Wind turbine accidents: A data mining study

Sobhan Asian, Gurdal Ertek, Cagri Haksoz, Sena Pakter, Soner Ulun

Wind Turbine Accidents: A Data Mining Study

Sobhan Asian, Member, Gurdal Ertek, Cagri Haksoz, Sena Pakter, Soner Ulun

Abstract— While the global production of wind energy is increasing, there exists a significant gap in the academic and practice literature regarding the analysis of wind turbine accidents. Our paper presents the results obtained from the analysis of 240 wind turbine accidents from around the world. The main focus of our paper is revealing the associations between several factors and deaths and injuries in wind turbine accidents. Specifically, the associations of death and injuries with the stage of the wind turbine's life cycle (transportation, construction, operation, and maintenance) and the main cause factor categories (human, system/equipment, and nature) were studied. To this end, we conducted a detailed investigation that integrates exploratory and statistical data analysis and data mining methods. The paper presents a multitude of insights regarding the accidents and discusses implications for wind turbine manufacturers, engineering and insurance companies, and government organizations.

Index Terms—Wind energy, Wind power generation, Accidents, Data Mining, Data analysis.

I. INTRODUCTION

The world demand for energy is expected to grow by more than two-thirds over the period 2011-2035 [1]. This demand will be met by a combination of nonrenewable (coal, fossil fuel, nuclear) and renewable (wind power, hydropower, solar energy, biomass, biofuel, geothermal) energy sources. The share of renewable energy sources in total power generation is expected to rise from 20% in 2011 to 31% in 2035, and renewables are expected to eventually surpass gas and coal and become the primary energy source in the world [1]. This global trend for the increasing usage of renewable energy is motivated mainly by the undesired global climate change due to carbon emissions as well as the depletion of fossil fuels. Furthermore, perceived notion of sustainability of renewable energy sources is driving governments to introduce legislations that promote use of renewable energy [2].

Wind energy has a long history [3], and is currently among the leading sources of renewable energy in terms of production capacity [4]. According to 2013 market statistics released by The Global Wind Energy Council (GWEC), the cumulative global wind energy capacity more than tripled in 6 years [5]. The cumulative installed wind energy capacity in the USA has increased more than 22-fold between 2000 and 2012 [6].

While wind energy industry and the installation of wind turbines are growing, the drawbacks of wind energy are not always considered and evaluated. One particular problem with wind energy is wind turbine accidents. Wind turbine accidents include a multitude of ways in which wind turbines fail due to mechanical problems, nature, or humans. In this paper we use the term "wind turbine accident" to

Sobhan Asian is with School of Business IT and Logistics, College of Business, RMIT University, 124 La Trobe Street, Melbourne, VIC 3000, Australia. (e-mail: sobhan.asian@rmit.edu.au).
Gurdal Ertek is with the Rochester Institute of Technology - Dubai, Dubai Silicon Oasis, Dubai, UAE (e-mail: gurdalertek@gmail.com).
Cagri Haksoz is with the School of Management, Sabanci University, 34956, Turkey (e-mail: cagrihaksoz@sabanciuniv.edu).
Sena Pakter is with Garanti Bank, Levent Nispetiye Mah. Aytar Cad. No:2 Beşiktaş 34340, Istanbul, Turkey (e-mail: senapakter@gmail.com).
Soner Ulun is with the Nanyang Technological University, 639798, Singapore (e-mail: soner001@ ntu.edu.sg).

describe any event involving a commercial wind turbine that was sufficiently noteworthy that it was reported in the public news media; it includes events where there was an injury or fatality, or where the wind turbine suffered significant damage, or both. Our literature review show that while there are several academic studies that primarily focus on the mechanical aspects of wind turbine accidents, the literature has fallen short of systematically analyzing wind turbine accidents (except report [7]).

There is a significant gap of knowledge and insights throughout the world with regards to wind turbine accidents. Specifically, there does not seem to exist any research that investigates the wind turbine accidents throughout the world and associates these accidents with cause factors and the stage of the wind turbine's life cycle. Investigating these two specific types of associations constitute the focus of our paper. Our main motivation to conduct a comprehensive analysis of wind turbine accidents is the significance of their occurrence as well as the variety of negative impacts they impose. They can result not only in technical failures and financial losses, but also and more importantly, human deaths and injuries.

To the best of our knowledge, one of the reasons for shortage of research on wind turbine accidents is the lack of publicly available data. While wind turbine manufacturers, owners, and contractors collect data about their operations, including data on accidents, they do not publicly share most of this data, especially the accident data. The reason for keeping these data private might be not only due to confidentiality, but also for preserving a positive public perception of wind energy [8]. Industry organizations, such as American Wind Energy Association (AWEA) also have not made a significant collection of data on wind turbine accidents publicly available.

As of January 2016, the most extensive data available on the Internet on wind turbines accidents was published by the Caithness Windfarm Information Forum (CWIF) [9], a UK-based grassroots organization opposing wind turbine installations. When we conducted our data collection in late 2013, the CWIF list contained more than 1400 wind turbine accidents. As of January 2015, the list contained more than 1500 accidents. While the list is impressive in magnitude, the quality and reliability of the list is questionable because of the following reasons: 1) Most of the web links to the news sources are not valid, and some of the accidents appear in multiple lines of the data. 2) In spite of containing much more magnitude of data, the data available in other online sources also exhibit similar deficiencies.

Given the growth of the wind turbine industry, and considering the lack of academic as well as industry research, we inspired to perform the first such study and contribute to the literature. To this end, we carried out a rigorous search of the news on wind turbine accidents (with confirmed references to the news sources) and implemented a variety of data analysis techniques to provide with critical and impactful insights on the topic. One innovation of the paper is the fact that a well-planned data mining approach and process has been applied for the first time in the wind turbine accidents literature. Furthermore, the applied data mining process has been documented in detail within the paper, so that future studies would benefit from an initial methodological benchmark, enabling them to propose methodological improvements, as well as novel empirical findings.

There are two critical and fundamental concepts in our paper, which shape the structure of our methodology and analysis: First, the stage of the wind turbine's life cycle, at which the accident took place. Figure 1 displays four possible stages when an accident may occur, namely, during the transportation, construction, operation, and maintenance. Second, the cause of the wind turbine accident, namely nature, system/equipment, and human (Figure 1). We investigate the association between these two categories of factors and two major effects (outcomes), i.e., Death and Injury (Figure 1). Thus, the main hypotheses of our paper are as follows:

Hypothesis 1. There exists association between deaths and predictor attributes (factors).
Hypothesis 2. There exists association between injuries and predictor attributes (factors).
These hypotheses are tested using formal statistical methods.

Fig. 1. The cause-effect relationship and stages where an accident occurs.

The remainder of the paper is organized as follows: Section 2 provides a brief review of some relevant literature as the background. Section 3 discusses the methodologies used in the data analysis, including the flowchart of the process for statistical hypothesis testing. Section 4 describes the data collection and cleaning process and describes the collected data. Section 5 presents the analysis and results. It begins with the exploratory analysis of the data, and continues with the application of statistical tests and data mining methods. Section 6 discusses the discovered insights. Finally, Section 7 concludes the paper and suggests future research directions.

II. Literature

Wind turbine accidents and wind energy risks have drawn certain attention of the academic community [10]-[20]. However, our review of the literature showed that none of existing studies have done a comprehensive analysis of the associations between factors (predictive attributes), and Death and Injury [14]-[16]. Furthermore, none of the papers we found in the literature combine formal statistical methods with data mining approach to analyze a dataset that contains multiple accidents.

In this section, we first briefly review the existing work on the analysis of individual wind turbine

accidents. Then, we discuss studies that analyze multiple wind turbine accidents. This is followed by a review of works that conduct risk analysis regarding wind turbines, farms, or electricity grids. Finally, we survey the application of a specific type of data mining, namely text mining, on wind energy and turbines.

A. Analysis of Individual Wind Turbine Accidents

There are a number of studies which focus on the failure of a single wind turbine. These studies, in chronological order, are as follows:

[16] reports the post-disaster inspection of a collapsed wind turbine during a typhoon in Taiwan. The study presents fresh insights into the causes of wind turbine failure, as well as lessons for the future. The authors also include a summary of 62 accidents of tower collapse that occurred between 1997 and 2009. However, the paper does not provide an analysis of the mentioned 62 accidents. The study draws insightful conclusions and generalized guidelines that should be considered by practitioners in the wind turbine industry.

[17] presents the fracture analysis of a wind turbine main shaft. The study determines that high stress concentrations were the cause behind the fracture.

[18] analyzes the failure of a large turbine blade. The study identifies the material and mechanical reasons behind the failure.

The above studies analyze a single turbine, mainly from a mechanical engineering or materials science perspective. We, on the other hand, analyze the outcomes of multiple accidents, and the association between cause factors and outcomes (Death and Injury).

B. Analysis of Multiple Wind Turbine Accidents

The most extensive report on multiple wind turbines is “Handboek Risicozonering Windturbines” [7], a handbook on wind turbine accidents, published in Dutch. The handbook was originally developed on the order of and updated annually for NOVEM (Netherlands Agency for Energy and the Environment). It aims at presenting the procedures for the risk assessment of wind turbines, and provides detailed statistics for different types of risks for wind turbines. The handbook firstly categorizes the different kind of failures of turbines (referred to as “scenarios”) that should be considered in a risk analysis. Then, the handbook presents the occurrence frequency for each scenario, based on the analysis of over 200 severe incidents and accidents in Denmark, Germany and the Netherlands. Braam and Rademakers [19] provide a general description for the project and the handbook in English.

While the mentioned handbook is extensive, it does not address theoretical research questions that we answer in our paper. Furthermore, the data that we collected and analyzed covers not only three countries, but the globe. Finally, our data covers 240 accidents, which is more than what the handbook covers.

A study that analyzes multiple accidents is presented in Yasuda et al. [20]. The authors focus on wind turbine blade incidents and present a new classification of such incidents. The authors also classify lightning damages and their possible causes, as well as recommending countermeasures.

C. Risk Analysis Regarding Wind Turbines and Farms

We encountered two recent sample studies where risk analysis is conducted in a more general context. Similar studies can be found by referring to the references listed in these two studies. De Andrade Vieira and Sanz-Bobi [21] introduce a new method for estimating the health condition of components of a wind turbine based on real-time sensor data, which enables the rescheduling of planned maintenance. The contribution of their developed method is the maintenance of the wind turbine at lower cost. Gonzalez et al. [22] introduce a novel approach to the problem of optimal design of wind farms (selection of the

turbines location, turbine type, and hub height) including decision making under risk. However, compared to our paper where the risk of accidents is the core of research, the main focus of Gonzalez et al. [22] is the uncertainty from wind direction and speed.

D. Text Mining for Wind Energy and Wind Turbines

Text mining refers to the application of data mining methods to text data. In the literature, text mining has been applied to wind energy industry and wind turbine systems in a few ways:

1. First, it is used to summarize the reasons of technical development constraints and suggest the research directions needed to be emphasized. For instance, the study in [23] discovers the key factors limiting the wind turbine scaling by mining textual reports, standards, and journals.
2. Second, text mining is applied to risk management by extracting information from the textual service records of wind turbines. For example, the inventions in [24] and [25] propose risk management systems with document classification capability for wind turbine service reports.
3. Lastly, text mining is used to identify technology trends and the promising technologies for technology transfer [26]-[28].

III. Methodology

A. Exploratory Data Analysis

Data analysis techniques can be grouped into three categories: Exploratory, Descriptive, and Predictive. The main goal in exploratory data analysis, which is implemented in our paper, is to obtain basic insights into the data. Exploratory data analysis includes the use of graphical techniques such as histograms, pie charts, geographical displays, besides basic summary tables. In our study, we start our data analysis with the graphical techniques and especially the mosaic display.

B. Statistical Hypothesis Testing

In empirical research, statistical hypothesis testing is the conventional form of supporting or refuting proposed hypotheses. In our analysis, we use three principal types of hypothesis tests within a unified process (Figure 2): Goodness of fit test, sample mean comparison tests, and correlation tests [29].

The Shapiro-Wilk goodness-of-fit test suggests whether a data sample follows normal distribution [30]. This is a crucial information needed for the proper selection of the “comparison of means” test. The parametric t-test or the nonparametric Mann-Whitney test [31] is used for testing whether two data samples have same mean values. The parametric ANOVA or the nonparametric Kruskal-Wallis test is used for testing whether the mean of any sample among a group of samples (more than two data samples) is different from the others.

If the normality of the involved samples (in the comparison of means tests) is rejected with a high confidence level (test resulting in a low p-value), then nonparametric methods are used. Parametric tests are used only if all the samples follow the normal distribution [29].

Correlation tests that we employ are Pearson’s Chi-Square test [32] for two numerical attributes, and Fisher’s test [33] for two categorical attributes. In both of these tests, a low p-value suggests a significant association between the two selected attributes (a low p-value suggests that it is highly unlikely that the correlation would be zero). In our analysis, we selected the threshold p-value to be 0.05. The process followed is shown as a flowchart in Figure 2.

C. Ranking of Attributes

We employ the information gain (Kullback–Leibler divergence) measure to rank the importance of the

attributes, when determining the occurrence of death and injuries. Information gain of an attribute A is the information gained about a response X based on observation of the values that A takes [34]. The information gain concept is used in information sciences to obtain a ranking among attributes [34], based on how much they help in the prediction of values of the response attribute. The higher the information gain value, the more information the attribute provides for predicting the response. In the context of our study, the attributes with the highest information gain values can be thought as those attributes that help us most in understanding and predicting whether death or injury will occur as a result of an accident.

D. Classification Trees

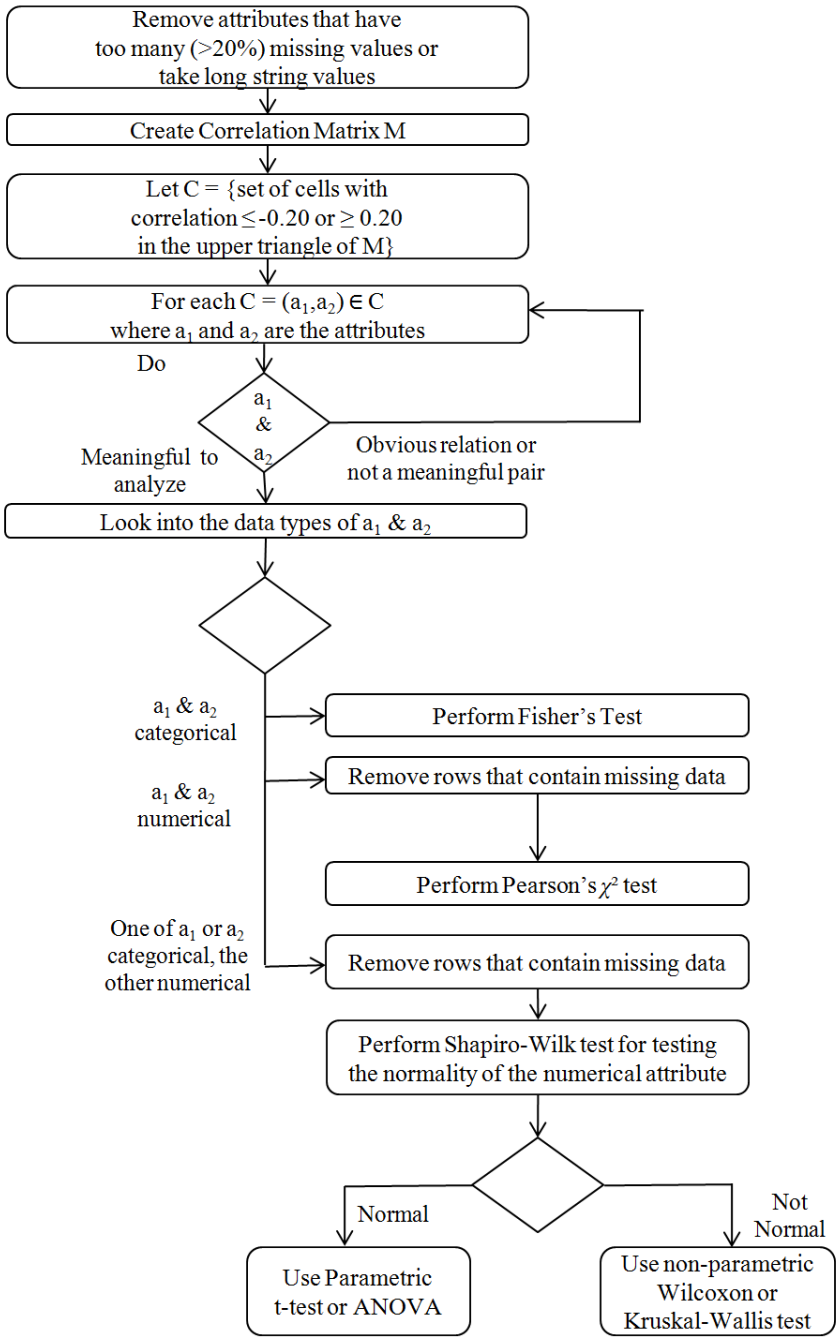

Fig. 2. Flowchart of the process for statistical hypothesis testing

Using classification tree models, one can summarize rule-based information about classification as trees. In classification tree, each node is split (branched) according to a criterion. Then, a tree is constructed with a depth until all the rules are displayed on the graph under a stopping criterion. At each level, the attribute that creates the most increase compared with the previous level is observed. The algorithms for decision tree analysis are explained in [35]. In classification trees, identifying the nodes that differ noticeably from the root node are important, because the path that leads to those nodes (represented as the antecedent of a rule) tells us how significant changes are observed in the subsample compared with

the complete data. By observing the shares of slices and comparing with the parent and root nodes, one can discover classification rules and insights.

E. Classification Analysis

In classification analysis, the dataset is divided into two groups, namely, the learning and test datasets. Classification algorithms, also referred to as classifiers (or learners), use the learning dataset to learn from data and predict the class attributes in the test dataset. The prediction success of each classifier is measured through a variety of performance measures, two of which will be used in this study: Classification accuracy (CA) is the percentage of correctly predicted cases in the test dataset. Area under curve (AUC) corresponds to the area under the ROC curve (which will be discussed in detail later) and is a measure of prediction quality [36]. We applied in our study the following classification algorithms (classifiers), which are among the best-known classifiers in the data mining field: Logistic Regression, k-Nearest Neighbor (kNN), Classification Trees, Support Vector Machines (SVM), and C4.5 [37].

F. Data Mining Process

The data mining model is constructed in the Orange software [38]. The data mining process contains four main types of analysis, namely ranking analysis, classification analysis, classification tree, and mosaic display, applied on two models (Model 1 and Model 2).

IV. Data

A. Data Collection

The accident news dataset in this study was collected over a 9-month period, scanning the Ebscohost and Lexis Nexis databases and also using Google as the search engine. All publicly available newspaper or magazine reports were considered for selection. The search keywords were “wind turbine accidents” and “wind turbine failures”. The search results were read and selected articles were checked by a graduate student. The main selection criteria were whether there was an impact on humans or the wind turbine. While reading the text of each news, only very certain statements describing specific outcomes were considered, and vague statements were ignored.

In total, more than 5,000 search results were scanned, more than 2,000 were read, and 247 were found highly related and were read in detail. Eventually, 216 news were found to directly report 240 wind turbine accidents, which were included in the dataset and analyzed in detail. Data on these 240 accidents was structured as a database table, containing the attributes explained below. All the original news articles, the word processor files that highlight the attribute fields in the data, and the structured database are well documented and are available upon request.

B. Data Cleaning

During the analysis of the news articles, it was firstly observed that some articles were either duplicates of other more extensive ones or were irrelevant to our study. These articles were removed from the data.

Data cleaning involved not only the verification and the validation of the data, but also the identification of missing values. While constructing the accidents dataset (Table 1), to the maximum possible extent, the data cells with missing values were eliminated through conducting additional search on the Internet. Specifically, search was conducted for finding the values of the attributes PowerOfWindFarm, Onshore/Offshore, TurbineModel, Manufacturer, PowerOfTurbine, Location, and Country.

C. Data Description

Selected columns in the constructed database, which contains the accident characteristics for the 240

accidents, are given below:

Accident No: Unique integer identification number for each accident (e.g. 1).

Country: Country where the accident took place (e.g. Denmark).

Turbine Model: Model of the wind turbine at which the accident took place; includes the manufacturer name, and the model name/code/number, and power (e.g. “Vestas, V80-2.0 MW”).

Manufacturer: The company that manufactured the wind turbine (e.g. Vestas).

Power of Turbine (kW): Power of the wind turbine in kW (e.g. 2000), where 1 MW=1000 kW.

Power of Wind Farm (kW): Total power of the wind farm in which the wind turbine is located.

Death: Tells whether human death has occurred because of the accident; takes binary values (e.g. 0). It takes the value of 1 when death occurs.

Injury: Tells whether human injury has occurred because of the accident; takes binary values (e.g. 0). It takes the value of 1 when injury occurs.

Fire: Tells whether fire has occurred because of the accident; takes binary values (e.g. 0). It takes the value of 1 when fire occurs.

Mechanical: Tells whether mechanical damage has occurred because of the accident; takes binary values (e.g. 0). It takes the value of 1 when mechanical damage has occurred.

Structural Break: Tells whether a structural break has occurred because of the accident; takes binary values (e.g. 1). It i takes the value of 1 when structural break has occurred.

TABLE I
DISTRIBUTION OF REASONS FOR ACCIDENTS CAUSED BY HUMANS

Cause	Count
Human (other)	23
Human (transportation)	18
Human (negligence)	4
Human (wrong action)	4
Human (interference in control systems)	2
Human (fall)	1
Human (heart attack)	1
Human (plane crash)	1

Affected Humans: Tells whether the accident has affected humans in the form of death or injury (e.g. 0). The value for this attribute is computed as the maximum of the values of the Death and Injury attributes.

Affected System/Equipment: Tells whether the accident has affected the turbine system or equipment (e.g. 1). The value for this attribute is computed as the maximum of the values of the Fire, Mechanical, StructuralBreak, and TransportAccident attributes.

Transport Accident: Tells whether the accident was a transport accident; takes binary values (e.g. 0). It takes the value of 1 when the accident was a transport accident.

Affected Component: All the major components affected because of the accident, summarized as a string (e.g. “Blade”). This string can contain more than one item, such as “Tower, Blade”.

Cause: Tells the particular cause of the accident (e.g. “Human (interference in control systems)”)

Cause Category: Tells the general cause category of the accident. Takes one of the following values: “Human”, “Nature”, “System/Equipment”.

Onshore/Offshore: Tells whether the wind turbine is located onshore (inland) or offshore (in sea). Takes one of the values of “OnShore” or “OffShore”.

EventOccurrence: The state of the wind turbine when the accident occurred. Takes one of the following values: “During construction”, “During maintenance”, “During operation”, “During transportation”.

Accident Year: Year in which the accident took place (e.g. 2002).

Accident Month: Month in which the accident took place (e.g. 11).

Accident Day: Day in which the accident took place (e.g. 4).

ANALYSIS AND RESULTS

In this section, we present the analysis of the constructed wind turbine accidents database using the introduced methodologies. The two processes that we apply are the statistical process (Figure 2), and the data mining process. The analysis has been conducted using five methods, namely, exploratory analysis, hypothesis tests, ranking analysis, classification tree analysis, and classification analysis.

D. Exploratory Data Analysis

Firstly, the values of different attributes (columns) were investigated. The Accident Year ranges from 1980 until 2013, except for two earlier accidents. The powers of the wind turbines mentioned in the news peak around certain points, such as 500 kW, 1500 kW and 2000 kW. These capacities are mainly because of the wind turbine capacities available in industry, where 500 kW, 1500 kW and 2000 kW are standard capacities, and many new wind turbine projects aim at developing turbines at these capacities. In the dataset, Danish wind turbine manufacturer Vestas is the wind turbine brand with the most accidents and GE coming as the second. USA has the largest number of wind turbine accidents, followed by Germany, China, and Australia. These statistics are consistent with the distribution of wind turbine installations.

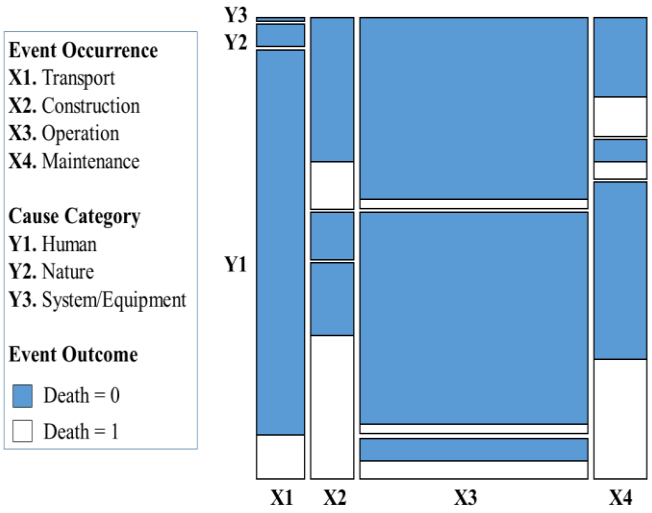

Fig. 3. Mosaic Plot Showing the Effect of Event Occurrence and Cause Category on Death.

TABLE II

DISTRIBUTION OF REASONS FOR ACCIDENTS CAUSED BY NATURE

Cause	Count
Nature (strong wind)	32
Nature (lightning strike)	9
Nature (storm)	4
Nature (other)	3
Nature (cyclone)	2
Nature (tornado)	2
Nature (cold)	1
Nature (due to collision)	1
Nature (strong wind, lightning strike)	1
Nature (strong wind, snow)	1
Structural (bolt failure)	1
Structural (smashed barge)	1

TABLE III

DISTRIBUTION OF REASONS FOR ACCIDENTS CAUSED BY SYSTEM/EQUIPMENT

Cause	Count
Mechanical	25
Mechanical (electrical)	8
Mechanical (faulty material)	5
Mechanical (due to collision)	4
Mechanical (material fatigue)	2
Mechanical (brake system failure)	1
Mechanical (cracks on blade)	1
Mechanical (failed transformer)	1
Mechanical (fire)	1
Mechanical (insufficient glue on blades)	1
Mechanical (lack of automatic braking system)	1
Mechanical (loose connections between the transformer's connection bars and the power cables from the generator circuit breaker)	1
Mechanical (low voltage ride through capability)	1
Mechanical (not properly secured foundation bolts)	1
Mechanical (platform collapse at construction site)	1

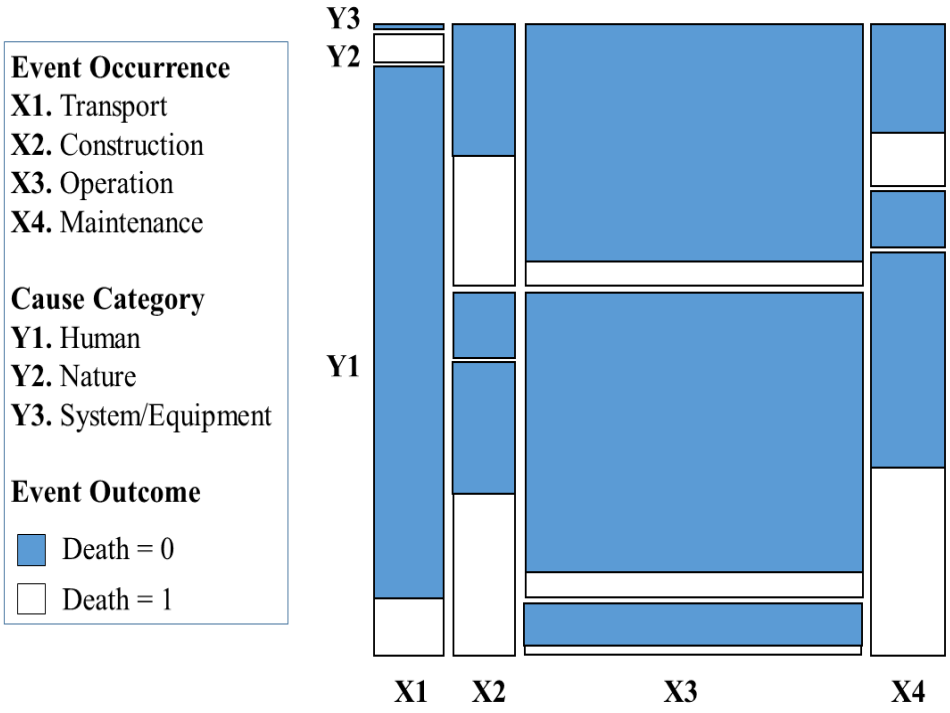

Fig. 4. Mosaic plot that shows the effect of Event Occurrence and Cause Category on Injury.

Table 1 suggests that a human caused accident mostly occurs during transportation (18 accidents).

Table 2 suggests that natural causes are mostly related to strong wind (32 accidents) and lightning strikes (9 accidents). Even though the reasons for strong wind and lightning strike are categorized under natural causes, these may also be interpreted as indirect human-related causes. However, in this paper, we classify these causes as natural causes.

System/Equipment is also seen as a major cause for accident with its sub-causes mostly related to electric causes, material fatigue, and faulty material (Table 3). This analysis shows that not only the design, but also the maintenance and operation of a wind turbine are important. Electric problems may be attributed to not only the design of the system, but also to the electricity grid and the problems associated with it.

We analyzed the Distribution of values for the attributes AffectedHumans and AffectedSystem/Equipment. According to the results, wind turbine accidents mostly affect the system and equipment.

The distribution of values for the attribute AffectedComponents suggest that the case of a wind turbine accident, the components blade, tower, nacelle have the highest chances of being affected. When EventOccurrence is analyzed, it is revealed that accidents occurred overwhelmingly during operation.

The mosaic plot displays the stages of accident occurrence on the x axis, while the causes of accidents (human, nature, and system/equipment) are shown on the y axis. The width of the columns on the x axis and the height of the blocks on the y axis are proportional to the number of accidents in each category or cause, so the area of each of the rectangles represents the total number of accidents that meet its two criteria. Several patterns can be observed from the mosaic plot in Figure 3 for accidents and deaths (outcomes denoted by color).

First, let us summarize our findings from Figure 3 for accidents, regardless of whether they resulted in death or not (regardless of the color in the mosaic plot).

• It is seen that accidents during operation (the area above the label "X3. Operation") are more than the sum of accidents in the other three stages (transport, construction, maintenance).

Furthermore, the figure also illustrates which Cause Category is most influential in each stage.

• During transportation, the Cause Category is overwhelmingly Human.
• During construction, the cause categories System/Equipment and Human are much more influential than Nature.
• During operation, Nature is the most influential Cause Category, followed by System/Equipment.
• During maintenance, the most important cause category is Human.
• Most deaths occur during the Construction and Maintenance of the wind turbine.

The mosaic plot shown in Figure 4 is similar to Figure 3, and yields insights for the distribution of accidents and death occurrences in those accidents. However, Figure 4 shows the effect of Event Occurrence and Cause Category on Injury, rather than Death as displayed in Figure 3.

Let us summarize our findings from Figure 4 for injuries (white-colored regions denoting occurrence of injuries).

• During Transportation, the Cause Category that results in the most deaths is Human. However, percentagewise, the effect of Nature on Injury is the highest. All the cases during Transportation where Nature was the Cause Category, resulted in Injury=1.
• During construction, the pattern is exactly the same as in Figure 3. However, during operation, the most influential Cause Category is System/Equipment, both in quantity and percentage. This pattern for the Operation stage is different compared with that of Death.
• Finally, during Maintenance, all injuries occur because of the System/Equipment or Human. None of the accidents during Maintenance occur due to Nature.

E. Statistical Analysis

In our statistical analysis we will be exploring the relations between the predicted attributes of Death and Injury, and a set of predictor attributes. The first step was to compute the correlation matrix between all attributes, so that we could observe all such relations, and apply appropriate statistical tests of significance for the most promising relations. To this end, cells (pairs of attributes) of the correlation matrix which were found to have correlation values ≤ -0.20 or ≥ 0.20 were selected. Detailed statistical analysis was conducted for 12 of these 26 cells, while 14 of them could not be analyzed in detail because of too many categorical values, being too obvious or not being meaningful. Table 4 presents the detailed information on the hypothesis tests for these 12 cells. The table shows the pairs of attributes selected for the correlation tests, the corresponding correlation values, the statistical tests performed for each attribute pair, the resulting p-values (p-values less than the threshold p-value of 0.05 suggest statistically significant correlations) and the test results (+ means that the correlation observed between the two attributes is statistically significant at the selected p-value threshold of 0.05). As a result, statistically significant correlations were found between 10 out of 12 pairs of attributes, as can be read from the last column of Table 4. Table 5 presents the interpretation of the test results.

In Table 5, an important observation is for Test 5 (row 5), which is "There is association: Injury rate is lowest when the cause is nature induced (compared with System/Equipment or Human as the Cause Category)." This shows that our preliminary Hypothesis 2 that there may be a difference among the various causes (Nature, System/Equipment, Human) on how they affect Injury, is indeed statistically supported.

Tables 5 and 6 do not include an analysis of the effect of the various causes on Death, because the correlation value was not in the range [-0.20, 0.20].

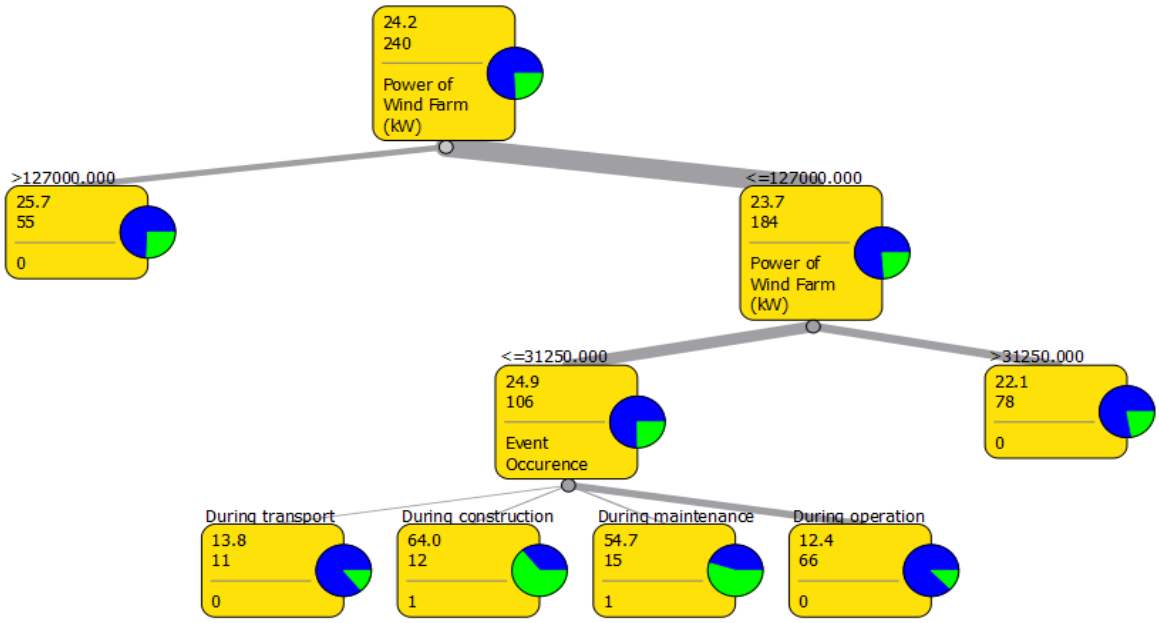

Fig. 5. Classification tree graph for Model 1, where Death is predicted.

F. Ranking of Predictor Attributes

TABLE IV

STATISTICAL TESTS PERFORMED AND THE RESULTING P-VALUES

Test No	Attribute1	Attribute2	Correlation	TestPerformed	p-value	Result
1	Structural Break	Death	0.39	Fisher's Test	2.07E-09	+
2	Structural Break	Injury	0.28	Fisher's Test	1.6E-05	+
3	Manufacturer	Mechanical	0.22	Fisher's Test	1.0000	-
4	Injury	Power of Turbine (kW)	0.20	Mann-Whitney test	0.0433	+
5	Cause Category	Injury	-0.22	Fisher's Test	0.0050	+
6	Cause Category	Power of Turbine (kW)	-0.24	ANOVA	0.0202	+
7	Mechanical	Death	-0.26	Fisher's Test	4.44E-06	+
8	Manufacturer	Structural Break	-0.26	Fisher's Test	0.6946	-
9	Death	Accident Year	-0.26	t-test	0.0013	+
10	Cause Category	Power of Wind Farm (kW)	-0.29	Kruskal-Wallis test	0.05	+
11	Event Occurrence	Power of Turbine (kW)	-0.32	Kruskal-Wallis test	0.05	+

The next analysis is the ranking of the predictor attributes, based on the information they provide in predicting Death or Injury. This analysis is important, since it helps us prioritize, among a multitude of attributes, the ones that potentially have the highest impact on the predicted attribute. To this end, two models have been constructed based on the same data mining process. The first model (Model 1) focuses on the occurrence of deaths, while the second model (Model 2) focuses on the occurrence of injuries.

The predictor attributes are Accident Month, Accident Day, Accident Year, Country, Event Occurrence, Onshore/Offshore, Power of Turbine (kW), and Power of Wind Farm (kW). The number of rows (corresponding to accidents) is 240. The predicted class attribute is Death, taking value of 1 (human death) or 0 (no human death) in the first model (Model 1), and Injury in the second model (Model 2).

TABLE V

INTERPRETATION OF THE RESULTS OF STATISTICAL TESTS

Test No	Test Result
1	There is association: Less death when structural break
2	There is association: Less injury when structural break
3	No relation (when only Vestas and GE are considered)
4	There is difference: Higher power turbines in case of injury
5	There is association: Injury rate is lowest when the cause is nature (compared with System/Equipment or Human as the Cause Category).
6	There is difference: Higher power turbines when cause category is human.
7	There is association: Less death when mechanical
8	No relation (when only Vestas and GE are considered)
9	There is difference: Accident year is less when death (More recent years when no death)
10	There is no difference
11	There is difference: Higher power turbines when the accident is during construction or maintenance, compared with during operation.
12	There is association: More injuries during construction or maintenance, compared with during operation or transport.

The results of ranking for Model 1 are displayed in Table 6, where the attributes are sorted according to their information gain values. Information gain is a measure of how much information is gained from a predictor attribute with respect to predicting a response attribute. The column titled Values tells the number of distinct discrete values that the attribute takes, where C denotes categorical attributes (which cannot be used in prediction).

TABLE VI

RANKING OF ATTRIBUTES FOR PREDICTING THE BINARY VALUE OF DEATH

Rank	Attribute	Values	Information Gain
1	Event Occurrence	4	0.234
2	Country	25	0.156
3	Onshore/Offshore	2	0.109
4	Power of Turbine (kW)	C	0.098
5	Accident Month	C	0.089
6	Accident Day	C	0.062
7	Power of Wind Farm (kW)	C	0.060
8	Accident Year	C	0.030

Table 6 shows that Event Occurrence is the most important predictor attribute, with almost double the information gain value of the next attribute, Country. Therefore, Event Occurrence, in other words, the stage of wind turbine, is the attribute that provides the most predictive information on whether a human

death occurs. Other attributes that follow include Onshore/Offshore, Power of Turbine, Accident Month, Accident Day, and Power of Wind Farm. The information gain value halves in the next attribute (Accident Year) that follows Power of Wind Farm, suggesting a large gap in the information provided by the first seven attributes and the last one. Therefore, the first seven attributes should be considered before the eighth one and those that come after. In Model 2, the same ranking analysis was conducted with the same eight predictors, but this time with Injury as the predicted class attribute. Table 7 shows the results of this analysis. The rank of Power of Turbine is now much higher, at the top of all the other attributes. The rank of Power of Wind farm is also higher ranked. In predicting Death, Power of Turbine and Power of Wind Farm do not play as much importance, while in predicting Injury, these two attributes make an important contribution. Country is still the second most important predictor. Event Occurrence is still important in predicting Injury, but ranks as the third most important predictor attribute, rather than first as in predicting Death. The rank of the attribute Onshore/Offshore is also different in Tables 6 and 7. In predicting Death, the Onshore/Offshore attribute of the wind turbine is important (ranked as the third most important predictor attribute), while it is the least important predictor in predicting Injury.

TABLE VII
RANKING OF ATTRIBUTES FOR PREDICTING THE BINARY VALUE OF INJURY

Rank	Attribute	Values	Information Gain
1	Power of Turbine (kW)	C	0.114
2	Country	25	0.093
3	Event Occurrence	4	0.068
4	Power of Wind Farm (kW)	C	0.048
5	Accident Year	C	0.030
6	Accident Month	C	0.011
7	Accident Day	C	0.003
8	Onshore/Offshore	2	-0.022

Table 7 shows that Power of Turbine is the most important predictor attribute for injury, with almost double the information gain value of the third attribute, Event Occurrence. Therefore, Power of Turbine is the attribute that provides the most predictive information on whether a human Injury occurs. The information gain value also almost halves in the next attribute that follows AccidentYear, suggesting a large gap in the information provided by the first five attributes and the remaining ones. The data mining process can thus be modified to include only the first five attributes in Table 7 as predictors of Injury.

G. Classification Tree Analysis

In the classification tree analysis, information gain was used as the attribute selection criterion in the split in the tree Only the first seven attributes of Table 6 were included as predictors while predicting whether Death occurs (Death=1) or not. The results of the classification tree analysis for Model 1 are displayed in Figure 5. Each node (little box) represents the percentage of observations with the target class attribute value (Death) and also the count. Each pie shows the distribution of the values of the target class attribute.

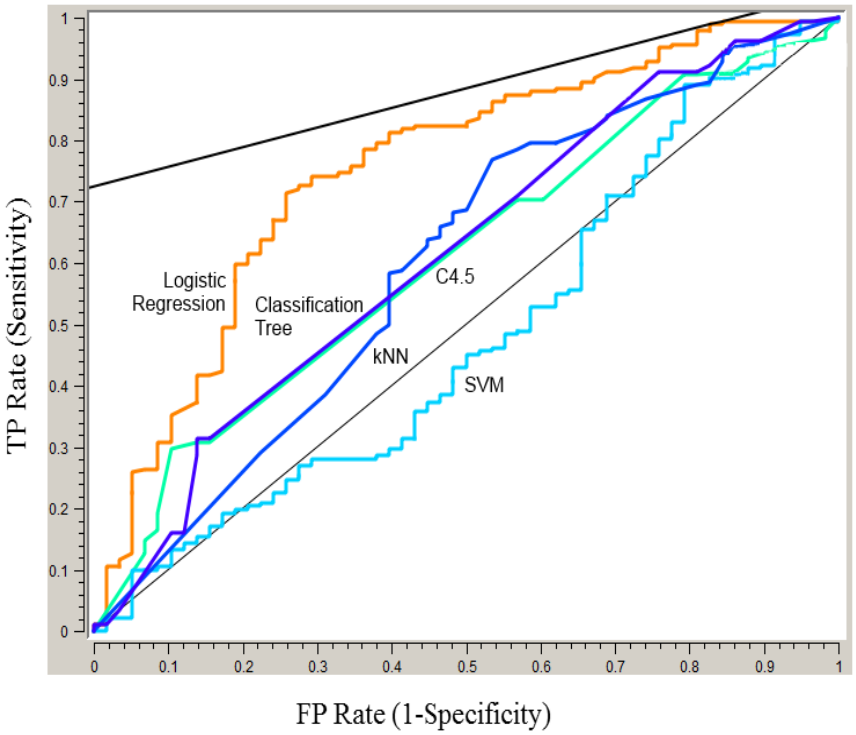

Fig. 6. ROC analysis for Death.

In the analysis of classification trees (Figure 5), visually identifying the nodes that differ noticeably from the root node are important, because the path that leads to those nodes (represented as the antecedent of a rule) tells us how significant changes are observed in the subsample compared with the complete data. By observing the shares of slices and comparing with the parent and root nodes, one can discover classification rules and insights. While the first split (according to the value of information gain) is based on Power of Wind Farm, this does not create a significant change in slice shares. The most significant change from the root node occurs based on the third split, is based on the attribute Event Occurrence. Deaths are much less frequent during transportation and operation, while they are much more frequent during construction and maintenance (clearly, a larger share of the light-colored slice compared with the root).

The classification tree analysis did not yield any insights for Model 2, where Injury was predicted. This means that none of the five attributes from Table 7 that were put into Model 2 provided enough information to create a significantly different split of the sample into subsamples.

H. Classification Analysis

The final analysis of the data is the classification analysis. The task in classification analysis is to predict the predicted attribute with a high classification accuracy. The ultimate goal is to be able to predict the class values of the predicted attribute in new cases. To this end, the data is systematically split into training and testing datasets, the training dataset is used to “teach” the classification algorithms (or shortly “classifiers”) about the data, and the performance of the classification algorithms is tested using the test dataset.

TABLE IX

EVALUATION OF THE PERFORMANCE OF VARIOUS CLASSIFICATION ALGORITHMS FOR PREDICTING THE BINARY VALUE OF INJURY.

Classifier	CA	AUC
Logistic regression	0.829	0.777
kNN	0.817	0.669
Classification Tree	0.850	0.500
SVM	0.850	0.500
C4.5	0.850	0.500

TABLE VIII

EVALUATION OF THE PERFORMANCE OF VARIOUS CLASSIFICATION ALGORITHMS FOR PREDICTING THE BINARY VALUE OF DEATH.

Classifier	CA	AUC
Logistic regression	0.763	0.758
SVM	0.750	0.728
Classification Tree	0.742	0.574
C4.5	0.738	0.565
kNN	0.642	0.605

The most popular metric used in measuring the quality of the results obtained by classification algorithms is “classification accuracy”, which is the percentage of observations in the test data set that are classified correctly. In our case, the classification is performed for Death and Injury, respectively. The goal is to predict whether Death or Injury will occur in a particular wind turbine accident. Tables 8 and 9 present the results of classification analysis. Among the five classifiers applied, Logistics Regression gives the best results for both models.

Figure 6 shows the receiver operating characteristic (ROC) curves for the first model. The ROC curve plots the true positive rate (TP-Rate) on the y-axis against the false positive rate (FP-Rate) on the x-axis, as a discrimination threshold is varied. The classifier predicts the class of the particular case in the testing dataset as "positive" (for example, predicting Death=1 in Model 1), if the function value for that classifier exceeds the discrimination threshold. TP-Rate refers to the percentage of cases which are correctly predicted to have positive class values (for example, cases which have Death=1 in Model 1 and have been correctly predicted as such by the classifier). FP-Rate refers to the percentage of cases which are predicted as positive, but are actually not positive (for example, cases with Death=0 in Model 1, that have been predicted as Death=1 by the classifier). Every single point on the ROC curve for a certain classifier (for example, logistic regression) reflects the (x,y)=(FP-Rate, TP-Rate) value pair corresponding to a particular value of the discrimination threshold. ROC curves with greater areas under the curve (AUC), which are closer to the upper left corner in the plot, correspond to better classifiers.

• In Model 1, it is possible to achieve a classification accuracy of at most 76.3%, using logistic regression. Logistic regression is a specific type of regression which is applicable in classification analysis, as logistic regression can be used to predict values of a categorical attribute (such as Death and Injury).

When the ROC curves in Figure 6 corresponding to logistic regression and SVM (Support Vector Machines) are compared, it is observed that logistic regression has a larger AUC value. Also, the ROC curve for SVM is mostly below the y=x line, showing that it results in a low TP-rate for the same FP-rate. Therefore, logistic regression is the most appropriate classifier to predict the occurrence of Death.

- In Model 2, for predicting Injury, the classification accuracy (CA) of the classifiers classification tree, SVM, and C 4.5 are the highest, reaching 85%. However, analyzing the confusion matrix reveals that these three predictors classify none of the Injury=1 cases correctly (The confusion matrix is a matrix that shows the distribution of correct and erroneous predictions; Each column of the matrix represents the observations in a predicted class, while each row represents the observations in an actual class). Obtaining a high value for CA, despite zero success in correctly classifying Injury=1 cases is interesting. This result is because of the high percentage of cases with Injury=0. So, even though CA is a good measure, it should be considered together with the confusion matrix and ROC curves.
- Logistic regression classifier, on the other hand, does classify some of the Injury=1 cases correctly. This is also revealed in the ROC curve (not given as a figure), where the AUC for logistic regression is the highest, followed by that of kNN. Therefore, logistic regression is the most appropriate classifier to predict the occurrence of the Injury, as well.

V. Conclusions and Future Work

For the first time in the literature, our research investigates the contents of news articles on wind turbine accidents to come up with multi-faceted insights and new knowledge. Specifically, we studied the association between the characteristic attributes of wind turbine failures and the outcomes of death and injury. A particular emphasis was on two factors, namely the stage of the wind turbine's lifecycle, and the cause of the accident. In the modeling and data collection phases of our research, a critical issue was the valid selection of the cause and effect categories. These selections have been tediously carried out through consulting with a well-known professor in the field, who was responsible of the design and development of a national wind turbine for Turkey in a research project which involved more than 100 researchers.

Some of the insights that have been obtained, as well as their implications, can be summarized as below:

1) Human caused accidents mostly occur due to human errors in transportation. Possible novel practices can include the rehearsal of the route and/or use of virtual reality simulators before the actual transportation is executed.

2) Natural causes are mostly related to strong wind and lightning strikes. Considering the fact that continuous improvements are made on wind turbine designs, we hypothesize that high rates of accidents for lightning strikes in our data can be due the accidents in earlier make turbines (we do not have data on the make year of turbines).

3) Major causes of accidents within the category of Systems/Equipment are electric causes, material fatigue, and faulty material.

4) In wind turbine accidents, blade and tower have the highest probability of being affected. During construction the cause categories System/Equipment and Human are much more influential than Nature.

5) During maintenance, the most important cause is also Human.

6) In the accidents during Construction, if the cause category is System/Equipment or Human, the probability of Death is higher than 0.5.

7) Most deaths occur during the Construction and Maintenance of the wind turbine.

8) During Maintenance, the number of accidents (rather than the probability of accidents) is highest when the Cause Categories are Human and System/Equipment.

9) During Transportation, percentagewise, the effect of Human on Injury is highest.

10) Our paper has established the statistically significant associations between all the factors and Death & Injury (Tables 4 and 5).

11) When predicting the possible occurrence of Death, the most information is gained from EventOccurrence, that is, the stage of the wind turbine's lifecycle. Other informative attributes are listed in Table 7.

12) When predicting the possible occurrence of Injury, the most information is gained from PowerOfTurbine. From Table 4, it can be seen that the correlation is positive. Thus larger turbines are more likely to lead to injuries. Other informative attributes are listed in Table 8.

13) When predicting the possible occurrence of Death given that an accident of the type we have defined has occurred, one should use the logistic regression classification method, rather than other methods. For our test dataset, this method predicted Death with a classification accuracy of 0.763.

14) When predicting the possible occurrence of Injury given that an accident of the type we have defined has occurred, one should again use the logistic regression classification method, rather than other methods. For our test dataset, this method predicted Injury with a classification accuracy of 0.829.

One important limitation and threat to the validity of our study is regarding the collection of the data and selection of the relevant news. The data that we collected is not complete, but is just a sample obtained through Internet by the Google search engine. Our assumptions were that the significant accidents made it to the news and were indexed by Google search engine with a somewhat high ranking. Google search engine utilizes sophisticated natural language processing algorithms as well as the Page-rank and other algorithms to obtain a ranking among the search results. For example, the search term "wind turbine accident" results in approximately 300,000 results. We scanned through only the first 5,000 of these results. Therefore, our data is not complete and is only a sample. As in every study where sampling from a population is carried out, there is the risk that our sample may not in fact be a random sample that represents the true population.

Future research on the topic can work with larger document collections, not necessarily coming from publicly available news articles, but maybe also from industry, NGO (non-governmental organization) and government sources, such as regulation bodies. Other research, from a methodological perspective, includes the automatic identification of documents that report particular outcomes, such as death and injuries by using data mining techniques such as classification.

As the wind turbine industry is growing, we believe that the stakeholders in the industry, as well as government organizations and the academic community, should put more emphasis on collecting and analyzing data on wind turbine accidents. Our study has provided a multitude of insights and also has outlined some possible suggestions regarding wind turbine accidents. These insights can be guidelines for a variety of studies and best practices to be developed for and implemented in the wind turbine industry.

REFERENCES

[1] WEO-2013, World Energy Outlook 2013, International Energy Agency, Nov. 2013. Available: http://www.worldenergyoutlook.org/publications/weo-2013/

[2] C. Beggs, "Energy: Management, Supply and Conservation". Routledge, London. 2009.

[3] J. K. Kaldellis, and D. Zafirakis, "The wind energy (r) evolution: A short review of a long history," Renewable Energy, vol. 36, no. 7, pp. 1887-1901, Jul. 2011.

[4] Renewables 2014 Global Status Report, REN21 Renewable Energy Policy Network for the 21st Century, Nov. Feb. 2014. Available: http://www.ren21.net/Portals/0/documents/Resources/GSR/2014/GSR2014_full%20report_low%20res.pdf

[5] Global Wind Statistics 2013, Global Wind Energy Council, 2014. Available: http://www.gwec.net/wp-content/uploads/2014/02/GWEC-PRstats-2013_EN.pdf

[6] 2012 Wind Technologies Market Report, Energy.gov., 2013. Available: http://www.energy.gov/wind-report
[7] C. J. Faasen, P. A. L. Franck, A. M. H. W. Taris, Handboek Risicozonering Windturbines, Sep. 2014.
[8] E. Malnick, and R. Mendick, 1,500 accidents and incidents on UK wind farms", The Telegraph, Dec. 2011. Available: http://www.telegraph.co.uk/news/uknews/8948363/1500-accidents-and-incidents-on-UK-wind-farms.html
[9] Caithness Windfarm Information Forum. Available: http://www.caithnesswindfarms.co.uk
[10] B. Kerres, K. Fischer and R. Madlener, "Economic evaluation of maintenance strategies for wind turbines: A stochastic analysis," IET Renewable Power Generation, vol. 9, no. 7, pp. 766-774, Sep. 2015.
[11] H. Badihi, Y. Zhang and H. Hong, "Wind turbine fault diagnosis and fault-tolerant torque load control against actuator faults," IEEE Trans. Control Systems Technology, vol. 23, no. 4, pp. 1351-1372, Jul. 2015.
[12] X. Chen, C. Li, and J. Xu. "Failure investigation on a coastal wind farm damaged by super typhoon: A forensic engineering study," Journal of Wind Engineering and Industrial Aerodynamics, vol. 147, pp. 132-142, Dec. 2015.
[13] X. Chen, and J. Z. Xu. "Structural failure analysis of wind turbines impacted by super typhoon Usagi," Engineering Failure Analysis, vol. 60, pp. 391-404, Feb. 2016.
[14] M. Ashrafi, H. Davoudpour, and V. Khodakarami, "Risk assessment of wind turbines: Transition from pure mechanistic paradigm to modern complexity paradigm", Renewable and Sustainable Energy Reviews, vol. 51, pp. 347-355, Nov. 2015.
[15] X. Jin, W. Ju, Z. Zhang, L. Guo, and X. Yang, "System safety analysis of large wind turbines," Renewable and Sustainable Energy Reviews, vol. 56, pp. 1293-1307, Apr. 2016.
[16] J. S. Chou, and W. T. Tu, "Failure analysis and risk management of a collapsed large wind turbine tower," Engineering Failure Analysis, vol. 18, no. 1, pp. 295-313, Jan. 2011.
[17] Z. Zhang, Z. Yin, T. Han, and A. C. Tan, "Fracture analysis of wind turbine main shaft," Engineering Failure Analysis, vol. 34, pp. 129-139, Dec. 2013.
[18] X. Chen, W. Zhao, X. L. Zhao, and J. Z. Xu, "Preliminary failure investigation of a 52.3 m glass/epoxy composite wind turbine blade," Engineering Failure Analysis, vol. 44, pp. 345-350, Sep. 2014.
[19] H. Braam, L.W.M.M. Rademakers, "Guidelines on the Environmental Risk of Wind Turbines in the Netherlands," Global Wind Energy Conference, Paris, 2002. Available: ftp://130.112.2.101/pub/www/library/report/2004/rx04013.pdf
[20] Y. Yasuda, T. Fujii, K. Yamamoto, N. Honjo, and S. Yokoyama, S, "Classification of wind turbine blade incidents regarding lightning risk management," In Lightning Protection (ICLP), 2014 IEEE International Conference, pp. 986-991, Oct. 2014.
[21] R. J. de Andrade Vieira, and M. A. Sanz-Bobi, "Failure risk indicators for a maintenance model based on observable life of industrial components with an application to wind turbines," IEEE Trans. Reliability, vol. 62, no. 3, pp. 569-582, Sept. 2013.
[22] J. S. González, M. B. Payan,, and J. M. Riquelme-Santos, "Optimization of wind farm turbine layout including decision making under risk," IEEE Systems Journal, vol. 6, no. 1, pp. 94-102, Feb. 2012.
[23] N. Srikanth, "Material adoption practices of wind industry and its effect on product scaling trends," 2011 IEEE Technology Management Conference (ITMC), San Jose, CA, pp. 679-705, Jun. 2011.
[24] S. Vittal, G.A. Curtin, K. Manar, and P. Shrivastava, "Risk management system for use with service agreements," US Patent: US2012/0053983 A1, 2012.
[25] K. Manner and S. Vittal, "Risk management system for use with service agreements," US Patent: US2012/0053984 A1, 2012.

[26] H. Park, J. J. Ree, K. Kim, "Identification of promising patents for technology transfers using TRIZ evolution trends," Expert Systems with Applications, vol. 40, pp. 736-743, Feb. 2013.
[27] D. Kucuk, Y. Arslan, "Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles," Renewable Energy, vol. 62, pp. 484-489, Feb. 2014.
[28] T. Daim, I. Iskin, X. Li, C. Zielsdorff, A. E. Bayraktaroglu, T. Dereli, A. Durmusoglu, "Patent analysis of wind energy technology using the patent alert system," World Patent Information, vol. 34. No. 1, pp. 37-47, Mar. 2012.
[29] W. J. Conover, "Practical nonparametric statistics", Wiley. 1999.
[30] S. S. Shapiro, and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol. 52, no. 3/4, pp. 591-611, 1965.
[31] H. B. Mann, and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," The Annals of Mathematical Statistics, 18 (1), pp. 50-60, 1947.
[32] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," Philosophical Magazine Series 5, vol. 50, no. 302, pp. 157–175, 1900.
[33] R. A. Fisher, "On the interpretation of $\chi 2$ from contingency tables, and the calculation of P," Journal of the Royal Statistical Society, vol. 85, no. 1, pp. 87–94, 1922.
[34] S. Kullback, and R. A. Leibler, "On information and sufficiency," Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79–86, 1951.
[35] L. Rokach, O. Maimon, "Data mining with decision trees: theory and applications," World Scientific Pub Co Inc., 2008.
[36] J. Han, M. Kamber, and J. Pei, "Data Mining: concepts and techniques," organ Kaufmann; 3 edition, Jul. 2011.
[37] E. Alpaydin, "Introduction to Machine Learning," The MIT Press, Cambridge, MA., 2010.
[38] J. Demšar, B. Zupan, G. Leban, and T. Curk, "Orange: From Experimental Machine Learning to Interactive Data Mining," in J.-F. Boulicaut, F. Esposito, F. Giannotti, & D. Pedreschi, (Eds.) Knowledge Discovery in Databases PKDD, Springer, 2004.

Sobhan Asian received his Ph.D. from School of Mechanical and Aerospace Engineering at Nanyang Technological University in 2014. His research was fully supported by a scholarship award from the Agency for Science, Technology and Research (A*Star). After finishing his Ph.D., Sobhan worked as a Systems Consultant in Dematic S.E.A, which is a leading global Intralogistics Automation company. In 2016, Sobhan joint School of Business IT and Logistics, RMIT, Melbourne, Australia. His research interests include supply chain risk management, business analytics, and operations management.

Gurdal Ertek is an Assistant Professor at Rochester Institute of Technology - Dubai, UAE. He received his Ph.D. from School of Industrial and Systems Engineering at Georgia Institute of Technology, Atlanta, GA, in 2001. His research areas include data visualization and mining, as well as warehousing and supply chain management.

Cagri Haksoz holds a Ph.D. in Operations Management from New York University, Stern School of Business. He is currently a professor of supply chain, risk, and operations management at Sabanci University, School of Management, Istanbul, His major research expertise focus on risk intelligence, supply chain risk, adaptive decision making, and Modern Silk Road supply chains. He has authored, coauthored, co-edited four books titled Managing Supply Chains on the Silk Road (2011), Global Perspectives: Turkey (2012), Risk Intelligent Supply Chains (2013), and *İ*pek Yolunda Tedarik Zinciri Yönetimi (2014).

Sena Pakter received her Bachelor of Science (B.S.) in Manufacturing Systems Engineering from Sabanci University, Istanbul, Turkey, in 2013. She received her Master of Science (M.Sc.) degree in Data Analytics from the University of Warwick, Coventry, UK. She currently works as a Credit Strategy Associate at Garanti Bank, Istanbul, Turkey.

Soner Ulun received his B.Sc. and M.Sc. degree in Mechatronics Engineering from Sabanci University in 2011 and 2013 respectively. Currently, he is is a Ph.D. student at Nanyang Technological University, Singapore. His research interests include machine vision, multirobot systems, and motion control.