

Article

Automated Customer Complaint Processing for Water Utilities Based on Natural Language Processing—Case Study of a Dutch Water Utility

Xin Tian ¹, Ina Vertommen ¹, Lydia Tsiami ^{1,2,*}, Peter van Thienen ¹ and Sotirios Paraskevopoulos ^{1,3}

- ¹ Department of Sustainability & Transitions, KWR Water Research Institute, Groninghaven 7, 3430 BB Nieuwegein, The Netherlands; xin.tian@kwrwater.nl (X.T.); ina.vertommen@kwrwater.nl (I.V.); peter.van.thienen@kwrwater.nl (P.v.T.); sotirios.paraskevopoulos@kwrwater.nl (S.P.)
- ² Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Iroon Politechniou 5, Zografou, 157 80 Athens, Greece
- ³ Department of Watermanagement, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
- * Correspondence: lydia.tsiami@kwrwater.nl

Abstract: Most water utilities have to handle a substantial number of customer complaints every year. Traditionally, complaints are handled by skilled staff who know how to identify primary issues, classify complaints, find solutions, and communicate with customers. The effort associated with complaint processing is often great, depending on the number of customers served by a water utility. However, the rise of natural language processing (NLP), enabled by deep learning, and especially the use of deep recurrent and convolutional neural networks, has created new opportunities for comprehending and interpreting text complaints. As such, we aim to investigate the value of the use of NLP for processing customer complaints. Through a case study about the Water Utility Groningen in the Netherlands, we demonstrate that NLP can parse language structures and extract intents and sentiments from customer complaints. As a result, this study represents a critical and fundamental step toward fully automating consumer complaint processing for water utilities.

Keywords: artificial intelligence; customer complaint processing; natural language processing; water sector



Citation: Tian, X.; Vertommen, I.; Tsiami, L.; van Thienen, P.; Paraskevopoulos, S. Automated Customer Complaint Processing for Water Utilities Based on Natural Language Processing—Case Study of a Dutch Water Utility. *Water* **2022**, *14*, 674. <https://doi.org/10.3390/w14040674>

Academic Editors: Christos Makropoulos and Panagiotis Kossieris

Received: 10 December 2021

Accepted: 16 February 2022

Published: 21 February 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water utilities often consider customer complaints to be a valuable source of information for identifying system malfunctions and improving their services. Traditionally, complaints are handled by phone operators. The essential information abstracted from verbal communication can only be recorded concisely. This can occasionally result in misunderstandings and requires water utilities to maintain an adequate number of telephone operators to handle incoming calls, particularly during rush hours or in the event of a malfunction affecting a larger area. In recent years, an increasing number of (Dutch) water utilities have put an online system in place for customers to easily submit complaints (see Appendix A). This enables information to be stored in a more organized and efficient manner. This is a critical step in advancing the digitalization of the water sector, since the information provided by customers is often more detailed than short-hand transcriptions made by phone operators. Additionally, the accuracy of this information is significantly improved, particularly with regard to names, addresses, and zip codes. While complaints tend to be collected digitally, their content still needs to be processed manually—e.g., extracting critical information, classifying the major issue, and responding to the customer with a solution.

Natural Language Processing (NLP) is a subfield of computer science that aims to automatically process and produce human language content. NLP can be divided into

Natural Language Generation (NLG), which focuses on automatic text generation (e.g., the automatic generation of responses [1,2]), and Natural Language Understanding (NLU), which aims to achieve machine reading comprehension (e.g., syntactic and semantic sentence analysis and discourse and dialogue structure) [3,4]. NLP can be used for a variety of tasks and applications such as machine translation, information retrieval, dialog and question answering, and opinion mining. Traditionally, NLP deals with texts in a rule-based way [5] (e.g., regular expressions) or is simply coupled with a machine learning algorithm (e.g., decision trees or support vector machines) [6,7], which demands a significant degree of domain expertise and a considerable number of manual operations [8]. Recently, with the rapid development of deep learning and the use of word vectorization (i.e., word embedding) [9], NLP has entered a new era, as deep learning models have alleviated the need for labelled datasets and enabled researchers to train models using large-scale corpora of texts [1,10]. One prominent example is the Bidirectional Encoder Representations from Transformers (BERT) model, based on the so-called Transformer neural network architecture, which has more than 300 million parameters [11] and can effectively consider the context of words by examining the words that precede and follow them [12]. Today, models comparable to BERT are among the most useful tools used to perform NLP tasks [2,13,14].

Across multiple sectors, NLP plays a valuable role in the automation of text processing, particularly about handling customer reports and complaints. Recent studies show that NLP is used, for instance, (i) by power companies to auto-label the (sub)classes of customer complaints in order to improve service quality [13] and (ii) by banks to identify negative sentiments from social media posts to provide quality responses to customers [15]. In doing so, NLP enables more efficient operations, reduces operating costs, increases customer satisfaction, etc. Despite the value of NLP, as demonstrated by the applications in a variety of sectors, the water sector, which has to deal with thousands of customer enquiries and complaints every year, has not explored the potential value of NLP to automate customer complaint processing yet. Therefore, we aim to identify the value of NLP in processing customer complaints about water problems in this case study by investigating whether an NLP model can understand the syntactic meanings of words in a particular context, classify a customer complaint correctly, identify the customer's emotion in the complaint, and recognize the intent and request mentioned in the complaint.

This case study is organized as follows. Section 2 describes the materials, including the study area, customer complaint data, and the NLP tasks and models. Section 3 presents the syntactic and semantic analyses. In Section 4, the advantages and limitations of the NLP models are discussed, together with some ideas for their future improvement. Our conclusions are given in Section 5.

2. Materials and Methods

2.1. Customer Complaints about Drinking Water Issues

Our study focuses on the Province of Groningen in the northern Netherlands (Figure 1), where drinking water is supplied by the Water Utility Groningen (in Dutch: Waterbedrijf Groningen, referred to as WBG hereafter). WBG is one of the Dutch water utilities that have begun collecting customer complaints digitally. Data collection and management have noticeably improved as a result of the deployment of a database and a well-designed website (an example is provided in Appendix A, Figure A1). When submitting complaints, customers need to select a category for the problem, describe the issue in detail, and include their address or customer number so that the water utility can quickly track the reported problem.

With 597,000 inhabitants spread over an area of 2960 km², WBG produces 44 million m³ drinking water per year and distributes it via its 5000 km-long water distribution networks. We retrieved a database of 4730 customer complaints collected by WBG via telephone, email, and webpage between February 2013 and January 2021 (i.e., ≈590 complaints per year). This database is a subset of the full complaint datasets with detailed problem descriptions, which can be used for NLP.



Figure 1. The ten water utilities in the Netherlands. The study area is the Water Utility Groningen (in Dutch: Waterbedrijf Groningen) in the north of the Netherlands. Source: <https://www.vewin.nl/sector-in-beeld> (accessed on 17 December 2021).

Figure 2 shows the geographical distribution of customer complaints in the province. The majority of complaints came from densely populated cities of the province, such as the cities of Groningen (c.a., 231,000 inhabitants) and Veendam (c.a., 28,000 inhabitants).

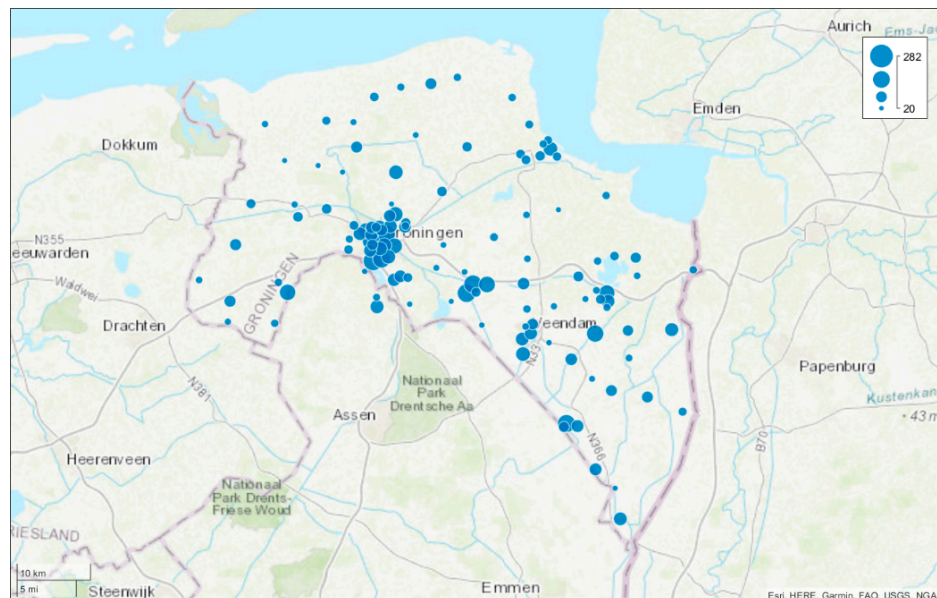


Figure 2. The spatial distribution of customer complaints in the province of Groningen between February 2013 and January 2021. The size of the blue nodes indicates the number of reports within this period.

Figure 3 illustrates some basic statistics about customer complaints: (a) The number of reports received per year reflects the possible range of complaints over years, ranging from 300 to 800. Moreover, it implies that water utilities need to consider/enhance their flexibility in allocating the number of employees to handle customer complaints. (b) Mondays are

commonly the busiest day of the week, with the number of complaints frequently double or triple that received on Fridays. This is because non-emergency complaints submitted on weekends are only processed (and thus registered) on the following business day—i.e., Monday or Tuesday (if Monday happens to be a national holiday). (c) December to February received significantly more complaints (50+ per month) than the rest of the year (20–50 per month). Last but not least, a peak of approximately 150 complaints (roughly five per day) can be seen in August 2018, which was the result of multiple payment issues that month.

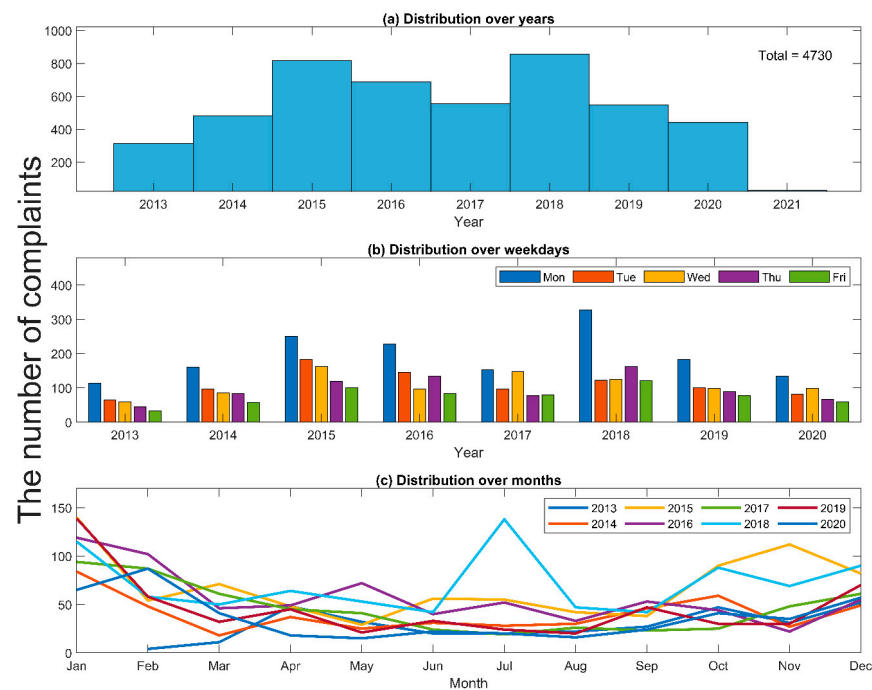


Figure 3. Basic statistics for customer complaints at WBG, including (a) the number of customer complaints received per year, (b) the total number of customer complaints on each day of the week, and (c) the total number of complaints in each month of the year. Complaints were received between February 2013 and January 2021, so the year 2021 shows a small number of complaints.

2.2. Natural Language Processing

Complaints from customers are a type of natural language data. Regardless of the language in which the complaint is written (in our case, Dutch), linguists and computer scientists have been looking for an interactive way of teaching computers how to understand the content of a conversation and then assist us in analyzing and processing (colossal) amounts of natural language data in a variety of domains. NLP emerged naturally as a result of these efforts. From symbolic NLP (1950s–1990s) to statistical NLP (1990s–2010s), state-of-the-art (neural) NLP techniques have entered a new era, benefiting greatly from the rapid development of deep learning and the information explosion. In this study, we focus solely on the latest development of NLP techniques and, more importantly, how they can be used to help the water industry automate and simplify consumer complaint processing.

Customers frequently include the following information in their complaints: greetings, a description of the situation, and a request for help and support from the water utility. Often, the latter two are semantically and contextually related. For instance, a customer reported: *There is a leakage from my pipe. Please fix it asap.* Note that we use *Italic font* to denote texts derived directly from the original or translated complaints. We should be able to infer that the pronoun *it* in the second sentence refers to *leakage/pipe* from the first sentence and that the situation is rather urgent based on the word *asap*. As a result, we aim to carry out three analyses in an order—namely, lexical analysis (Section 2.2.1), syntactic analysis (Section 2.2.2), and discourse analysis (Section 2.2.4), which is based on word vectorization (Section 2.2.3).

2.2.1. Lexical Analysis

Lexical analysis [16], as the first and most important step of NLP, aims to parse a sentence by interpreting words. To begin, words are tokenized, which involves breaking a sentence into smaller units of words (commonly referred to as tokens). Following that, words are lemmatized. In other words, they are transformed into their corresponding word stems. For instance, *am*, *are*, *is*, *was*, *were*, and *will be* all derive from the same root verb, *to be*. Though these words have distinct spellings and grammatical functions in a conversation, these distinctions introduce complexity when processing texts with identical meanings. As a result, regardless of the grammatical aspect of the word, they are commonly characterized as their stem words. The tokenization and lemmatization of words are often based on pre-defined rules for a particular language database (e.g., [17]). Finally, we also need to remove stop words, which are a collection of frequently used words that carry little meaning for textual data (e.g., *the*, *in*, *a*). To process texts in a variety of languages, we need to adopt suitable stop words. In this study, we use an open-sourced stop word list, including about 300 stop words in Dutch [11].

2.2.2. Syntactic Analysis

The syntactic analysis [18] aims to parse the functions of words in a sentence based on the lemmatized tokens. It is critical to distinguish between, for instance, *book* as in *book a room* and *book* as in *read a book*. By carrying out so-called part of speech (POS) analysis [19], a trained algorithm typically looks for the type/function of the word. POS is a linguistic term that refers to the classification of words, which includes nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, interjections, numerals, articles, determiners, and symbols. By doing so, machines are able to determine that *book* is a (transitive) verb in the first example (because of the following nominal object) and a noun in the second one (because of the preceding article and verb).

After implementing the POS analysis, it is also necessary to parse the dependency of words in a sentence, which is accomplished by statistically examining the word association and the sentence structure [20]. The dependency parsing process enables the algorithm to discern essential information as humans do, such as the root of the sentence, the principal (nominal) subject, and occasionally an open clausal complement (i.e., the verb following the modal verb), as well as the descriptive information. In this study, we adopt a word-based dependency grammatic analysis rather than a constituent-based one [21]. As the names imply, the word-based method makes it easier to express the relationship between each word in a sentence. When conducting a dependency analysis, we typically regard the verb, particularly when transitive verbs and their associated objects are involved, as the central component of the sentence (so-called headwords). The dependent word, or the so-called child word, functions as an modifier of the headword. Additionally, prepositions and their associated objects, which can also convey important information about time and space, also need to be considered. For example, a user may specify that the leakage occurs *in the basement* or *beneath the water sink*. The dependency of words can often be displayed in a table or by a dependency tree [22].

2.2.3. Vectorization of Words

While humans comprehend a language via words, computers comprehend it via numbers. Especially in NLP, word vectors (also referred to as word embeddings) are used to enable computers to understand language. Word vectors are a series of real numbers representing words as a vector in a high-dimensional geometric space. Word vectors facilitate the arithmetic operations of words. In other words, by using the l2-norm of two word vectors (i.e., $\|wv_1 - wv_2\|_2$), we can easily calculate the geometric distance between two words. The smaller the l2-norm value is, the more semantically similar the two words are. The word vectorization provides a way to map all words into a high-dimensional space that is suitable for machine learning algorithms to classify or cluster texts. In other words, word vectors are the actual connector between linguistics and NLP. To obtain the

word vectors, we can either train a computationally expensive model based on a large-sized self-defined corpus or use a pre-trained model that performs well on a particular corpus. With the advancement of deep learning (DL) and NLP models, there are now multiple accurate pre-trained models accessible for general use. In this study, we adopted an open-sourced corpus incorporating non-scientific and conversational texts in Dutch for word vectors [23]. Based on the distance of word vectors, we could assess the similarity of complaints. Therefore, when receiving new complaints, one could find a similar processed complaint and respond to customers in a similar way.

2.2.4. Discourse Analysis

Vectorized words enable us to infer the emotions of customers from their texts, which is also called sentiment analysis. This is a useful reference for water utility staff when assessing customers' attitudes and level of satisfaction. Specifically, the water utility can also track the total sentiment score year after year to assess the service satisfaction level (or more broadly, carry out an evolution of the politeness of the public). There are two ways to implement sentiment analysis. We can manually mark a considerable number of examples with a score between -1 (absolutely dissatisfied) and 1 (absolutely satisfied). Then, using the input (vectorized words in a sentence) and output (scores), a regression model (e.g., a support vector machine) can be trained to predict the sentiment score of new complaints. The second way to do this is to employ a pre-trained model in which words have been annotated with sentimental labels based on a large corpus. For instance, in this research we adopt the open-sourced model and corpus provided by TextBlob, which supports multiple languages, including English and Dutch [24]. The use of a pre-trained model is preferable because it lowers the need for skilled staff to label sentiments for thousands of sample complaints.

In addition to sentiment analysis, another key component of discourse analysis is intent recognition. Although not every NLP project involves intent recognition, it is vital when we need to interactively respond to a conversation (e.g., when a chatbot is deployed). Due to the domain-specific nature of intent recognition, we have to specify the categories of training samples manually (e.g., Appendix B). With the defined categories, the intent recognition problem can be effectively transformed into a typical machine learning classification problem—i.e., predicting the category to which a new complaint belongs.

2.3. Tools and Software

In this study, we utilized two Python packages, Spacy and Rasa, for conducting NLP tasks. Spacy is an open-source software library for advanced NLP implemented in the programming languages Python, R, and Cython [25]. Spacy makes use of a convolutional neural network model based on the library Thinc [26] and a transition-based approach [27]. Spacy is employed in this study to perform the majority of NLP tasks, except intent recognition. The latter was dealt with by Rasa. Rasa is an open-source contextual AI built based on Spacy [28]. Rasa provides flexibility in creating customized and automated interactions between humans and machines, which aligns with our goal of customizing intent recognition. Rasa (version 3.0+) uses a new state-of-the-art lightweight, multitask transformer architecture for NLU: Dual Intent and Entity Transformer (DIET). DIET is a multi-task transformer architecture that is able to handle both intent classification and entity recognition together. Readers can refer to the architecture of the DIET model for more details [29].

3. Results

In this section, we present the findings of applying NLP to process customer complaints, aiming at understanding the extent to which machines can automate textual message processing.

3.1. Lexical and Syntactic Analysis of Customer Complaints

The following example complaint is used in subsequent sections to explain how to process the complaint using NLP techniques. Since the complaints were written in Dutch, we added an English translation where necessary to make the text more comprehensible to the reader. However, note that all analyses were performed on Dutch texts, rather than English ones (Box 1).

Box 1. An example complaint received by the water utility WBG.

Original customer complaint in Dutch:
 Geachte heer/mevrouw, Sinds het wisselen van de watermeter is de druk vrij laag. Vooral met douchen is dit hinderlijk. Hierbij een vriendelijk verzoek om dit probleem te verhelpen.
 Translated complaint in English:
 Dear sir or madam, since the replacement of the water meter, the water pressure is rather low, especially when taking a shower. Hereby we kindly request you to fix this problem.

We first conducted a lexical (word) analysis (introduced in Section 2.2.1) to identify the lemma of each word. As we can see from Table 1, the distinction between original words and lemmas occurs with verbs, which often have multiple grammatical aspects. For example, the word *is* corresponds to the lemma *zijn* (*to be*). Next, stop words are identified and denoted as ‘TRUE’ in the column ‘is_stop’ of Table 1. Additionally, punctuation marks in sentences, such as commas and full stops, are also annotated, as shown in the column ‘is_punct’ of Table 1.

After the lexical analysis of words, we processed their dependency and the POS (introduced in Section 2.2.2). The POS is shown in the column ‘pos’ of Table 1. Note that every algorithm or model has errors. In this example text, 2 out of 32 words were labeled with a wrong POS—namely, *wisselen* and *vrij*. *Wisselen* (*replace*) can work both as noun (as in our example) or verb. *Vrij* (*rather*) functions as an adverb and is used to describe the extent to which the pressure is low (as in our example). It can also mean *free* used as an adjective to describe time or a product. The algorithm failed to assess the POS for *vrij* and *wisselen*, mainly due to the coverage of similar samples in the training dataset and the order of the words in sentences. In Dutch, inversion, which means that the order of the verb and the subject in a sentence are switched, is frequently used. We also tested an un-inverted sentence, *De druk is vrij laag* (*The pressure is rather low*), in which *vrij* can be correctly labeled as an adverb for its POS. Although errors are sometimes inevitable, they can be further analyzed together with word dependencies, as discussed in the next paragraph.

We applied dependency parsing (introduced in Section 2.2.2) to extract the grammatical structure of a sentence—i.e., the grammatical relationship between the head words and child words. According to the dependency tree shown in Figure 4, the customer began the sentence with a signal adverb *hierbij* (*hereby*), followed by the request (*verzoek*), which was described by a determiner and an adjective modifier. Next, the request was followed by detailed information about resolving (*verhelpen*) the problem (*probleem*). By observing the orientation of arcs, the connection of words is presented. Furthermore, we extracted critical information from the sentence by examining nouns and verbs, and sometimes adverbs, as well as their dependencies. In this instance, the primary structure of the sentence recognized by the model is *hereby* (signal adverb) -> *request* (noun) -> *solve* (verb) -> *problem* (noun). This is a key step in machines extracting the underlying meaning of a sentence and is also necessary for subsequent analyses, such as comparing sentence similarities or evaluating sentence sentiments. A detailed list of dependency labels can be found in the following reference [30].

Table 1. Original text and tokenized text with lemma, part of speech, stop words, and punctuation. Note that the column ‘translation’ is only added to facilitate reading by non-Dutch readers. The translation is not needed for the lexical and syntactic analysis of the complaints.

Number	Original Text	Translation	Lemma	Is_Stop	Is_Punct	POS	DEP
1	Geachte	Dear	geacht	FALSE	FALSE	ADJ	adjectival modifier
2	heer/mevrouw	sir/madam	heer/mevrouw	FALSE	FALSE	NOUN	nominal subject
3	,	,	,	FALSE	TRUE	SYM	punctuation
4	Sinds	Since	sinds	TRUE	FALSE	ADP	case marking
5	het	the	het	TRUE	FALSE	DET	determiner
6	wisselen	change	wisselen	FALSE	FALSE	VERB	modifier of nominal
7	van	of	van	TRUE	FALSE	ADP	case marking
8	de	the	de	TRUE	FALSE	DET	determiner
9	watermeter	water meter	watermeter	FALSE	FALSE	NOUN	modifier of nominal
10	is	is	zijn	TRUE	FALSE	VERB	copula
11	de	the	de	TRUE	FALSE	DET	determiner
12	druk	pressure	druk	FALSE	FALSE	NOUN	nominal subject
13	vrij	rather	vrij	TRUE	FALSE	ADJ	adverbial modifier
14	laag	low	laag	FALSE	FALSE	ADJ	ROOT
15	.	.	.	FALSE	TRUE	SYM	punctuation
16	Vooral	Mostly	vooral	TRUE	FALSE	ADV	adjectival modifier
17	met	with	met	TRUE	FALSE	ADP	case marking
18	douchen	showering	douchen	FALSE	FALSE	AUX	oblique nominal
19	is	is	zijn	TRUE	FALSE	VERB	copula
20	dit	this	dit	TRUE	FALSE	PRON	nominal subject
21	hinderlijk	annoying	hinderlijk	FALSE	FALSE	ADJ	ROOT
22	.	.	.	FALSE	TRUE	SYM	punctuation
23	Hierbij	Hereby	hierbij	FALSE	FALSE	ADV	ROOT
24	een	a	een	TRUE	FALSE	DET	determiner
25	vriendelijk	friendly	vriendelijk	FALSE	FALSE	ADJ	adjectival modifier
26	verzoek	request	verzoek	FALSE	FALSE	NOUN	parataxis
27	om	for	om	TRUE	FALSE	ADP	marker
28	dit	this	dit	TRUE	FALSE	DET	determiner
29	probleem	problem	probleem	FALSE	FALSE	NOUN	object
30	te	to	te	TRUE	FALSE	ADP	marker
31	verhelpen	resolve	verhelpen	FALSE	FALSE	VERB	clausal modifier of noun (adjectival clause)
32	.	.	.	FALSE	TRUE	SYM	punctuation

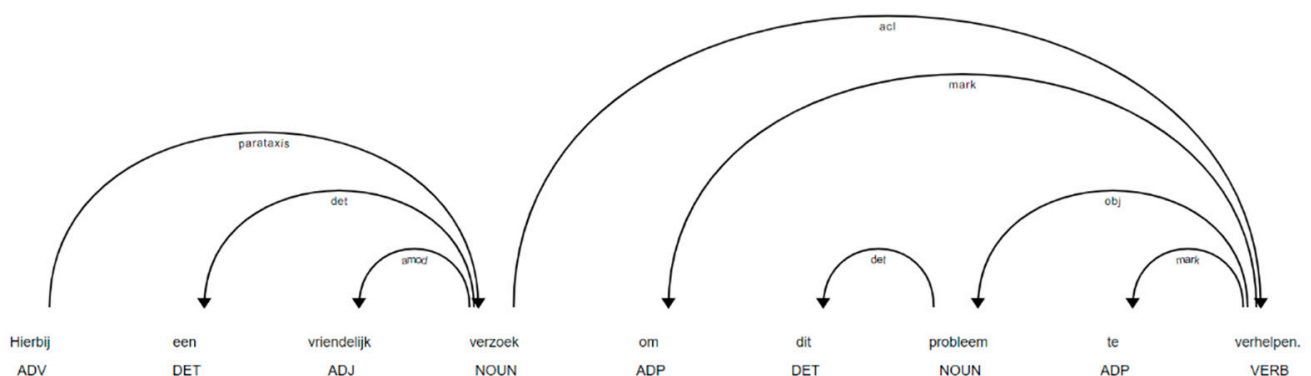


Figure 4. A dependency tree of a customer complaint, where the customer kindly requested the water utility WBG to resolve their issue of low water pressure. The dependency tree depicts the grammatical relationship between words in the sentence. English translation of this sentence: “Hereby a friendly request to solve this problem”.

3.2. Similarity and Sentiment Analyses of Customer Complaints

We vectorized each word into a 300-long vector using a pre-trained NLP model for Dutch texts (introduced in Section 2.2.3). The word vectors first enable the calculation of the complaint similarity, which is mathematically defined as the geometric distance of two vectors mapped in the 300-dimensional space. Based on the word vector, one can easily find similar complaints in the complaint database, with a similar example being shown below in Box 2 (similarity = 0.9). This similar complaint has been previously addressed by a WBG employee, so the response to this complaint can be used to reply to the example complaint (which is assumed to be a new complaint needing processing by WBG staff). In other words, when responding to new complaints, we can relate back to formerly enclosed cases that are similar to the new cases. In doing so, this increases the efficiency of complaint handling to a semi-automated level. When a significant number of processed instances accumulate, NLG models are expected to produce answer emails automatically. Although NLG is not included in this study, we explore it as a possible future research direction in Section 4.2.

Box 2. A complaint from the database that is similar to the example complaint shown in Box 1.

Translated complaint in English:
 Dear Sir/Madam,
 We have been living at xxx since October 2020.
 What strikes us about the drinking water supply compared to our previous home (xxx) is that the water pressure is much lower and that the capacity also leaves something to be desired. When two or more users use tap water simultaneously, the pressure drops sharply. For example, taking a shower and at the same time a toilet being flushed or the washing machine is turned on. Can you indicate whether this can also be remedied?
 I would like to hear from you.
 Yours sincerely,

By using word vectors, we also conducted a sentiment analysis to extract customers' emotions and level of satisfaction (introduced in Section 2.2.4). Figure 5 shows the distribution of sentiment scores between -1 and 1. The shape is roughly that of a Gaussian distribution, with a slightly longer tail in the negative part (score < 0) and a high peak at the value 0.1. This shows that the majority of customers used a neutral tone when reporting their water problems in complaints. By comparing extremely positive and negative tones, we discovered that the number of negative voices slightly outnumbered the number of positive voices. A small number of customers expressed great dissatisfaction in their complaints (e.g., score < -0.75, also see an example in Box 3). This is frequently the result of an unresolved earlier issue or an unaccepted bill for a large amount. Water utilities may need to pay close attention to these complaints to assess whether the case has been resolved well or whether their services can be improved. When we have a longer history of complaints (e.g., >10 years), we can also study the interannual trend of sentiment scores and determine whether the trend meets the expectation of the water utility for their service improvement.

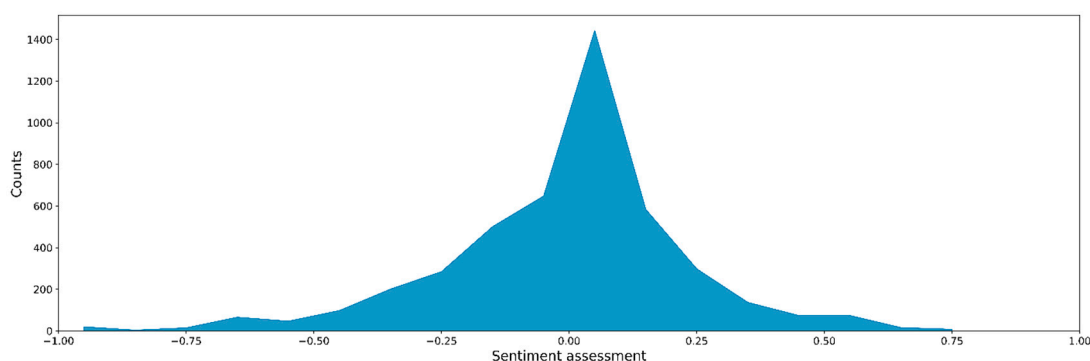


Figure 5. Sentiment score distribution of collected complaints.

Box 3. An example of a complaint with a sentiment score of -1 .

Translated complaint in English:

Good day, I got really angry this afternoon about the way your staff handled our assignment. We requested to move the water meter one meter from the kitchen to the basement. But this assignment costs us way too much.

The majority of the collected complaints can be classified based on their intents by NLP (introduced in Section 2.2.4). We used 5-folder cross validation to test the model performance. The accuracy, precision, and F1-score of the test dataset (30% of the total dataset) were 0.825, 0.802, and 0.79, respectively. These performance indicators imply that the intents of over 80% of complaints can be correctly categorized when applying the model to analyze new texts. Uncategorizable complaints often concern a minor topic, which is not listed above and lacks a sufficient number of examples to support it (e.g., a customer submitted a letter with an apology for using an aggressive tone in an earlier complaint). On the other hand, we did not consider processing a message involving multiple topics in this study. This is because our model does not aim to solve a multi-intent problem.

The recognized intent can assist water utility staff in processing and tracking the reported issue. According to the intent categories listed in Appendix B, the issue with a recognized intent can proceed to a particular department within the water utility, which is in charge of this type of issue, potentially reducing the amount of time spent by staff who must perform this categorization task manually.

We do not attempt to list the contents of all complaints because intent recognition is a practical task in NLP in which each application needs to be handled particularly (e.g., the definition of intents). Here, we aim to show the potential value of applying intent recognition to detect requests from users. In addition, there are no generic solutions that can be applied in all circumstances. Instead, algorithm and model developers must customize models practically based on the actual requirements. For instance, many water utilities are also responsible for urban flood management and need to address complaints regarding flooded/damaged properties. In this instance, a category pertaining to flooding and a sufficient number of associated examples must be added to the training dataset for an intent recognition model. In contrast to the previous tasks dealt with in this study (sentiment analysis), for which a pre-trained model can be used, intent recognition requires building a new NLP model from scratch, as shown in this case study.

4. Discussion

4.1. How Can Water Utilities Currently Benefit from the Latest NLP Techniques

NLP is a field of research that largely relies on DL to enable computers to understand natural languages. Using the latest NLP approaches, we can extract the grammatical structure of a phrase, identify the most meaningful parts of speech (e.g., transitive verbs and their objects), and distinguish intents derived from texts. A critical component that enables this is the conversion of words into machine-readable numeric vectors based on the selected corpus. As we aimed to process conversational texts rather than scientific terms, this study adopted a large general-purpose corpus trained on Wikipedia.

NLP can potentially reduce the time taken to manually process complaints. For instance, using machine-aided information, a water utility employee who is responsible for email replies can quickly grasp the main topic of a complaint by only checking the summarized key information (see Section 3.1), understand the main request of the complaint by reviewing the recognized intent (see Section 3.2), and respond by referring to a similar complaint that has been processed earlier. However, we should also notice that this also demands considerable efforts relating to collecting samples and feedback from users during the process of training the NLP model. Water utilities may need to consider the tradeoff between the effort expended in training a model and the benefit of using a pre-trained model.

4.2. How Will Water Utilities Benefit More from NLP in the near Future

We can anticipate that NLG will complete the cycle of NLP in facilitating interactive communication between customers and machines in the near future. Although NLG is not investigated in this study and is still being developed in NLP research, we can expect it to be a beneficial tool in assisting us in understanding different situations. The real understanding of a problem also requires additional information from customers, often more than once. Therefore, a chatbot equipped with NLP and NLG may be expected to be more effective at gathering detailed information. In future practical applications, we expect that NLU will be able to comprehend key information such as the time, address, or multiple intents of customers, while NLG can generate messages confirming information, inquiring about the nature of the problem from the customer, or accepting/declining the proposed request. This is a way to standardize the complaint description by including precise time, location, problem narrative, and other relevant activities. By doing so, lots of (repeated) work can be performed by machines, which can, for instance, enable the optimal assignment of personnel within a water utility.

However, for the entire system to function properly, we also need the support of other databases and models. For instance, we need a database that stores user profiles (with personal information securely maintained by water utilities) to quickly identify recent maintenances that can lead to this reported issue. Additionally, we also need real-time information about ongoing or planned projects in the management area so that we can track whether the problem is related to the project and assess the number of customers affected in the area. Moreover, if a frequently asked questions (FAQ) database has been established based on historical complaints and responses, the NLP should be able to search for a query that is similar to a previous one and reply referring to the earlier response. Additionally, vice versa, the FAQ database should also be updated automatically with new questions that appear often. Finally, a model should determine the nature of the problem and find the appropriate technician from the internal employee staff database who is responsible for performing the specific type of work required for the reported particular area.

Nonetheless, we also need to accept that NLP is not a perfect solution for all problems (at least at this stage). Due to the complexity of natural languages, expressions may be difficult to comprehend even for native speakers, let alone machines. As a result, we continue to require assistance from experienced staff capable of resolving complex difficulties. It is also necessary to include a section inquiring about the satisfaction of a customer with the automatic response. If customers are not satisfied, this implies that the NLP model encountered a case with fewer examples available in the past that would help the model to cope with new cases. Therefore, a human assistant can excel at this task. This process is also referred to as human-in-the-loop or active (machine) learning. Ideally, the more instances we use to train the model, the less expert assistance is required. However, it is evident that not all future situations are comparable to previous ones. Thus, it is always advisable to have a human assistant available for 'unexpected' cases during the early stage of implementing NLP models, but they will need to deal with a decreasing number of cases as the model matures. Human-addressed instances can be stored in the FAQ database, becoming new samples for training the model. Last but not least, this system can also be used to instruct new water utility employees, allowing them to become familiar with business issues and standardize their responses to customers.

5. Conclusions

We demonstrated that natural language processing, as an interdisciplinary field combining linguistics and deep learning, is an effective tool for automating text processing. When applied to a case study involving customer complaints about urban water-related problems collected by the water utility WBG, NLP models were able to carry out multiple tasks to enable more efficient operations to replace repetitious and laborious work that generally has to be conducted by human operators. Our analyses yielded the following conclusions:

1. Via a lexical analysis, it was determined that NLP can remove unnecessary words and symbols and determine the lemmas of essential words in customer complaints. This implies that current NLP techniques are capable of analyzing words in sentences and extracting key information in terms of words.
2. Via a syntactic analysis, it was determined that NLP can address the relationship between words as well as the functionality of words. Therefore, syntactic analysis is more useful when attempting to extract structured information—for example, for identifying the detailed information (e.g., location, time) of a water problem or a request.
3. Both lexical and syntactic analyses can be undertaken manually by humans. However, recent advances in deep learning enable machines to (partially) perform these activities in place of humans. The key component is word vectorization, which represents words as numerical vectors in a high-dimensional space. Using machine-readable vectors, similar information can be detected; sentiments can be analyzed; and, more importantly, intents can be recognized.
4. This study presents a fundamental investigation of applying NLP to automate text processing for the water sector. With the anticipated research outcomes of NLP in the future, NLP will be able to automate text processing further and more deeply, including text generation, interaction with humans, and connection to digital databases.

Author Contributions: Conceptualization: X.T., I.V. and P.v.T.; methodology: X.T. and S.P.; validation: X.T. and I.V.; formal analysis: X.T.; resources: I.V.; writing: X.T., L.T., I.V., P.v.T. and S.P.; visualization: X.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the joint research program of the Dutch and Flemish water utilities (BTO).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to confidentiality agreements with the Dutch Water Utilities that funded this research project, the supporting data and code are not available to share (due to private information of clients). Text examples used for training the NLP model have been specially processed and shared in Appendix B.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

BERT	Bidirectional Encoder Representations Transformer
DIET	Dual Intent and Entity Transformer
DL	Deep Learning
FAQ	Frequently Asked Questions
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
POS	Part of Speech
WBG	Waterbedrijf Groningen (Water Utility Groningen)

Appendix A

Complaint form
Fields marked with a * are mandatory fields.

1. Describe your complaint

Subject *
Payments

Description *

Appendix
Select Files

2. Your data

First name *

Surname *

E-mail address *

telephone number

3. Address to which your complaint relates

Zip code *

House number * **Addition**

Clientnumber

Figure A1. An example of a complaint form designed by the WBG for collecting customer complaints (adapted from the website of WBG: <https://waterbedrijfgroningen.nl/>, accessed on 17 December 2021). Subject categories include payment, water meter reading, water meter replacement, malfunction, construction of main lines, moving houses, maintenance, connections, and planned projects.

Appendix B

Below we show selected examples used for training the NLP model, details about the full dataset can be found in English (https://xintian7.github.io/Publication-NLP-2022/complaints_sample_english.txt, accessed on 17 December 2021) and in Dutch (https://xintian7.github.io/Publication-NLP-2022/complaints_sample_dutch.txt, accessed on 17 December 2021). Note that the problem statement may sound not natural to a native English speaker due to the language difference. However, we attempt to keep the original word order and show their direct translation below. Also note that these examples can be complete sentences or just a semantic component (e.g., noun phrases and verb phrases).

Training dataset I—request for repair or replacement

- (1) Please replace the meter.
- (2) With this letter, I want to make an appointment for the installation of a new water meter.
- (3) would like to have a new meter installed

Training dataset II—request for investigation about payment

- (1) I would like to see this refunded to my account
- (2) This is a repair of the water company itself, so these costs are not on me.
- (3) For the above reasons, I hereby object to the imposed administration costs and do not agree to payment.

Training dataset III—request for investigation of water leakage/noise

- (1) Our water meter beeps very annoyingly.
- (2) I have a lot of water at the meter behind the front door.
- (3) There is also a leakage from the main water tap.

Training dataset IV, request for investigation of water pressure issue

- (1) a fairly low water pressure
- (2) reduced water pressure
- (3) We currently have no water.

Training dataset IV, request for investigation of water quality issue

- (1) The water from the tap was brown and sandy.
- (2) As of today, tap water smells and tastes strange.
- (3) The water has a strange metallic (copper) smell.

References

1. Borg, A.; Boldt, M.; Rosander, O.; Ahlstrand, J. E-mail classification with machine learning and word embeddings for improved customer support. *Neural Comput. Appl.* **2021**, *33*, 1881–1902. [CrossRef]
2. Dong, L.; Spencer, M.C.; Biagi, A. A Semi-supervised Multi-task Learning Approach to Classify Customer Contact Intents. In Proceedings of the 4th Workshop on e-Commerce and NLP, Bangkok, Thailand, 1–6 August 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 49–57.
3. Liu, X.; He, P.; Chen, W.; Gao, J. Multi-Task Deep Neural Networks for Natural Language Understanding. *arXiv* **2019**, arXiv:1901.11504.
4. Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; Bowman, S.R. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. *arXiv* **2018**, arXiv:1804.07461.
5. Elina, M.; Elina, M.; Santaholma, M.; Issco, E.T.I.T.I.M. Grammar sharing techniques for rule-based multilingual NLP systems Grammar sharing techniques for rule-based multilingual NLP systems. In Proceedings of the 16th Nordic Conference of Computational Linguistics (NODALIDA), Tartu, Estonia, 25–26 May 2007.
6. Tootooni, M.S.; Pasupathy, K.S.; Heaton, H.A.; Clements, C.M.; Sir, M.Y. CCMapper: An adaptive NLP-based free-text chief complaint mapping algorithm. *Comput. Biol. Med.* **2019**, *113*, 103398. [CrossRef]
7. Afzal, N.; Sohn, S.; Abram, S.; Scott, C.G.; Chaudhry, R.; Liu, H.; Kullo, I.J.; Arruda-Olson, A.M. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *J. Vasc. Surg.* **2017**, *65*, 1753–1761. [CrossRef] [PubMed]
8. Qiu, X.P.; Sun, T.X.; Xu, Y.G.; Shao, Y.F.; Dai, N.; Huang, X.J. Pre-trained models for natural language processing: A survey. *Sci. China Technol. Sci.* **2020**, *63*, 1872–1897. [CrossRef]
9. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013-Workshop Track Proceedings, Scottsdale, AZ, USA, 2–4 May 2013; pp. 1–12.
10. Hake, P.; Rehse, J.R.; Fettke, P. Toward Automated Support of Complaint Handling Processes: An Application in the Medical Technology Industry. *J. Data Semant.* **2021**, *10*, 41–56. [CrossRef]
11. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
12. Kalyan, K.S.; Rajasekharan, A.; Sangeetha, S. AMMUS: A Survey of Transformer-based Pretrained Models in Natural Language Processing. *arXiv* **2021**, arXiv:2108.05542.
13. Tang, X.; Mou, H.; Liu, J.; Du, X. Research on automatic labeling of imbalanced texts of customer complaints based on text enhancement and layer-by-layer semantic matching. *Sci. Rep.* **2021**, *11*, 11849. [CrossRef]
14. Prabhu, S.; Mohamed, M.; Misra, H. Multi-class Text Classification using BERT-based Active Learning. *arXiv* **2021**, arXiv:2104.14289.
15. Kumar, A.; Dabas, V. A social media complaint workflow automation tool using sentiment intelligence. In *Lecture Notes in Engineering and Computer Science, Proceedings of The World Congress on Engineering, London, UK, 29 June–1 July 2016*; Newswood Limited: Hong Kong, China; pp. 176–181.
16. Hanks, P. *Lexical Analysis: Norms and Exploitations*; MIT Press: Cambridge, MA, USA, 2013; ISBN 0262018578.
17. SpaCy Guides: Linguistic Features. Available online: <https://spacy.io/usage/linguistic-features#language-data> (accessed on 11 February 2022).
18. Koopman, H.; Sportiche, D.; Stabler, E. *An Introduction to Syntactic Analysis and Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2013; ISBN 1118470478.
19. Petrov, S.; Das, D.; McDonald, R. A universal part-of-speech tagset. *arXiv* **2012**, arXiv:1104.2086.
20. Yamada, H.; Matsumoto, Y. Statistical Dependency Analysis with Support Vector machines. In Proceedings of the 8th International Conference on Parsing Technologies, Nancy, France, 23–25 April 2003.
21. Johansson, R. Dependency-Based Semantic Analysis of Natural-Language Text. Ph.D. Thesis, Lund University, Lund, Sweden, 2008.
22. SpaCy Guides: Visualizers. Available online: <https://spacy.io/usage/visualizers> (accessed on 11 February 2022).
23. SpaCy Trained Pipelines: Dutch. Available online: <https://spacy.io/models/nl> (accessed on 11 February 2022).
24. Github Resources Textblob-nl Sentiment Model. Available online: <https://github.com/gvisniuc/textblob-nl> (accessed on 11 February 2022).
25. SpaCy. Available online: <https://spacy.io/> (accessed on 11 February 2022).
26. Thinc: A Refreshing Functional Take on Deep Learning, Compatible with Your Favorite Libraries. Available online: <https://github.com/explosion/thinc> (accessed on 11 February 2022).
27. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 9 June 2016; pp. 260–270. [CrossRef]

28. Rasa. Available online: <https://rasa.com/> (accessed on 11 February 2022).
29. Introducing DIET: State-of-the-Art Architecture That Outperforms Fine-Tuning BERT and Is 6X Faster to Train. Available online: <https://rasa.com/blog/introducing-dual-intent-and-entity-transformer-diet-state-of-the-art-performance-on-a-lightweight-architecture/> (accessed on 11 February 2022).
30. SpaCy. Dependency Tree. Available online: https://github.com/clir/clearnlp-guidelines/blob/master/md/specifications/dependency_labels_old.md (accessed on 11 February 2022).