



# A Model for Analysis of Environmental Accidents Based on Fuzzy Logic

## Case Study: Exxon Valdez Oil Spill

Ana Claudia Golzio<sup>(✉)</sup> and Mirelys Puerta-Díaz

São Paulo State University, Av. Hygino Muzzi Filho, 737, Mirante, Marília,  
SP CEP 17525-900, Brazil

**Abstract.** This research aims to present a fuzzy-logic-based conceptual model for environmental accidents analysis, to reveal corporate social responsibility initiatives by companies responsible for the disasters. We studied one of the biggest environmental man-made disasters in history, the one that occurred on March 24, 1989 in Prince William Sound, Alaska when the oil tanker Exxon Valdez spilled 10.8 million gallons of American crude oil. The data was collected from the online database of the newspaper The New York Times for the timespan 03/24/1989–09/01/2017. As a central point of the research, we investigate ethical issues based on the mapping of an ethical vocabulary carried out in the corpus of the analyzed documents. The results show that the proposed model can be replicated, after some adjustments, to verify actions in accordance with the principles of corporate social responsibility for other environmental accidents.

**Keywords:** Fuzzy Rules-Based System · Corporate social responsibility · Data mining

## 1 Introduction

Environmental accidents can result in a series of damages to the population, leaving irreparable marks on the planet. In this context, the agent who caused the accident (directly or indirectly) may be asked by several sources to take measures to mitigate the damage. These sources could be legal, that is, agents could be legally penalized and forced to take corrective measures, or, due to some type of social pressure, agents could be led to reverse damages and generate benefits, favoring the quality of life of the affected population and the preservation of nature. The latter characterizes what is known as corporate social responsibility (CSR).

The second half of the 20th century witnessed a long debate on corporate social responsibility and, in recent decades, a renewed interest has arisen with the proposal of new concepts and interdisciplinary relations to study this topic [2, 3, 6, 7].

The echo in the interest of the scientific community on the topic is linked to the drastic increase in criticism of the business system. In the last decade, national and international companies have faced increasing demands due to transparency problems

and the occurrence of accounting, legal scandals and reputation failures or corporate collapses, such as the famous cases of Enron, WorldCom and Tyco in the United States.

Critical public opinion has led companies that operate in the environment to continually improve their social, environmental and ethical performance. Increased media interest and public pressure also reinforced stakeholder awareness [12]. In this scenario, corporate social responsibility goes beyond being seen as a potential competitive advantage for companies and it starts to be considered as a real strategic need [11].

The growing interest in CSR was accompanied by the substantial publication of CSR articles in journals on management, focusing on ethical and environmental issues [13]. In practice, CSR theories have four dimensions: profits, political performance, social demands and ethical values [3]. With regard to the latter dimension, theories are based on the ethical responsibilities of companies with society.

CSR aims to raise ethical and environmental issues within and outside organizations, insisting that organizations adopt socially responsible and sustainable attitudes, aiming at protecting the environment and the ethical regulation of the conduct of their members.

Awareness and attribution of the reasons for CSR are essential aspects that need to be evaluated, as they represent the main drivers of the type of assessment that society will make of the company and the crisis situation [5].

Although there has been a parallel growth in scientific publications that analyze this phenomenon, there are few critical approaches based on methods of data mining and fuzzy logic to identify the assignment and recognition of corporate social responsibility. Fuzzy logic is the logic underlying approximate, rather than exact, reasoning modes [15] and, therefore, it can be considered adequate for analyses involving subjective properties or inaccurate attributes, absent in the research that is traditionally conducted on the topic of interest to CSR.

This research aims to fill this gap that exists in the traditional literature, by examining the capacity of corporate response when facing environmental accidents based on the processing of massive databases on public opinion. Considering issues about environmental disasters, we believe it is interesting to find out if there was an initiative, over time, to assume corporate social responsibility by companies that caused major environmental disasters.

The main objective of this work is to propose a model, which gathers analyses of an ethical vocabulary and a Fuzzy Rules-Based System, to predict possible ethical and social responsibility responses given by companies that have caused environmental disasters reported in the media.

Our research is based on a case study and the model is applied to the environmental accident that occurred on March 24, 1989, in which the oil tanker Exxon Valdez, one of the ships belonging to the Exxon Mobile Company, ran aground in the north of Prince William Sound, in Alaska, spilling 42 million liters of oil and contaminating 1,990 km of coastline.

We believe that the conceptual model proposed here can be used as a basis for understanding the dynamics of other environmental accidents, as well as identifying ethical actions for assigning or recognizing corporate social responsibility in companies based on pressure by the public opinion.

This article is organized into four sections. In the first section the methodological issues are addressed. The second section presents the details of the model developed to deal with the problem and the pre-processing of the collected data, describing in detail each of its phases, as well as its application in the studied environmental accident. In the next section, the questions related to the validation of the model are described. In the final section, the results obtained in the previous sections and the scope of application of the model are discussed.

## 2 Methodology

The proposed research is descriptive and combines qualitative and quantitative methodological approaches based on theoretical assumptions of corporate social responsibility with special emphasis on the dimension of ethical values proposed by [3].

### Data Collection

As the main aim of our research is to propose a conceptual model based on fuzzy logic, our data was collected from the online news database of the newspaper The New York Times (NYT). The choice of this source of data is justified by the prestige, influence and extensive international academic reference to the New York Times newspaper, which was also portrayed in the scientific literature. For instance, in [4], the authors explain their preference of this data source over other newspapers, an aspect that also fits the purposes of our research, stating that one of the advantages of its use is the free access to the abstracts.

In addition, our research is part of the Project *Understanding Language and Opinion Dynamics using Big Data*, which specifically focuses on the study of the dynamics of opinion and the evolution of language based on data emitted by this kind of media.

The search and retrieval of data were conducted using the *Article Search API* to obtain news available on the developer platform (<https://developer.nytimes.com/docs/articlesearch-product/1/routes/articlesearch.json/get>). The use of the API's function *get/articlesearch.json* allowed filtering the search by date of publication of the news and keywords. With the application of filters it was possible to reduce the existing computational cost in handling large amounts of data thus making subsequent processes manipulable through current hardware and software.

The considered parameters were the following ones:

- begin\_date: 19890324;
- end\_date: 20170901
- facet\_fields: null
- Query: exxon
- Sort<sup>1</sup>: null

The retrieved news included the word 'exxon' in the title field, keywords and lead paragraph, the choice of this term allows to apply a more comprehensive search of the

<sup>1</sup> The parameter allows retrieving by relevance criteria or by the most recent publication data or by the oldest published on the NYT platform.

news about the company responsible for the accident and thus guarantees the retrieval of the entire universe of data in the temporary window under analysis. The considered temporary window is viable to study the dynamics of public opinion about the event, since the collection allowed to retrieve a representative data set from the date of the accident until the year 2017, including a total of 28 years of media coverage.

Following these search criteria, a total of 2001 news items were retrieved and incorporated into the file named 'DB Exxon 1989–2017.txt' (data set available at: <https://doi.org/10.6084/m9.figshare.13555661> as Attachment 1). Other criteria on the characteristics of the data records of the retrieved news are included in the 'Data collection, filtering and pre-processing' section.

### Methods and Techniques

In this work, we apply fuzzy system techniques, specifically designing a fuzzy rules-based system, to meet the research objective. Basically, fuzzy logic was applied as an attempt to make rational decisions in an environment of inaccurate information [16]. The research was initiated through the presentation of a conceptual model that includes in one of its phases the application of a fuzzy model in massive databases. Alternatively, the conceptual model based on fuzzy logic uses inference rules that allow adaptation to various contexts, especially those that involve the treatment of some degree of uncertainty present in the data analysis.

An important phase of the present work is based on the identification of the presence of an ethical vocabulary for the assignation and recognition of corporate social responsibility. The authors of the present research developed a categorization of this vocabulary into three main groups, described below:

- a) Concepts of assignation of responsibility,
- b) Concepts of recognition of responsibility and
- c) Neutral concepts (whose role is defined due to the proximity of the terms of assignation or recognition, that is, they can launch both functions depending on the context).

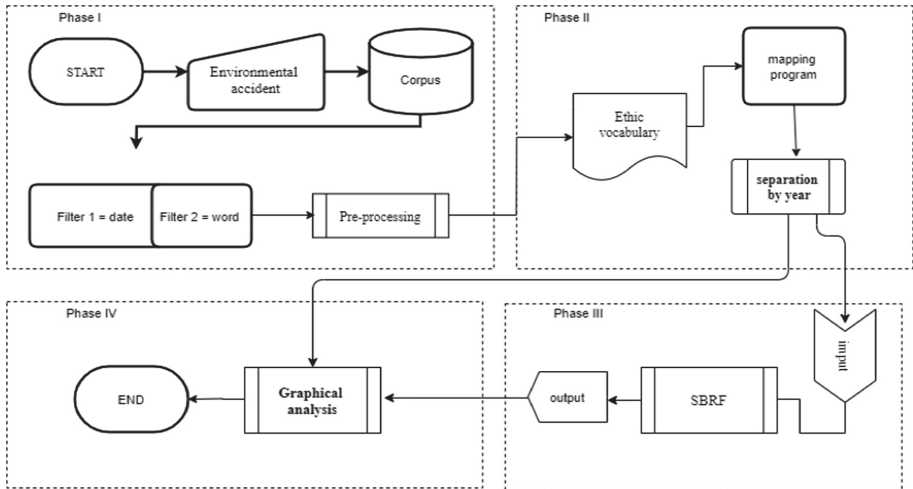
A description of how this vocabulary was mapped will be further explored in the next section, following the criteria (a)–(c) of categorization.

## 3 Development and Application of the Model

In the construction phase of the model, we developed a workflow that represents the model and is divided into four phases, shown in the Fig. 1 below.

We analyzed in detail the viability of each stage of the model, considering that the ability to abstract and understand the dynamics of public opinion in relation to companies that caused major environmental disasters is an important point that should guide the methodology. The database used to verify the hypothesis was composed of news collected from online newspapers, because, although they do not directly reflect public opinion, they indicate the way the accident affected the social environment.

Each stage of the model will be described below, indicating the application in the environmental accident caused by the ship Exxon Valdez concurrently.



**Fig. 1.** Environmental accident analysis model (Source: elaborated by the authors).

### Phase 1 - Data Collection, Filtering and Pre-processing

*Collection:* The environmental accident to be analyzed and the news database used in the analysis.

#### Collection Application:

Environmental accident to be analyzed: Oil spill by the tanker Exxon Valdez.

News database: Online corpus of the newspaper The New York Times (NYT).

*Filtering:* The model uses two types of filters, one by time period and the other by keyword. The purpose of applying the filters is to delimit the universe of analysis and ensure that the analysis is conducted on the largest number of news related to this environmental accident.

#### Filtration Application:

Filter by period: Only news published in the period from March 24, 1989, to September 1, 2017, were selected.

Keyword filter: Only news items containing the word (“exxon”) were selected.

After applying the filters, a total of 2001 news were retrieved from the NYT online database. Each news item received an identification number (ID) from 0 to 2000.

*Pre-processing*<sup>2</sup>: the metadata was filtered and some Natural Language Processing (NLP) methods were also applied.

Metadata filtering: the collected data were used to map the ethical vocabulary and to prepare the fuzzy system.

<sup>2</sup> The scripts applied in the processing were considered those published in the methodology published on the website <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/>. These were optimized and adapted according to the context of the research data and interests.

### Application of Metadata Filtering:

At first, it was necessary to remove irrelevant values and metadata for research purposes (“web\_url”, “print\_page”, “source”, “multimidia”, “headline”, “keywords”, “document\_type”, “news\_desk”, “type\_of\_material”, “\_id”, “word\_count”, “uri”), that is, some information about the organization of the data in the file that is not necessary for the application of the model. The fields considered were: “headline”, “lead\_paragraph”, “pub\_date”, “snippet” and “abstract”.

Metadata and unnecessary data were removed from the 2001 news selected from the NYT online database. The records were left only with the metadata and values of the fields that are relevant for the processing of the natural language and subsequent application of the mapping of the ethical vocabulary and the fuzzy system. It is important to mention that each news item was identified with the field “\_id” (denoted here by “ID”) that was assigned consecutively, starting at “0” until the value of “2000”.

1. PLN methods: this sub-phase included methods of tokenization, removal of stopwords, lemmatization and n\_grams and bigram in the news from the database.

The PLN methods were applied to the 2001 news retrieved and stored in the file ‘DB Exxon 1989–2017.txt’ (Attachment 1<sup>3</sup>).

Description of natural language processing at the morphological level:

- Tokenization: this technique was applied to transform a set of documents into a matrix, where each document is transformed into a list of words, and each generated word list will occupy a line in the matrix, using the Gensim library. The following symbols were removed: punctuation marks, isolated letters (a, e, o), numbers and special symbols present in the analyzed corpus.
- Removal of stopwords: the stopwords of common words in the language without seemingly relevant meaning, such as “I”, “that”, “for”, were removed using the NLTK library in the English version.
- Lemmatization, as part of the NLP at the morphological level, this technique is applied to transform the word in its “root” form.
- n\_grams: bigram was applied to identify those terms that appear together frequently, such as ‘social network’, etc.

After filtering and pre-processing, the data were used in the following phase of the model.

### **Phase 2 - Mapping the Ethical Vocabulary**

In the second phase, the mapping of the ethical vocabulary was done in the filtered and pre-processed database. Attachment 2 (available at <https://doi.org/10.6084/m9.figshare.13582898>) shows the deployment of the terms of the ethical vocabulary following the categorization described in the methodology:

- a) **Group 1** - Concepts of assignation of responsibility (23 concepts),
- b) **Group 2** - Concepts of recognition of responsibility (9 concepts) and

<sup>3</sup> Data set available at: <https://doi.org/10.6084/m9.figshare.13555661>.

- c) **Group 3** - Neutral concepts (whose role is defined due to the proximity of the terms of assignation or recognition, that is, they can perform both functions depending on the context) (6 concepts).

This phase was developed in four stages:

**First stage:** Identification of the news, in the filtered and pre-processed database, which contains the words of the ethical vocabulary.

**Second step:** Selection and ordering of only the news IDs that contain the words from the ethical vocabulary.

**Third stage:** The selected news that contained the terms from the ethical vocabulary were related to their respective terms of the ethical vocabulary and years of publication. With that, we present a table that contains the ID of the news, the year in which it was published and the words from the ethical vocabulary that are present in the news.

**Fourth stage:** The table developed in the third stage was divided by year, indicating the occurrence of words from the ethical vocabulary for each of the three groups of concepts (neutral, assignation and recognition).

#### Ethical Vocabulary Mapping Application:

**First and Second Stage:** In these stages we made a mapping and located those news (from the clean and pre-processed database) that contain words from the ethical vocabulary. This mapping was divided into three parts:

1. We identify the *Porter stem* of each word in the ethical vocabulary, using the website: <https://morphadorner.northwestern.edu/morphadorner/lemmatizer/example/>.
2. We searched the *Porter stem* of the words in the clean and pre-processed database using a searcher developed in Python [10] (Attachment 3 available at <https://doi.org/10.6084/m9.figshare.13567817>). In some cases, where the *Porter stem* changed some letters of the word in relation to the original word, we searched both versions, for example, in “duty”, the *Porter stem* turn back “duti”, so we seek “duty” and “duti”. The same happened with “penalty”, we searched “penalty” and “penalti” and with “culpability”, we seek “culpability” and “culpabl”. Our search program uses the resource *wildcard* to search for variations of the word in the ethical vocabulary. A *wildcard* is a symbol used to replace or represent one or more characters in the word. In particular, the asterisk “\*” is a *wildcard* is normally used at the end of a root word, when we want to search for variable endings of a root word. In our case we use the *Porter stem* of the word as the root of the search.
3. The search program, developed in Python, returns as a response a list with the ID of the news item that contains the searched word. We organized these data, relating the words of the ethical vocabulary searched with the ID of the news in which they appear. It is important to note that our search does not take into account the frequency of the word in the news, we only mark whether the word occurs in the news or not.

**Third and Fourth Stage:** In the third stage we related the data from the second stage to the year in which the news was published, that is, we showed the relationship between

the ID of the news, the year in which it was published and the words of the ethical vocabulary present in the news (where “1” indicates the presence of the word in the news and “0” indicates the absence). In the fourth step, we separated this information year by year (from 1989 to 2017) and calculated the occurrence of words from the ethical vocabulary for the group of neutral concepts, for the group of assignation concepts and for the group of recognition concepts. The table in Attachment 4 available at <https://doi.org/10.6084/m9.figshare.13567838> illustrates how the calculation was made.

In Attachment 4<sup>4</sup>, for example, we calculated that in the year 2000, we recorded 12 news items that have some words from the ethical vocabulary. Of these 12 news items, we have 276 possibilities of occurrences of words in the vocabulary belonging to the group of responsibility assignation, 108 possibilities of occurrences of words belonging to the group of responsibility recognition and 72 possibilities of occurrences of words from the group of neutral concepts. For each group of concepts, we counted the respective occurrences of the words. Thus, we have 4 occurrences of assignation concepts among the 276 possible occurrences for this group, that is 1%, we have 7 occurrences of recognition concepts among the 108 possible for this group, that is 6% and we have 4 occurrences of neutral concepts. among the 72 possible occurrences for this group, which also means 6%.

The same calculation was made for the other years analyzed by the proposed model. The data from all 4 stages of mapping the ethical vocabulary were organized in tables (Attachment 4 (See footnote 4) and Attachment 5<sup>5</sup>), and used in the next stages of application of the model. The data from the mapping of the ethical vocabulary were used to construct the variables of the fuzzy system, described below .

### Phase 3 - Fuzzy Rules-Based System

Lotfi Zadeh, in the 1960s [14], suggested an alternative set theory, less rigid than usual. This theory was called the fuzzy set theory.

In this theory, proposed by Zadeh, the change from pertinence to non-pertinence is gradual and not abrupt. Thus, in the fuzzy sets, for each element of the discourse universe, we have a corresponding degree of relevance in the fuzzy set that is given by a real number between 0 and 1. With the fuzzy sets, the possibility of interpreting non quantitative and vague, also increasing the need to seek mechanisms for inferences from these data. One of these mechanisms was the Fuzzy Rules-Based System<sup>6</sup> (FRBS).

A FRBS is a system that uses fuzzy logic to produce outputs from fuzzy inputs. In these systems the linguistic variables<sup>7</sup> they play a fundamental role, since the linguistic terms, translated by fuzzy sets, are used through a rule base to obtain a fuzzy inference relation in which an output is produced for each system input.

<sup>4</sup> Available at <https://doi.org/10.6084/m9.figshare.13567838>.

<sup>5</sup> Available at <https://doi.org/10.6084/m9.figshare.13567871>.

<sup>6</sup> Due to its multidisciplinary nature, the fuzzy rules-based system is known by several other names, such as “fuzzy rules-based inference system”, “expert fuzzy system”, “fuzzy model” and “controller logical fuzzy”.

<sup>7</sup> A *linguistic variable* is a variable whose values are fuzzy set names. For example, the temperature of a given process can be a linguistic variable assuming low, medium, high, etc. These values can be described by fuzzy sets.

In short<sup>8</sup>, a FRBS consists of four components connected, they are:

- 1) Fuzzification module;
- 2) Rule base module;
- 3) Inference module;
- 4) Defuzzification module.

**Application of FRBS:**

The FRBS, called “Mars” (Model for social responsibility analysis or *Modelo de análise de responsabilidade social* in Portuguese), was developed using Mamdani’s inference method and implemented using Matlab software, through the “Fuzzy Logic Designer” toolbox.

**Fuzzification and Defuzzification**

The FRBS has three input variables, they are: “assignment”, denoted by “a” (which refers to the occurrences of concepts of assignment of responsibility), “recognition”, denoted by “r” (which refers to occurrences of concepts of recognition of responsibility) and “neutral”, denoted by “n” (which refers to occurrences of neutral concepts). These data are obtained in the previous phase by mapping the ethical vocabulary, based on the tables in Attachment 4 (See footnote 4) and Attachment 5 (See footnote 5), and were used here in the construction of the system variables (Table 1).

**Table 1.** Definition of membership functions and characteristics of input variables.

Definition of the membership functions of the input variables: neutral (n), recognition (r) and assignment (a)			
Universe of discourse: [0; 12]			
Notation	Type	Delimiters	Terms
x1	Trapezoidal	[0 0 3 5]	Low
x2	Trapezoidal	[3 5 7 9]	Medium
x3	Trapezoidal	[7 9 12 12]	High

We present the membership function (trapezoidal type) of each one of these input variables. As they are identical in the three cases: neutral (n), recognition (r) and assignment (a). The values of the input variables are mapped by their respective membership functions that were constructed based on the occurrences (measured in percentage) of the words of the ethical vocabulary in the corpus.

The input variables were classified as “low”, “medium” and “high”, according to an equidistant partition of the speech universe, taking into account the results of the mapping of the ethical vocabulary. The universe of discourse was defined considering the maximum value obtained in the table in Attachment 5 (See footnote 5), in relation to the three variables (recognition, assignment and neutral).

<sup>8</sup> For more information about a FRBS, see [1].

We have in Mars an output variable called “indicator\_SR”, denoted by “i”. We present the membership functions (of the triangular type<sup>9</sup>) of the output variable in Table 2 below:

**Table 2.** Definition of the membership functions and the characteristics of the output variable.

Definition of the output variable membership function: indicator_SR (i)			
Notation	Type	Delimiters	Terms
z1	triangular	[-1 -1 -0,5]	assigned
z2	triangular	[-0,5 0 0,5]	no_evidence
z3	triangular	[0,5 1 1]	recognized

The universe of discourse of the output variable was defined considering the 3 values: -1, 1, 0. They refer, respectively, to the idea that “assigned” is a negative measure for the company responsible for the environmental disaster, “recognized” is a positive measure and 0 means that a value could not be assigned. The output variable was classified as “assigned”, “no\_evidence” and “recognized”, according to the results of the mapping of the ethical vocabulary and from an analysis in the database. The intervals of the triangular functions were constructed considering the idea of “around”, that is, we classified as “assigned” the values around -1, as “recognized” the values around 1 and “no\_evidence” the values around 0.

The defuzzification<sup>10</sup> method used was the LOM (last of maximums) because it showed a better adherence to our proposal to approach the problem.

**Rules and Inference**

The rules were built after a careful analysis of the database and the results of the mapping of the ethical vocabulary, taking into account the following premises:

1. If (r > a) and (n δ r), then (no\_evidence)
2. If (r > a) and (n > r), then (recognized)
3. If (r < a) and (n δ r), then (no\_evidence)
4. If (r < a) and (n > r), then (assigned)
5. If (r = a), then (no\_evidence)

We need at least 27 rules so that all combinations of terms for the input variables were attended, that is, the number of rules we built was  $3 \wedge 3 = 27$  rules.

The fuzzy rules used in Mars are defined in the table in Attachment 6<sup>11</sup>.

<sup>9</sup> Triangular-type functions were chosen, as it was the one that presented the best result, according to what was expected in the model.

<sup>10</sup> For more information on other defuzzification methods, see [1].

<sup>11</sup> Available at <https://doi.org/10.6084/m9.figshare.13582988>.

The inference was made based on the Mamdani method, for more details on the Mamdani method, we recommend the book by Barros and Bassanezi [1]. The Mars code developed on Matlab can be found in Attachment 7<sup>12</sup>.

### Phase 5 - Graphical Analysis of the Results

The results obtained in the previous phases were interpreted graphically. Based on the data from the mapping of the ethical vocabulary, it was possible to analyze the behavior of the ethical vocabulary over time, as indicated in Attachment 8<sup>13</sup>.

The figure in Attachment 8 (See footnote 13) indicates the dynamics, over time, of the ethical vocabulary found in the analyzed corpus. We identified that the year 1991 was the year with the highest peaks of terms of assignation or recognition of responsibility (present in the NYT news about the ecological disaster). In 1991, there was a greater presence of words such as *damage*, *accusation* and *guilt* in the news, which leads us to infer that NYT's journalistic discourse in 1991 had a predominance of the responsibility assignation vocabulary. This inference is also justified based on the annual analysis of the mapping of the ethical vocabulary in the news, since the year 1991 appears in the table in Attachment 5 (See footnote 5) (in the column of concepts of assignation of responsibility) with the second highest value of the average 0.048007 (5%) in relation to other years. This behavior in relation to the concepts of assignation even increased in 1992 to 0.058385 (6%).

Coincidentally, in the year 1991 the NYT reported that a judge accepted the Exxon Pact and the Exxon Valdez spill lawsuits ended with a \$900 million payment by Exxon over 10 years to a trust fund administered by three state officials and three federal officials [9]. In addition, the Exxon company was responsible for an additional payment of US \$ 100 million, necessary for the recovery of the affected area.

It is interesting to note that the term *compensation* (concept of assignation of responsibility) reached a peak in 2006, which may be related to the proximity of the deadline of 15 years to request a new financial compensation to the civil agreement of Exxon Corporation with the Department of Justice and the state of Alaska, valued at 900 million dollars (federal and state lawyers requested 92 million more for the recovery of the affected ecosystem).

In 2010, the debate about this environmental disaster started again after the Deep-water Horizon oil spill, which was an industrial disaster that started on April 20, 2010 in the Gulf of Mexico, in the Macondo Prospectus, operated by the BP company. This year registered a higher average and percentage value in neutral concepts (those that can play both the assignation and recognition role depending on the context) 0.116667 (12%).

This new oil spill was echoed in the international press with an emphasis on the NYT, as new information on the consequences of the 1989 Exxon spill was published; an example of this was the news about Exxon's new precautionary culture, which allowed the company to socialize opinions on the safety issues linked to the environmental disaster at Prince William Sound. On the other hand, that same year new information emerges about the difficulties that hinder the progress of the cleaning due to the lack of research and rules that hinder new tests in the affected area [8].

<sup>12</sup> Available at <https://doi.org/10.6084/m9.figshare.13567883>.

<sup>13</sup> Available at <https://doi.org/10.6084/m9.figshare.13583042>.

Other inferences can be made considering the crossing of the data obtained from the mapping of the ethical vocabulary over time, thus facilitating new discoveries in the NYT journalistic discourse on the phenomenon under study.

The developed FRBS was applied, considering the data provided in the table in Attachment 5 (See footnote 5). We show, in Attachment 9<sup>14</sup> an example of application of Mars in Matlab.

The example in Attachment 9 (See footnote 14) shows that in 1991, when we have a percentage of 5% of neutral concepts, 3% of recognition concepts and 5% of assigned concepts, Mars returns the value of Indicator\_SR as  $-1$ . This means that, in 1991, the NYT news about Exxon addressed issues related to the assignment of responsibility to the company in relation to the environmental disaster caused by Exxon Valdez. We observed that the FRBS result agrees with the analysis made previously from the results obtained in the mapping of the ethical vocabulary.

The results of applying Mars in all years collected in the corpus are described in the table in Attachment 10<sup>15</sup>.

In Attachment 11<sup>16</sup> we graphically show the result of the application of Mars in the corpus, the diameter of the circles reflects the occurrence of concepts from the ethical vocabulary in the year (column “News with concepts from the ethical vocabulary (%)” in Attachment 5 (See footnote 5)) and the colors green, red and purple represent, respectively, the classification made: “no\_evidence”, “assigned” and “recognized”.

We note, both in the table in Attachment 10 (See footnote 15) and in the graph (Attachment 11 (See footnote 16)) that Mars shows that in the years 1991, 1992, 1994, 1995, 2005, 2006, 2008 and 2012 the news published about Exxon by the NYT, highlights, in the majority, a certain assignment of social responsibility to the company due to the Exxon Valdez accident. While it was only in 1999 that we had some prominence in the news published about the recognition, by the company, of its responsibility in relation to the oil spill in Alaska. We also note that in most of the years analyzed in the corpus, we do not have enough evidence to conclude the content of the news published by the NYT about Exxon.

## 4 Model Validation

A scientific model must be reliable, that is, it must be possible to be replicated and validated. The validity of the model must be external and internal. Internal validity concerns the analysis of the causality relationship in the model independently of subjective factors, that is, we must answer the question: Can we obtain the same type of relationship between the elements by other means, in addition to those used in a specific experiment? External validity is related to the generalizability of the model.

In the proposed model, reliability and external validity are guaranteed, considering that the proposed model is described in detail and can be easily replicated for analysis of other environmental disasters, after some adjustments of the data, mainly with regard

<sup>14</sup> Available at <https://doi.org/10.6084/m9.figshare.13583081>.

<sup>15</sup> Available at <https://doi.org/10.6084/m9.figshare.13583123>.

<sup>16</sup> Available at <https://doi.org/10.6084/m9.figshare.13583216>.

to the input variables of Mars. The internal validity will be evaluated by comparing the results obtained in Mars with a manual classification of the specialist (expert) list (in: “no\_evidence”, “assigned” and “recognized”), of 30% (10 random years) of the 29 years analyzed in the corpus. The individual classification of each article in the 10-year period can be found in Attachment 12<sup>17</sup>. We present the results of the validation in Table 3 below:

**Table 3.** Most frequent responses annually from the expert assessment on the nature of the news.

Year	Number of news with ethical vocabulary	Expert assessment	Results in the Mars	Comparison
1989	41	no_evidence	no_evidence	equal
1990	42	no_evidence	no_evidence	equal
1991	48	assigned	assigned	equal
1994	23	assigned	assigned	equal
1996	8	no_evidence	no_evidence	equal
1999	15	no_evidence	recognized	different
2000	12	no_evidence	no_evidence	equal
2008	9	assigned	assigned	equal
2014	5	no_evidence	no_evidence	equal
2017	7	no_evidence	no_evidence	equal

Although the error rate seems high (10%) when analyzed year by year, we observed that 30% of the years analyzed were tested. We have a total of 518 news items with some term from the ethical vocabulary (according to Attachment 4 (See footnote 4)) and in total we manually classify 210 news items (referring to the 10 years tested), which represents 40% of the result of mapping the ethical vocabulary. The year 1999 presents only 15 news items and the application of the model presented an error in the classification of 4 of them. From these data, we can infer that the model had an error in 4 out of 210 news classified manually, reducing the error rate to less than 2%. Considering the large number of tested news and the complexity involved in the analysis in question, we consider that the model performs well in the classification of news.

## 5 Analysis of Results and Some Comments

After completing the application of the model, here we present a brief analysis of the obtained results as well as a discussion of the scope of application and the limitations of the proposed model.

According to our model, in the years 1991, 1992, 1994, 1995, 2005, 2006, 2008 and 2012 the news published about Exxon brings information on topics related to the assignation of social responsibility to Exxon for the environmental disaster in Alaska.

<sup>17</sup> Available at <https://doi.org/10.6084/m9.figshare.13583555>.

Regarding the events that occurred in those years, we highlight that in the year 1991 a federal judge accepted a package of US\$ 1 billion in criminal and civil agreements to close the state and federal lawsuits against Exxon Corporation. Subsequently, in 1992, transcripts of telephone conversations were released among employees of the oil industry during the response to the environmental disaster. A federal jury in 1994 imposed a penalty of \$ 5 billion in punitive damages to 34,000 fishermen and other Alaskans, but the American public opinion found the sums imposed on Exxon to be totally inadequate to prevent it from being penalized in the future. In 1995, Exxon announced a donation of US \$ 5 million for the protection of the tiger habitat in the wild, which, in our opinion, is a strategy to manipulate the public opinion, since the company was accused of not being fast enough during the oil spill. Lawrence G. Rawl, who led the review of Exxon's operations in 1980 and shaped the company's response to the Exxon Valdez oil spill, died at the age of 76 in 2005. A year later, the federal appeals panel reduced the punitive damages from \$ 4.5 billion against Exxon Mobil to \$ 2.5 billion; arguing that the company's negligent conduct was unintentional. In 2008, the United States Supreme Court further reduced the claim to \$ 500 million. In 2012, the book "Private Empire" was published, revealing the true extent of the ExxonMobil corporation's power in American and foreign politics.

Through the facts that occurred, we realized that the classification obtained through our model as "assigned" is related to the relevant events in the same period.

We also highlight that, during the reading of the database, for the manual classification of the specialist in relation to the corpus, in the 210 analyzed news, only 21 of them were classified manually as "recognized", which means that only 10% of the news who have some term in the ethical vocabulary indicate the theme of recognizing social responsibility. The news classified manually as "no\_evidence" was 118, that is, more than half of the classified news did not have enough evidence to be classified as dealing with the topic of assignation or recognition of social responsibility, and the news classified as "assigned" was 71 of the 210 tested, which represents 34% of them.

The results of the annual analysis of the model show that 3% of the years that were analyzed indicate themes related to the recognition of social responsibility by Exxon; 28% refer to themes surrounding the assignation of responsibility for causing the accident to Exxon, and 69% do not have enough evidence to indicate whether the most popular texts about the accident of the year point to assignation or recognition of social responsibility.

Despite comparing different magnitudes, we noticed that the results are similar in both types of analysis (by year and by news). This result shows that, despite the complexity of an analysis of this kind in which the model is applied considering isolated words in the text and not the semantic aspect, the proposed model proved to be adequate to develop this type of analysis. We also highlight the advantage of the high potential for adapting this model to the study of assigned and recognized social responsibility in relation to other environmental accidents.

**Acknowledgments.** We would like to thank professors Mariana Claudia Broens, Jose Arthur Quilici Gonzalez and Maria Eunice Quilici Gonzalez for the helpful supervision and indications during the fundamental stages of this research. This paper was developed with the collaboration of the team of project "Understanding opinion and language dynamics using massive data", financed

by The São Paulo Research Foundation - FAPESP (process number: 2016/50256-0), to whom we thank the opportunity to conduct this research. Ana Claudia Golzio also received support FAPESP (process number: 2019/08442-9) during the development of this paper, and Mirelys Puerta-Díaz received funding from Coordination for the Improvement of Higher Education Personnel (CAPES) - Financial Code 001.

## References

1. Barros, L.C., Bassanezi, R.C.: Tópicos de Lógica Fuzzy e Biomatemática. Unicamp/IMECC (2015)
2. Campbell, J.L.: Why would corporations behave in socially responsible ways? An institutional theory of corporate social responsibility. *Acad. Manag. Rev.* **32**(3), 946–967 (2007)
3. Garriga, E., Melé, D.: Corporate social responsibility theories: mapping the territory. *J. Bus. Ethics* **53**(1–2), 51–71 (2004). <https://doi.org/10.1023/B:BUSI.0000039399.90587.34>
4. Hicks, D., Wang, J.: The New York times as a resource for mode 2. *Sci. Technol. Hum. Values* **38**(6), 851–877 (2013). <https://doi.org/10.1177/0162243913497806>
5. Janssen, C., Sen, S., Bhattacharya, C.B.: Corporate crises in the age of corporate social responsibility. *Bus. Horiz.* **58**(2), 183–192 (2015). <https://doi.org/10.1016/j.bushor.2014.11.002>
6. Matten, D., Moon, J.: “Implicit” and “explicit” CSR: a conceptual framework for a comparative understanding of corporate social responsibility. *Acad. Manag. Rev.* **33**(2), 404–424 (2008)
7. McWilliams, A., Siegel, D.: Corporate social responsibility: a theory of the firm perspective. *Acad. Manag. Rev.* **26**(1), 117–127 (2001)
8. NYT: Since Exxon Valdez, Little Has Changed in Cleaning Oil Spills. Print edition, p. A23, 25 June 2010. <https://www.nytimes.com/2010/06/25/us/25clean.html?searchResultPosition=5>
9. NYT: Judge Accepts Exxon Pact, Ending Suits on Valdez Spill. Print edition, 9 October 1991. <https://www.nytimes.com/1991/10/09/us/judge-accepts-exxon-pact-ending-suits-on-valdez-spill.html>
10. Python Software Foundation: Gensim 3.8.3: project description (2020). <https://pypi.org/project/gensim/>
11. Story, J., Neves, P.: When corporate social responsibility (CSR) increases performance: exploring the role of intrinsic and extrinsic CSR attribution. *Bus. Ethics A Eur. Rev.* **24**(2), 111–124 (2015). <https://doi.org/10.1111/beer.12084>
12. Våland, T., Heide, M.: Corporate social responsiveness: exploring the dynamics of “bad episodes.” *Eur. Manag. J.* **23**(5), 495–506 (2005). <https://doi.org/10.1016/j.emj.2005.09.005>
13. Visser, W.: Corporate social responsibility in developing countries. In: *The Oxford Handbook of Corporate Social Responsibility* (2008). <https://doi.org/10.1093/oxfordhb/9780199211593.003.0021>
14. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
15. Zadeh, L.A.: Fuzzy logic. *Computer* **21**(4), 83–93 (1988). <https://doi.org/10.1109/2.53>
16. Zadeh, L.A.: Is there a need for fuzzy logic? *Inf. Sci.* **178**(13), 2751–2779 (2008). <https://doi.org/10.1016/j.ins.2008.02.012>