Fault Log Text Classification Using Natural Language Processing And Machine Learning For Decision Support

2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA) | 978-1-6654-9334-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/SKIMA57145.2022.10029587

Darlington-Njoku Chidinma A.
School of Computing and Engineering
University of Gloucestershire
Gloucestershire, United Kingdom
chidinmadnjoku@gmail.com

Bhupesh Kumar Mishra
School of Computing and Engineering
University of Gloucestershire
Gloucestershire, United Kingdom
bmishra@glos.ac.uk

Will Sayers
School of Computing and Engineering
University of Gloucestershire
Gloucestershire, United Kingdom
wsayers@glos.ac.uk

***Abstract* - In recent years, various industries have been on the quest to derive new knowledge and information from the data they produce. When these data are well utilised, they can create frameworks for improving business processes, product quality, and services. However, more often, data are in unstructured and semi-structured data formats. Because of this, the discovery of critical issues within textual data becomes challenging. In the past few years, the adoption of natural language prepossessing (NLP) and machine learning (ML) techniques are increasingly becoming popular for exploring knowledge within text documents that could help decision-makers and experts to solve business challenges and improve their business processes and systems. This research is being experimented with NLP and ML on the fault log of a UK-based commercial MRO (Maintenance, Repair, and Overhaul) provider in the Aerospace Industry to support decision-making. The first stage systematically leverages text analysis to extract valuable information from many customers' fault notifications, compares its similarity with the expert's maintenance action, and then classifies them into three categories which are Modification, Replacement, and No-fault-found. In the second phase, the extracted features get fed into the machine learner to categorise and predict future faults diagnosis in commercial aircraft' FQIS (Fuel Quantity Indicating System) to automate troubleshooting, support maintenance operations, and improve decision-making in MRO services.**

Keywords—MRO (Maintenance, Repair and Overhaul); Natural Language Processing; Machine Learning; Classification; Text Mining; Aerospace Industry

I. Introduction

Over the years, a significant concern for many businesses is how to empower their experts to capture, transfer, derive, and comprehend the information inherent in such data [1]. In the aerospace industry, terabytes of data get generated from the critical systems modern aircraft are fitted with, and aircraft status is collected through the utilisation of sensor technology and the data generated from the repair shops, inventory systems, and different regulatory bodies. This avalanche of data is useless if its value is not unleased; hence, the aviation industry has to adopt storage, analysis, understanding, and translation solutions into meaningful MRO (Maintenance, Repair and Overhaul) measures using complex machine learning and artificial intelligence models [2]. Airlines companies are increasingly demanding improved reliability of components, superior performance, optimised parts management, heightened safety standards, and increased maintenance turn-around time. The unpredictability and fluctuations in the repair time of commercial aviation are caused by several factors, which also bring about a severe delays in repair services. These factors include maintenance, repair, and overhaul services. These, therefore, made MRO providers turn to data analytics, and machine learning techniques for intelligent automation, thereby depicting its application as the transformational tool for enhanced efficiency in MRO operations and improved customer experience.

This paper exploited NLP and ML-based text classification techniques on the historical Fault Log data. The standard approach adopted in this study was the extraction of the keyword and key phrase for the analysis of the fault notification and maintenance action documents, similarity measurement of these two documents, then classification models were built to classify faults and predict the category of maintenance action. The performance of the classification algorithms applied was evaluated using the F1 score as an effective quantitative metric and the results obtained indicated that the best-performing classifier was Convolutional Network, which achieved the highest level of training accuracy of 99% for class prediction of Maintenance Action and 93% for the Fault Notification features. The outcome demonstrates the reliability of the presented work in real-world aviation organisation applications in providing support to expert systems and decision-makers.

The rest of this paper is arranged thus: section 2 introduces the methodology of text analysis, similarity measurement, and machine learning classifier schemes; section 3 explains the implementation of text mining, text analysis, similarity measurement, and the NLP framework; section 4 presents the outcome and section 5 has the discussion and conclusion.

II. Methodology

Recognising fundamental concerns within textual data could greatly assist decision-makers and experts manage their prospective tasks more effectively. Automatic text classification (or categorisation) provides the blueprint for organising relevant information into relevant sets. This section, therefore, focuses on text classification workflow, which is made up of different stages, as illustrated in fig. 1, comprised of data collection, data pre-processing, exploratory text analysis, text pre-processing, classification, and performance evaluation.

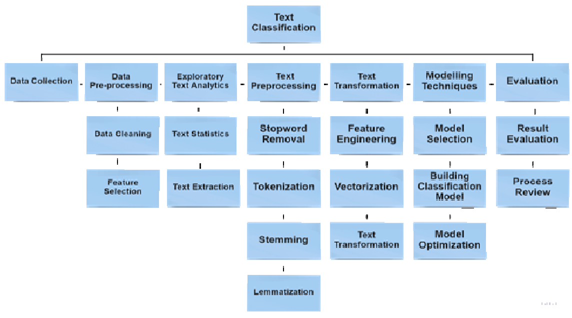

Fig. 1. Text Classification Workflow

A. *Data Acquisition*

The amount of text data being generated daily is increasing, and there is a pragmatic need for structuring them because they cannot be understood or processed by computers. Through data acquisition, meaningful data samples are collected to analyse and explore the hidden knowledge.

B. *Data pre-processing*

The documents used in text classification tasks are made up of thousands of values that are useful and not useful to the study, hence the need to eliminate non-useful values to reduce complications during experimentation [3]. This phase is a vital part of most text-mining projects. Its sequence of actions implemented includes data cleaning, keyword selection, feature extraction

C. *Text Mining*

It is challenging to handle the vast data organisations, and individuals produce. Conventional data mining tools do not have the capability of handling text data as much time is spent on information retrieval. Hence, text mining is essential which extract useful information and discovering patterns from textual data. Weiss et al. [4] illustrated that text mining and its interrelationship with other fields such as information retrieval, NLP, data mining, and ML. Text mining can be applied in every human endeavour field to analyse natural language texts in semi-structured and unstructured formats using different techniques [5].

D. *Text Pre-processing*

The implementation of text classification can only be done on pre-processed data as raw data consists of errors and noises. The text pre-processing technique eliminates noise existing in text and gets it ready for knowledge discovery operations; in most cases, information from the initial source is transformed into a usable form [6]. The text pre-processing task involves a sequence of actions; the general operations include tokenisation, a technique of splitting the words in the text document into distinctive components to become values for diverse NLP operations. Other steps like stop word removal are used for the removal of irrelevant words (for instance, is, of, an, Etc.), Similarly, case conversion, lemmatisation, and stemming, are used for the reduction of terms to their root.

E. *Text Analysis*

Most of the data generated in most business domains, in this case, the aerospace industry, are digital and unstructured; therefore, it is unavoidable that textual data overshadows statistical or numerical data [7]. This further draws on the need for text analytics, which is a part of text mining capable of bringing structure to text data by converting text to numbers. Also, it helps with pattern identification as the quality of the analysis depends on the data structure; it, in turn, influences the quality of decisions made [8].

F. *Text Representation*

The consideration of how to represent the text document must be done before fitting the model into the classifier. Comprehensive studies have been done to suggest diverse text representation techniques as many challenges are encountered during text classification operations. Harish et al. (2010) explain that the first challenge is the difficulty of capturing natural language complex ideas and high-level semantics from a few keywords. The second challenge is dealing with the significant number of documents with contents and quality of high dimensionality and varying lengths. These challenges have placed the need for transformation into a standard format. Thus, it is necessary to adopt an efficient document representation model to develop an effective classification technique.

G. *Text Similarity Measurement*

Text similarity measurement plays a significant role in tasks like document clustering, text classification, information retrieval, and machine translation. It assigns to documents real numbers between 1 and 0, a zero value indicates the complete dissimilarity between the documents, and one means that the documents are similar either lexically (similar character sequence) or semantically (having the same theme) [9]. The initial and essential part of the task is to find the similarity between words in a sentence or document, then represent the document with a vector which is used chiefly where each element signifies the value of a matching character within the document.

H. *Text Classification*

The demand for automated text processing methods, including text classification, has been brought about by the rise in data's velocity, volume, and veracity [10]. It is an operation where the documents are classified based on supervised knowledge, thus identifying text classification as an essential point of focus in the text mining field. Furthermore, it is the task of automatically assigning one or several predefined category labels (or classes or topics) to a given text written in a natural language according to its similarity to a previously labelled corpus used as a reference set [11,12]. Therefore, the automatic categorisation of text documents into one or more delineated classes is the objective of text classification tasks [13]. Text classification is one of the most crucial parts of supervised machine learning as it focuses on categorising or labelling textual units like sentences, queries, paragraphs, and documents. The supervised text classification for assigning labels to classes, the unsupervised text classification for drawing inferences from data into various groups that are not labelled, and the semi-supervised text classification technique that combines both supervised and unsupervised learning.

Nevertheless, this study focuses on text classification and the application of the supervised text classification approaches with the process flow presented in fig. 2.

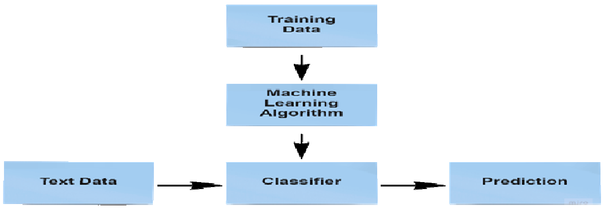

Fig. 2. Illustration of Text Classification Modelling Process Flow

In the supervised text classification method, the classifier learns from the training data to allocate predetermined category labels to new subjects [13,14]. There are many supervised text classification methods, including Naïve Bayes, Logistic Regression, Linear Regression, Random Forest, K-Nearest Neighbour, support Vector machine;

I. Performance Evaluation

This phase is vital for the measurement of the model's performance. Generally, in machine learning operations, the classifier's accuracy is evaluated using cross-validation. Another method of evaluating the classifier's performance is the confusion matrix, which is usually in the form of a matrix. The concept is to compare the actual values to the predicted values. On the contrary, there are various evaluation techniques for text categorisation tasks and supervised learning algorithms. Accuracy estimation being the easiest, precision, recall, and F-measure are the most prevalently used for performance evaluation [15].

III. IMPLEMENTATION

The implementation of this project is divided into eight phases. Each step is implemented in a sequential order to achieve higher classification accuracy.

A. Data Acquisition

This study analyses historical data for parts of commercial aircraft from 1997 to 2022 retrieved from a fault log for the repair shop of the MRO provider. It is information on maintenance that was filled by the maintenance team and is a record of fault notifications reported by customers on different parts in the fuel quantity indicating the system and the maintenance action carried out on them. The historical failure data for parts of the commercial aircraft from 1997 to 2022 was retrieved from a fault Log of the repair shop of the MRO provider presented in Table I; the part description has been altered for data privacy. It is a record of fault notifications for different parts in the FQIS (Fuel Quantity Indicator System) and the maintenance activities carried out on them.

Part Description	Faullt Notification	Maintenance Action
Part D	indicator fault of temperature	xxx replaced on xxx board. unit test to rev 2
Part B	repair	replaced unit xxxx with ####. software update not embodied. unit to cmm xxx
Part E	component to be replaced on 92vu. possible afff contermination	no fault found. no modification required by customer. unit final tested to xxxx
Part D	loss of component on rh wing; unit lost. attached xxxx	unit tested to xxxx no fault found.
Part B	returned for special investigation	unit repaired tested and released to cmm xxxx
Part C	xxxx level sensing	seal intact. faultlogs downloaded. carried out incoming tests- passed. all test passed. faultlogs show no fault shows no faults. reported fault unrelated to unit, unit opened for inspection -satisfactory.
Part A	Data recording, read class 3 faults. fuel fqis status. no data from adriu 1 34-12-34, further investigation	seals intact. faultlogs downloaded. carried out incoming test-passed all tests. faultlogs show no faults. reported fault unrelated to unit, unit opened for inspection -satisfactory.
Part B	repair	software upgrade not required. fitted now comp. re-tested satisfactory.

TABLE I. SAMPLE OF RAW DATASET

B. Data Pre-processing

Different types of issues were identified in the raw dataset, which was incompleteness such as fault notification not detailing the behaviour of part failure, missing values-for example- omitted report of maintenance activities carried out in correspondence to the fault mode detected and misspelling of part's description causing duplication and affecting the count of unique elements. These observed challenges require a clean-up, data restructuring, and other sequences of actions in the feature engineering phase.

Feature Engineering: The raw data was made up of several messages that documented the series of intercommunication between the customer and the service provider throughout the cases. It was in a free-text form and was compiled in an MS Excel file containing 27,615 records of fault notification and corresponding maintenance action. The dataset had eight features: the date of the notice, customer name (name of the customer reporting the failed part), Part Description (the name of the part in the fuel quantity indicating system), Part Number (the manufacturer's part identifier code), Fault Notification (the behaviours of the part as reported by the customer), Status, Serial Number, and maintenance action (the fault mode detected, and service carried out by the maintenance team). The data contained ten years of record for 952 unique parts, fault notification feature, 27,615 entries, and maintenance action had 27,615 records.

2)Feature Selection: Initial treatment of the data was performed, which involved the selection of three attributes: Part Description, Fault Notification, and Maintenance Action, which are essential for this study. On the other hand, the other five attributes, date of notification, customer name, Part Number Status, and Serial Number, were excluded from the observation. Each observation was utterly read for the labelling of the feature, and search terms such as 'overhaul', 'repair', 'modify', 'shelf life', Etc. were identified within the texts and used as the keyword representation for the selection and grouping process. The keywords manually identified within the text are used to obtain a count of the actions in that category.

3)Feature Extraction: 5 parts have been chosen for this study following the expert's opinion from the organisation. The five selected parts had 3648 observations recorded against them. Based on expert opinion, non-relevant features have been excluded from this study, leaving the new dataset with only the three pertinent features: part description, fault notification, and maintenance action.

4)Target Labelling: This phase of feature engineering involved the creation of the target column. For the accuracy of categorisation, each fault notification was manually observed against its corresponding maintenance action. A summary of eight actions was identified to be relevant in the text document: 'repair', 'replacement', 'scrap', 'modification', 'no-fault found', 'record closed', 'returned to customer/vendor', and 'others' others' (inadequate information recorded). A target label column named 'Class' was created to categorise both features as it is vital for the training of the machine learning classifier. Upon further examination, three categories, ' modification', 'replacement', and 'no-fault-found', were found to be the most predominant. There has been a huge imbalance existed in the dataset, was seen in the previous feature extraction section hence another selection was made using the three classes as the criteria. 500 samples were selected from each class to create a new balanced dataset of 1500 observations for the classification model.

5) Text Preprocessing

Since many irregularities were observed in the text, the data set has been cleaned up using the natural language tool kit (NLTK). The text pre-processing started with case transformation from uppercase strings to lowercase strings. Then, the noise in the data, which are the punctuations and special characters, was removed using the regular expression, the tokenisation was implemented to split the words into separate tokens, and some were removed (for example, words like on, for, the Etc.) as they are among the stop words and therefore nonessential for the task. The downside of the stop word removal on the fault notification and maintenance action documents is the exclusion of a word like "no", which adds meaning to the "no-fault found", "no reason on paperwork", and "no reason given" key phrases. Stemming and lemmatisation, using the Wordnetlemmatizer from NLTK, have also been performed for the derivation of the tokens' root word. For instance, words such as 'coating' and 'coated' in the data were used, stemmed, and lemmatised to become 'coat'. This operation is presented in table II; it helps capture the main information in the document and reduces the data's dimension for the efficiency of the classification task.

Raw Data	New Components Fitted; Front Unit, and Push Button
After Case Transformation	new components fitted; front unit, and push button
After Noise Removal	new components fitted front unit and push button
After Tokenisation	new, components, fitted, front, unit, and, push, button
After Stemming and Lemmatisation	new, compon, fit, front, unit, and, push, button
After Stop Word Removal	new, compon, fit, front, unit, push, button

TABLE II. SUMMARY OF PRE-PROCESSING RESULT

C. Text Statistics

The keyword analysis is implemented to investigate the distribution of the data; in addition to that, the results in Table III are used to prove the effectiveness of the pre-processing task. The statistics of the documents obtained before and after the pre-processing are different, showing that the pre-processing data operations which included removing unnecessary words and characters within the text, successfully removed all the values that are not relevant to the text classification task. The maximum number of words for both documents, as shown in Table III was reduced from 1506 obtained for the fault notification document before pre-processing to 162 and for the maintenance action document from 986 to 126 after the pre-processing task was performed.

Before Pre-Processing	Fault Notification	Maintenance Action
Maximum Number Of Words	1506	986
Minimum Number Of Words	3	23
Average Number Of Words	64	173
Count Of Unique Words	1379	1295
After Pre-Processing		
Maximum Number Of Words	162	126
Minimum Number Of Words	0	3
Average Number Of Words	8	21
Count Of Unique Words	1349	1143

TABLE III. SUMMARY STATISTICS OF DATA DISTRIBUTION

1) Similarity Measurement and Word Representation

Before performing the similarity measurement, it is essential to represent the text data mathematically, so the TF-IDF vectorisation method was employed because of its ability to effectively calculate the similarity between two texts using the Cosine similarity measure. In this implementation phase, the reference documents have been compared against each other to derive a similarity score. This is an essential phase because it is implemented to determine how both documents match. Furthermore, this operation has answered the correlation between the customers' fault complaints and the expert action as a result.

Given the two input documents, the fault notification and the maintenance action documents, experimentation was done automatically using a cosine similarity measure to derive a score between 0 and 1 that measures the relationship between documents using the cosine similarity measure. The measurement was done in two phases Firstly, it measured the similarity between the fault notification document and the maintenance action document as a whole, splitting both documents and grouping them by their class to measure the similarity of each sub-document by its equivalent in the same class.

The outcome for the first stage, as shown in Table IV with a value of 0.0701 that is far below 1, demonstrates that there is no significant closeness between the fault notification document and the maintenance action document and also indicates that there is a difference between customers' complaints and the maintenance activities carried out on failed parts. To further check for the possibility of improved performance, both documents were split and compared using the three categories. Table V shows that for all of the categories, the similarity score obtained was very low, indicating the absence of a match between the documents and, in essence, a mismatch between the MRO customer's reason for returning a part and the expert's activities carried out on it. In conclusion, since all the similarity scores derived are either zero or close to zero, they strongly indicate that the text documents are not matching.

Document	Similarity Score
Maintenance Action: Fault Notification	0.0701

TABLE IV. TEXT SIMILARITY MEASUREMENT SCORES FOR THE MAINTENANCE ACTION AND FAULT NOTIFICATION DOCUMENTS

Category	Document	Similarity Score
Modification	Maintenance Action: Fault Notification	0.001
No-Fault-Found	Maintenance Action: Fault Notification	0.007
Replacement	Maintenance Action: Fault Notification	0.002

TABLE V. SIMILARITY SCORES FOR THE THREE CLASSES OF MAINTENANCE ACTION AND FAULT NOTIFICATION DOCUMENTS

D. *Text Classification*

This stage aims to develop a text classification model that can learn from the historical data and automatically classify and predict the category of maintenance action accurately, then provide answers on how this technique can be helpful for expert action and decision-making. After the pre-processing data phase, the final dataset of 1500 samples had three features, the part description, fault notification, and the maintenance action with one column of labelled data; hence the text classification project became a multi-class classification task. The data is split into 80% for training and 20% for testing, thus giving 1200 for training and 300 for testing.

For the classifier's training, eight classification algorithms consisting of Multinomial Naïve Bayes, Logistic Regression, SDG SVM Classifier, Random Forest, Decision Tree, Deep learning Functional and Sequential Models, and the Convolutional Network (CNN) were implemented. Among these models, CNN has appeared as the model with the highest accuracy. CNN has been implemented in Keras. Rectified Linear Unit (ReLU) has been used as the hidden layer. The SoftMax activation function has been applied as the output layer because of the multi-class labels, and the Global max pooling 1D layer at the output. Finally, the model was compiled, and 832,851 parameters were available for training.

IV. PERFORMANCE EVALUATION

The CNN outperformed all the classification algorithms with the highest level of accuracy of 94% for the Fault Notification and 99% for the Maintenance Action features. For most of the classifiers, hyperparameter tuning was performed, for example, changing the number of epochs and adding padding to make the input length equal to a maximum of 200 for deep learning. For overall accuracy, the Convolutional Model is the best algorithm for the classification of the maintenance action document at 99% and an F1 score of 94%, and for the fault notification document with an accuracy of 94% and an F1 score of 0.34 (which is not too far from the 0.37 score achieved by Naïve Bayes that has a shallow level of accuracy at 37%).

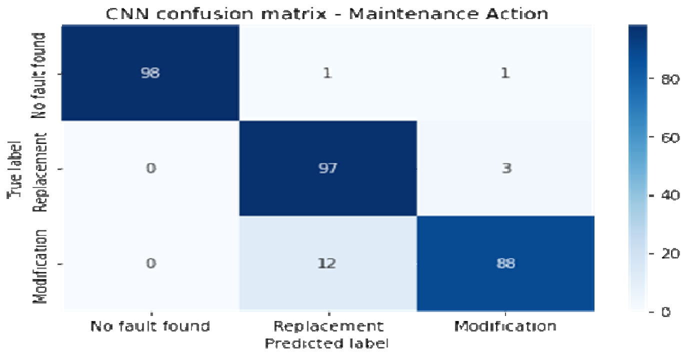

Fig. 3. Confusion Matrix For Maintenance Action using Convolutional Network

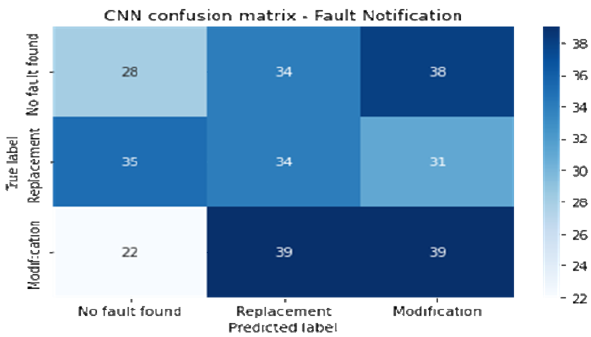

Fig. 4. Confusion Matrix For Fault Notification using Convolutional Network

V. CONCLUSION

In this study, text analysis has been performed using different NLP techniques to discover the most frequently occurring words, keywords, and key phrases and compare different ML algorithms to transfer learning and appraise their performances. In addition, an NLP framework is presented with details that can be recommended for classifying fault notification and maintenance action into various classes of predicted expert action for similar use cases. The classification accuracy using CNN indicated it has the highest training accuracy of 99% for class prediction with the maintenance Action feature and 93% for the Fault Notification feature 94% outperforming all the other classifiers. Data sensitivity was exposed through this framework alongside the reliability of Natural Language applications and Text Classification techniques to provide support to expert systems and decision-makers in the specialised MRO (Maintenance, Repair, and Overhaul) field. Despite implementing this framework using the Aerospace industry MRO as a case study, it is also applicable to most text-based documents in diverse organisations and industries as it is beneficial for saving decision-making time through intelligent automation.

Some challenges encountered during this project are inconsistencies in data entry in terms of abbreviations and typographical errors, potentially affecting the classification model's training and accuracy of prediction. Furthermore, the annotation was done manually, and the possibility of bias exists as further feature engineering was not achievable given the time constraint and data limitations. In the future, studies would explore automatic annotation approaches for data labelling and multi-label classification techniques for improved performance. The possibility of auto-annotation of the data would also be explored. In addition, designing a web-based application on the suggested framework would be investigated with the possibility of more data acquisition.

VI. References

[1] T. Wang and H. Chiang, "Fuzzy support vector machine for multiclass text categorization," *Information Processing & Management*, vol. 43, no. 4, pp. 914–929, 2007.

[2] B. Shukla, I. Fan, and I. Jennions, "Opportunities for Explainable Artificial Intelligence in Aerospace Predictive Maintenance," in *PHM Society European Conference*, vol. 5, no. 1, pp. 11–11, 2020.

[3] T. Wang and H. Chiang, "Fuzzy support vector machine for multiclass text categorization," *Information Processing & Management*, vol. 43, no. 4, pp. 914–929, 2007.

[4] S. M. Weiss, N. Indurkhya, T. Zhang, and F. Damerau, *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media, 2010.

[5] M. Allahyari *et al.*, "Text Summarization Techniques: A Brief Survey," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017, doi: 10.14569/ijacsa.2017.081052.

[6] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 20–38, Mar. 2019, doi: 10.1016/j.ijresmar.2018.09.009.

[7] N. Sadman, S. Tasneem, A. Haque, M. M. Islam, M. M. Ahsan, and K. D. Gupta, "'can nlp techniques be utilized as a reliable tool for medical science?'building a nlp framework to classify medical reports," in *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2020, pp. 0159–0166.

[8] M. Thangaraj and M. Sivakami, "Text classification techniques: A literature review," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 13, p. 117, 2018.

[9] M. Vijaymeena and K. Kavitha, "A survey on similarity measures in text mining," *Machine Learning and Applications: An International Journal*, vol. 3, no. 2, pp. 19–28, 2016.

[10] M. Thangaraj and M. Sivakami, "Text classification techniques: A literature review," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 13, p. 117, 2018.

[11] S. M. Weiss, N. Indurkhya, T. Zhang, and F. Damerau, *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media, 2010.

[12] J. VillenaRomán, S. ColladaPérez, S. LanaSerrano, and GonzálezCristóbal, José Carlos, "Hybrid approach combining machine learning and a rulebased expert system for text categorization," 2011.

[13] A. I. Kadhim, "Survey on supervised machine learning techniques for automatic text classification," *Artificial Intelligence Review*, Jan. 2019, doi: 10.1007/s10462-018-09677-1.

[14] A. Géron, *Handson machine learning with ScikitLearn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2019.

[15] Kowsari, Jafari Meimandi, Heidarysafa, Mendu, Barnes, and Brown, "Text Classification Algorithms: A Survey," *Information*, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.