www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

INFLUENCE OF PRE-PROCESSING METHODS ON THE AUTOMATIC PRIORITY PREDICTION OF NATIVE-LANGUAGE END-USERS' MAINTENANCE REQUESTS THROUGH MACHINE LEARNING METHODS

SUBMITTED: July 2023
REVISED: February 2024
PUBLISHED: March 2024
EDITOR: Žiga Turk
DOI: 10.36680/j.itcon.2024.006

Marco D'Orazio, Prof.,
DICEA, Università Politecnica delle Marche, via brecce bianche, Ancona, Italy;
m.dorazio@staff.univpm.it

Gabriele Bernardini, Dr.,
DICEA, Università Politecnica delle Marche, via brecce bianche, Ancona, Italy;
g.bernardini@staff.univpm.it

Elisa Di Giuseppe, Prof.,
DICEA, Università Politecnica delle Marche, via brecce bianche, Ancona, Italy;
e.digiuseppe@staff.univpm.it

SUMMARY*: Feedback and requests by occupants are relevant sources of data to improve building management, and building maintenance. Indeed, most predictable faults can be directly identified by occupants and communicated to facility managers through communications written in the end-users' native language. In this sense, natural language processing methods can support the request identification and attribution process if they are robust enough to extract useful information from these unstructured textual sources. Machine learning (ML) can support assessing and managing these data, especially in the case of many simultaneous communications. In this field, the application of pre-processing and ML methods to English-written databases has been widely provided, while efforts in other native languages are still limited, impacting the real applicability. Moreover, the performance of combinations of methods for pre-processing, ML and classification classes attribution, has been limitedly investigated while comparing different languages. To fill this gap, this work hence explores the performance of automatic priority assignment of maintenance end-users' requests depending on the combined influence of: (a) different natural language pre-processing methods, (b) several supervised ML algorithms, (c) two priority classification rules (2-class versus 4-class), (d) the database language (i.e. the original database written in Italian, the native end-users' language; a translated database version in English, as standard reference). Analyses are performed on a database of about 12000 maintenance requests written in Italian concerning a stock of 23 buildings open to the public. A random sample of the sentences is supervised and labelled by 20 expert annotators following the best-worst method to attribute a priority score. Labelled sentences are then pre-processed using four different approaches to progressively reduce the number of unique words (potential predictors). Five different consolidated ML methods are applied, and comparisons involve accuracy, precision, recall and F1-score for each combination of pre-processing action, ML method and the number of priority classes. Results show that, within each ML algorithm, different pre-processing methods limitedly impact the final accuracy and average F1-score. In both Italian and English conditions, the best performance is obtained by NN, LR, SVM methods, while NB generally fails, and by considering the 2-class priority classification scale. In this sense, results confirm that facility managers can be effectively supported by ML methods for preliminary priority assessments in building maintenance processes, even when the requests database is written in end-users' native language.*

KEYWORDS*: Machine learning, natural language processing, building maintenance, facilities management, text mining, bi-lingual analysis.*

REFERENCE*: Marco D'Orazio, Gabriele Bernardini, Elisa Di Giuseppe (2024). Influence of pre-processing methods on the automatic priority prediction of native-language end-users' maintenance requests through machine learning methods. Journal of Information Technology in Construction (ITcon), Vol. 29, pg. 99-116, DOI: 10.36680/j.itcon.2024.006*

COPYRIGHT*: © 2024 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.*

1. INTRODUCTION

When an unpredictable fault occurs in a building, a maintenance process starts entailing a corrective action. In large and distributed buildings organizations, such as hospitals, universities and other public administrations, the number of contemporary corrective actions can be relevant, even if preventive maintenance approaches are in use (Dzulkifli *et al.*, 2021; Ferreira *et al.*, 2021; Hong *et al.*, 2015; Rampini and Cecconi, 2022; Razali *et al.*, 2020; Shalabi and Turkan, 2020). In these contexts, the maintenance process is usually managed by facility management (FM) contractors.

Maintenance requests are usually produced by end-users, i.e. building occupants, through short textual communications typically transmitted by e-mail to the technical staff in charge of the process (Bortolini and Forcada, 2020; Dzulkifli *et al.*, 2021). The technical staff then manages the requests, by usually assigning a priority, a work category, the needed staff or similar information, useful to create a "link" between the requests and the staff workers with the necessary skills, in order to avoid delays in interventions, to reduce safety risks and long interruptions of critical services (Chanter and Swallow, 2007; Sanni-Anibire *et al.*, 2021).

The number of contemporary maintenance requests can be very relevant with respect to the employed technical staff, and can involve different buildings and systems located in different places (D'Orazio *et al.*, 2022; Fotovatfard and Heravi, 2021; Sourav Das Adhikari *et al.*, 2019). Therefore, introducing automatic procedures and tools able to support FM is essential to guarantee a timely and efficient maintenance process.

For this purpose, Machine Learning (ML) methods have been recently applied in this field (Loyola, 2018; Rampini and Cecconi, 2022; Žižka *et al.*, 2019), due to their ability to discover hidden knowledge or to threaten information starting from Computerized Maintenance Management Systems (CMMSs), which store databases of information on maintenance processes (Bortolini and Forcada, 2020; Chen *et al.*, 2019; Pärn *et al.*, 2017; Pishdad-Bozorgi *et al.*, 2018; Shalabi and Turkan, 2020). Each record in the CMMS databases contains unstructured information (as the requests produced by the building end-users) and labels attributed by the technicians, such as "priority", "type of work", "starting time", "management and execution time" and "ending time". Considering that part of the recorded information is in textual form, text-mining methods have been proposed to extract information from CMMS, to improve the management process (Dzulkifli *et al.*, 2021) and to reduce the time necessary to organize and perform corrective actions (Çınar *et al.*, 2020; Yan *et al.*, 2020).

In a comprehensive review, Pärn et al. describe the recent advancements in this field, underlining efforts, future trends and main challenges, which are essentially focused on decision-making process optimization and sustainability (Pärn *et al.*, 2017)(Pan and Zhang, 2021). Different BIM-based frameworks have been also proposed to locate and detect which space in a building can be more critical for maintenance needs (Shalabi and Turkan, 2020), and to automatize the scheduling of FM work orders (Chen *et al.*, 2018). In this sense, some authors reviewed the attempts focused on developing automated approaches to enhance the work order execution process(Matarneh *et al.*, 2019) (Bortoluzzi *et al.*, 2019; McArthur *et al.*, 2018) and to schedule FM work orders to help prioritize the assignment of maintenance activities (Bortoluzzi *et al.*, 2019; McArthur *et al.*, 2018).

The automation of the attribution process through ML (scheduling and staff assignment, resource optimization) has been analysed by several authors (El-Dash, 2007; Gutjahr and Reiter, 2010; Wu and Sun, 2006). Text mining algorithms have been introduced for different applications in the construction field (Baek *et al.*, 2021; Bugalia *et al.*, 2022; Leoni *et al.*, 2024) and also employed to analyse the relationship between occupants' complaints and building performance (Goins and Moezzi, 2013). Mo et al. (Mo *et al.*, 2020) propose ML methods to automatically assign the proper staff, by mining the text of end-users' maintenance requests contained in CMMS (Mo *et al.*, 2020). Bortolini et al. propose a text-mining method to recognize words expressing different levels of "urgency" in a dataset of end-users' maintenance requests(Bortolini and Forcada, 2020). Similarly, Gunay et al. underline the necessity to improve the automatic recognition of the fault and the required staff (Gunay, Shen and Newsham, 2019). Marocco et al. proposed text-mining methods to extract information from textual maintenance requests, such as the room where the fault happens and the system involved (Marocco and Garofolo, 2021).

Similar methods have been proposed to analyse and extract other information from CMMS databases (Gunay, Shen and Yang, 2019), such as those related to faults of HVAC components (Shalabi and Turkan, 2020). Fault frequency analysis was also performed through the recognition of keywords contained in the WOs (Yang *et al.*, 2018), while anomaly detection was proposed using neural-based learning methods (Du *et al.*, 2017). Sentiment analysis methods have been also introduced in this field (D'Orazio *et al.*, 2022).

Despite the recent advancements, maintenance request management in buildings still remains a manual and time-consuming task for several organizations (Bouabdallaoui *et al.*, 2020). Indeed, despite their recent advancements, ML methods are not easily generalizable and applicable to different contexts(Rampini and Cecconi, 2022). Most of the works addressed to extract information from end-user's maintenance requests and to predict the required priority or category only rely on the English language, eventually based on the English translation of databases written in other languages (Sala *et al.*, 2022).

Nevertheless, the language used by end-users to exchange information and require corrective actions is usually the native one, thus it can be different from English. Moreover, the end-users' maintenance requests are generated in different working and technical contexts, thus relying on several potential *Thesauri* (Guyot *et al.*, 2010). At the same time, requests are provided by occupants with different technical knowledge, thus potentially affecting the final quality of communications. In view of the above, the vocabulary and the structure of each request can be influenced by the context, by the specific knowledge of the users, and by the language.

For these reasons, ML extensions to other languages apart from English should be also provided in the building maintenance field, as previously performed by works in other contexts (Carroll *et al.*, 2024; Khan *et al.*, 2023; Mercha and Benbrahim, 2023), such as legal domain (Bellandi *et al.*, 2024a; Licari and Comandè, 2024a) and human health (Valdez *et al.*, 2023). In this sense, the influence of different pre-processing methods (Parisi *et al.*, 2021), applied before the use of ML methods, should be assessed (D'Orazio *et al.*, 2022; Gunay, Shen and Newsham, 2019; Kim *et al.*, 2022; McArthur *et al.*, 2018). Evaluating the performance of combined applications of pre-processing and ML methods could then support FM in different Countries, by making contractors aware of the advantages and limitations of their application to the corpus of requests written in the end-users' native language. Moreover, the possibility of using communications in the native language rather than in English can improve the quality of the collected data (Bugalia *et al.*, 2022).

Given the above, this work investigates how the end-user's request language can affect the accuracy of ML methods for the automatic detection of the priority of maintenance interventions, considering the combination of four pre-processing methods, five ML algorithms and two priority classification rules.

To this end, a dataset of e-mails written to require corrective maintenance actions by the end-users of a stock of 23 buildings has been collected for 34 months. This dataset is written in Italian, the end-users' native language. The same dataset, translated in English, has been presented in a previous work of the authors, to investigate the effectiveness of five consolidated ML methods on priority assignment depending on pre-processing methods and priority classification scales (D'Orazio *et al.*, 2023). In particular, that work showed that the best combination of pre-processing/ML/priority classes is represented by the original dataset (i.e. no pre-processing action)/Neural Networks or Decision Tree/binary classes (i.e.: low or high priority).

In this contribution, the assessment is performed by combining different pre-processing, ML and priority classification methods on the original dataset written in Italian. To the best of the authors' knowledge, although previous research dealt with natural language processing and ML applications to databases written in Italian, these dealt with other domains (i.e. mainly the legal one) (Bellandi *et al.*, 2024b; Licari and Comandè, 2024b) This work then represents the first application of pre-processing and ML techniques on a wide dataset in the context of building maintenance.

This dataset has been labelled (by priority) by expert annotators. Then, four different pre-processing methods are utilized on this corpus, characterized by the progressive reduction in the number of unique words in the requests. Due to their recognized efficiency in the field (Baek *et al.*, 2021; Çınar *et al.*, 2020; McArthur *et al.*, 2018; Mo *et al.*, 2020; Žižka *et al.*, 2019), Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Neural Network (NN) and Decision Tree (DT) methods have been applied as ML algorithms in combination with the pre-processing methods. Finally, the ability of ML methods has been then checked on labelled sentences by considering two different priority scales (2-class, 4-classes). Finally, the outcomes have been compared to those coming from the analogous analysis on the dataset translated into English and reported in the reference work (D'Orazio *et al.*, 2023), to assess if the application of pre-processing techniques can alter the final priority assignment for native language applications.

2. DATASET AND METHODS

2.1 Dataset

The dataset used in this work includes about 12,000 end-users' maintenance requests, collected in 34 months in 23 buildings used for didactic, research and administrative activities, hosting the University "Politecnica delle Marche", Italy. The total gross floor area of the buildings is about 152,000 m^2, and more than 16,000 occupants are generally hosted, including students, teaching staff and other workers.

End-users' maintenance requests are written in Italian, the native language of the building's occupants. The dataset, translated into English, has been already analysed with sentiment and emotion analysis techniques (D'Orazio et al., 2022) and with selected ML methods (D'Orazio *et al.*, 2023).

Maintenance activities are outsourced to a general contractor, named "GETEC", which performs planned ordinary maintenance tasks and periodical inspections on the building systems but does not employ maintenance predictive approaches. Extraordinary maintenance activities (i.e. building envelope renovation, HVAC replacement, etc.) are not part of this process and are directly managed by the technical staff of the university. The maintenance process entails the e-mail communication of the fault by the end-user and the translation of the request into a "Work Order" (WO) by the technical staff, using a CMMS. In a typical day, the number of contemporary requests can entail up to (or even more) 30 WO (and then interventions) on the buildings.

Each WO includes both structured and unstructured information: date; e-mail text of the end-user (in natural language); category; expected action to perform. In detail, the technical staff manually attributes the category, the expected maintenance action, the needed staff with adequate skill and the necessary equipment, thus allowing the effective plan and start of the corrective action. The technical staff defined seven main intervention categories, following a schematization based on the standard ASTM E 1557 (Charette and Marshall, 1999; Kula and Ergen, 2018; Systems et al., 2007): 1) building components, i.e. walls, doors, windows, floors, stairs; 2) dialer alarm, comprising the alarm system; 3) electrical, i.e. lighting/power systems, LAN and WLAN connection; 4) elevator; 5) fire, i.e. fixed and moveable equipment; 6) HVAC, i.e. heating, ventilation, and cooling units, pipes; 7) plumbing. A second classification level for each of these main categories allows the detection of the specific component (e.g. "door", in the case of the "building component" category). A third level combines the second level class with the presumed action to perform, to assign the staff with adequate skill and the necessary equipment (e.g. "handle replacement", for a "door").

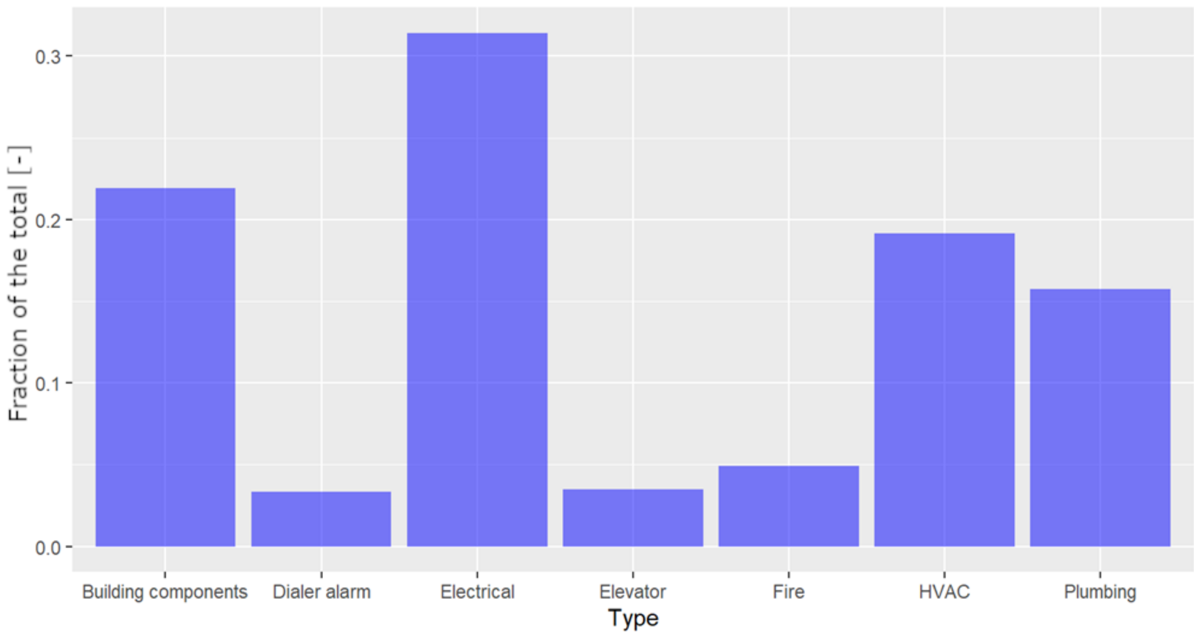

Figure 1: Overview of collected end-users' maintenance requests by main intervention category.

Figure 1 shows the collected end-users' requests in the dataset by intervention category. About one-third of the total number of requests is due to electrical faults, probably depending on the high number of installed appliances. About one-fifth of the requests relate to building components (principally doors and windows). The number of requests due to faults in HVAC and Plumbing components is similar. The number of requests in the Fire category is limited, due to the preventive maintenance measures performed for fire safety, according to the national legislation. Moreover, the number of requests related to elevators is limited, due to their low number.

2.2 Methods

Figure 2 resumes the methodological framework followed in this work. Specific tasks have been performed for priority score attribution, pre-filtering/pre-processing, ML application, thus obtaining the specific performance indicators (reported in Section 3). Thus, the comparison of performance between the results obtained on the dataset in different languages (English, Italian) (D'Orazio *et al.*, 2023) is provided in Section 4.

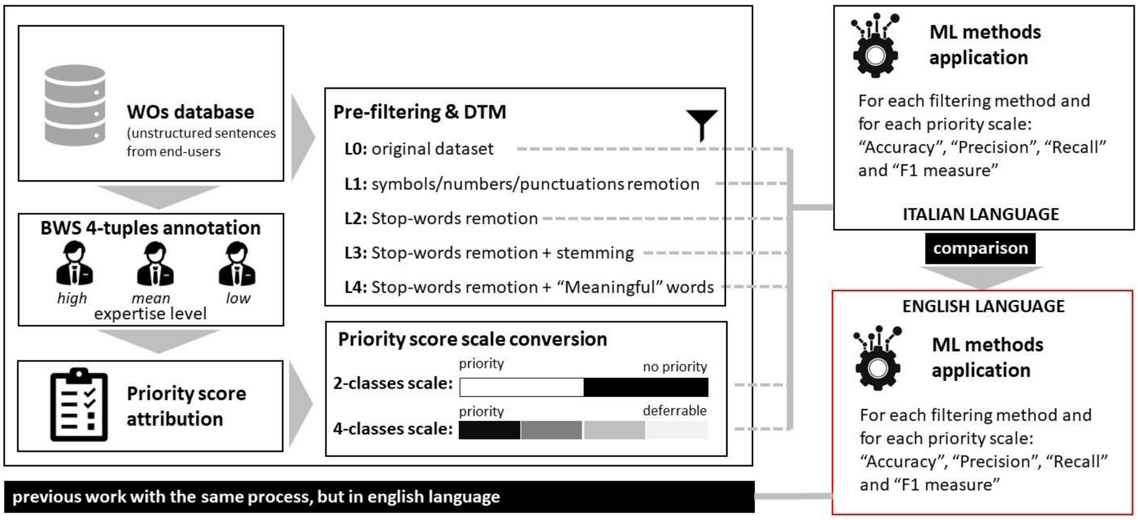

Figure 2: Overview of the methodological framework.

The original WOs database does not comprise priority annotations. Then, as a first step of the work, a randomly extracted set of sentences (corresponding to end-users' maintenance requests) was labelled with priority scores by 20 annotators. The attribution of the priority scores is based on the Best-Worst Scaling (BWS) method (Mohammad, 2018). 150 distinct 4-tuples were randomly generated through a "bwstuples" *python* script (http://valeriobasile.github.io/), to ensure that each sentence was repeated 5 times in different 4-tuples. The ability of the extracted sentences to represent the whole dataset has been checked through a term frequency (TM) analysis, to guarantee that at least 90% of the most frequent words of the original dataset are contained in the extracted sentences.

Annotators were divided into three groups, depending on the expertise level (high, mean, low) in the field of building maintenance (D'Orazio et al., 2022). 12 high-level, 5 mean-level and 12 low-level annotators participated in the HMA task. On each 4-tuple, each annotator defined the "Best" (high priority) and the "Worst" (low priority) case. The final score for each sentence was calculated by counting the number of times the sentence was chosen as best/worst by each annotator, divided by the number of times the sentence appeared, i.e. 5 (Kiritchenko and Mohammad, 2017; Louviere et al., 2015). As a result, each sentence obtained a different priority score by each annotator, on a scale from -1 to 1.

To check the concordance between annotators, a correlation analysis has been performed. The mean score attributed to each sentence has been calculated and correlated with the scores attributed by all annotators. The correlation was performed through the *Spearman* method, after a *Shapiro-Wilkinson* test, which revealed the non-normality of the sample (Pagano, 2012). Shapiro-Wilkinson and Correlation tests were performed through *R statistics-rel.4.1* and "stats" and "ggally" packages. Annotators whose labels revealed a limited correlation with the mean of the group (*Spearman rho* < 0.8) were excluded from the later stages of the work. The final agreement between the remaining annotators has been checked through *Krippendorf's alpha* method, verifying an *alpha value* threshold > 0.67 (Schmidt et al., 2018). *K-alpha* statistical test was performed using *R statistics - rel. 4.1*.

The labelled dataset, composed of these 150 extracted sentences, has been then pre-filtered by using 4 pre-processing methods, before applying ML methods. Sentences are written in Italian, and include paragraphs expressing what happened and where, and the corrective action required. Each paragraph then consists of nouns, verbs, adjectives, adverbs, spaces, symbols, numbers, and punctuations. Some of the words are clearly related to the problem (i.e. wall, door, locked, broken), while others are complementary (i.e. please, the, with, by), but necessary to lexically construct the sentence. In this sense, the pre-processing of the 150 extracted sentences has been performed since each word in the sentence could be a potential predictor for ML methods, even if not significant in the specific context. Thus, starting from the original extracted dataset of 150 sentences (L0), 4 different pre-processing approaches were followed as shown in Table 1: (L1) remotion of symbols, numbers and punctuations from L0; (L2) remotion of stop-words from L1; (L3) stemming of L2; (L4) selection of meaningful words (2-3 words for each sentence) by expert human annotators from L2. The number of potential predictors has been then intentionally reduced from L0 to L4. Table 1 also includes examples of the pre-processing actions for each of the final datasets. Text-mining methods were applied to pre-process the text through *R statistics-rel. 4.0* "TM, Quanteda, stringr, tyditext, dplyr, spacyr" packages. "Spacyr" is the interface to the "spacy" python package and to the "it_core_news_sm" model. This package has been used to perform a POS#tagging analysis of the text.

Table 1: Approaches used to pre-process the original dataset L0.

		Pre-processed dataset code				
Pre-processing action	**Example**	**L0**	**L1**	**L2**	**L3**	**L4**
Symbols, numbers, punctuations remotion	e.g. "°","$"		X	X	X	X
Remotion of stop-words (words that do not carry significant meanings)	e.g. "il", "per favore" *"the", "please" (English transl.)*			X	X	X
Stemming (reduction to the inflected word)	e.g. "rifiuto","rifiuti" "waste","wastes","wasted" *(English transl.)*				X	
Meaningful words selection (manual selection of the most meaningful words by the annotators)	e.g. "incendio", "urgente" *"fire", "urgent" (English transl.)*					X

At the end of pre-processing, a specific document term matrix (DTM) has been built for each method (form L0 to L4), through *R statistics-rel.4.0* and "TM package". A DTM is a matrix where rows are composed of the 150 sentences labelled by the 20 annotators, while columns are composed of the words contained in the dataset. When a word appears in a sentence, the corresponding element of the DTM reports a number (0,1,2, … n), expressing the number of times that the word appears in the sentence. Then, all the DTMs (L0, L1, L2, L3, L4) have the same row length but a different number of columns, since the number of predictors (words) is different. The last column of each DTM contains the priority scores attributed by the expert annotators.

Before applying ML methods, priority scores (which range in a discrete scale from -1 to 1, with steps of 0.2) have been converted in a categorical scale, considering, alternatively, a 2-class priority scale (high, low) and a 4-class priority scale (high, mean, low, deferrable). The threshold between the classes in the 2-class priority scale is 0. The limits between the four classes in the 4-priority scale are: -0.5; 0; 0.5.

Five supervised ML methods, i.e. Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Neural Network (NN) and Decision Tree (DT) (Baek et al., 2021; Mo et al., 2020), have been chosen and applied, according to (D'Orazio *et al.*, 2023). The "classification learner" module of *Matlab R2021b* has been used to apply the selected methods. A 5-fold cross-validation approach has been adopted. Each round of the 5-fold cross-validation involved the random partitioning of the original dataset into a training set and a testing set. The average cross-validation error has been used as a performance indicator.

For each combination of ML method, pre-processing method and priority scale, a confusion matrix was generated and results assessed according to the following main performance indicators (Gonçalves et al., 2013; Ribeiro et al., 2016):

- the "Accuracy", calculated as the number of elements correctly classified with respect to the total number of elements;
- the "Precision", calculated as the ratio between the number of elements correctly classified and the total predicted in each class;
- the "Recall", calculated as the ratio between the number of elements correctly classified and the number of known elements in each class;
- the "F1-score, calculated as the harmonic mean between both precision and recall.

The higher the "accuracy" and the "F1-scores", the better the ML method works. The whole process has been applied to both the original dataset of Italian sentences, and that one translated into English (whose results are provided in the authors' previous work in (D'Orazio *et al.*, 2023)).

The "accuracy" and "F1-scores" obtained in the two works have been compared to test the influence of the specific language (Italian or English) on the results, depending on the pre-processing/ML/priority classes combination (Section 4).

3. RESULTS: NATIVE LANGUAGE DATASET ANALYSIS

This section concerns the results obtained on the original Italian dataset and the native language of the end-users, by first reporting the outcomes of priority score attribution (Section 3.1) and pre-processing methods application(Section 3.2). Then, accuracy and F1-scores are provided to assess the impact of pre-processing methods on the ML, for 2-class and 4-class priority scales (Section 3.3).

3.1 Priority score attribution

The correlation analysis was performed to check the concordance between annotators. Table 2 then reports the Spearman's rho coefficients obtained by comparing the mean score attributed to each sentence by each annotator, with the scores attributed by all annotators. It can be noticed that some annotators attributed very different priority scores with respect to the mean of the group. In particular, Spearman's rho coefficients for annotators "14" and "15" are characterized by a negative sign. These annotators probably inverted the meaning of "best" and "worst". The Spearman's rho coefficient for the annotator "5" is very low. The results obtained with the other annotators show a good agreement, with Spearman's rho coefficients in the range 0.71-0.90. To reduce the influence of the annotations characterized by low Spearman's rho coefficient values (<0.8), all the related results have been discarded. Accordingly, the work of only 15 annotators was effectively selected. The results of the correlation analysis on the work of these 15 annotators are shown in Figure 3, where the last column is the Spearman's Rho coefficient for each annotator's work with respect to the mean score.

Table 2: Correlation analysis. Spearman rho coefficients obtained by comparing the mean score attributed to each sentence by each annotator with the scores attributed by all annotators, considering the Italian dataset.

	Annotator									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
rho	0.80	0.80	0.71	0.90	0.11	0.76	0.80	0.82	0.85	0.85
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
rho	0.87	0.85	0.81	-0.85	-0.78	0.87	0.87	0.84	0.90	0.87

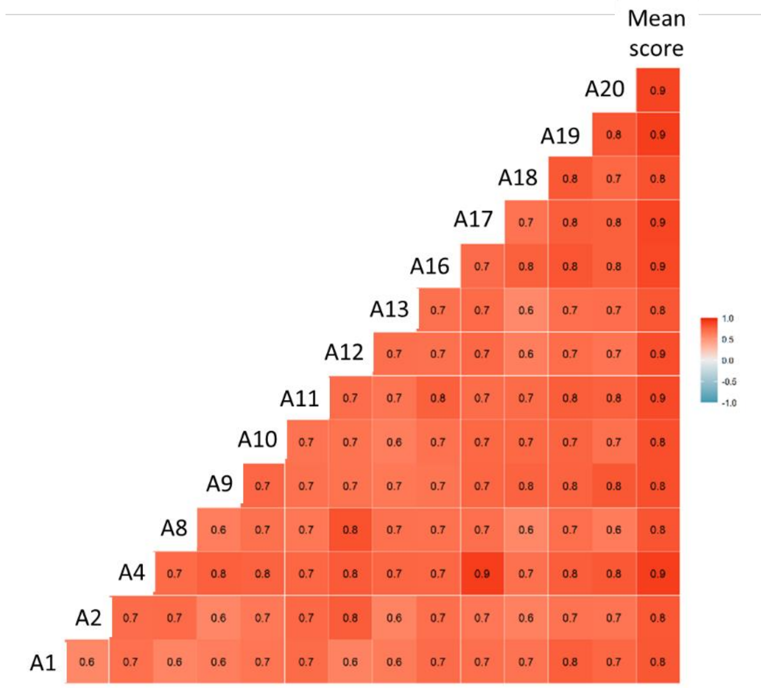

Figure 3: Correlogram describing the results of the correlation analysis performed on the scores attributed by the selected 15 annotators (Spearman's Rho > 0.8) with respect to the mean score (last column).

To check the annotators' agreement, the Krippendorff's alpha test has been additionally performed. The test shows that there is an acceptable, even not optimal, confidence level among the selected 15 annotators (k-alpha = 0.707). Therefore, we did not assume the mean value as a unique priority score for each sentence, but we preferred to label each sentence with the different priority scores attributed by each annotator, also according to previous works (D'Orazio *et al.*, 2022; Mohammad, 2018).

3.2 Pre-processing method

Figure 4 shows the result of the POS#tagging process performed on the original Italian-written dataset. The dataset is principally composed of nouns, verbs and adjectives, expressing what happened, where happened, and the related action required.

Pre-processing actions progressively reduced the number of words contained in each dataset, as shown in Table 3. As expected, the strongest reduction in the number of unique words with respect to L0 refers to L4 (which essentially relies on meaningful words), while lower differences have been retrieved between L2 (L1+ Stop-words) and L3 (L2 + stemming). The reduction in potential predictors (i.e. the words) could then impact the accuracy of different ML methods.

Table 3: Number of words for each pre-processed dataset, and % coverage with respect to the L0 dataset.

Pre-processed dataset code	Number of unique words	% of unique words with respect to L0
L0	1074	100%
L1	847	78.8%
L2	697	64.9%
L3	571	53.1%
L4	318	29.6%

CC BY

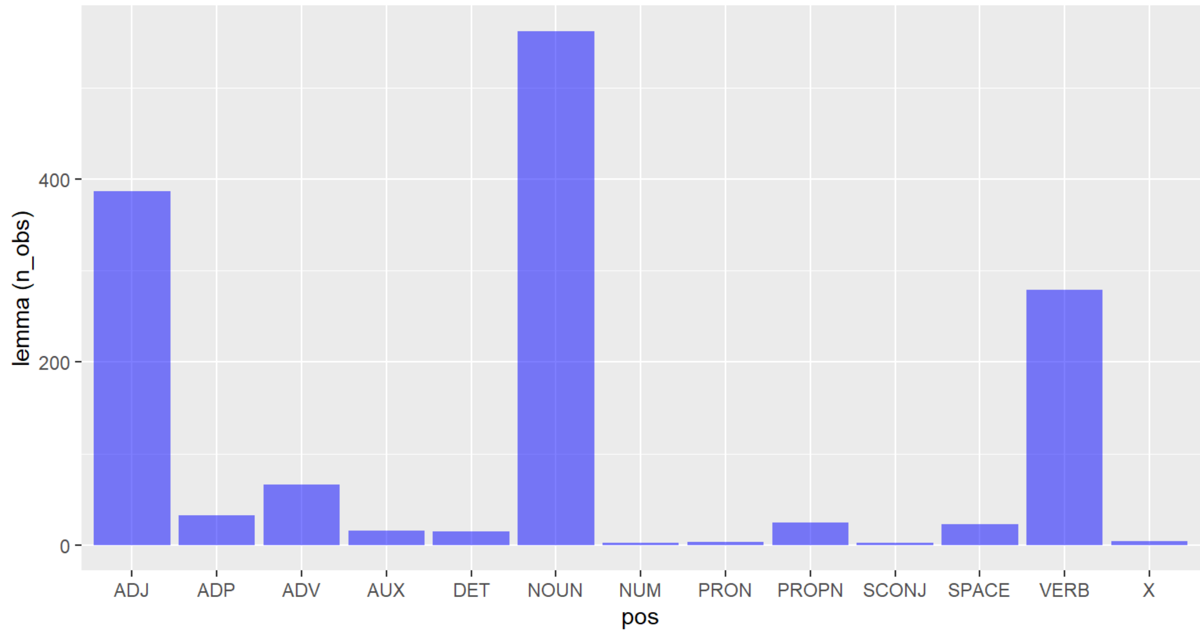

Figure 4: POS#tagging process results for the original Italian dataset (L0): ADJ: adjectives; ADp: adposition; ADV: adverb; AUX: auxiliary; DET: determiner; NOUN: noun; NUM: numeral; PRON: pronoun; PROPN: proper noun; SCONJ: subordinating conjunction; SPACE: space; VERB: verb; X: other.

3.3 Impact of pre-processing on ML methods: accuracy and F1-score

Table 4 and Table 5 provide the values of accuracy and F1-score for the combinations of pre-processing and ML methods, respectively considering the 2-class and 4-class priority scales, for the original dataset in Italian. The complete overview of the confusion matrices elaborated in this work is proved within the electronic supplementary materials.

It is noteworthy that, within a certain ML method, different pre-processing methods limitedly impact the final accuracy and average F1-score. Indeed, their percentage variations due to pre-processing methods are on average around 3% for results related to the 2-class priority scale (Table 4), and 5% for those of the 4-class priority scale (Table 5). A slightly higher impact of pre-processing methods, even if always very limited, can be found only in DT method, where L4 provides accuracy reduction by 4% and 8%, and F1-score reductions by 9% and 23%, respectively for 2-class and 4-class priority scales.

Results also show that NB method generally fails in predicting the priority class with reference to all the pre-processing methods and priority scales. This outcome is confirmed by the low accuracy values (under 0.56 for the 2-class priority scale and 0.34 for the 4-class priority scale) and the inability to evaluate F1-score. "Not assessable values" are reported by "-" in Table 4 and Table 5 for NB since this method concentrates all the predictions in a single priority class, thus substantially differing from the dataset distribution and nullifying the significance of the indicators.

Conversely, considering the 2-class priority scale (Table 4), NN, LR, SVM methods provide the best (and similar) results, with average accuracy values of 0.82 and F1-scores of 0.81. DT method shows a lower performance (with average accuracy values of 0.71 and F1-scores of 0.68). Also concerning the 4-class priority scale (Table 5), the best performance is again obtained by NN, LR, SVM methods, however with lower accuracy and F1-scores (on average both around 0.58).

In general, it can be noticed that the 2-class priority scale (Table 4) provides better accuracy and F1-score than the 4-class priority scale (Table 5). Results obtained for the 2-class priority scale show an average accuracy of 0.66 (versus 0.45 for the 4-class priority scale) and an average F1-score of 0.78 (versus 0.54 for the 4-class priority scale).

It is also interesting to observe that, for both 2-class and 4-class priority scales, F1-score related to the "high priority" classes (Class 1) are generally higher than those obtained for the "lower priority" classes. Thus, the proposed approach could support the technical staff by pointing out the "high" priority requests in a reliable manner, and it could allow them to focus on the check of a more limited number of end-users' communications.

Table 4: Accuracy and F1-score values depending on the pre-processing and ML combination, considering the 2-class priority scale (class 1: high; class 2: low). "Not assessable values" due to the fact the ML method fails in prediction when applying a certain pre-processing method are shown by"-".

ML: indicators and priority classes		**Pre-processed dataset code**				
		L0	**L1**	**L2**	**L3**	**L4**
2-class	*Overall accuracy*	*0.67*	*0.62*	*0.67*	*0.67*	*0.65*
DT	Accuracy	0.72	0.70	0.71	0.72	0.69
	F1-score: Class 1	0.78	0.77	0.78	0.78	0.77
	F1-score: Class 2	0.61	0.59	0.60	0.61	0.52
LR	Accuracy	0.82	0.81	0.82	0.82	0.81
	F1-score: Class 1	0.84	0.84	0.84	0.84	0.79
	F1-score: Class 2	0.79	0.79	0.79	0.79	0.79
NB	Accuracy	0.56	-	0.56	0.56	0.56
	F1-score: Class 1	-	-	-	-	-
	F1-score: Class 2	-	-	-	-	-
NN	Accuracy	0.81	0.82	0.81	0.82	0.81
	F1-score: Class 1	0.83	0.84	0.83	0.79	0.83
	F1-score: Class 2	0.79	0.79	0.79	0.79	0.79
SVM	Accuracy	0.82	0.81	0.82	0.82	0.81
	F1-score: Class 1	0.83	0.84	0.84	0.84	0.83
	F1-score: Class 2	0.79	0.79	0.79	0.80	0.79

Table 5: Accuracy and F1-score values depending on the pre-processing and ML combination, considering the 4-class priority scale (class 1: high; class 2: mean; class 3: low; class 4: deferrable).

ML: indicators and priority classes		**Pre-processed dataset code**				
		L0	**L1**	**L2**	**L3**	**L4**
4-class	*Overall accuracy*	*0.45*	*0.44*	*0.46*	*0.46*	*0.44*
DT	Accuracy	0.47	0.48	0.49	0.49	0.45
	F1-score: Class 1	0.57	0.58	0.56	0.58	0.59
	F1-score: Class 2	0.55	0.55	0.56	0.55	0.55
	F1-score: Class 3	0.24	0.24	0.39	0.31	0.02
	F1-score: Class 4	0.31	0.35	0.33	0.42	0.26
LR	Accuracy	0.59	0.59	0.59	0.59	0.59
	F1-score: Class 1	0.69	0.70	0.67	0.69	0.70
	F1-score: Class 2	0.63	0.62	0.63	0.62	0.63
	F1-score: Class 3	0.44	0.42	0.42	0.44	0.42
	F1-score: Class 4	0.60	0.59	0.60	0.59	0.59
NB	Accuracy	0.34	-	0.34	0.34	0.34

	F1-score: Class 1	-	-	-	-	-
	F1-score: Class 2	-	-	-	-	-
	F1-score: Class 3	-	-	-	-	-
	F1-score: Class 4	-	-	-	-	-
NN	Accuracy	0.58	0.58	0.58	0.57	0.58
	F1-score: Class 1	0.71	0.70	0.70	0.70	0.71
	F1-score: Class 2	0.60	0.59	0.60	0.57	0.58
	F1-score: Class 3	0.45	0.45	0.44	0.43	0.45
	F1-score: Class 4	0.59	0.59	0.59	0.58	0.60
SVM	Accuracy	0.59	0.58	0.59	0.58	0.58
	F1-score: Class 1	0.70	0.70	0.72	0.70	0.71
	F1-score: Class 2	0.60	0.60	0.62	0.58	0.60
	F1-score: Class 3	0.45	0.45	0.43	0.45	0.44
	F1-score: Class 4	0.61	0.60	0.59	0.60	0.59

4. DISCUSSION

This section first discusses how the natural language of maintenance databases influences the performance of pre-processing methods coupled with ML algorithms for the automatic priority prediction of maintenance requests (Section 4.1). To this end, results on the Italian (native) language database, shown in Section 3, are compared to those obtained on the English-translated database and reported in (D'Orazio *et al.*, 2023). Then, novelties and impacts of this work are discussed from both an academic and practical perspective (Section 4.2). Finally, limitations and future works are also addressed (Section 4.3).

4.1 Influence of maintenance database language

The analysis of the impact of end-users' request language (Italian versus English) on the priority assignment has been organized into two levels according to the results section outcomes: (a) pre-processed datasets; (b) priority score attribution comparison; (c) impact of pre-processing on ML methods according to the performance indicators.

Concerning (a), the application of pre-processing methods essentially leads to a similar reduction in the number of unique words for both Italian and English databases. The initial difference in the number of words for L0 is essentially due to the translation of requests from Italian (1074) to English (958). The application of L1 has the same impact on the percentage of unique words with respect to L0 due to the removal of symbols, numbers, and punctuations, which are essentially the same in both languages. Stop-word remotion in L2 is similar due to the English translation of words and phrasal expressions. Finally, L3 and L4 are characterized by the close values in the percentage of unique words due to the application of the same process on the remaining unique words in the requests.

Concerning (b), the mean priority score of each sentence in both datasets attributed by HMA, ranging from -1 (highest priority) to 1 (lowest priority), is reported in the scatterplot of Figure 5. The linear correlation shown in Figure 5 shows R^2= 0.6175. This result suggests a direct correlation among priorities attributed to sentences in the original Italian and English-translated datasets, but the trend coefficient is lower than 1, meaning that priority scores in the original Italian dataset are generally higher than those in the English-translated dataset. It could be argued that the annotators are influenced by the subtleties or sense of urgency provided within the request, since Italian is their mother tongue. Figure 6 shows the distribution of the residuals at the end of the fitting process. Residuals are higher for negative and neutral labelled sentences (< 0), thus confirming the possibility of differences among the process with sentences expressed by original (Italian) and translated language. Nevertheless, the application of the process in both languages seems to lead to comparable outcomes, thus encouraging the replication of the methodology in another non-English corpus to confirm the general trend similarities.

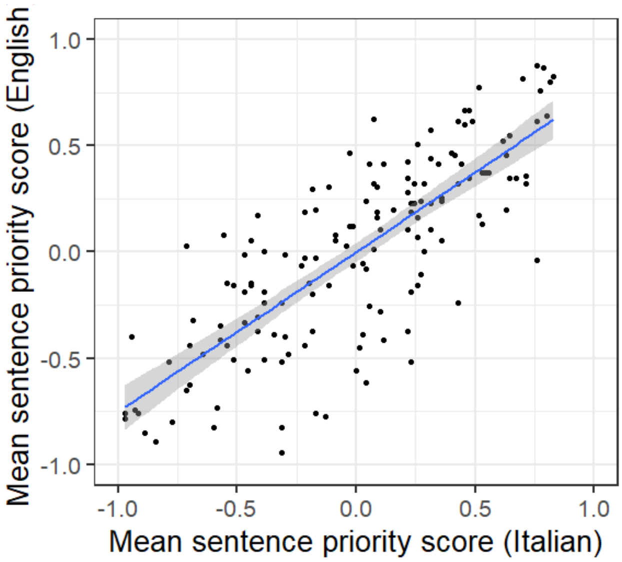

Figure 5: Mean sentence priority score correlation: Italian Vs English datasets.

Concerning (c), the results on accuracy and F1-score for the Italian database (presented in Table 4 and Table 5) are generally in line with those previously provided by the work for the English-translated database (D'Orazio *et al.*, 2023). In particular, for both the assessed languages, the best results are obtained with the binary priority classification, while the low impact of pre-processing methods (indicators differences <1) is confirmed. The highest accuracy values are noticed especially for high-priority classes, regardless of the pre-processing method. Furthermore, regardless of the pre-processing methods and the number of priority classes, only NB fails predictions.

The accuracy obtained with the two databases is comparable for all the pre-processing/ML methods/priority class combinations since differences are lower than 4%, except for DT. Indeed, in the Italian dataset, DT is generally less performing than expected according to the English-translated database, with an average accuracy decrease of 14% in the 2-class priority scale and by 20% in the 4-class priority scale. This outcome is probably due to overfitting problems, especially generated for the Italian database, and to the related domination of certain priority classes. It could be argued that the annotators have been more able to cover the wider range of priorities in the 4-class scale being affected by the analysis of communications written in their mother tongue.

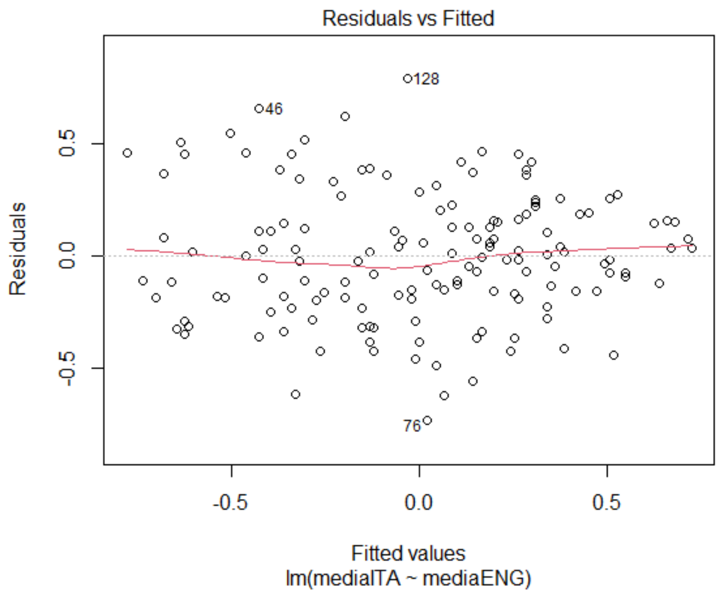

Figure 6: Residuals distribution: Italian Vs English datasets.

cc BY

4.2 Novelties and impacts of the research

The first academic novelty of the work is mainly related to the application of pre-processing and ML methods to a building maintenance database written in Italian (users' native language). Other works already applied natural language processing to corpus written in Italian, but, to the best of the authors' knowledge, they were essentially related to the legal domain, as well as they did not include the assessment of the impact of pre-processing methods (Bellandi *et al.*, 2024b; Licari and Comandè, 2024b).

Moreover, from an academic perspective, the work validates the reliability of the methodological approach and its performance by comparing results obtained in the two languages. Indeed, the comparison demonstrates that the performance of the combination of ML algorithms and pre-processing techniques, given different priority classification rules, is similar for the two databases. In particular, in both cases, results confirm the efficiency of ML in detecting the most urgent requests.

As a practical consequence, this result suggests that facility managers could take advantage of the application of pre-processing and ML methods even applied to the original corpus written in the user's native language, thus avoiding any translation operation.

ML methods can then be implemented in automatic preliminary filtering of maintenance requests, and pre-processing levels established depending on the related algorithms running speed. Therefore, the priority assignment process by the technical staff can be speeded up.

It is worth noting that the accuracy of ML methods increases while adopting 2-class priorities, and, in particular, higher values of accuracy are generally related to requests automatically assigned to the "high priority" class. From an operational point of view, the technical staff can hence check these "high priority" requests, reading them in their native language, to plan the related interventions. The additional check of requests could be ideally limited to the "non-priority" class, in view of the lower level of detected accuracy, thus reducing the complexity and quantity of contemporary requests that could be manually verified by the staff. In both cases, the possibility to read end-users' communication written in their native language rather than in non-native (i.e. English) language can improve the assessment process and make the staff capable to better understand the details and subtleties of maintenance requests. In this sense, this possibility increases the capabilities of collected data and reduces mistakes and errors in the possible manual check, as underlined by previous works on other construction sector fields (Bugalia *et al.*, 2022; Carroll *et al.*, 2024).

4.3 Limitations and future works

The outcomes of this work encourage the application of ML methods to other buildings' operation and maintenance tasks, even when considering a corpus of textual data written in non-English languages. Nevertheless, some operative limitations also encourage future research to confirm the reliability of results. In fact, the work limits the analysis to a single case study, related to the context of university buildings, focusing on priority assessment tasks. Moreover, this work is limited to the analysis of the same dataset in two languages, and thus other different databases should be explored using the same proposed methodological workflow, to ensure larger comparisons of results.

Given the case study, maintenance requests are related, and thus limited, to the technological and management specificities of the assessed scenario, as well as to its intended uses. In this sense, application to different building stocks should be encouraged, due to the possibility of specific lexicons and *Thesauri* used in a given context by end-users to request maintenance activities.

This work limits the comparison to Italian versus English-translated databases. Thus, applications to other languages are needed, to expand this work insights to different Countries. In addition, the translation to other non-English languages of the original database could be performed to better compare outputs on the same requests corpus.

Moreover, this work relies on a limited range of end-users' typologies, i.e. university staff, teachers, other research and technical staff and students. Lexicons and *Thesauri* are also part of the end-users' language as well. The end-users' needs and level of knowledge of the buildings and systems could be different depending on their role, thus pointing out several recurrent conditions in requests. The influence of occupant level of expertise on the maintenance needs and the mother tongue conditions on communications of end-users should be also assessed, to detect if some barriers exist, also due to the level of knowledge of the (technical) language itself.

Moreover, geographical and social issues could similarly affect the communication and the style of the requests, by varying the composition of requests in L0 dataset. Thus, the reduction of the unique words due to pre-processing could have different trends in view of the communication style for a certain native language. From an operational perspective, such an issue could also affect the computational timing of pre-processing/ML in future application contexts, in the case of large databases with a significant number of contemporary requests. In this sense, this work does not investigate the efficiency of combined pre-processing/ML methods under these operational tasks, especially considering the calculation timings of the assessed techniques. Further efforts towards proper pre-processing actions are needed, maybe distinguishing related methods by language, especially when combined with ML approaches.

Finally, previous limitations on language and occupant typology underline how the aforementioned issues could appear essential, especially in all buildings where users come from different Countries and Regions, thus suggesting the extension of future works toward such a direction.

5. CONCLUSIONS

The automatic recognition of the priority of a maintenance request can be very useful for the technicians involved in the maintenance of large building stocks, such as hospitals, universities, and other public and private large organizations. Indeed, these require the daily management of tens or hundreds of contemporary requests, due to unexpected failures that can occur, even if planned and preventive maintenance strategies are planned. A certain number of requests can be related to safety issues or to the interruption of critical services, both requiring immediate interventions. Current common approaches to the problem imply that technical staff manually analyses the requests, and then assigns priorities to the interventions, losing useful time in tasks that are not directly connected to planning and actuation of the corrective actions.

Machine Learning methods can support the building technical staff in the timely management of maintenance requests sent by the occupants, who, being final end-users' of buildings, can directly suffer from failures in building components and systems. Since requests are generally sent by e-mail, they are unstructured textual communications, thus needing specific methodologies to automatically extract useful information for failure identification, priority and staff assignment. Most of the approaches investigate end-users' request datasets written in English, while assessment analysis considering native language is still an ongoing but needed research task.

This work then adopts consolidated combinations among pre-processing methods, ML, and priority classification rules, to evaluate their impact on accuracy predictions considering a native language (Italian) database of about 12,000 end-users' requests coming from a stock of 23 buildings for 34 months. Previous research already provided insights into the English version of the database (D'Orazio *et al.*, 2023).

Technicians with different expertise levels have been firstly engaged in human manual annotation of a randomly extracted dataset of sentences, with the aim of generating a supervised dataset of end-users' maintenance requests which include priority assignment. Human manual annotation is hence performed on sentences written in the technicians' native language.

The extracted dataset, in which each sentence is hence labelled by priority, has been pre-processed using different techniques, which can also reduce the number of terms (and thus of ML predictors) to 30% of the original sentence words. Five consolidated ML methods, that are Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Neural Network (NN), have been considered for each pre-processing method, in combination with 2-class and 4-class priority classification rules.

The obtained results are in line with the ones previously obtained on the English-translated database. They generally confirm that several ML methods (in the specific case, LR, SVM, NN) can properly perform priority assignments of end-users' maintenance requests in Italian and English. In both cases, the accuracy of ML methods increases while adopting 2-class priorities, while pre-processing limitedly affects the results. Although results could be affected by the application to a specific case study and, mainly, to a specific native language, this work points out the promises of pre-processing/ML/priority classification methods, regardless of the language of end-users' requests. Therefore, applications to other contexts, in terms of native language and building context are encouraged. At the same time, future applications to other automatic assignment tasks in the context of end-users' maintenance requests could be also provided, such as those on activity or workers' identification.

ACKNOWLEDGEMENTS

This research has received funding from the project Vitality – Project Code ECS00000041, CUP I33C22001330007 - funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.5 - 'Creation and strengthening of innovation ecosystems,' construction of 'territorial leaders in R&D' – Innovation Ecosystems - Project 'Innovation, digitalization and sustainability for the diffused economy in Central Italy – VITALITY' Call for tender No. 3277 of 30/12/2021, and Concession Decree No. 0001057.23-06-2022 of Italian Ministry of University funded by the European Union – NextGenerationEU.

REFERENCES

Baek, S., Jung, W. and Han, S.H. (2021), “A critical review of text-based research in construction: Data source, analysis method, and implications”, *Automation in Construction*, Elsevier B.V., Vol. 132 No. August, p. 103915, doi: 10.1016/j.autcon.2021.103915.

Bellandi, V., Bernasconi, C., Lodi, F., Palmonari, M., Pozzi, R., Ripamonti, M. and Siccardi, S. (2024a), “An entity-centric approach to manage court judgments based on Natural Language Processing”, *Computer Law & Security Review*, Vol. 52, p. 105904, doi: 10.1016/j.clsr.2023.105904.

Bellandi, V., Bernasconi, C., Lodi, F., Palmonari, M., Pozzi, R., Ripamonti, M. and Siccardi, S. (2024b), “An entity-centric approach to manage court judgments based on Natural Language Processing”, *Computer Law & Security Review*, Vol. 52, p. 105904, doi: 10.1016/j.clsr.2023.105904.

Bortolini, R. and Forcada, N. (2020), “Analysis of building maintenance requests using a text mining approach: building services evaluation”, *Building Research & Information*, Routledge, Vol. 48 No. 2, pp. 207–217, doi: 10.1080/09613218.2019.1609291.

Bortoluzzi, B., Efremov, I., Medina, C., Sobieraj, D. and McArthur, J.J. (2019), “Automating the creation of building information models for existing buildings”, *Automation in Construction*, Elsevier, Vol. 105 No. August 2018, p. 102838, doi: 10.1016/j.autcon.2019.102838.

Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L. and Bennadji, B. (2020), “Natural language processing model for managing maintenance requests in buildings”, *Buildings*, Vol. 10 No. 9, pp. 1–12, doi: 10.3390/BUILDINGS10090160.

Bugalia, N., Tarani, V., Kedia, J. and Gadekar, H. (2022), “Machine learning-based automated classification of worker-reported safety reports in construction”, *Journal of Information Technology in Construction*, Vol. 27, pp. 926–950, doi: 10.36680/j.itcon.2022.045.

Carroll, P., Singh, B. and Mangina, E. (2024), “Uncovering gender dimensions in energy policy using Natural Language Processing”, *Renewable and Sustainable Energy Reviews*, Vol. 193, p. 114281, doi: 10.1016/j.rser.2024.114281.

Chanter, B. and Swallow, P. (2007), *Building Maintenance Management*, edited by Chanter, B. and Swallow, P.*Building Maintenance Management: Second Edition*, Blackwell Publishing Ltd, Oxford, UK, doi: 10.1002/9780470692011.

Charette, R.P. and Marshall, H.E. (1999), “UNIFORMAT II Elemental Classification for Building Specifications, Cost Estimating, and Cost Analysis, NISTIR 6389, Gaithersburg, MD: National Institute of Standards and Technology”, p. 109.

Chen, K., Chen, W., Li, C.T. and Cheng, J.C.P. (2019), “A BIM-based location aware AR collaborative framework for facility maintenance management”, *Journal of Information Technology in Construction*, Vol. 24 No. July, pp. 360–380, doi: 10.36680/j.itcon.2019.019.

Chen, W., Chen, K., Cheng, J.C.P., Wang, Q. and Gan, V.J.L. (2018), “BIM-based framework for automatic scheduling of facility maintenance work orders”, *Automation in Construction*, Elsevier, Vol. 91 No. August 2017, pp. 15–30, doi: 10.1016/j.autcon.2018.03.007.

Çınar, Z.M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M. and Safaei, B. (2020), “Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0”, *Sustainability*, Vol. 12 No. 19, p. 8211, doi: 10.3390/su12198211.

D’Orazio, M., Bernardini, G. and Di Giuseppe, E. (2023), “Automated Priority Assignment of Building Maintenance Tasks Using Natural Language Processing and Machine Learning”, *Journal of Architectural Engineering*, Vol. 29 No. 3, doi: 10.1061/JAEIED.AEENG-1516.

D’Orazio, M., Di Giuseppe, E. and Bernardini, G. (2022), “Automatic detection of maintenance requests: Comparison of Human Manual Annotation and Sentiment Analysis techniques”, *Automation in Construction*, Vol. 134, p. 104068, doi: 10.1016/j.autcon.2021.104068.

Du, M., Li, F., Zheng, G. and Srikumar, V. (2017), “DeepLog: Anomaly detection and diagnosis from system logs through deep learning”, *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 1285–1298, doi: 10.1145/3133956.3134015.

Dzulkifli, N., Sarbini, N.N., Ibrahim, I.S., Abidin, N.I., Yahaya, F.M. and Nik Azizan, N.Z. (2021), “Review on maintenance issues toward building maintenance management best practices”, *Journal of Building Engineering*, Elsevier Ltd, Vol. 44 No. July, p. 102985, doi: 10.1016/j.jobe.2021.102985.

El-Dash, K. (2007), “Assessing Human Resource Management in Construction Projects in Kuwait”, *Journal of Asian Architecture and Building Engineering*, Vol. 6 No. 1, pp. 65–71, doi: 10.3130/jaabe.6.65.

Ferreira, C., Silva, A., de Brito, J., Dias, I.S. and Flores-Colen, I. (2021), “The impact of imperfect maintenance actions on the degradation of buildings’ envelope components”, *Journal of Building Engineering*, Elsevier Ltd, Vol. 33 No. June 2020, p. 101571, doi: 10.1016/j.jobe.2020.101571.

Fotovatfard, A. and Heravi, G. (2021), “Identifying key performance indicators for healthcare facilities maintenance”, *Journal of Building Engineering*, Vol. 42, p. 102838, doi: 10.1016/j.jobe.2021.102838.

Goins, J. and Moezzi, M. (2013), “Linking occupant complaints to building performance”, *Building Research & Information*, Routledge, Vol. 41 No. 3, pp. 361–372, doi: 10.1080/09613218.2013.763714.

Gonçalves, P., Araújo, M., Benevenuto, F. and Cha, M. (2013), “Comparing and combining sentiment analysis methods”, *COSN 2013 - Proceedings of the 2013 Conference on Online Social Networks*, pp. 27–37, doi: 10.1145/2512938.2512951.

Gunay, H.B., Shen, W. and Newsham, G. (2019), “Data analytics to improve building performance: A critical review”, *Automation in Construction*, Elsevier, Vol. 97 No. June 2018, pp. 96–109, doi: 10.1016/j.autcon.2018.10.020.

Gunay, H.B., Shen, W. and Yang, C. (2019), “Text-mining building maintenance work orders for component fault frequency”, *Building Research & Information*, Routledge, Vol. 47 No. 5, pp. 518–533, doi: 10.1080/09613218.2018.1459004.

Gutjahr, W.J. and Reiter, P. (2010), “Bi-objective project portfolio selection and staff assignment under uncertainty”, *Optimization*, Taylor & Francis, Vol. 59 No. 3, pp. 417–445, doi: 10.1080/02331931003700699.

Guyot, J., Falquet, G. and Teller, J. (2010), “Incremental development of a shared Urban ontology: The urbamet experience”, *Electronic Journal of Information Technology in Construction*, Vol. 15 No. February, pp. 132–139.

Hong, T., Koo, C., Kim, J., Lee, M. and Jeong, K. (2015), “A review on sustainable construction management strategies for monitoring, diagnosing, and retrofitting the building’s dynamic energy performance: Focused on the operation and maintenance phase”, *Applied Energy*, Elsevier Ltd, Vol. 155, pp. 671–707, doi: 10.1016/j.apenergy.2015.06.043.

Khan, W., Daud, A., Khan, K., Muhammad, S. and Haq, R. (2023), “Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends”, *Natural Language Processing Journal*, Vol. 4, p. 100026, doi: 10.1016/j.nlp.2023.100026.

Kim, Y., Bang, S., Sohn, J. and Kim, H. (2022), "Question answering method for infrastructure damage information retrieval from textual data using bidirectional encoder representations from transformers", *Automation in Construction*, Elsevier B.V., Vol. 134 No. October 2021, p. 104061, doi: 10.1016/j.autcon.2021.104061.

Kiritchenko, S. and Mohammad, S.M. (2017), "Best–Worst scaling more reliable than rating scales: A case study on sentiment intensity annotation", *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, Vol. 2 No. August, pp. 465–470, doi: 10.18653/v1/P17-2074.

Kula, B. and Ergen, E. (2018), "Review of Classification Systems for Facilities Management", *BIM-Classification*, No. September, pp. 12–14.

Leoni, L., BahooToroody, A., Abaei, M.M., Cantini, A., BahooToroody, F. and De Carlo, F. (2024), "Machine learning and deep learning for safety applications: Investigating the intellectual structure and the temporal evolution", *Safety Science*, Vol. 170, p. 106363, doi: 10.1016/j.ssci.2023.106363.

Licari, D. and Comandè, G. (2024a), "ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain", *Computer Law & Security Review*, Vol. 52, p. 105908, doi: 10.1016/j.clsr.2023.105908.

Licari, D. and Comandè, G. (2024b), "ITALIAN-LEGAL-BERT models for improving natural language processing tasks in the Italian legal domain", *Computer Law & Security Review*, Vol. 52, p. 105908, doi: 10.1016/j.clsr.2023.105908.

Louviere, J.J., Flynn, T.N. and Marley, A.A.J. (2015), *Best-Worst Scaling: Theory, Methods and Applications*, Cambridge University Press.

Loyola, M. (2018), "Big data in building design: A review", *Journal of Information Technology in Construction*, Vol. 23 No. October 2017, pp. 259–284.

Marocco, M. and Garofolo, I. (2021), "Operational text-mining methods for enhancing building maintenance management", *Building Research and Information*, Vol. 49 No. 8, pp. 893–911, doi: 10.1080/09613218.2021.1953368.

McArthur, J.J.J., Shahbazi, N., Fok, R., Raghubar, C., Bortoluzzi, B. and An, A. (2018), "Machine learning and BIM visualization for maintenance issue classification and enhanced data collection", *Advanced Engineering Informatics*, Elsevier, Vol. 38 No. October 2017, pp. 101–112, doi: 10.1016/j.aei.2018.06.007.

Mercha, E.M. and Benbrahim, H. (2023), "Machine learning and deep learning for sentiment analysis across languages: A survey", *Neurocomputing*, Vol. 531, pp. 195–216, doi: 10.1016/j.neucom.2023.02.015.

Mo, Y., Zhao, D., Du, J., Syal, M., Aziz, A. and Li, H. (2020), "Automated staff assignment for building maintenance using natural language processing", *Automation in Construction*, Vol. 113, p. 103150, doi: 10.1016/j.autcon.2020.103150.

Mohammad, S.M. (2018), "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words", *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, Vol. 1, pp. 174–184, doi: 10.18653/v1/p18-1017.

Pagano, R.R. (2012), *Understanding Statistics in the Behavioral Sciences*, Cengage Learning.

Parisi, F., Fanti, M.P. and Mangini, A.M. (2021), "Information and Communication Technologies applied to intelligent buildings: a review", *Journal of Information Technology in Construction*, Vol. 26, pp. 458–488, doi: 10.36680/j.itcon.2021.025.

Pärn, E.A., Edwards, D.J. and Sing, M.C.P. (2017), "The building information modelling trajectory in facilities management: A review", *Automation in Construction*, Elsevier B.V., Vol. 75, pp. 45–55, doi: 10.1016/j.autcon.2016.12.003.

Pishdad-Bozorgi, P., Gao, X., Eastman, C. and Self, A.P. (2018), “Planning and developing facility management-enabled building information model (FM-enabled BIM)”, *Automation in Construction*, Elsevier, Vol. 87 No. October 2017, pp. 22–38, doi: 10.1016/j.autcon.2017.12.004.

Rampini, L. and Cecconi, F.R. (2022), “Artificial intelligence in construction asset management: a review of present status, challenges and future opportunities”, *Journal of Information Technology in Construction*, Vol. 27, pp. 884–913, doi: 10.36680/j.itcon.2022.043.

Razali, M.N., Jamaluddin, A.F., Abdul Jalil, R. and Nguyen, T.K. (2020), “Big data analytics for predictive maintenance in maintenance management”, *Property Management*, Vol. 38 No. 4, pp. 513–529, doi: 10.1108/PM-12-2019-0070.

Ribeiro, F.N., Araújo, M., Gonçalves, P., André Gonçalves, M. and Benevenuto, F. (2016), “SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods”, *EPJ Data Science*, Ribeiro et al., Vol. 5 No. 1, pp. 1–29, doi: 10.1140/epjds/s13688-016-0085-1.

Sala, R., Pirola, F., Pezzotta, G. and Cavalieri, S. (2022), “NLP-based insights discovery for industrial asset and service improvement: an analysis of maintenance reports”, *IFAC-PapersOnLine*, Vol. 55 No. 2, pp. 522–527, doi: 10.1016/j.ifacol.2022.04.247.

Sanni-Anibire, M.O., Zin, R.M. and Olatunji, S.O. (2021), “Machine learning - based framework for construction delay mitigation”, *Journal of Information Technology in Construction*, Vol. 26, pp. 303–318, doi: 10.36680/j.itcon.2021.017.

Schmidt, T., Burghardt, M. and Dennerlein, K. (2018), “Sentiment annotation of historic German plays: An empirical study on annotation behavior”, *CEUR Workshop Proceedings*, Vol. 2155 No. January, pp. 47–52.

Shalabi, F. and Turkan, Y. (2020), “BIM–energy simulation approach for detecting building spaces with faults and problematic behavior”, *Journal of Information Technology in Construction*, Vol. 25, pp. 342–360, doi: 10.36680/j.itcon.2020.020.

Sourav Das Adhikari, Steve Hoffman and Britta Lietke. (2019), *Six Emerging Trends in Facilities Management Sourcing*.

Systems, B., Systems, B., Systems, B., Sav-, N., Systems, B., Systems, B., Systems, B., *et al.* (2007), “ASTM E1557 - 09(2015) Standard Classification for Building Elements and Related Sitework—UNIFORMAT II”, Vol. 09 No. Reapproved, doi: 10.1520/E1557-09R20E01.2.

Valdez, D., Soto-Vásquez, A.D. and Montenegro, M.S. (2023), “Geospatial vaccine misinformation risk on social media: Online insights from an English/Spanish natural language processing (NLP) analysis of vaccine-related tweets”, *Social Science & Medicine*, Vol. 339, p. 116365, doi: 10.1016/j.socscimed.2023.116365.

Wu, M.C. and Sun, S.H. (2006), “A project scheduling and staff assignment model considering learning effect”, *International Journal of Advanced Manufacturing Technology*, Vol. 28 No. 11–12, pp. 1190–1195, doi: 10.1007/s00170-004-2465-0.

Yan, H., Yang, N., Peng, Y. and Ren, Y. (2020), “Data mining in the construction industry: Present status, opportunities, and future trends”, *Automation in Construction*, Elsevier, Vol. 119 No. May, p. 103331, doi: 10.1016/j.autcon.2020.103331.

Yang, C., Shen, W., Chen, Q. and Gunay, B. (2018), “A practical solution for HVAC prognostics: Failure mode and effects analysis in building maintenance”, *Journal of Building Engineering*, Vol. 15 No. October 2017, pp. 26–32, doi: 10.1016/j.jobe.2017.10.013.

Žižka, J., Dařena, F. and Svoboda, A. (2019), *Text Mining with Machine Learning*, *Text Mining with Machine Learning*, doi: 10.1201/9780429469275.