directions. A separate supplement document provides further details on data collection and preparation, the literature, the developed analytics framework, the contents of datasets, process workflows, and the analysis results.

2. Literature Review

2.1. Individual Wind Turbine Accidents and Failures

The presented research constructs and analyzes a dataset of hundreds of accidents worldwide. However, it is essential to understand how and why individual accidents occur. A multitude of research papers in the field analyze and present individual incidents in detail, as well as discussing verified or possible causes. Some selected studies are as follows: the works in [9,10] study blade failure accidents in Taiwan and Japan, respectively; in [11] examined foundation failure in Japan; in [12] considered gear and motor failures that occurred in a wind farm in Japan; in [13] studied accidents at a farm due to lightning; in [14] investigated a nacelle failure at a wind farm in Japan; in [15] analyzed a turbine tower collapse incident in Taiwan; in [16,17] studied failures due to power grid trip-off in China; and in [18] examined an accident in which a worker fell off a turbine. While each of these studies analyzed in detail an actual accident, the possible types of accidents are much more varied, and thus, the listed research is only a mere sample of the work on the topic.

2.2. General Overview of Accidents and Risk Analysis

Once an understanding is gained of how and why individual wind turbine accidents can or may occur, the next step would be to develop an overall understanding of patterns through the analysis of multiple accidents. This stream of research can be classified into three groups as described below.

A major and extensive stream of research in the field is the damage caused to birds by wind turbine accidents [19]. Any such incidents were not included in our research, as the main theme of our research is the impact of wind turbine accidents on humans.

2.2.1. Analyzing Risks without Analyzing Statistics on Multiple Accidents

These studies describe and analyze the risks faced in a wind farm that could lead to possible accidents, but did not analyze any accident data. For example, [20] considered and quantified the societal risk associated with energy systems, including wind turbines, using reliability and failure rate data, while [21,22] focused on occupational risks and safety hazards and methods to manage such risks.

2.2.2. Statistics on Accidents

These studies were based on the collected data and considered multiple types of accidents. Some of these studies have reported the statistical analysis of collected historical accident data. The studies in this group did not apply any data mining methods beyond

tabular or visual methods. Studies that include the systematic application of more advanced data mining methodologies are discussed in Section 2.4.

In the second group, [23,24] compared the risk of energy accidents (through frequency, scope, severity of accidents, and normalized risk) across a spectrum of modern low-carbon energy systems (hydroelectric, nuclear, wind, and others) and their consequences. Accordingly, the risk profiles of each system are presented. While hydroelectric energy results in the most severe accidents (fatalities), and nuclear energy is the most expensive system, wind energy contributes to the majority of accidents. In addition to the previous analyses in [23,24], [25] examined 4450 energy accidents from 1800 to 2018 across a variety of energy systems and estimates up to 1.72 million possible deaths in 2040, identifying the root causes and activities that resulted in the most deaths and financial damage.

While [23–25] analyze all sources of energy, the following studies in this group focused specifically on wind energy:

The work in [26] is a collection of presentations from a conference organized by the IEA on energy technology, research, development, and deployment (RD&D) of wind turbine systems. The presentations include a rich collection of information on the reliability and maintenance requirements of wind turbines, as well as statistics on wind turbine failures. Researchers call for a joint shared database of operations and maintenance activities, as well as cost-effective maintenance and failure assessment.

The work in [27] summarized several databases from Europe and the USA containing data on wind turbine reliability and failure mechanisms. The role of factors such as turbine components and types of turbines on parameters such as downtimes and failure rates are evaluated along with the difficulties and advantages of combining such data from across the world.

The work in [28] analyzed and discussed wind energy accident risks and their adverse effects. The authors quantified the societal risk for wind energy (based on accident category, frequency, and consequences) using accident data of 1093 cases from 1995 to October 2011, sourced from the CWIF.

The authors in [29] conducted a fault tree analysis (FTA) of wind turbine accidents through expert consultation. This study investigates the effects of failures on safety to the public and quantifies the frequencies of component failures and failure mechanisms of 209 accidents in Europe, selected from CWIF. The authors cite loss of blade or its parts as the most common type of failure and mention shortage of data and lack of expertise as significant challenges.

The work in [30] identified, categorized, and statistically summarized wind turbine accidents, categorized under natural disasters, structural, human, and system issues. The researchers analyzed data across different time periods, collected from three sources, identifying the underlying causes of different accident types: 10 years (2001–2010) of domestic data from a Korean ministry; 445 cases from 4 years (2004–2007) of data from the New Energy and Industrial Technology Development Organization (NEDO) database, and 1443 cases from 15 years (2000–2014) of data from CWIF.

The work in [31] analyzed 1614 wind turbine accident incidents from CWIF, until September 2014, together with failure data. Detailed statistics were computed based on the nature, frequency, causes, and outcomes of accidents. Failure rates and downtimes were analyzed to identify the significant influences of factors such as wind turbine design complexity and location. The major categories analyzed include blade failure, structural failure, and fire. These insights can be used to improve the safety features and procedures of wind turbines.

The work in [32] discusses the state of the global wind industry as of 2016, discussing its challenges and drawbacks, common causes of wind turbine failures, and future R&D orientation. CWIF was used as the source for accident data. The authors highlight the significance of system reliability and levelized cost of energy (LCOE).

The work in [33] compiled and presented statistical summaries (based on accident causes and consequences) of the data collected by CWIF on 2744 wind turbine accidents from the 1980s to September 2020.

2.2.3. Statistics on a Single Type of Accident or Single-Cause Category

These studies also analyze risks and conduct statistical analyses, but they consider only a single-type, or single-cause, category. For example, [34,35] studied historical wind tower structural failures and collapses (sourced from CWIF) to explore causes, collapse mechanisms, and techniques to mitigate such risks. The work in [36] focused on accidents related to fire based on four selected cases and examined the common causes and protection methods.

2.3. Mechatronics for Maintenance and Monitoring

An important and technical stream of research on wind turbine accidents is on the mechanical, electrical, and electronic aspects. Studies in this stream discuss risk management through fault diagnosis, maintenance, and monitoring. This research can be grouped into two groups: the first group on the overall stability and health of wind turbines, and the second group on the monitoring and maintenance of a specific component or mode of failure.

2.3.1. General Health of the Wind Turbine

These studies focus on the general health monitoring of wind turbines for the management of risks and improvement of maintenance procedures.

The work in [37] presents a review of the literature that focuses on the assessment of wind turbine performance and efficiency. In an integrative approach, the research compares the methods and metrics, such as the capacity factor, failure rate, and mean down time, used in different projects and initiatives.

The work in [38] comprehensively evaluated health monitoring systems for offshore wind turbines. These methods include SCADA and CMS, as well as monitoring systems for different components. Systems and methods for various aspects of monitoring, such as data, feature identification, and safety improvement, are discussed.

The work in [39] reviewed methods for improving the maintenance and conducting inspection of wind farms.

The work in [40] studied the management of maintenance operations for offshore wind turbines (OWTs) through the study of accidents and the role of human influences in those accidents. Resilience engineering (RE) concepts are integrated into a model to manage the risks involved while considering human and organizational aspects.

The work in [41] evaluated several current procedures used to detect and forecast faults, as well as methods used to maintain the reliable operation of wind turbines.

2.3.2. Monitoring Specific Modes of Failure

These studies focused on the monitoring of specific modes of failure: in [42] discussed blade failure monitoring; in [43] studied lightning protection mechanisms; in [44] analyzed wind turbine generator trip-off failures and methods to avoid them; in [45] analyzed the risks of ice and blade parts being thrown by wind turbines and their mitigation; in [46] studied crane safety and methodology to improve crane operations in wind farm construction; in [47] focused on the management of risks related to transportation and installation of offshore wind turbines; and in [48] analyzed the risks of maintenance ship collisions at offshore wind farms.

2.4. Application of Advanced Data Analytics Methods

The papers summarized in this section are a sample of the research that applies text mining, machine learning, or other data analytics methods for wind energy. Studies in this section can be analyzed in two groups: the first group is those for monitoring and

predicting component failures, and the second group consists of other studies. Our research also applied data analytics techniques to extract comprehensive insights into wind energy. Thus, it falls into this category and, specifically, within the second group.

2.4.1. Monitoring and Component Failure Prediction

These studies applied data analytics, especially machine learning techniques, to improve the condition monitoring of wind turbines and to predict faults in different components.

The work in [49] proposed a methodology to predict wind turbine defects at three different levels, including their occurrence, criticality, and type. The performance of different data mining methods was analyzed using real-world data from the field.

The work in [50] developed a methodology based on deep neural networks to predict lubricant pressure in gearboxes, which provides higher prediction accuracy compared to five other methods. Supported by a control chart of prediction errors, the framework can monitor log data and predict gearbox failures before they occur.

The work in [51] applied text mining to analyze wind turbine service history (operation and maintenance (O&M)) reports and identified failure-associated words and terms. The authors analyzed data from wind turbines with generator or gearbox issues and applied decision tree and random forest classifiers to identify contextual words in failure situations. The ultimate objective of this research was to automatically detect possible failures. The work in [51] is related to our research because of the use of text mining; while [51] focuses on O&M reports for a subset of possible types of accidents for operational planning, our research considers news accidents for all possible types of accidents for strategic planning.

The work in [52] reviewed the literature on the application of machine learning (ML) in the monitoring of wind turbines. The authors reviewed and categorized the models into 144 research papers. The authors identified the most common data sources and model techniques and presented guidelines for selecting the most applicable ML techniques.

The work in [53] proposed a diagnosis approach that uses deep learning to predict gearbox errors. Although deep learning models by definition require large training datasets, the system described in [53] is reported to achieve high prediction accuracy even with small data samples.

The work in [54] utilized deep convolutional generative adversarial networks (DCGANs) to supervise the quality of wind turbine generator bearings. Validated through a real wind turbine dataset, the method was designed to recognize atypical conditions in bearings.

The work in [55] introduced a methodology to detect specific wind turbine defects using transfer learning algorithms. The advantages of the new transfer-learning method are presented.

2.4.2. Other Applications of Advanced Analytics

These studies apply data analytics and methods to various aspects of wind turbines, such as the prediction of wind power. We used data mining methods to predict death and injury incidences from text data in wind turbine accident news reports, and our study thus falls into this category, both with respect to the type of data collected and the application of data mining methods.

The work in [56] presents a review of the literature on the application of data analytics for wind energy. This paper describes different analytics techniques and applications, especially in forecasting wind power. Different models used in very short-term, short-term, medium-term, and long-term wind power estimations are compared, and the better performing models are identified.

The work in [57] considered the volatile nature of wind and developed a statistical hybrid wind power forecast technique (SHWIP) based on dynamic clustering and linear regression, which performs better with less data compared to benchmark models. The

applicability of the model was demonstrated using real-world data and observations from Turkey.

In [58], the authors developed a model for the short-term estimation of wind power, with the ultimate goal of optimizing power and managing energy storage. The effectiveness of the developed model was presented through a case study.

The studies closest to our work are [5,8], as these two studies both analyze wind turbine accident news (where failures are also considered as accidents).

In [8], based on a tabular dataset of accident news, the authors analyzed the relationship between two major factors and two major responses, effects, and outcomes. The first factor is the stage of the life cycle of the wind turbine at which the accident occurred, and the second factor was the cause of the wind turbine accident, namely, nature, system and equipment, or humans. The two outcomes were the occurrence of death, the occurrence of injury, or a combination of the two. The authors employed Pearson's Chi-square test and Fisher's test to compute correlations and the information gain (Kullback–Leibler divergence) measure to evaluate the significance of the factors with respect to affecting death and injury. In addition to applying multiple classification algorithms, the authors also employed mosaic plots and classification tree plots to visually discover insights.

In [5], based on a text dataset of accident news, the authors employed text analytics and unsupervised machine learning methods of clustering and multidimensional scaling (MDS) to generate fresh insights regarding wind turbine accidents. Many of these insights revealed the relationship between country, month of the year, and the nature of the accident, such as the foundation of turbines failing frequently in December in Germany. In our study, we also employed unsupervised machine learning methods similar to those in [5], yet the methods employed in our study, namely, topic analysis and association mining, were applied for the first time in the analysis of wind turbine accident news.

2.5. Health Impact

The present research analyzed the factors behind deaths and injuries, similarly to [5,8,23–25,28,30,31,33–35] and other studies. Thus, it is related to the impact of wind turbines on public health. The work in [59] presents a review of the scope of the literature on the health impacts of wind turbines by analyzing 84 articles and identifying commonalities. Some of the analysis observations include the increase in publications since 2012 and considerable focus on research related to annoyance and noise elements.

3. Methodology

3.1. Data Analytics

Data analytics refers to the analysis of data with the goal of extracting patterns, discovering actionable insights, and constructing knowledge. Data analytics is also termed data mining or data science, even though slight differences exist, and these terms comprise a wide collection of techniques, ranging from visualization and summary statistics to advanced machine learning models [5,8,60–62]. The data analytics methods applied in a project are selected based on the nature of the data, knowledge of the domain, the questions to be answered, and the experience of the analyst [62].

In this study, tabular analysis (cross-tabulation) was used for summary statistics, and topic modeling and association mining were used to discover insights from a text collection of wind turbine accident news from around the world.

3.2. Machine Learning

Machine learning (ML) refers to computer algorithms that learn from data through training. Besides identifying patterns and determining insights, ML can be used to produce predictions and enhance the efficiency of systems [63]. These algorithms, which have expanded significantly in quantity and quality over recent decades, constitute the core of artificial intelligence (AI) and its applications. ML techniques are broadly classified as unsupervised, supervised, and reinforcement learning. In unsupervised learning (e.g.,

clustering, association mining, and topic modeling), learning aims to discover patterns or structures in a dataset without considering any target attribute. In supervised learning (e.g., regression and classification), training is performed using a target attribute, where each data point is accompanied by a target value, which can be numerical (in regression) or categorical (in classification). In reinforcement learning, similar to conditioning, the model attempts to maximize the reward signal through multiple trial-and-error efforts. By interacting with its environment several times and learning each time, the operation of the reinforcement algorithm is enhanced [64].

In this study, the unsupervised machine learning techniques of association mining and topic modeling were successfully applied to the collected text data.

3.3. Text Analytics

Text analytics (text mining) is the process of discovering new knowledge from unstructured texts. The process includes cleaning and preparing the source text data as input, processing the text data to structure it, and analyzing the structured data through data analytics techniques [5].

In this study, text analytics was conducted by following the common procedures of preparing the documents, obtaining document vectors (quantification of terms in documents) in terms of term frequency–inverse document frequency (TF-IDF) values, and conducting further analyses such as word cloud and topic modeling. The workflow for the implemented text analytics process is fully provided and described in the supplementary document.

3.4. Tabular Analysis

Tabular analysis (cross-tabulation, cross-tab analysis, and contingency table) was used to quantify the relationships between categorical attributes and other attributes through frequency tables. Numerical attributes can also be inputted into tabular analysis by transforming them into categorical attributes [62]. Tabular analysis is a highly useful and popular exploratory data analysis method used to compare subgroups within a group of data.

3.5. Word Cloud

A word cloud or a tag cloud, used as an exploratory visualization, represents the frequent terms in text data, with the size of the term indicating its relative frequency and importance [65,66]. Word cloud analysis was used in the present study to enable an overall understanding of the relevant terms in accident news.

3.6. Topic Modeling

Topic modeling is an unsupervised learning technique that algorithmically analyzes and identifies prominent topics from a given collection of unlabeled text data. Latent Dirichlet allocation (LDA), developed in [67], is a widely used probabilistic topic model that was selected and applied in the present research. In LDA, documents are thought of as a collection of different topics, where each topic is a group of associated words (terms) [67]. The technique provides the identified topics as the main output, the most important words in each topic, and their weights. The technique also outputs, for each document in the text collection, the weights of the identified topics for that document. Topic modeling has proven useful in automatically detecting topics from text collections (corpi) in various application domains, including the analysis of text collections relating to accidents [65,68–72].

3.7. Association Mining

Association mining is an unsupervised machine learning technique used to discover association patterns between different attributes or items. The input data can be inherently transactional (e.g., items in a purchase transaction at a supermarket) or can be transformed

into a transactional format. To this end, numerical attributes can be discretized to represent them as transaction items.

The two main outputs of association mining are frequent itemsets (sets of items) and association rules, characterized by performance metrics of support, confidence, and others. Frequent itemsets are combinations of items that appear frequently. Association rules are IF-THEN rules that describe the relationship between items, in the form of A⇒B, interpreted as "IF antecedent A THEN consequent B". Association rules can be either positive (associations that are actually observed) or negative (associations that would be expected but not observed) [61].

Of the important metrics used to assess the significance of association mining results, two of note are support and confidence. Support is the percentage of transactions in which the itemset is encountered in all transactions. Confidence of a rule A⇒B is the conditional probability of observing the consequent item or itemset B in a transaction, given that the antecedent item or itemset A is also present [61]. Applying threshold limits to confidence and support values helps to filter for the most significant results and focuses on their analysis.

While a multitude of faster algorithms have been recently developed, the standard and most popular algorithm for association mining is the Apriori algorithm [73], which efficiently computes itemsets that have support and confidence values above given thresholds.

3.8. Decision Tree Analysis

A decision tree (classification tree) analysis is a data mining technique in which a categorical target attribute is characterized in terms of input attributes in the form of a tree and associated rules [8]. Since the decision tree analysis did not yield conclusive results, further explanation of this technique and its application in the research are presented in the Supplement.

4. Developed Framework

This section describes the process steps and methods of the data mining framework that was custom-developed in this research and successively applied for the analysis of wind turbine accident news. The framework, presented in Figure 2, integrates the various methods of text processing and data analytics introduced in Section 3 as a unified framework, such that it can be applied to similar datasets. As shown in the legend of Figure 2, each type of data analysis is represented with a parallelogram, datasets and databases are shown with cylinders, process steps are shown with rectangles, and analysis results are shown with cornered ellipsoids.

4.1. Data Collection and Preparation

The first step of the developed framework was the collection and cleaning of the data, followed by the preparation of a tabular dataset consisting of core metadata attributes in columns, news in rows, and values in cells (Dataset A), as well as a text collection, also referred to as a corpus (Dataset B).

4.2. Tabular and Visual Analysis

As an exploratory analysis, tabular analysis (cross-tabulation) was performed with pivot tables on the tabular Dataset A, followed by visualization of the results. For focused results, each analysis filtered only the top few values of the relevant categorical attribute, where the filtered values accounted for the bulk of the observations.

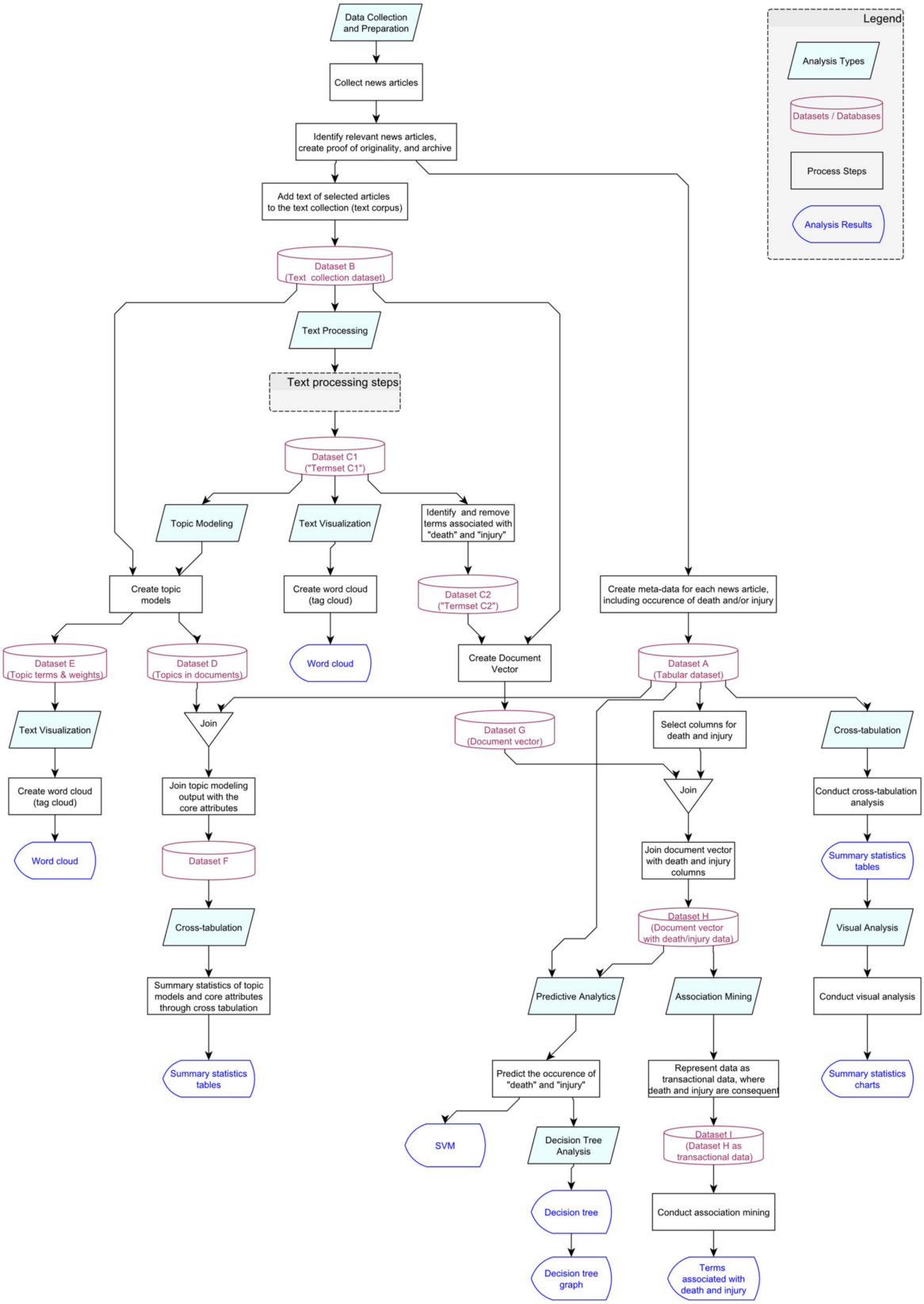

Figure 2. Flowchart representation of the developed and implemented data analytics framework.

4.3. Text Processing

Text processing was performed on the text collection of all the news from Dataset B.

In the initial steps of text processing, punctuation, characters with fewer than three characters, and terms consisting of digits were removed. Furthermore, text was converted to a lower case, stop words were removed, and words in the remaining text were stemmed.

In the later steps of text processing, a bag of words (terms) was derived, together with term frequencies. At this stage, certain filtering criteria can be used to focus on the more relevant terms in the documents and to reduce the number of terms.

An essential filtering criterion adopted from graph theory is *the minDegree* [74]. In our research context, *minDegree* indicates the minimum threshold value for the number of documents in the text corpus where a term appears. In our study, only the terms appearing in at least seven documents (1% of the text collection) were filtered after the initial pre-processing steps and used in further analysis.

After initial cleaning of the data (whose steps are described below) and creation of a bag of words, two sets of terms were prepared for further analysis. The first set of terms, referred to as Termset C1 (Dataset C1), included all frequent terms (except irrelevant terms and standard terms directly associated with "wind" and "energy") and was used for text visualization with word cloud and topic modeling (LDA). The second set of terms, referred to as Termset C2 (Dataset C2), further excluded any terms that could be directly linked to death or injury. This second set of terms was used to find the relation between the remaining words in the text and the occurrence of death or injury through association mining and decision trees.

In the next step of text processing, a document vector was created for the dataset of terms. At this stage, a metric of term frequency, for example, the term frequency–inverse document frequency (TF-IDF), can be calculated to create vectors for each document. Relative term frequency is a measure of how frequently a term occurs in a document relative to the total number of terms contained in that document. Consider a document with 1000 terms in which a certain term occurs 50 times; then, its relative frequency can be calculated as 50/1000 = 0.05 [5]. The inverse document frequency (IDF) indicates the value of a term in the entire document corpus. While there are different formulae and definitions in the literature for IDF, the formula for smooth IDF in the KNIME software, which was used in this research, is

$$\text{IDF} = \log\left(1 + \frac{\text{Total number of documents}}{\text{Number of documents containing the specific term}}\right) \quad (1)$$

The product of TF and IDF is referred to as TF-IDF and helps to measure the significance of a term in a document within the complete corpus or set [75]. The values of the term frequency metric can then be used for further text processing and analysis.

For association mining and predictive analytics, the document vector had to be created using the text collection Dataset B only for the terms in Termset C2 (Dataset C2). This document vector was created as Dataset G. Then, the document vector Dataset G was augmented to include information on the occurrence of death and injury, resulting in Dataset H. A total of two derivations of Dataset H were used: Dataset H1 contained the IsThereInjury column only, whereas Dataset H2 contained the IsThereDeath column only. For association mining, Dataset H was converted into a transactional format as in Dataset I.

4.4. Text Visualization

The frequent terms in the text collection (Dataset B) were visualized through word clouds (tag cloud), where the size of each term represents its frequency and weight, which were read from Datasets C1 and E. A select number of the most frequent words can be filtered through visualization to enable easier interpretation.

4.5. Topic Modeling

Topic modeling using the LDA (Latent Dirichlet allocation) algorithm [67] was applied to detect topics from the dataset. The number of topics and the number of words in each topic were selected as parameter values. The output containing each document's mapping

to an assigned topic and the probability of each document belonging to a certain topic were produced as Dataset D. A list of terms allotted to each topic, along with term weights, was also generated as Dataset E. Using this information, a word cloud was created for each topic, with the size of the term representing its weight in the topic. By combining this information, the main themes of the extracted topics, which represent the recurring ideas in the text collection analyzed in this study, were identified. Furthermore, through a merged dataset F, the distributions of topics over the core attributes of the accident news (IsThereInjury, IsThereDeath, Country, and others) were investigated.

4.6. Association Mining

Association mining was performed with Dataset I using the Apriori algorithm, considering the values of the target attribute as the consequent B of the A $\Rightarrow$ B rules. Frequent itemsets and association mining results were generated, and their support and confidence values were above the chosen threshold values. Association rules were filtered to list only those where "Death/NoDeath" and "Injury/NoInjury" were in the consequent.

4.7. Decision Tree Analysis

Decision tree analysis was conducted using Dataset H to gain insights into the terms associated with deaths and injuries.

4.8. Predictive Analytics

Predictive analytics were applied using both Dataset A and Dataset H to predict the occurrence of death and injury based on tabular and text data. Specifically, supervised machine learning methods for support vector machines (SVM) and decision trees were selected for application.

5. Data

5.1. Data Collection and Data Cleaning

The dataset analyzed in this study was a collection of 721 news reports from 2010 to 2019. The data were constructed using two main sources: the news cited in the CWIF dataset and results from Google search engine using the search terms "wind turbine accidents" and "wind turbine failures". Starting with an initial collection of more than 1200 news articles, unrelated, duplicate, and unverifiable news was eliminated. Any negative-impact incidents, accidents, failures, or breakdowns during any stage of their life cycle were considered, and only news reports that provided a certain level of detail were included. Multiple other considerations and decisions were made when selecting the most reliable data. Many interesting cases, which were not included but are worth mentioning, are presented exhaustively for the first time in the literature within the Supplementary Materials to this paper.

This analysis is dependent on publicly available accident news. While the availability of a larger corpus of news would most likely improve the quality of results, there is a potential threat to validity that applies to this research, as well as similar research where accident news is analyzed, in that the validity of any research study depends on the reliability of the data, and the extent to which the sampled data represent the population. Without full access to all wind turbine accidents and their details, it is impossible to prove that the data used in the present study or related studies truly represent all wind turbine accidents. Such a full dataset cannot practically be constructed by an independent research team, because it would not be possible to convince all wind turbine manufacturers to share all their data on accidents, let alone convince even one manufacturer. The data used in the present study is thus valid and reliable in itself because of tedious efforts in data acquisition and cleaning, yet it cannot be proven to be truly representative of all accidents. The mentioned limitation is applicable to all the research based on wind turbine accident news available on the Internet (Section 2.2 and [5,8]), and is a potential threat to validity.

All the included news, as inherent to the nature of reporting, reported accidents that took place before the date of the news. In some cases, the accident occurred well before the news reports. This would have been a threat to the validity of our research if our analysis was focused on dates. However, our analysis was conducted regardless of the year or month of the accidents and focused on the nature of the accidents and their impacts on human life, thereby eliminating this possible threat to validity.

All news in languages other than English were translated using Google Translate one final time in May 2021, and the complete data analysis was conducted again. This was needed because of the significant improvements in the quality of Google Translate translations in the timeframe of the research.

Because the present research analyzes the text content of news topics, another potential threat to validity is the possible inaccuracies in automated machine translations. Human versus algorithmic assessment are two approaches for assessing the quality of machine translations [76]. For the data used in the study, posterior qualitative human inspection was conducted to assess the quality of translations, and the translations were found to be accurate. In addition to the high translation quality for the overall context, frame of reference, and meaning, the translations of technical terms—which form the basis of the analysis— were found to be particularly accurate. As an illustration of the accuracy of the translations, a sample of an unedited excerpt from German to English can be found in Section 6.5.3, which can be observed to be accurate and meaningful.

A drawback of using Google Translate is that the results are not reproducible [77]. However, the basis of the analysis in the present paper was simply the occurrence frequencies of accident-related terms, and thus errors related to sentence structure, grammar, framing, and sentiment mapping would not affect the results. Furthermore, various studies on Google Translate's performance report high quality scores overall [78,79], particularly for translations from the German language [80], even when translation errors other than the errors for term translations are included. In conclusion, the possible effects of machine translation errors can be considered minimal in the present study.

Accident news reports for 2020 were excluded from the dataset for three main reasons: First, the data were revised and finalized one final time in June 2020 to discover any new accidents from 2019 that could have been omitted in earlier searches. In other words, the data collection was completed in mid-2020 to ensure that as many of the 2019 accidents as possible were included. Second, as the global COVID-19 pandemic had effects on the economy and energy consumption, the data for 2020 would be affected by the new world under the COVID-19 pandemic, different from those in the decade 2010–2019. Third, a significant amount of time and effort over more than a year was required to develop, test, and implement the framework and analyze and interpret the data.

Metadata (data about data) from these news reports were compiled in a structured database table (Dataset A). Details such as location, offshore vs. onshore, and phase of the life cycle were populated manually into Dataset A by reading every news report in detail and understanding its content, as well as carrying out further research online to complement any missing information. For example, in many cases, the web page cited in CWIF was removed, so the original or moved news source was searched for online using the title and other terms. To serve as proof of originality and enable scientific reproducibility [81], the text of all news, as well as the screenshots of the source web pages, were fully archived and will be made available to readers upon contacting the authors.

Some of the texts in the text collection were edited such that only the wind turbine accident under consideration was reflected, to the extent possible. Previous wind turbine accident history or wind turbine incidents in other locations mentioned within the same news text were removed so as not to affect text mining results (in cases where it was not possible to distinctly distinguish accident news, the original text was retained as such). In many cases, the removed text has already been reflected in the analysis of other news texts. For the cases in which there were more than three or four sentences describing another incident, it was included as a separate piece of news. For the cases in which a single news

article cites multiple wind turbines in a region affected by similar weather conditions at the same time, the article was retained in its original form.

5.2. Data Attributes

The attributes in the structured dataset (Dataset A) are the following:

- ID: Unique identifier for each accident;
- Year, month, day: year, month, and day of accident reporting;
- Location 2: Specific location (city, town, or village) of the turbine where the accident occurred;
- Location 1: State, province, county, and territory in which the accident occurred;
- Country: Country where the accident occurred;
- Language: Language (English, or German, etc.) in which news is reported;
- IsThereInjury: Indicates whether injury to a human occurred or not (takes the value "Injury" or "No Injury");
- IsThereDeath: Indicates whether death of a human took place or not (takes the value "Death" or "No Death");
- Offshore Onshore: Location of the turbine with respect to land (takes the values "Offshore" or "Onshore");
- PhaseOfLifeCycle: Stage of the life cycle of the wind turbine during which the accident took place (takes the values "Transportation," "Construction," "Operation," "Maintenance"). Deciding on the phase of the life cycle in which the accident occurred was a major challenge. Descriptions of the categories that guided this decision are provided in the Supplement;
- Mode of transportation when the accident took place, in case the PhaseOfLifeCycle was "Transportation" stage (takes values "Road," "Water," or "Rail");
- Weblink: Link to the webpage of the news;
- IsLinkActive 1, 2, 3, 4: Indicates whether the webpages are currently working at different time points in the project. IsLinkActive3 refers to the state of May 2021;
- IsAllDataAvailable: Indicates whether the complete set of supporting documents, including screenshots of the original webpage and the corresponding text, are available in the collected and archived dataset. The value is "Yes" for all accidents;
- IsThereInjuryOrDeath: Indicates whether either injury to or death of a human took place (takes the value "Injury or Death" or "No Injury or Death"). If any of these two outcomes was observed, the value is "Yes";
- IsFullText: Indicates whether the complete text of the news is available. The possible values are "FullText," "TruncatedText," "VideoSource," or "PhotoSource". In all cases, the collected and archived data have proof of originality and contain details of the accident. For the research, only "FullText" news was used;
- IsOriginalSource: Indicates whether the news is an original text or a news aggregator;
- DerivedFrom: Indicates the ID of the original news if the text had to be edited. For the instances in which it was not possible to identify the attribute value (e.g., while determining the PhaseOfLifeCycle, OffshoreOnshore location, or IsThereInjury), the respective field was left blank.

6. Analysis and Results

This section details the application of the framework, which is presented in Section 4, using the data collected and created during our research. The KNIME (https://knime.org, accessed on 7 October 2021) analytics platform was the primary software used throughout the study.

6.1. Data Preparation

A text collection (corpus) of 721 items of news, covering a decade of accidents between 2010 and 2019, was created over three years of labor and archived as text files and

screenshots. Furthermore, a tabular dataset (Dataset A), where each row represents news in the text collection (Dataset B), was constructed by carefully reading each news item.

6.2. Tabular and Visual Analysis

The initial analysis in this research was the analysis of the tabular dataset of the core attributes of news (Dataset A). To this end, tabular analysis and visualization were conducted. The analysis was conducted over all the years.

The first analysis focused on the distribution of accidents with respect to countries, as shown in Figure 3. News reporting accidents in the United States constitute the majority of the news in the dataset. The USA, Germany, the UK, Canada, and Australia account for almost 90% of the total reported cases in the dataset. There is a bias in the dataset towards English-speaking countries, due to the research being conducted and reported in the English language and the search terms also being in the English language.

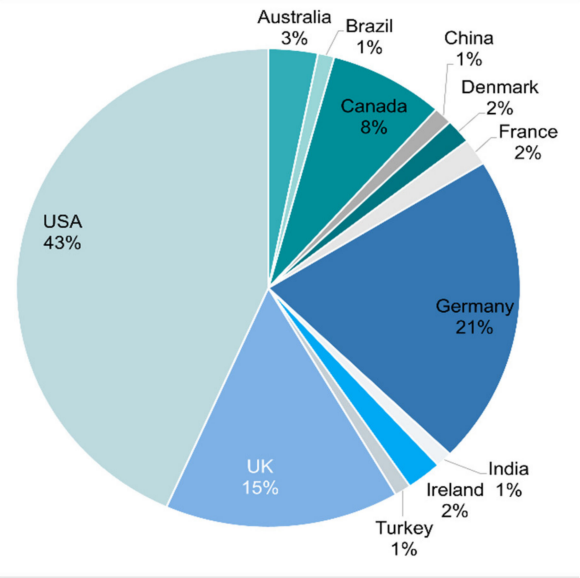

Figure 3. Distribution of accidents across top twelve countries.

The second analysis focused on the frequency of accident news with occurrences of death and injury across the top 12 countries, as shown in Table 1. It can be observed that 42.86% and 28.57% of accidents in Brazil and 37.50% and 25.00% of accidents in China resulted in deaths and injuries, respectively, with notably higher percentages compared to other countries. Table 1 reveals that many countries had no deaths or injuries reported in the accident news (in five cases overall, injury could not be determined).

Table 1. Percentage of death and injury in accidents, across countries.

	Percentage of Death	Percentage of Injury
Brazil	42.86%	28.57%
China	37.50%	25.00%
India	14.29%	14.29%
Turkey	14.29%	14.29%
Ireland	13.33%	6.67%
Germany	7.19%	8.63%
UK	6.00%	9.00%
USA	4.86%	5.56%
Canada	2.04%	8.16%
Australia	0.00%	19.05%
Denmark	0.00%	0.00%
France	0.00%	8.33%

In the collected data, the number of news articles from Brazil, China, India, and Turkey was considerably lower than in other countries. While Figure 3 shows the composition of the news collection, these countries were included. For the sake of completeness, the same countries, despite having very little news for them, are also included in Table 1. However, it should be noted that because of the small sample size (seven or eight items of news for each of these countries), the death and injury statistics for these countries may not be accurate. Moreover, these items of news, having made their way to the media and the Internet, mostly in English, may have higher chances of reporting events with extreme consequences, such as deaths and injuries. Hence, the proportion of deaths and injuries compared to the number of news articles may be higher for them. It is important to consider these possible biases when reading Table 1.

The third analysis was on a possible association between the phase of the wind turbine's life cycle and the frequency of accidents. Figure 4 reveals that, for the top 12 countries in the dataset, more accidents occurred during the operation (65.51% of accidents in these countries) and transportation (15.21% of accidents in these countries) phases, compared to the other phases of construction and maintenance. The life cycle phase could not be identified clearly for some of the news (1.51% of the accidents in these countries).

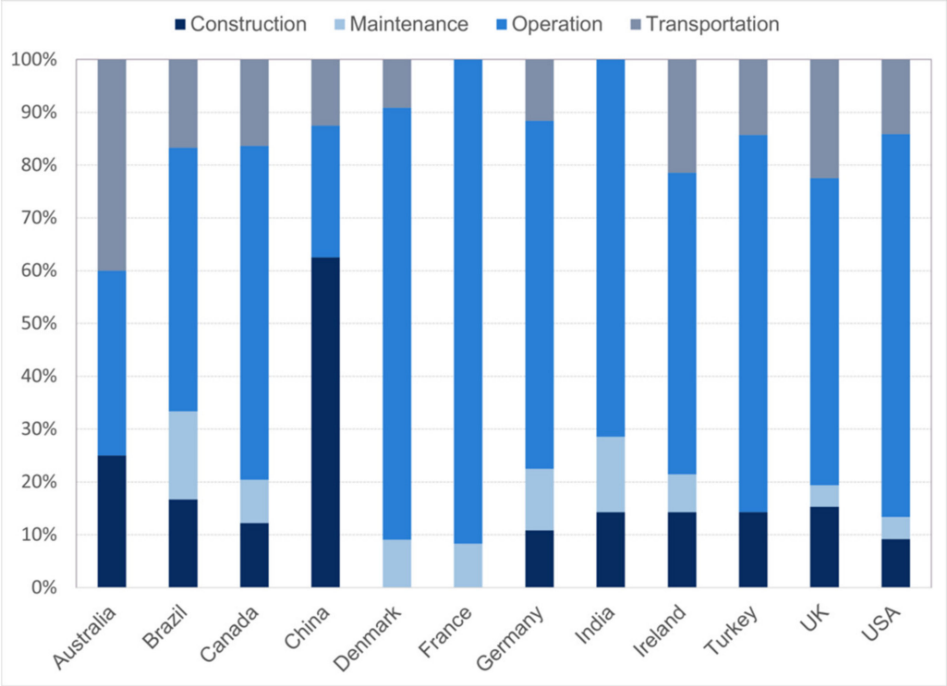

Figure 4. Association between the phase of the wind turbine's life cycle and the frequency of accidents in the top twelve countries.

The fourth analysis focused on the association between death and injury, and the phase of the life cycle. As tabulated in Table 2, one can observe that, when computed for accidents within each phase, a higher proportion of deaths and injuries occurred during construction (28.75%, and 30%, respectively) and maintenance (19.15%, 29.79%, respectively). In 11 cases, the phase of the life cycle could not be identified clearly, and hence, the deaths and injuries (27.27% and 63.64% of the uncategorized phase, respectively) in these accidents could not be classified within any particular phase.

The fifth analysis examined the association between death and injury, and location (offshore vs. onshore) of the turbine. Table 3 shows that a higher proportion of offshore turbine accidents resulted in deaths and injuries (10.87% and 19.57%, respectively) compared to accidents at onshore turbines (5.79% and 7.42%, respectively).

Table 2. Association between death and injury, and phase of life cycle.

Phase of Life Cycle	Total Cases	No. of Cases with Injury	Percent of Cases with Injury	No. of Cases with Death	Percent of Cases with Death
Construction	80	24	30.00%	23	28.75%
Maintenance	47	14	29.79%	9	19.15%
Operation	477	2	0.42%	2	0.42%
Transportation	106	12	11.32%	7	6.60%
Uncategorized phase	11	7	63.64%	3	27.27%
Total	**721**	**59**	**8.18%**	**44**	**6.10%**

Table 3. Association between death and injury, and location of the turbine.

Location	Total Cases	No. of Cases with Injury	Percent of Cases with Injury	No. of Cases with Death	Percent of Cases with Death
Onshore	674	50	7.42%	39	5.79%
Offshore	46	9	19.57%	5	10.87%
Uncategorized location	1				
Total	**721**	**59**	**8.18%**	**44**	**6.10%**

Considering the complete dataset, the percentages of deaths and injuries were 6.10% and 8.18%, respectively.

These summary statistics, based on the recent decade of accidents, can be used in many ways, including where to focus on reducing accidents and in the calculation of rates for insurance.

6.3. Text Processing

Text analytics, including topic modeling, have been performed using the text collection (corpus), Dataset B. After completing pre-processing and the construction of a bag of words (terms), only the terms appearing in at least seven documents (1% of the text collection, *minDegree* = 7) were filtered and used for further analysis. Irrelevant terms, such as human names and meaningless ASCII characters, as well as standard and obvious terms such as "wind turbine," were identified and removed during data processing to create Termset C1 (Dataset C1) of frequent, meaningful, and non-obvious terms. Termset C1, consisting of 1527 terms, was used as the filtered bag of words in topic modeling.

To prepare the data for predictive analytics and association mining, terms related to death and injury were removed from Termset C1 to create Termset C2 (Dataset C2), consisting of 1505 terms. This was a necessary step, as such words directly related to death and injury would distort the results of predictive analytics. The terms removed in creating Termsets C1 and C2 are provided in the Supplement.

After creating Termset C2, a document vector was constructed by calculating the TF-IDF value for each term (in Termset C2) in each document, resulting in Dataset G (document vector).

In the applied text analytics workflow, it is possible that some terms containing the same meaning, but different expressions, may not be recognized by automatic computer mining, leading to the problem of omission and potential threats to validity. However, in the dataset used in this study, such term sets were very small in number and percentage. An inspection by the authors of term sets with similar meanings revealed 20 such sets. Furthermore, while the terms in each set may contain practically the same meaning, there may also be subtle differences in meaning between them. Because these word sets constitute only 3% of all the words, the analysis was conducted by treating them as separate terms.

6.4. Text Visualization

In our analysis, 100 terms from Termset C2 were selected to construct the word cloud in Figure 5. From Figure 5, one can quickly identify the frequently occurring relevant terms,

such as "fire," "blade," "damage," "caus," "road," "oper," and "compani". These terms, at a very high level, represent the key terms in wind turbine accident news from 2010 to 2019.

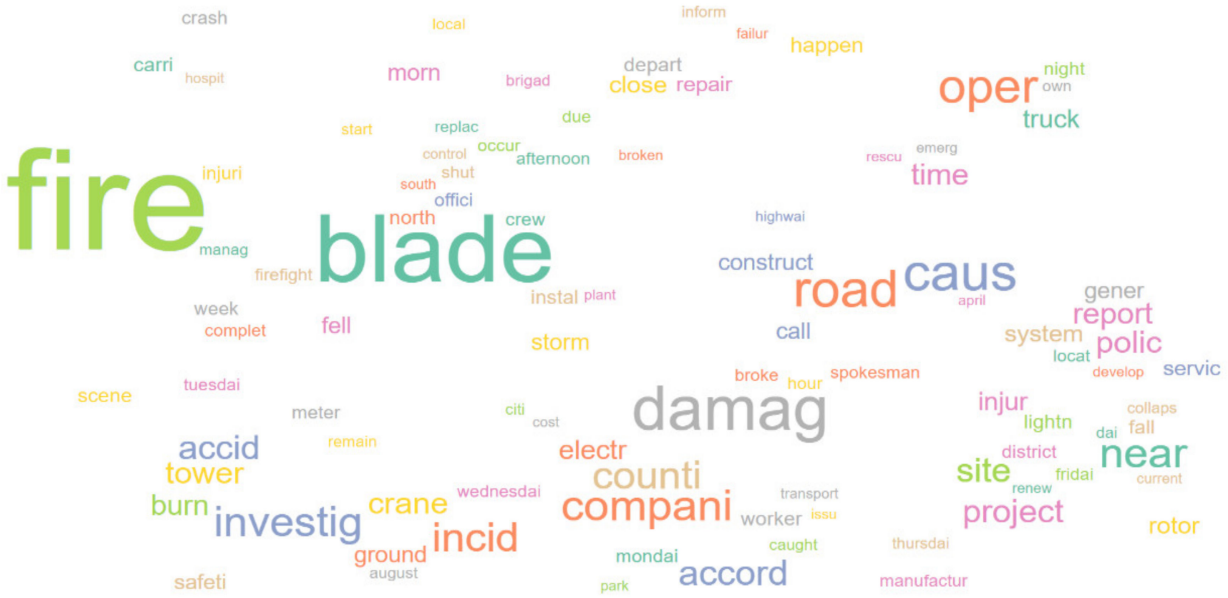

Figure 5. Frequently occurring relevant terms in Dataset B, visualized as a word cloud.

6.5. Topic Modeling

Topic modeling was applied to the text collection (Dataset B), which was updated to include only the terms in Termset C1, which excludes irrelevant terms, infrequent terms (which appear in less than 1% of documents), and standard terms (directly related to wind energy and turbines).

The LDA algorithm [67] was applied for topic modeling to identify 10 topics with 20 words each, as shown in Figure 6.

The choice of the number of topics is a critical choice when applying the LDA algorithm. To this end, the elbow method, which is extensively used in clustering [82], can be used to detect a suitable value for the number of topics. In the elbow method, the sum of squared errors (or a similar error measure) is plotted against the number of topics to identify where significant inflection points ("elbows") appear. The core idea is to identify the parameter value after which only diminishing returns are obtained. When the elbow method was applied to the dataset (Figure 7), notable elbows were discovered at 2, 6, 10, 13, and 17. A high number of topics could lead to extremely granular themes, while a lower number may not capture distinct themes effectively. To this end, the default value of 10 in KNIME's LDA node, which also appeared as an inflection point in Figure 7, was determined to be an appropriate parameter value for the dataset at hand.

To evaluate the quality of the LDA topic model, a metric used extensively in similar studies is the log likelihood of the latent variables in LDA [83,84]. To this end, in our analysis, we obtained the log-likelihood values over the iterations of the LDA algorithm ran [85] and plotted them against the iteration count (Figure 8). As can be observed from Figure 8, the log-likelihood values are significantly improved in the first 100 iterations and converge to an upper bound value between (–7.1, –7.0). The maximization of the log likelihood implies the maximization of the likelihood function, which implies that the obtained probability density function (PDF) is most likely to generate the observed data [86].

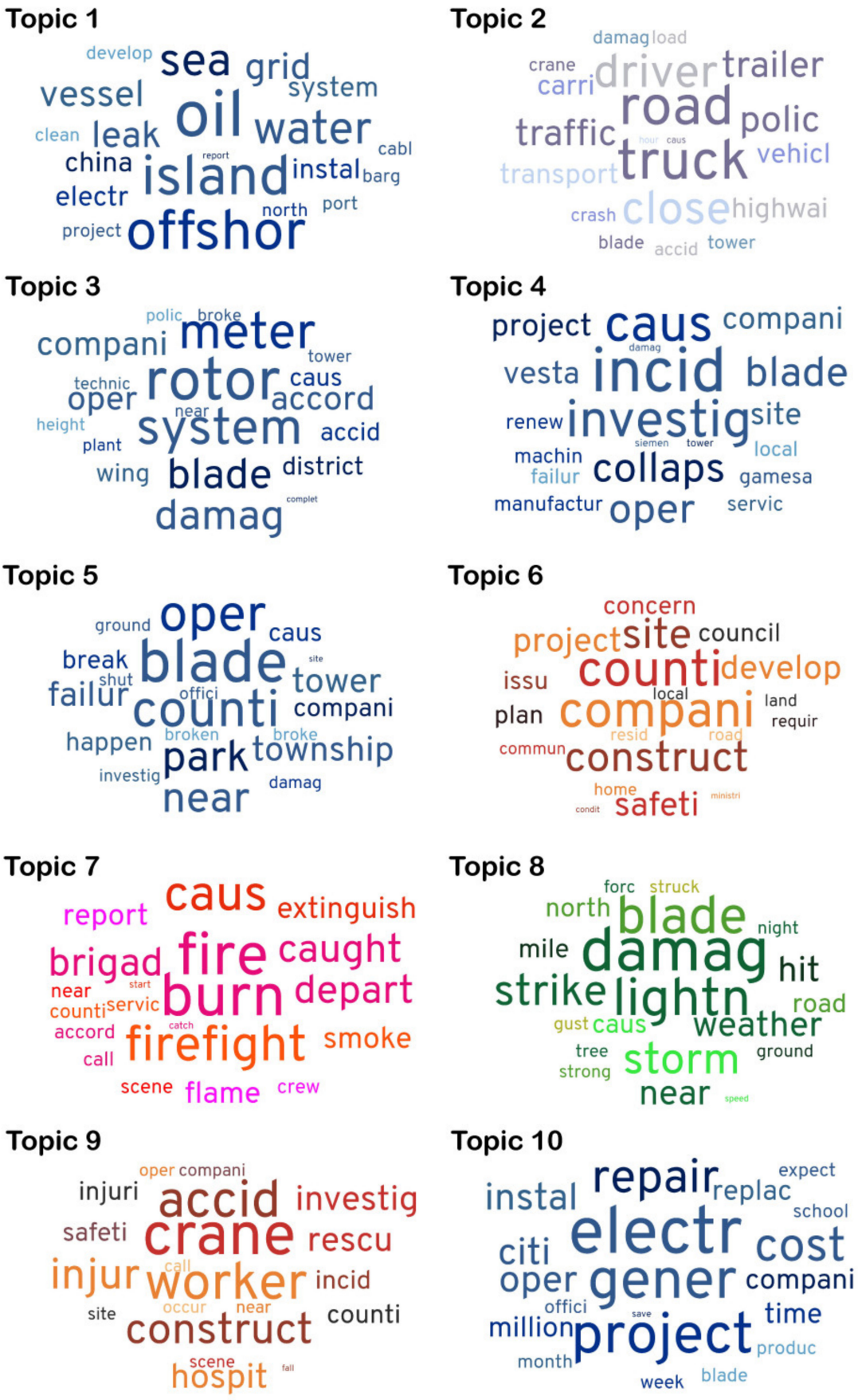

Figure 6. Word cloud visualization of topic modeling results.

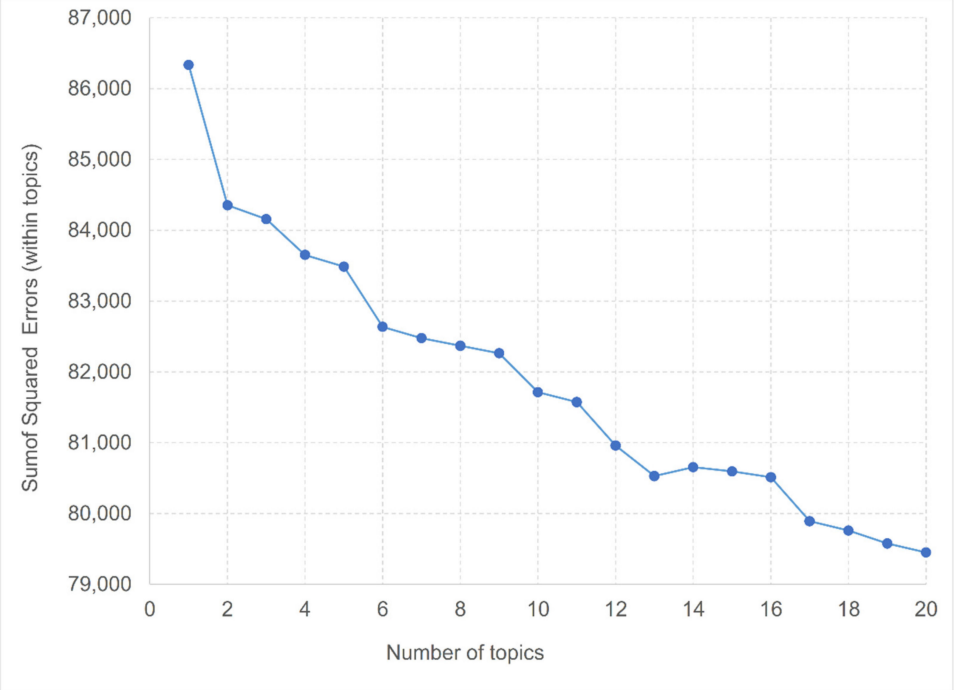

Figure 7. Elbow plot for selecting the number of topics in the LDA algorithm.

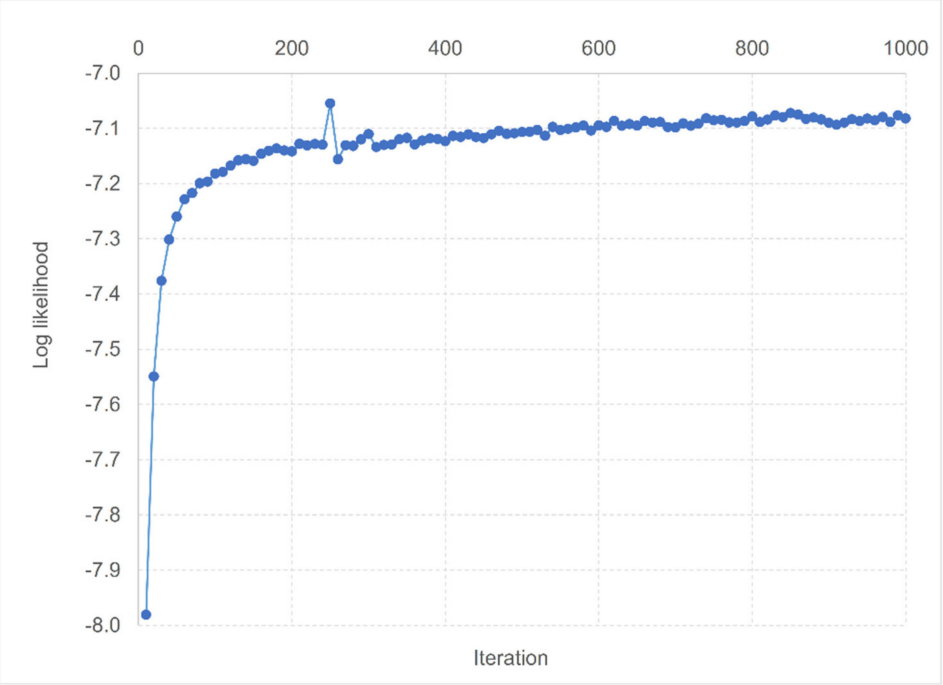

Figure 8. Log likelihood of the latent variables in LDA vs. iterations of the LDA algorithm.

In addition to assessing the model's quality with a metric, visual assessment of the model's quality could prove insightful. To this end, the t-distributed stochastic neighbor embedding (t-SNE) method [87] was implemented in KNIME [88] with *perplexity* = 30 to represent the data in two-dimensional space (Figure 9). In Figure 9, similar accident news is modeled as nearby points, and dissimilar news is modeled as distant points. Furthermore, the topics determined by the algorithm are denoted by the color on the plot. Figure 9 suggests that the topics selected by the LDA algorithm are in accordance with the two-dimensional mapping of the data through t-SNE. While the value of *perplexity* =

30 was used in Figure 9, the same consistency pattern was observed with other values of perplexity.

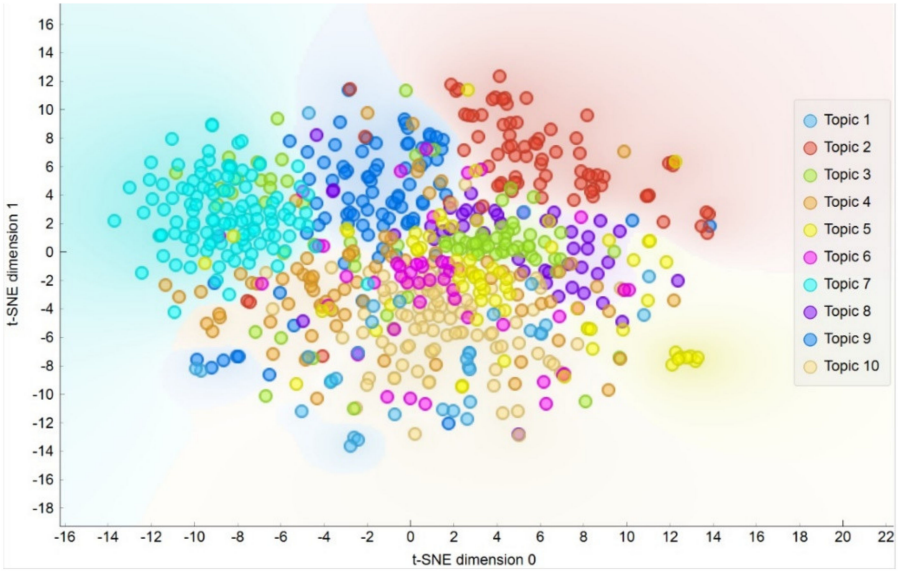

Figure 9. Mapping of the documents to two-dimensional plane with the t-distributed stochastic neighbor embedding (t-SNE) method. The topic labels obtained through the LDA algorithm are denoted with color.

The themes and accident characteristics of these 10 topics are interpreted in this section, with the interpretations provided in "quotes" in the beginning of each topic's discussion. The terms in the word cloud for each topic in Figure 6 are shown in bold font below. Furthermore, for each topic, example news reports with a weight of >0.85, along with excerpts, are provided as examples. While the interpretations and descriptions for a topic may not be completely representative of every news assigned to that topic, they still provide a general characterization and profiling of the news assigned to each topic.

It must be noted that a news report may not only contain details about the specific accident, but may also mention other information, including, but not restricted to, history of the wind turbine or company, information about the city and stakeholders, the general nature of the wind industry, or even other local news. This can be considered as noise in the data affecting the topic modeling analysis results and a possible threat to the validity of the research.

An interpretation and description of the topics in Figure 6 are as follows.

6.5.1. Topic 1

"**Reports** about **offshore projects developed in sea** or **water** that could be facing accidents with **barges**, other **vessels, cables**, or at **ports**. Some of the news reports talk about cases that affected the **electricity** given to the **grid** or those that required the **cleaning** of **oil leaks**". (This topic covers a mixture of news spanning different themes, which do not strongly belong to other topics.)

Accident news reports that display a high weight for this topic include 2018-0065, 2015-0034, and 2014-0204. The excerpts below from the accident news report 2018-0065 (which has a weight of 0.911 for this topic) can serve as a representation of the content in this topic.

"Power was restored Saturday evening to about 22,000 Maui customers who experienced an outage earlier in the day, Maui Electric Co. spokesperson Shayna Decker said Sunday. At around 3:55 p.m. Saturday, about 22,000 customers in parts of Upcountry, Lahaina, Kahului, Haiku and East Maui lost power when wind energy from an independent wind farm on the island suddenly dropped and affected frequency of energy on the

electrical grid. Decker said this caused the grid to automatically shed loads on various circuits to protect the system from damage" [89].

6.5.2. Topic 2

"**Traffic accidents** during **transport**, especially those involving **trucks** and **trailers carrying loads** of **tower** and **blade** components. Many of these **accidents** or **crashes** were due to **drivers** and resulted in **damage** (especially to other **vehicles**), **road closure** (especially **highways**), and police intervention. Some of the accidents were **caused** by **cranes**".

Accident news reports that display a high weight for this topic include 2017-0137, 2012-0035, and 2015-0089. The excerpts below from accident news reports 2017-0137 (which has a weight of 0.985 for this topic) can serve as a representation of the content in this topic.

"Police say traffic diversions will remain in place until heavy lifting equipment can recover the vehicle. A road in southern Scotland has been completely blocked after an accident involving a lorry towing part of a wind turbine. The incident happened at about 22:45 on Monday on the A713 Castle Douglas to Ayr road just north of Parton. A Daf lorry, part of a convoy heading to the Brockloch wind farm at Carsphairn, left the roadway. Nobody was injured in the crash but the route is expected to be closed for a considerable length of time" [90].

6.5.3. Topic 3

"**Damage** to **wing, blade, tower** during **operations**, **caused** by **technical** problems, including **rotor** and other **system** components that were **broken**, and in some cases were blown some **meters** away from the original **plant**. These **accidents** sometimes involved **police** from the **district**".

Accident news reports that display a high weight for this topic include 2013-0001, 2018-0207, and 2018-0117. The excerpts below from accident news report 2013-0001 (which has a weight of 0.971 for this topic), translated from German to English, can serve as a representation of the content in this topic.

"A piece of one of the wings on one of the wind turbines near Schäcksdorf broke off," says Ute Neumann, managing director of the Drahnsdorf agricultural cooperative, on the reporter's phone. . . . The fact that the wing tip of a rotor blade breaks off is an exceptional event that has not yet occurred on any of the wind turbines operated by them, explains Andreas Ehrenhofer, Managing Director of Teut Windprojekte GmbH in Berlin. The damage occurred during the Xaver hurricane. At this point in time, the systems had already been switched off. "The gondolas then turn in the main wind direction in order to give the storm as little attack surface as possible," explains the engineer. . . . The wind turbine remains inoperative until the rotor blade has been replaced" [91].

6.5.4. Topic 4

"Typically, **operational incidents causing** damage and which required **investigations** on **site** and **servicing** of the **machines**. Many of these accidents took place in **projects** where **manufacturers** or **companies** such as **Vestas, Gamesa, and Siemens** were mentioned, resulting in **blade failures** and even **tower collapse**".

Accident news reports that display a high weight for this topic include 2014-0137, 2012-0031, and 2016-0092. The excerpts below from accident news report 2014-0137 (which has a weight of 0.987 for this topic) can serve as a representation of the content in this topic.

"On Jan. 26, a Vestas blade failure occurred at a wind farm in Northern Jutland, Denmark. Vestas spokesperson Matthew Whitby tells NAW that the incident involved a single V90-3.0 MW wind turbine, which uses 44-m-long blades. Two of the blades were damaged as a result of the incident, but no one was injured . . . "As a result of the incident, the [affected] turbine was shut down," he says. "Vestas service technicians are present at the site, and an investigation into the cause of the incident is under way"" [92].

6.5.5. Topic 5

"**Failures** such as **broken blades or tower damage caused** during **operations**, especially at **sites** in or **near counties, townships** or **parks**, resulting in **shut**down and **official investigation** from the **company**".

Accident news reports that display a high weight for this topic include 2016-0034, 2010-0208, and 2012-0201. The excerpts below from accident news report 2016-0034 (which has a weight of 0.987 for this topic) can serve as a representation of the content in this topic.

"When D'Eon, who lives near the Pubnico Point Wind Farm, came outside to investigate what he heard, he saw that one of the blades on a turbine was "in distress". The blade was bending, says D'Eon, who says he heard the thunder-like sound at around 5 p.m. on March 19 . . . "The next steps include conducting an investigation into the cause of the blade issue—that is already underway—and getting the necessary equipment to the site to remove the damaged blade. I don't have the timing on that at this point," he said Sunday. The site has been secured" [93].

6.5.6. Topic 6

"Mostly assigned to **safety issues** and **concerns** in wind turbine **project development** and **construction** sites in **counties** where **communities reside**. Due to effects on assets/resources, such as **land**, **roads**, and **homes**, **project planning** by the **company** and **local council requirements** are affected and may involve **ministries**".

Accident news reports that display a high weight for this topic include 2016-0073, 2017-0151, and 2015-0045. The excerpts below from accident news report 2016-0073 (which has a weight of 0.858 for this topic) can serve as a representation of the content in this topic.

"The owners of farmland used to build the Pilot Hill Wind Farm in Iroquois County are suing the wind farm's developers, arguing that they are still owed money for access roads built on their property and for damage done to their land during construction . . . After the project was sold by Vision Energy to EDF, construction of turbines and access roads began on the Haleys' and others' property. According to the lawsuit, workers contracted by EDF dug a large borrow pit in the middle of the Haleys' farm—without the Haleys' permission or knowledge—and the excavated soil was used to create roadways over ditches to allow heavy machinery to cross into their farmland. The Haleys further allege that the digging of the borrow pit—and the subsequent filling of the pit with water—affected the farm's watershed, causing infrastructure damage to some 80 to 100 acres . . . In a statement, EDF spokesperson Sandi Briner said: "EDF Renewable Energy develops renewable energy projects in a responsible fashion, taking into account key stakeholders in the process of adhering to all local requirements and guidelines. The local community was extensively consulted in the development phase, and the construction was implemented according to a plan that had consents from all affected parties. If and when variances from a project's approved plans arise, EDF Renewable Energy meets with all affected parties to discover the appropriate and best solution"" [94].

6.5.7. Topic 7

"**Fire** incidents **near counties** where components **caught flames** and **burned** with **smoke** on the **scene**. **Firefighter brigades** and **service crews** have led **to extinguishing** efforts".

Accident news reports that display a high weight for this topic include 2019-0032, 2019-0072, and 2019-0049. The excerpts below from accident news report 2019-0072 (which has a weight of 0.983 for this topic) can serve as a representation of the content in this topic.

"A wind turbine fire broke out at a North Fork winery recently, police said. According to Southold Town Police, the fire broke out on May 30 at 1:29 p.m. at Shinn Estate Vineyards on Oregon Road in Mattituck. Southold Town Police and Mattituck Fire Department responded to the actively burning wind turbine; the Mattituck Fire Department was able to extinguish the fire before it was able to cause any injury or further damage, police said" [95].

6.5.8. Topic 8

"Accidents **caused** by **weather** conditions such as **lightning strikes**, **storms**, **and strong gusts hitting** with **force** and **speed**, resulting in **damage** to **blades**, **roads**, and **trees**".

Accident news reports that display a high weight for this topic include 2012-0039, 2019-0055, and 2011-0124. The excerpts below from accident news report 2019-0055 (which has a weight of 0.961 for this topic) can serve as a representation of the content in this topic.

"The north country woke up to thunderstorms Wednesday morning and apparently the lightning damaged a windmill blade in Lewis County. A picture of the damage was taken by the folks at Moser's Mapleridge Farm on Wilson Road in the town of Denmark. They told 7 News they believe the lightning made a hit on the windmill and weakened it. It was a little later, as the turbine was spinning, that the top piece fell off" [96].

6.5.9. Topic 9

"**Accidents** during activities such as the **construction** of wind turbines on-**site**, resulting in **workers** being **rescued** and **hospitalized** with **injuries**. Some of these **incidents** involved **crane operations** and **workers** who **fell**. **Investigations** were carried out on whether **safety** procedures were followed at the **scene**".

Accident news reports that display a high weight for this topic include 2017-0136, 2016-0119, and 2017-0048. The excerpts below from accident news report 2016-0119 (which has a weight of 0.968 for this topic) can serve as a representation of the content in this topic.

"At 12:15 p.m. Tuesday, the Dodge County Sheriff's Office responded to a report of a man who had fell 200 feet off of a wind turbine at N12048 West Line Road just outside the village of Brownsville. The initial call stated that the man, who was wearing a rescue device, fell fast from wind tower No. 38. Initial investigation showed the man fell 50 feet and was wearing a harness that slowed his fall, according to Capt. Trace Frost of the Dodge County Sheriff's Office. He was conscious and talking, and sustained injuries on his lower legs" [97].

6.5.10. Topic 10

"Typically assigned to news reporting **electricity generation** failures during **operation** of **installed** wind turbine **projects** that require **repair** or **replacement** of components such as **blades** by the **official company**. The news sometimes mentions the **time**frames (could be **weeks** or **months**) and the **cost** or amount (could be in **millions**) involved. Some of these incidents occurred at **schools** with turbines".

Accident news reports that display a high weight for this topic include 2018-0051, 2018-0068, and 2014-0214. The excerpts below from accident news report 2018-0051 (which has a weight of 0.994 for this topic) can serve as a representation of the content in this topic.

"The two wind turbines, overlooking the city from atop Haeckel Hill, will be decommissioned and the company is hoping to sell them off. "Turbines of this generation typically only last about 20 years. So they've reached pretty much their end of life," said Andrew Hall, CEO of Yukon Energy. The first turbine was installed in 1993, and the second, larger one went up in 2000. Hall says the larger one had a mechanical problem last year, and it was never fixed. "When we looked at the business case to repair it, based on the lifetime of the rest of the turbine, you know, the business case just wasn't there," . . . There are no immediate plans to replace the turbines, but Hall expects Yukon Energy's proposed Standing Offer Program—which would streamline the process for small, independent power producers to sell electricity to the grid—could change that" [98].

6.5.11. Cross-Tabulation Analysis of Topics

Topic modeling provides multiple results, including the topics, the weight of each word in each topic, the weight of each topic for each text document, and a single topic to which each document can be assigned. Having identified the topics, as a subsequent analysis to further characterize and profile the identified topics, the association between the assigned topics (Dataset D) and accident attributes of news (from Dataset A) can be

investigated. To this end, Dataset A, containing core attributes, was augmented by merging the AssignedTopic column from the topic modeling to create a new Dataset F to be analyzed with cross-tabulation, in which each row is a news report and each column is its attribute. The attributes in Dataset F are Filename, AssignedTopic, IsThereInjury, IsThereDeath, IsThereInjuryOrDeath, Country, Location1, OffshoreOnshore, and PhaseOfLifeCycle.

Table 4 displays the analysis of the assigned topics associated with the occurrence of deaths and injuries. Analyzing news within each topic shows, as highlighted with bold text, that Topic 9 is associated with a higher proportion of news related to deaths (38%) and injuries (52%), followed by Topic 1 (taking deaths and injuries together). Topics 4 and 2 also display a high proportion of news related to deaths and injuries compared to other news.

Table 4. Death and injury incidence in each topic.

	Injury or Death	No Injury or Death	Injury	No Injury	Uncategorized Injury Cases	Death	No Death
Topic 1	**12%**	**88%**	**3%**	**97%**	**0%**	**9%**	**91%**
Topic 2	**9%**	**91%**	**7%**	**91%**	**3%**	**3%**	**97%**
Topic 3	1%	99%	0%	100%	0%	1%	99%
Topic 4	**9%**	**91%**	**5%**	**95%**	**0%**	**5%**	**95%**
Topic 5	0%	100%	0%	100%	0%	0%	100%
Topic 6	2%	98%	2%	98%	0%	2%	98%
Topic 7	4%	96%	3%	97%	0%	1%	99%
Topic 8	0%	100%	0%	100%	0%	0%	100%
Topic 9	**86%**	**14%**	**52%**	**44%**	**4%**	**38%**	**62%**
Topic 10	0%	100%	0%	100%	0%	0%	100%

Table 5 displays the distribution of topics across the lifecycle phases of the wind turbine. Topic 9 is the most common during construction and maintenance, whereas topics 7, 10, 3, 5, and 4 are common during operation. Topic 2 was the most common one during transportation. Topic 9 is also the topic most associated with the (11) cases where the phase could not be classified.

Table 5. Topic incidence vs. life cycle phase (count).

	Construction	Maintenance	Operation	Transportation	Uncategorized Phase Cases	Total
Topic 1	4	2	18	9		33
Topic 2			5	**71**		76
Topic 3	3	6	**67**	1	1	78
Topic 4	11	5	**58**	3		77
Topic 5	3		**63**	2	1	69
Topic 6	13	1	28	6		48
Topic 7		7	**108**	1	1	117
Topic 8	2	1	44			47
Topic 9	**41**	**18**	5	12	8	84
Topic 10	3	7	81	1		92
Total	**80**	**47**	**477**	**106**	**11**	**721**

Table 6 displays the relationship between the topics and the location of the wind turbine (offshore vs. onshore). Topics 1, 4, and 9 were more associated with offshore turbines than the other topics. Although news related to onshore turbines can be assigned to any topic, they are most commonly assigned to topics 7, 10, 3, 2, and 9.

Table 7 displays topic distribution across countries with the highest number of accident news reports in the study. Topics in relation to countries yield multiple insights: for news from the USA, the most frequently assigned topics are Topics 10, 5, 7, and 2. For news from the UK, the most frequently assigned topics are Topics 4, 6, 8, 9, and 2. For Germany, Topic

3 is the most frequently assigned topic. For news from Australia, Topics 2, 9, 4, and 10 are the most frequent. Finally, for news from Canada, Topics 10 and 5 are more frequent.

Table 6. Topic incidence vs. location (count).

	Offshore	Onshore	Uncategorized Location Case	Total
Topic 1	**18**	14	1	33
Topic 2		**76**		76
Topic 3		**78**		78
Topic 4	**13**	64		77
Topic 5		69		69
Topic 6	2	46		48
Topic 7		**117**		117
Topic 8		47		47
Topic 9	**10**	**74**		84
Topic 10	3	**89**		92
Total	**46**	**674**	**1**	**721**

Table 7. Topic incidence vs. countries (percentage).

	Australia	Brazil	Canada	China	Denmark	France	Germany	India	Ireland	Turkey	UK	USA
Topic 1	5%	0%	2%	13%	9%	0%	6%	0%	7%	0%	7%	3%
Topic 2	**33%**	0%	8%	13%	0%	0%	9%	0%	7%	14%	**11%**	**13%**
Topic 3	0%	**29%**	0%	13%	**27%**	8%	**47%**	0%	0%	**29%**	0%	0%
Topic 4	**14%**	**43%**	10%	**25%**	**27%**	8%	5%	14%	**40%**	0%	**22%**	5%
Topic 5	0%	0%	**20%**	0%	0%	**17%**	0%	14%	0%	0%	4%	**18%**
Topic 6	0%	0%	16%	13%	0%	0%	0%	14%	20%	0%	**14%**	7%
Topic 7	10%	0%	4%	0%	18%	**42%**	**19%**	14%	0%	**43%**	9%	**17%**
Topic 8	5%	0%	4%	0%	0%	**17%**	3%	**29%**	0%	0%	**14%**	7%
Topic 9	**19%**	**29%**	10%	**25%**	9%	8%	12%	0%	**27%**	14%	**14%**	10%
Topic 10	**14%**	0%	**24%**	0%	9%	0%	0%	14%	0%	0%	5%	**22%**
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

6.6. Association Mining

Association mining was conducted on Dataset I using Borgelt's implementation of the Apriori algorithm [99]. Association rules were produced in the form of "IF antecedent, THEN consequent," i.e., A⇒B. The Apirori algorithm was ran to obtain association rules with at least 0.01% support and 20% confidence. Out of the 170,411 generated rules, 62 rules had "death" as the consequent and 109 rules had "injury" as the consequent. In other words, there were 62 rules in the form "IF antecedent THEN death" and 109 rules in the form "IF antecedent THEN injury".

These rules are displayed in Figures 10 and 11, respectively, where the x-axis denotes support, the y-axis denotes confidence, and the label denotes the antecedent. Figure 10 displays the association rules generated for death as a consequence, whereas Figure 11 displays the association rules generated for injury as a consequence.

According to Figure 10, terms with high y-axis values, such as "occup," "crush," "trap," "emploi," and "colleagu" were found to have a higher confidence with respect to having "death" as the consequent. Hence, if these terms occur in a document, there are higher chances that the news contains reporting of "death," compared to terms lower on the y-axis.

Similarly, Figure 11 shows that terms with high y-axis values, such as "airlift," "stabl," "helicopt," "crush," "rig," "serious" were found to have a higher confidence with respect to having "injury" as the consequent. Hence, if these terms occur in a document, there are higher chances that the news contains reporting of "injury," compared to terms lower on the y-axis.

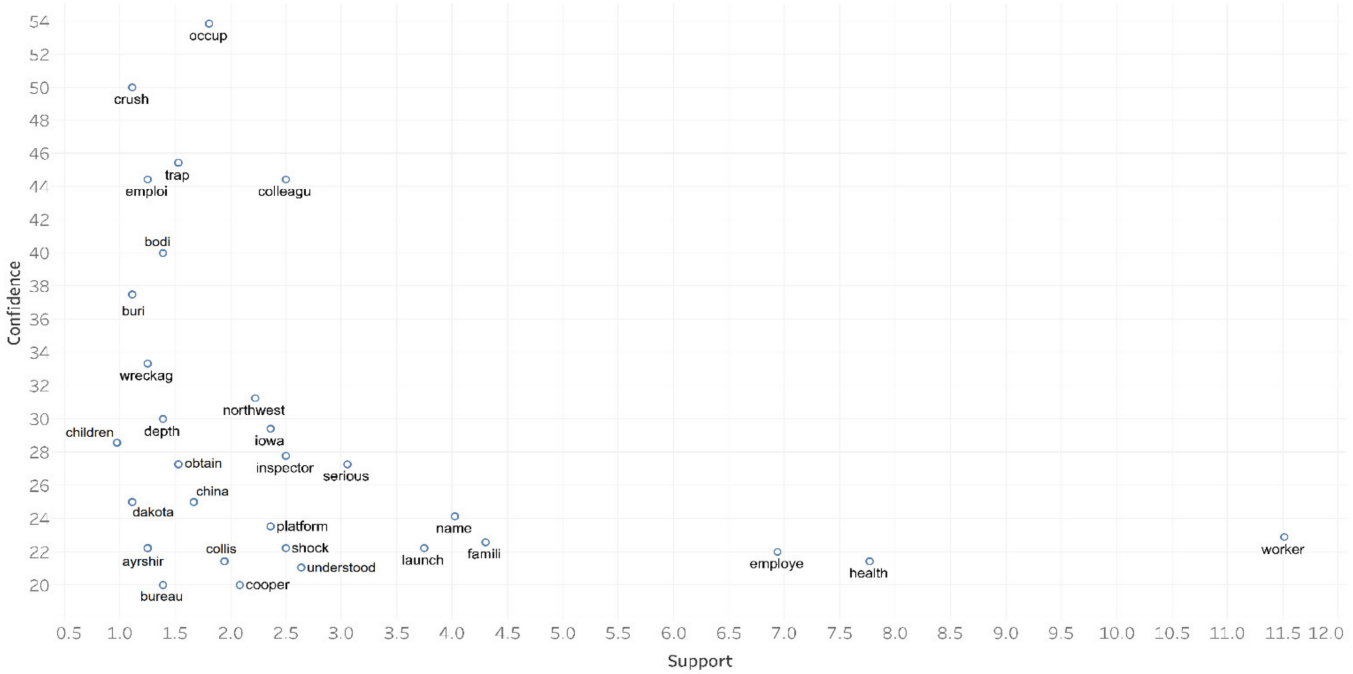

Figure 10. Association rules with death as consequent.

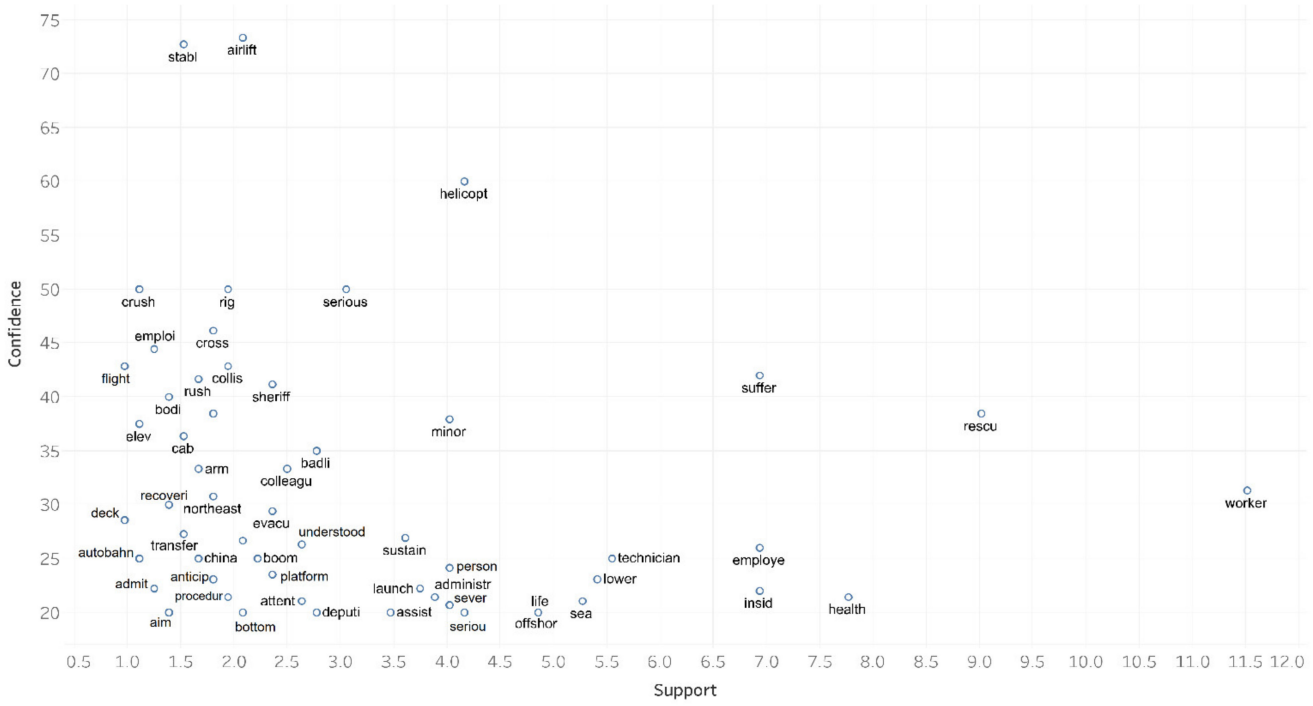

Figure 11. Association rules with injury as consequent.

Higher support of an association rule indicates that the terms in the rule appear together more frequently in the text corpus. For example, in Figure 10, terms such as "worker," "health," and "employe" appear frequently in the corpus with "death," even though the probability of these terms having death as a consequence (confidence, shown on y-axis) is lower. Similarly, in Figure 11, terms such as "health," "employe," and "insid" appear together relatively frequently in the corpus with "injury," even though the

probability of these terms having injury as a consequence (confidence, shown on y-axis) is lower.

In Figure 11, it is important to note the terms "worker," "rescu," and "suffer," which have both very high support values (x-axis), meaning that they appear very frequently together with "injury". Furthermore, the rules (IF "worker" THEN "injury"), (IF "rescu" THEN "injury"), and (IF "suffer" THEN "injury") have relatively high confidence values (y-axis).

6.7. Decision Tree Analysis

Decision tree analysis was conducted to identify the terms that would be associated with and contribute the most to predicting death or injury. The terms identified in this analysis were not insightful.

6.8. Predictive Analytics

In our research, predictive analytics did not yield significant or interesting results.

7. Conclusions and Future Work

In this study, we developed a text analytics framework for topic modeling of wind turbine accidents (Figure 2), which is novel for the literature on wind energy research. The applicability of the developed framework is demonstrated through the analysis of an extensive text collection (corpus) of wind turbine accident news between the years 2010 and 2019, using the text dataset collected during the research process.

There were two main types of data which were analyzed: Dataset A is a tabular dataset of core metadata attributes, where each row is a news report, and each column is an attribute of that news report. Dataset B is the text collection (corpus) cleaned during the process. Novel insights for the wind energy sector were generated in the research by applying a portfolio of data analytics techniques (Figure 2), including visual analysis, cross-tabulation, text analytics, and machine learning (ML). There were two text analytics techniques, which are also unsupervised ML techniques, namely topic modeling and association mining, which were applied for the first time in the wind turbine accident literature. The topic modeling results (Figure 6) revealed recurring themes and patterns in wind turbine accident news reports. The association mining results (Figures 10 and 11) revealed the support and confidence of rules in the form "IF antecedent THEN consequent". The application of these methods for the domain, which revealed hidden patterns and yielded fresh insights, can be used by stakeholders in the wind energy sector to make data-driven plans and decisions.

The present research can be linked to the Penta-helix model of innovation [100,101]. Evolving from the Triple- and Quadruple-helix models, the Penta-helix model is a framework used to understand how stakeholders from the government, private businesses, academic universities, society, and social entrepreneurs meet and interact in projects. Different stakeholders represent different perspectives, and the knowledge flows and innovations brought forth depend on the interactions between these stakeholders.

The wind energy sector can also be analyzed within the Penta-helix framework, with each stakeholder driven by diverse interests. While all the stakeholders have robust reasons to drive the adoption of green energy, society and social entrepreneurs focus more on the problems associated with green energy, including wind energy. The present research uses data analytics to examine the root causes of injuries and deaths related to wind energy, which can create a platform where fact-based discussions and interactions can be facilitated, and the different stakeholders can be supported in making fact-based informed decisions.

For future research, additional insights can be generated by applying the same framework to a more extensive collection of news. Similar to analyzing accident news, accident reports can also be analyzed using an approach similar to that used for accident news. As pointed out in [5], other research possibilities include developing machine learning

methodologies to automatically detect news reports that mention death and injury, as well as data mining mechanisms to automatically aggregate data.

As future research with respect to topic modeling, the method can be applied with different parameter values and value combinations, which may yield further insights. For example, instead of selecting 10 as the number of topics, as in the current research, another value can be selected and applied. Even better, outputs resulting from different values of this parameter can be compared, and more holistic insights can be gained.

Several preventive actions can be adopted by stakeholders to reduce the occurrence and negative outcomes of wind turbine accidents. Our research can be used as a reference to guide which preventive actions could be applied under which settings (country and location, etc.). Training programs (initial and refresher) administered at appropriate time intervals are recommended by [102] to manage technicians' skill decay. Furthermore, the sector may see more extensive use of robots to replace human labor, especially in the riskiest environmental settings identified in this research. A regular maintenance schedule incorporating suitable mechanisms to improve wind turbine reliability [103] will be beneficial in reducing accident frequencies.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/su132212757/s1, Figure S1: Distribution of accident news within the top states of the USA, Figure S2: Association between the location of the wind turbine (offshore vs. onshore) and frequency of accidents in the top twelve countries, Figure S3: Decision tree graph for predicting the occurrence of death, Figure S4: Decision tree graph for predicting the occurrence of injury, Figure S5: KNIME workflow for text processing, Figure S6: KNIME workflow for topic modeling, Figure S7: KNIME workflow for decision tree analysis, Figure S8: KNIME workflow for decision tree-based prediction modeling on Dataset B (Termset C2), using Dataset H, Figure S9: KNIME workflow for decision tree-based prediction modeling on Dataset A, Figure S10: KNIME workflow for SVM-based prediction modeling on Dataset B (Termset C2), using Dataset H, Figure S11: KNIME workflows for SVM-based prediction modeling on Dataset A, Figure S12: Screenshot of Dataset A, Figure S13: Screenshot of Dataset B, Figure S14: Screenshot of Dataset C1, Figure S15: Screenshot of Dataset C2, Figure S16: Screenshot of Dataset D, Figure S17: Screenshot of Dataset E, Figure S18: Screenshot of Dataset F, Figure S19: Screenshot of Dataset G, Figure S20: Screenshot of Dataset H, Figure S21: Screenshot of Dataset I, Table S1: Death and injury incidence in each topic (counts), TableS2: Topic incidence vs. countries (counts), Table S3: Topic incidence in the US states (counts), Video S1: title, and Video S1: Components of a wind turbine with three blades, which is a popular turbine design (https://bit.ly/3zphcUc, accessed on 7 October 2021).

Author Contributions: Conceptualization, G.E.; data curation, L.K.; formal analysis, L.K.; funding acquisition, G.E.; investigation, G.E. and L.K.; methodology, G.E. and L.K.; project administration, G.E.; resources, G.E.; software, G.E.; supervision, G.E.; validation, G.E. and L.K.; visualization, G.E. and L.K.; writing—original draft, G.E. and L.K.; writing—review and editing, G.E. and L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the United Arab Emirates University, Startup Grant Funding 31B127.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. The data used in this research will be uploaded at https://ertekprojects.com/wind-turbine-accidents/, accessed on 7 October 2021. The screenshots of the source web pages, which serve as evidence of originality, are not publicly available because of copyright concerns, but can be requested from the corresponding author.

Acknowledgments: This research was based on the Startup Grant Funding from the United Arab Emirates University. The authors also thank Eda Ülgen for her earlier help in data collection and to Dilara Erdoğan for her design of the wind turbine animation and illustration.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Overview, World Energy Outlook 2020, International Energy Agency. October 2020. Available online: https://www.iea.org/reports/world-energy-outlook-2020 (accessed on 28 September 2021).
2. Renewable Energy Explained, US Energy Information Administration (EIA). 22 June 2020. Available online: https://www.eia.gov/energyexplained/renewable-sources/ (accessed on 28 September 2021).
3. Renewables, World Energy Outlook 2019, International Energy Agency. November 2019. Available online: https://www.iea.org/reports/world-energy-outlook-2019/renewables#abstract (accessed on 28 September 2021).
4. Wind Powers America Amended Annual Report 2019, American Wind Energy Association (AWEA). 2019. Available online: https://www.awea.org/resources/publications-and-reports/market-reports/2019-u-s-wind-industry-market-reports/amr2019_public (accessed on 1 October 2020).
5. Ertek, G.; Chi, X.; Zhang, A.N.; Asian, S. Text Mining Analysis of Wind Turbine Accidents: An Ontology-Based Framework. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 3233–3241. [CrossRef]
6. Ertek, G.; Tunç, M.M.; Kurtaraner, E.; Kebude, D. Insights into the efficiencies of on-shore wind turbines: A data-centric analysis. In Proceedings of the 2012 International Symposium on Innovations in Intelligent Systems and Applications, Trabzon, Turkey, 2–4 July 2012; pp. 1–7. [CrossRef]
7. The Inside of a Wind Turbine, US Department of Energy. Available online: https://www.energy.gov/eere/wind/inside-wind-turbine (accessed on 26 November 2020).
8. Asian, S.; Ertek, G.; Haksoz, C.; Pakter, S.; Ulun, S. Wind Turbine Accidents: A Data Mining Study. *IEEE Syst. J.* **2017**, *11*, 1567–1578. [CrossRef]
9. Chou, J.-S.; Chiu, C.-K.; Huang, I.-K.; Chi, K.-N. Failure Analysis of Wind Turbine Blade under Critical Wind Loads. *Eng. Fail. Anal.* **2013**, *27*, 99–118. [CrossRef]
10. Uchida, T. Numerical Investigation of Terrain-Induced Turbulence in Complex Terrain by Large-Eddy Simulation (LES) Technique. *Energies* **2018**, *11*, 2638. [CrossRef]
11. Okawa, M.; Fujiyama, C.; Koda, Y.; Kado, M. Damage Investigation of Existing Onshore Wind Turbine Foundation. In Proceedings of the Grand Renewable Energy Organized by Japan Council for Renewable Energy, Yokohama, Japan, 17–22 June 2018; p. 133. [CrossRef]
12. Uchida, T.; Takakuwa, S. A Large-Eddy Simulation-Based Assessment of the Risk of Wind Turbine Failures Due to Terrain-Induced Turbulence over a Wind Farm in Complex Terrain. *Energies* **2019**, *12*, 1925. [CrossRef]
13. Yamamoto, K.; Sumi, S. Validations of Lightning Protections for Accidents at a Wind Farm. In Proceedings of the 2013 International Symposium on Lightning Protection (XII SIPDA), Belo Horizonte, Brazil, 7–11 October 2013; pp. 329–334. [CrossRef]
14. Uchida, T. Computational Fluid Dynamics (CFD) Investigation of Wind Turbine Nacelle Separation Accident over Complex Terrain in Japan. *Energies* **2018**, *11*, 1485. [CrossRef]
15. Chou, J.-S.; Tu, W.-T. Failure Analysis and Risk Management of a Collapsed Large Wind Turbine Tower. *Eng. Fail. Anal.* **2011**, *18*, 295–313. [CrossRef]
16. Liu, Y.; Li, Y.; Chi, Y.; Wang, Y. Analysis on a large scale wind turbines cascading trip-off accident in North China. In Proceedings of the 2013 IEEE Grenoble Conference, Grenoble, France, 16–20 June 2013; pp. 1–6. [CrossRef]
17. Sun, H.; Zhang, Z.; Lin, W.; Tang, Y.; Luo, X.; Wang, A. Analysis on serious wind turbine generators tripping accident in Northwest China power grid in 2011 and its lessons. *Power Syst. Technol.* **2012**, *36*, 76–80.
18. Dethleff, D.; Weinrich, N.; Seide, K.; Jürgens, C.; Stuhr, M. Accidental fall from an onshore wind turbine: Implications for offshore rescue using the example of a case study. *Notf. Rett.* **2016**, *19*, 591–597. [CrossRef]
19. Bielecki, A.; Ernst, S.; Skrodzka, W.; Wojnicki, I. The Externalities of Energy Production in the Context of Development of Clean Energy Generation. *Environ. Sci. Pollut. Res.* **2020**, *27*, 11506–11530. [CrossRef]
20. Lee, S.H.; Kang, H.G. Integrated Societal Risk Assessment Framework for Nuclear Power and Renewable Energy Sources. *Nucl. Eng. Technol.* **2015**, *47*, 461–471. [CrossRef]
21. Astiaso Garcia, D.; Bruschi, D. A Risk Assessment Tool for Improving Safety Standards and Emergency Management in Italian Onshore Wind Farms. *Sustain. Energy Technol. Assess.* **2016**, *18*, 48–58. [CrossRef]
22. Gul, M.; Guneri, A.F.; Baskan, M. An Occupational Risk Assessment Approach for Construction and Operation Period of Wind Turbines. *Glob. J. Environ. Sci. Manag.* **2018**, *4*, 281–298. [CrossRef]
23. Sovacool, B.K.; Andersen, R.; Sorensen, S.; Sorensen, K.; Tienda, V.; Vainorius, A.; Schirach, O.M.; Bjørn-Thygesen, F. Balancing Safety with Sustainability: Assessing the Risk of Accidents for Modern Low-Carbon Energy Systems. *J. Clean. Prod.* **2016**, *112*, 3952–3965. [CrossRef]
24. Sovacool, B.K.; Kryman, M.; Laine, E. Profiling Technological Failure and Disaster in the Energy Sector: A Comparative Analysis of Historical Energy Accidents. *Energy* **2015**, *90*, 2016–2027. [CrossRef]

25. Kim, J.; Ryu, D.; Sovacool, B.K. Critically Assessing and Projecting the Frequency, Severity, and Cost of Major Energy Accidents. *Extr. Ind. Soc.* **2021**, *8*, 100885. [CrossRef]
26. International Statistical Analysis on Wind Turbine Failures, 67nd IEA Topical Expert Meeting, CENER, International Energy Agency, March 2011. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.641.8834&rep=rep1&type=pdf (accessed on 28 September 2021).
27. Report on Wind Turbine Subsystem Reliability—A Survey of Various Databases, National Renewable Energy Laboratory, U.S. Department of Energy. June 2013. Available online: https://www.nrel.gov/docs/fy13osti/59111.pdf (accessed on 28 September 2021).
28. Moura Carneiro, F.O.; Barbosa Rocha, H.H.; Costa Rocha, P.A. Investigation of Possible Societal Risk Associated with Wind Power Generation Systems. *Renew. Sustain. Energy Rev.* **2013**, *19*, 30–36. [CrossRef]
29. Brouwer, S.R.; Al-Jibouri, S.H.S.; Cárdenas, I.C.; Halman, J.I.M. Towards Analysing Risks to Public Safety from Wind Turbines. *Reliab. Eng. Syst. Saf.* **2018**, *180*, 77–87. [CrossRef]
30. Kim, G.-S.; Jeong, J.-H. A Classification on the Causes of Wind Turbine Accidents. *J. Korea Soc. Power Syst. Eng.* **2015**, *19*, 76–81. [CrossRef]
31. Santos, F.P.; Teixeira, A.P.; Guedes Soares, C. Review of wind turbine accident and failure data. In *Renewable Energies Offshore*, 1st ed.; Guedes Soares, C., Ed.; Taylor & Francis Group: London, UK, 2015; pp. 953–959.
32. Konstantinidis, E.I.; Botsaris, P.N. Wind Turbines: Current Status, Obstacles, Trends and Technologies. *IOP Conf. Ser. Mater. Sci. Eng.* **2016**, *161*, 012079. [CrossRef]
33. Summary of Wind Turbine Accident Data to 30 September 2020, Caithness Windfarm Information Forum. 2020. Available online: http://www.caithnesswindfarms.co.uk/AccidentStatistics.htm (accessed on 28 September 2021).
34. Ma, Y.; Martinez-Vazquez, P.; Baniotopoulos, C. Wind Turbine Tower Collapse Cases: A Historical Overview. *Proc. Inst. Civ. Eng.-Struct. Build.* **2019**, *172*, 547–555. [CrossRef]
35. Chou, J.-S.; Ou, Y.-C.; Lin, K.-Y. Collapse Mechanism and Risk Management of Wind Turbine Tower in Strong Wind. *J. Wind. Eng. Ind. Aerodyn.* **2019**, *193*, 103962. [CrossRef]
36. Uadiale, S.; Urbán, E.; Carvel, R.; Lange, D.; Rein, G. Overview of Problems and Solutions in Fire Protection Engineering of Wind Turbines. *Fire Saf. Sci.* **2014**, *11*, 983–995. [CrossRef]
37. Pfaffel, S.; Faulstich, S.; Rohrig, K. Performance and Reliability of Wind Turbines: A Review. *Energies* **2017**, *10*, 1904. [CrossRef]
38. Lian, J.; Cai, O.; Dong, X.; Jiang, Q.; Zhao, Y. Health Monitoring and Safety Evaluation of the Offshore Wind Turbine Structure: A Review and Discussion of Future Development. *Sustainability* **2019**, *11*, 494. [CrossRef]
39. Shafiee, M.; Sørensen, J.D. Maintenance Optimization and Inspection Planning of Wind Energy Assets: Models, Methods and Strategies. *Reliab. Eng. Syst. Saf.* **2019**, *192*, 105993. [CrossRef]
40. Mentes, A.; Turan, O. A New Resilient Risk Management Model for Offshore Wind Turbine Maintenance. *Saf. Sci.* **2019**, *119*, 360–374. [CrossRef]
41. Gao, Z.; Liu, X. An Overview on Fault Diagnosis, Prognosis and Resilient Control for Wind Turbine Systems. *Processes* **2021**, *9*, 300. [CrossRef]
42. Shohag, M.A.S.; Hammel, E.C.; Olawale, D.O.; Okoli, O.I. Damage Mitigation Techniques in Wind Turbine Blades: A Review. *Wind. Eng.* **2017**, *41*, 185–210. [CrossRef]
43. Yamamoto, K.; Honjo, N. Latest Trends in Technologies for Sound Operation of Wind Turbines against Lightning. *Electr. Eng. Jpn.* **2018**, *205*, 3–7. [CrossRef]
44. Ye, X.; Lu, Z.; Qiao, Y.; Li, J.; Wang, F.; Luo, W. A Primary Analysis on Mechanism of Large Scale Cascading Trip-off of Wind Turbine Generators. *Autom. Electr. Power Syst.* **2012**, *36*, 11–17. [CrossRef]
45. Rastayesh, S.; Long, L.; Dalsgaard Sørensen, J.; Thöns, S. Risk Assessment and Value of Action Analysis for Icing Conditions of Wind Turbines Close to Highways. *Energies* **2019**, *12*, 2653. [CrossRef]
46. Kotb, M.; Elhelloty, A.; Shaaba, M. A Methodology to The Safe Operation of Mobile Crane in Erection of Wind Turbine. *Life Sci. J.* **2019**, *16*, 6–10. [CrossRef]
47. Bai, X.; Sun, L.; Sun, H. Risk Assessment of Hoisting Aboard and Installation for Offshore Wind Turbine. In Proceedings of the ASME 2012 31st International Conference on Ocean, Offshore and Arctic Engineering, Rio de Janeiro, Brazil, 1–6 July 2012; pp. 107–114. [CrossRef]
48. Presencia, C.E.; Shafiee, M. Risk Analysis of Maintenance Ship Collisions with Offshore Wind Turbines. *Int. J. Sustain. Energy* **2018**, *37*, 576–596. [CrossRef]
49. Kusiak, A.; Li, W. The Prediction and Diagnosis of Wind Turbine Faults. *Renew. Energy* **2011**, *36*, 16–23. [CrossRef]
50. Wang, L.; Zhang, Z.; Long, H.; Xu, J.; Liu, R. Wind Turbine Gearbox Failure Identification with Deep Neural Networks. *IEEE Trans. Ind. Inf.* **2017**, *13*, 1360–1368. [CrossRef]
51. Blanco, A.; Marti-Puig, P.; Gibert, K.; Cusidó, J.; Solé-Casals, J. A Text-Mining Approach to Assess the Failure Condition of Wind Turbines Using Maintenance Service History. *Energies* **2019**, *12*, 1982. [CrossRef]
52. Stetco, A.; Dinmohammadi, F.; Zhao, X.; Robu, V.; Flynn, D.; Barnes, M.; Keane, J.; Nenadic, G. Machine Learning Methods for Wind Turbine Condition Monitoring: A Review. *Renew. Energy* **2019**, *133*, 620–635. [CrossRef]
53. Saufi, S.R.; Ahmad, Z.A.B.; Leong, M.S.; Lim, M.H. Gearbox Fault Diagnosis Using a Deep Learning Model with Limited Data Sample. *IEEE Trans. Ind. Inf.* **2020**, *16*, 6263–6271. [CrossRef]

54. Chen, P.; Li, Y.; Wang, K.; Zuo, M.J.; Heyns, P.S.; Baggeröhr, S. A Threshold Self-Setting Condition Monitoring Scheme for Wind Turbine Generator Bearings Based on Deep Convolutional Generative Adversarial Networks. *Measurement* **2021**, *167*, 108234. [CrossRef]
55. Chen, W.; Qiu, Y.; Feng, Y.; Li, Y.; Kusiak, A. Diagnosis of Wind Turbine Faults with Transfer Learning Algorithms. *Renew. Energy* **2021**, *163*, 2053–2067. [CrossRef]
56. Colak, I.; Sagiroglu, S.; Yesilbudak, M. Data Mining and Wind Power Prediction: A Literature Review. *Renew. Energy* **2012**, *46*, 241–247. [CrossRef]
57. Ozkan, M.; Karagoz, P. A Novel Wind Power Forecast Model: Statistical Hybrid Wind Power Forecast Technique (SHWIP). *IEEE Trans. Ind. Inf.* **2015**, *11*, 375–387. [CrossRef]
58. Haghi, H.V.; Lotfifard, S.; Qu, Z. Multivariate Predictive Analytics of Wind Power Data for Robust Control of Energy Storage. *IEEE Trans. Ind. Inf.* **2016**, *12*, 1350–1360. [CrossRef]
59. Freiberg, A.; Schefter, C.; Girbig, M.; Murta, V.C.; Seidler, A. Health Effects of Wind Turbines on Humans in Residential Settings: Results of a Scoping Review. *Environ. Res.* **2019**, *169*, 446–463. [CrossRef]
60. Demiriz, A.; Ertek, G.; Atan, T.; Kula, U. Re-Mining Item Associations: Methodology and a Case Study in Apparel Retailing. *Decis. Support Syst.* **2011**, *52*, 284–293. [CrossRef]
61. Ertek, G.; Tunc, M.M. Re-Mining Association Mining Results Through Visualization, Data Envelopment Analysis, and Decision Trees. In *Computational Intelligence Systems in Industrial Engineering*; Kahraman, C., Ed.; Atlantis Press: Paris, France, 2012; Volume 6, pp. 601–622.
62. Abdallah, S.; Malik, M.; Ertek, G. A Data Mining Framework for the Analysis of Patient Arrivals into Healthcare Centers. In *Proceedings of the 2017 International Conference on Information Technology—ICIT 2017, Singapore, 27–29 December 2017*; ACM Press: Singapore; pp. 52–61. [CrossRef]
63. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; Adaptive Computation and Machine Learning Series; MIT Press: Cambridge, MA, USA, 2012; ISBN 9780262018258.
64. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; Adaptive Computation and Machine Learning Series; The MIT Press: Cambridge, MA, USA, 2018; ISBN 9780262039246.
65. Muguro, J.K.; Sasaki, M.; Matsushita, K.; Njeri, W. Trend Analysis and Fatality Causes in Kenyan Roads: A Review of Road Traffic Accident Data between 2015 and 2020. *Cogent Eng.* **2020**, *7*, 1797981. [CrossRef]
66. John, M.; Marbach, E.; Lohmann, S.; Heimerl, F.; Ertl, T. MultiCloud: Interactive Word Cloud Visualization for the Analysis of Multiple Texts. In Proceedings of the Graphics Interface 2018, Toronto, ON, Canada, 8–11 May 2018; p. 8. [CrossRef]
67. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
68. Roque, C.; Lourenço Cardoso, J.; Connell, T.; Schermers, G.; Weber, R. Topic Analysis of Road Safety Inspections Using Latent Dirichlet Allocation: A Case Study of Roadside Safety in Irish Main Roads. *Accid. Anal. Prev.* **2019**, *131*, 336–349. [CrossRef]
69. Zhong, B.; Pan, X.; Love, P.E.D.; Ding, L.; Fang, W. Deep Learning and Network Analysis: Classifying and Visualizing Accident Narratives in Construction. *Autom. Constr.* **2020**, *113*, 103089. [CrossRef]
70. Zhong, B.; Pan, X.; Love, P.E.D.; Sun, J.; Tao, C. Hazard Analysis: A Deep Learning and Text Mining Framework for Accident Prevention. *Adv. Eng. Inform.* **2020**, *46*, 101152. [CrossRef]
71. Kwayu, K.M.; Kwigizile, V.; Lee, K.; Oh, J.-S. Discovering Latent Themes in Traffic Fatal Crash Narratives Using Text Mining Analytics and Network Topology. *Accid. Anal. Prev.* **2021**, *150*, 105899. [CrossRef]
72. Ali, F.; Ali, A.; Imran, M.; Naqvi, R.A.; Siddiqi, M.H.; Kwak, K.-S. Traffic Accident Detection and Condition Analysis Based on Social Networking Data. *Accid. Anal. Prev.* **2021**, *151*, 105973. [CrossRef]
73. Agrawal, R.; Imieliński, T.; Swami, A. Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data—SIGMOD' 93, Washington, DC, USA, 25–28 May 1993; pp. 207–216. [CrossRef]
74. Gupta, K.; Potika, K. Fake News Analysis and Graph Classification on a COVID-19 Twitter Dataset. In Proceedings of the 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, UK, 23–26 August 2021; pp. 60–68.
75. Mishra, A.; Vishwakarma, S. Analysis of TF-IDF Model and Its Variant for Document Retrieval. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015; pp. 772–776. [CrossRef]
76. Rivera-Trigueros, I. Machine Translation Systems and Quality Assessment: A Systematic Review. *Lang. Resour. Eval.* **2021**, *53*, 57–86. [CrossRef]
77. Chan, C.-H.; Zeng, J.; Wessler, H.; Jungblut, M.; Welbers, K.; Bajjalieh, J.W.; van Atteveldt, W.; Althaus, S.L. Reproducible Extraction of Cross-Lingual Topics (Rectr). *Commun. Methods Meas.* **2020**, *14*, 285–305. [CrossRef]
78. Hidalgo-Ternero, C.M. Google Translate vs. DeepL: Analysing neural machine translation performance under the challenge of phraseological variation. In *Análisis Multidisciplinar del Fenómeno de la Variación Fraseológica en Traducción e Interpretación /Multidisciplinary Analysis of the Phenomenon of Phraseological Variation in Translation and Interpreting*; Mogorrón Huerta, P., Ed.; Universitat Jaume I: Castellón de la Plana, Spain, 2020; Volume 6, pp. 154–177.

79. Akyürek, A.F.; Guo, L.; Elanwar, R.; Ishwar, P.; Betke, M.; Wijaya, D.T. Multi-Label and Multilingual News Framing Analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020*; Association for Computational Linguistics: Stroudsburg, PA, USA; pp. 8614–8624.
80. Banitz, B. Machine Translation: A Critical Look at the Performance of Rule-Based and Statistical Machine Translation. *Cad. Tradução* **2020**, *40*, 54–71. [CrossRef]
81. Ivie, P.; Thain, D. Reproducibility in Scientific Computing. *ACM Comput. Surv.* **2018**, *51*, 63. [CrossRef]
82. Yuan, C.; Yang, H. Research on K-Value Selection Method of K-Means Clustering Algorithm. *J* **2019**, *2*, 226–235. [CrossRef]
83. Liu, B. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [CrossRef]
84. Kherwa, P.; Bansal, P. Topic Modeling: A Comprehensive Review EAI Endorsed Transactions on Scalable Information Systems. *EAI Endorsed Trans. Scalable Inf. Syst.* **2019**, *7*, e2.
85. Newman, D.; Asuncion, A.; Smyth, P.; Welling, M. Distributed Algorithms for Topic Models. *J. Mach. Learn. Res.* **2009**, *10*, 1801–1828.
86. Dunteman, G.H.; Ho, M.H.R.; Ho, M.H.R. Maximum Likelihood Estimation—SAGE Research Methods. 2006, Volume 145. Available online: https://methods.sagepub.com/book/an-introduction-to-generalized-linear-models/n5.xml (accessed on 25 October 2021).
87. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605. Available online: https://www.jmlr.org/papers/v9/vandermaaten08a.html (accessed on 25 October 2021).
88. Silipo, R. Is Zero Closer to Eight or to One? Available online: https://towardsdatascience.com/is-zero-closer-to-eight-or-to-one-c6392242b696 (accessed on 25 October 2021).
89. Power Restored to Maui Customers. 22 October 2018. Available online: https://www.mauinews.com/news/local-news/2018/10/power-restored-to-maui-customers/ (accessed on 10 June 2021).
90. Wind Turbine Lorry Crash Blocks A713 Near Parton. 7 February 2017. Available online: https://www.bbc.com/news/uk-scotland-south-scotland-38892444 (accessed on 10 June 2021).
91. Windrad Verliert Flügelspitze. 24 December 2013. Available online: https://www.lr-online.de/lausitz/luebbenau/windrad-verliert-fluegelspitze-34145302.html (accessed on 10 June 2021).
92. Vestas Investigates Blade Failure at Danish Wind Farm. 29 January 2014. Available online: http://nawindpower.com/vestas-investigates-blade-failure-at-danish-wind-farm (accessed on 10 June 2021).
93. Damaged Pubnico Point Wind Turbine: "Like Thunder Out of the Clear Blue Sky". 20 March 2016. Available online: http://www.thevanguard.ca/News/Local/2016-03-20/article-4472202/Damaged-wind-turbine%3A-Like-thunder-out-of-the-clear-blue-sky-says-Pubnico-Point-resident-describing-sound/1 (accessed on 10 June 2021).
94. Farmland Owners Sue Wind Farm Company. 27 June 2016. Available online: http://www.news-gazette.com/news/local/2016-06-27/farmland-owners-sue-wind-farm-company.html (accessed on 10 June 2021).
95. Wind Turbine Fire Breaks Out at Winery: Cops. 8 June 2019. Available online: https://patch.com/new-york/northfork/wind-turbine-fire-breaks-out-winery-cops (accessed on 10 June 2021).
96. Lightning Apparently Damages Windmill Blade. 4 September 2019. Available online: https://www.wwnytv.com/2019/09/04/lightning-apparently-damages-windmill-blade/ (accessed on 10 June 2021).
97. Man Falls from Wind Tower in Dodge County. 1 November 2016. Available online: https://www.fdlreporter.com/story/news/local/2016/11/01/man-falls-wind-tower-dodge-county/93123732/ (accessed on 10 June 2021).
98. Whitehorse's Wind Turbines Are Pretty Much Blown. 16 August 2018. Available online: https://www.cbc.ca/news/canada/north/whitehorse-wind-turbines-haeckel-hill-decommission-1.4786508 (accessed on 10 June 2021).
99. Borgelt, C. Frequent Item Set Mining. *WIREs Data Min. Knowl. Discov.* **2012**, *2*, 437–456. [CrossRef]
100. Calzada, I. Democratising Smart Cities? Penta-Helix Multistakeholder Social Innovation Framework. *Smart Cities* **2020**, *3*, 1145–1172. [CrossRef]
101. Caraka, R.E.; Noh, M.; Chen, R.C.; Lee, Y.; Gio, P.U.; Pardamean, B. Connecting Climate and Communicable Disease to Penta Helix Using Hierarchical Likelihood Structural Equation Modelling. *Symmetry* **2021**, *13*, 657. [CrossRef]
102. Lawani, K.; Hare, B.; Cameron, I. Scheduling Refresher-Based Rescue and Evacuation Training for Wind Turbine Technicians. In Proceedings of the The Canadian Society for Civil Engineering 5th International/11th Construction Specialty Conference, Vancouver, BC, Canada, 7–10 June 2015. [CrossRef]
103. Wind Turbine Generator Maintenance: What to Expect and Why. *POWER Magazine*, 1 December 2014. Available online: https://www.powermag.com/wind-turbine-generator-maintenance-what-to-expect-and-why/(accessed on 25 May 2021).