Journal of Autonomous Intelligence (2024) Volume 7 Issue 1
doi: 10.32629/jai.v7i1.874

ORIGINAL RESEARCH ARTICLE

SOSFloodFinder: A text-based priority classification system for enhanced decision-making in optimizing emergency flood response

Siti Husna Kamal[1], Azwa Abdul Aziz[1,*], Wan Azani Mustafa[2,3]

[1] *Faculty of Informatics & Computing, Universiti Sultan Zainal Abidin (UniSZA), Besut Campus, Terengganu 22200, Malaysia*

[2] *Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis (UniMAP), Pauh Putra Campus, Arau 02600, Malaysia*

[3] *Advanced Computing, Centre of Excellence (CoE), Universiti Malaysia Perlis, Arau 02600, Malaysia*

*** Corresponding author:** Azwa Abdul Aziz, azwaaziz@unisza.edu.my

ABSTRACT

Flooding is a significant concern in nations with frequent precipitation because it can instantly affect multiple regions simultaneously. Due to the unpredictability of their occurrence caused by rapid water level rise, it is challenging to predict such natural disasters accurately. During flooding, prompt rescue efforts are crucial for the affected population. Due to flooded highways and residences, rescue teams may have difficulty locating victims. This hinders the potentially perilous and time-consuming rescue operation. To address this problem, we propose a web-based system that integrates natural language processing (NLP) with global positioning system (GPS) functionality. The SOSFloodFinder system provides automatic classification priorities for text messages sent by flood victims, as well as their most recent or current locations. The classification of text based on priority enables efficient resource allocation during rescue operations. In conclusion, this system has the potential to reduce future flood-related fatalities. Additional research and development are necessary to thoroughly investigate this method's practical capabilities and effectiveness.

Keywords: text-based priority; natural language processing; GPS integration

ARTICLE INFO

Received: 3 July 2023
Accepted: 1 August 2023
Available online: 8 November 2023

1. Introduction

Malaysia, located in Southeast Asia, has a sweltering and humid tropical rainforest climate with year-round high temperatures and abundant precipitation[1]. While the country is not prone to tsunamis or typhoons, the annual monsoon season causes significant inundation along the east coast and East Malaysia. Floods are Malaysia's most prevalent natural calamity, and they pose a severe threat to the population[2].

Floods can result in a more significant number of deaths than other hazards associated with thunderstorms. Common flood-related causes of mortality include vehicles being carried away by floodwaters and electrocution. In addition, floodwaters can disseminate infectious diseases, transport dangerous substances, and pose electrical hazards[3]. Effective communication between flood victims and rescue teams is crucial during such disasters.

However, traditional communication channels frequently fail, disconnecting victims and responders and leaving them without vital information. As was observed during the flooding in Shah Alam in 2021, social media platforms such as Twitter and Facebook are utilized by individuals when they are unable to access emergency phone lines. However, limited time and resources restrict the ability of first responders to actively monitor and respond to numerous postings and inquiries. Consequently, there is a chance that communications will be lost amidst the constant flow of online content, especially considering the possibility of network disruptions.

Thus, this study aims to resolve the difficulties encountered during flood emergencies by proposing a web-based system that integrates text analytics algorithms and GPS functionality. This system allows rescue teams to monitor and locate individuals, even during nighttime operations, by utilizing text messages sent by flood victims that indicate their location. In addition, the system integrates a mechanism for prioritization based on the contents of these text messages, allowing for efficient resource allocation during rescue operations.

Moreover, by leveraging the effectiveness of text-based communication and cutting-edge technology, this system has the potential to substantially reduce the number of casualties during future flood disasters. Further research and development are necessary to thoroughly explore and optimize the practical capabilities of this innovative approach, thereby ensuring its successful application in actual emergencies. The structure of the subsequent sections of this paper is as follows: Section 2 provides a comprehensive literature review on text analytics, GPS, and natural disaster management-related research. In section 3, the methodology utilized in the creation and implementation of the text-based priority system is described in detail. Section 4 provides experimental results and evaluation metrics. Section 5 emphasizes the significance of text analytics in the domain of flood disasters by discussing the study's findings, limitations, and future directions. The conclusion of the study emphasizes the significance of text analytics in flood disasters.

2. Literature review

Flood disasters represent one of the most devastating and recurring natural hazards globally, leading to significant economic losses, loss of human lives, and environmental damage. It has long been recognized as one of the most frequent and destructive natural disasters, posing significant threats to communities, economies, and ecosystems worldwide. The relentless force of floodwaters, fueled by intense rainfall, snowmelt, storm surges, or the failure of water infrastructure, can wreak havoc on landscapes and inundate entire regions. With climate change exacerbating extreme weather events, including intense precipitation and rising sea levels, flood disasters have become more frequent and severe in recent decades.

The increasing frequency and severity of flood events around the world have underscored the urgent need for robust and efficient emergency flood systems. An emergency flood system is a comprehensive and integrated framework designed to address the challenges posed by flooding, safeguard human lives, protect infrastructure, and facilitate effective disaster response and recovery. These systems combine advanced technology, real-time data monitoring, community engagement, and coordinated responses from multiple stakeholders to mitigate the devastating impacts of flood disasters.

To mitigate the impact of flood during disaster several flood management system are developed for various objectives. For example to restore the network problems during flood disaster, Liu et al.[4] propose a two-stage emergency power supply optimisation model is suggested to increase the distribution network's resilience. Multiple island micro-networks are constructed by combining medium and low mobile emergency generators and network reconstruction, and power supply may be promptly restored according to load priority order. This is done in consideration of the vulnerability modelling of various equipment during flood disaster. One of major challenges of this system is the cost and practically of proposed solution.

In another work by Fan et al.[5] utilizes WebGIS technology for geological disaster emergency decision support system supported by location-based service and satellite communication technology. In order to reduce the loss of lives and property due to geological disasters while enhancing the safety and security capabilities of emergency rescue personnel, it realises business functions such as comprehensive management of geological disaster data, quick report of disaster and danger, three-dimensional desktop deduction of the emergency plan, monitoring and dispatching of rescue personnel. A study by Weerasinghe and Jayasena[6] uses a multimedia big data platform with a deep learning approach for flood emergency management. For mining purposes, it leverages multimedia data that is openly accessible, such as social media (Twitter and Facebook) data, satellite image data, crowdsourcing, and sensor network data. However, the most concern is the availability of network when disaster occurred.

This study focuses on text analytic and GPS technology to provide intelligence emergency response system based on monsoon season flood in Malaysia. A method known as text priority classification is proposed to provide ranking order of high to low critical of emergency request during the peak hour of flood disaster.

2.1. Text analytics

Text analytics analyses unstructured data and extracts valuable insights through pre-processing methods[7]. Text classification and text extraction are two of the most prominent techniques employed in text analytics. These methods have found widespread application in numerous scenarios related to natural disasters. In one notable academic publication proposed a framework named the “Twitter Situational Awareness” (TwiSA) which combined sentiment analysis and topic modeling to analyze flood-related tweets during the 2015 South Carolina flood using text analytics[7]. TwiSA offered real-time feedback, making it a cost-effective and efficient solution for disaster management. It successfully identified temporal patterns of public concerns, enabling disaster managers to develop a better SA and decision-making plans. However, limitations included potential noise and misinformation in social media data, and the focus on negative sentiments may overlook positive and neutral expressions.

These limitations are relevant when compared to other studies in disaster relief optimization, as demonstrated by the pioneering study proposed scheduling algorithms to match volunteers and flood victims during Hurricane Harvey[8]. While the scheduling approach effectively reduced waiting times for victims and maximized volunteer resource utilization, it relied on Twitter data, introducing potential biases and limited generalizability. Moreover, ethical concerns related to data privacy and coordination between agencies were not extensively addressed.

Similarly, another study focused on surveying Twitter users who requested rescue during Hurricane Harvey to identify their socioeconomic characteristics and post-disaster satisfaction levels with their area of residence[9]. While valuable in understanding user demographics and experiences, the survey-based approach suffered from survival bias and limited scalability. Furthermore, neither of these studies utilized automated priority classification or GPS integration, missing opportunities for real-time response, precise location determination, and personalized prioritization.

Another study proposed an automated set of services that integrated social media data with weather forecasts to enhance emergency response during extreme weather events and natural disasters[10]. By utilizing ensemble forecasting and event detection algorithms, the system provided qualitative feedback for meteorological models, detected emergency events in real-time, and extracted informative content from social media platforms. This approach demonstrated the potential to improve situational awareness for monitoring agencies, meteorological offices, and first responders, enhancing early warning capabilities and disaster response efforts. However, challenges related to data accuracy, volume, and privacy were acknowledged, indicating the need for further validation and addressing ethical concerns. While promising, this automated approach could also encounter noise and misinformation in social media data, similar to the TwiSA framework. Additionally, ethical considerations, such as data privacy and agency coordination, should be carefully addressed to ensure responsible and effective use of social media data in emergency response strategies. Consequently, text analytics is helpful in analysing numerous data, which always happens during emergency situations such as floods.

2.2. GPS

The United States government developed the global positioning system (GPS), a globally deployed satellite navigation system with multiple vital functions[11]. These functions include the precise detection of object locations and positions, navigation capabilities for traversing between different areas, continuous tracking of object movements, generation of comprehensive global maps, and transmission of exact time signals with an accuracy of 10 billionths of a second. It significantly impacted our daily lives by making our surroundings safer and more convenient. GPS technology is crucial in disaster management, specifically in monitoring and analyzing geographical data during emergencies[12].

The study by Fale et al.[13] addresses a mobile application designed to aid flood victims stranded during a flood event. The application is a communication platform between the victims and the rescue team, providing vital GPS location data to facilitate prompt rescue operations. The system uses the travelling salesman problem (TSP) algorithm to determine the quickest route for the rescue team to reach the user's location, thereby accelerating the rescue operation. The system's implementation of distinct folders for categorizing working, pending, and finished tasks as shown in **Figure 1** is notable.

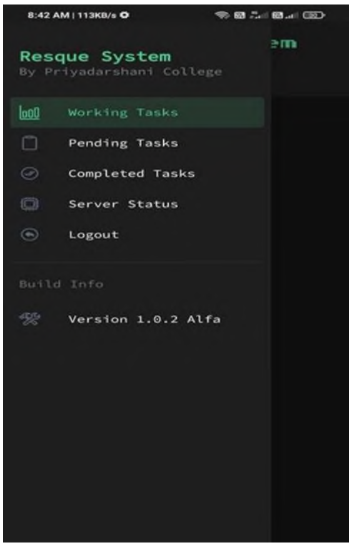


Figure 1. Rescue system's interface[13].

This organizational structure allows the rescue team to manage and monitor their designated tasks effectively, reducing the risk of task neglect.

Nonetheless, it is essential to recognize the limitations of this method when working with a large volume of data. Due to the examination and evaluation of each request that must be performed manually, the system may become less efficient and effective as time passes. This indicates the need for alternative strategies or enhancements to efficiently manage and prioritize requests.

One notable study proposed an Android-based application prototype that utilized Mobile Ad hoc Network (MANET) to provide communication and information required for search and rescue operations during disasters[14]. The application's peer-to-peer (P2P) communication approach, supported by GPS integration, allowed devices to share real-time location information and exchange text messages within a limited 50 m range. While the prototype demonstrated the feasibility of MANET in emergency scenarios, it had some limitations. Its limited range was mentioned as 50 m. While this may be suitable for certain scenarios, it may not be sufficient for larger disaster areas or situations where search and rescue teams are spread over a wider region.

Other than that, an article proposed MobiRescue, a human mobility-based rescue team dispatching system designed to optimize rescue operations during flooding disasters[15]. Utilizing machine learning techniques such as support vector machine (SVM), reinforcement learning (RL) and GPS, MobiRescue aims to maximize the total number of fulfilled rescue requests, minimize rescue teams' driving delays to request positions, and reduce the number of dispatched rescue teams. The system utilizes real-world human mobility data from Hurricane Florence for trace-driven experiments, demonstrating its superior performance compared to existing emergency vehicle dispatching methods and rescue team dispatching methods. MobiRescue's strengths lie in its innovative approach, leveraging SVM to predict the distribution of potential rescue requests based on disaster-related factors, and using RL for real-time dispatching guidance. The system optimizes rescue team routes by considering the impact severities in different regions and people's movement patterns during disasters.

Another article proposed a comprehensive earthquake emergency rescue command system that leverages the global position system (GPS) and geographic information system (GIS) technologies[16]. The system's key strengths include its ability to achieve accurate real-time location tracking and data visualization through electronic maps. Additionally, the integration of heuristic and improved genetic algorithms enables optimal path planning, leading to increased efficiency in rescue operations. The system's three-layer structure and functional modules, such as disaster monitoring, ruin's structure analysis, and management statistics, contribute to its versatility and applicability in various emergency scenarios. However, it might be challenging to implement and maintain due to its technical complexity, especially for organizations with limited IT expertise and resources.

3. Materials and method

Agile development, which is characterized by its iterative and incremental approach, has emerged as the preferred method for this study. This method is a combination of incremental and iterative processes in which the product is released in ongoing cycles and then tested and enhanced after each iteration.

Figure 2. Agile model.

As depicted in **Figure 2**, agile-based products are developed by dividing the entire production process into phases, allowing for quicker change execution and deployment. In contrast to the waterfall model, agile-based products do not necessitate specific deployment downtime periods. In this context, the six phases of agile development are requirements, design, development, testing, deployment, and review. By utilizing agile, development teams can avoid product defects by focusing on individual phases without being burdened by other or previous tasks, resulting in increased development process efficiency and adaptability.

On top of that, as shown in **Figure 3**, the proposed framework for this study consists of three primary stages.

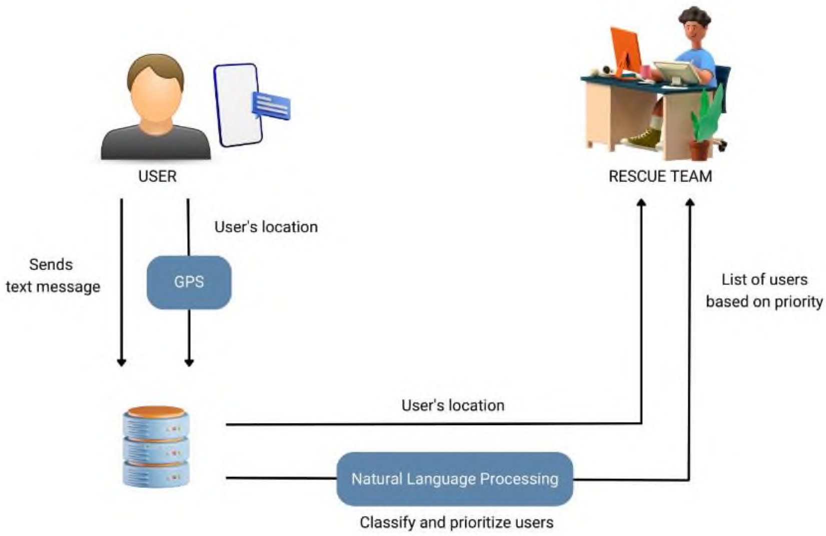


Figure 3. Framework.

In the initial stage, a user initiated the process by sending a text message including the location obtained via global positioning system (GPS) technology, which is then stored in a database for subsequent processing. Textual data within the database is analyzed using natural language processing (NLP) techniques known as text analytics in the second stage. This method seeks to categorize and prioritize users according to the content of their messages. In the final stage, the system generates a comprehensive report for the rescue team using the processed information, including the user's location and a list of users ranked by priority. This report enables the rescue team to respond to emergencies effectively by providing relevant information and facilitating quick decision-making. This framework integrates text message collection, text analytics, and priority-based user management to improve the efficacy as well as the effectiveness of emergency response operations.

The categorization procedure entails comparing the keywords provided by the rescue team and the words found in the user's messages. These keywords are further subdivided into three distinct categories, corresponding to the user categorization (high, medium, low) and indicating their respective priority levels. Users are classified according to a stringent set of criteria. The users are classified accordingly if the content messages contain only keywords from a single category (high, medium, or low). The user is automatically categorized as high when the content messages contain keywords from all categories. Users are classified as medium if their content messages comprise keywords from the medium and low categories. Moreover, an entity-relationship diagram (ERD) of this system described the system's entities and their relationships as seen in **Figure 4**.

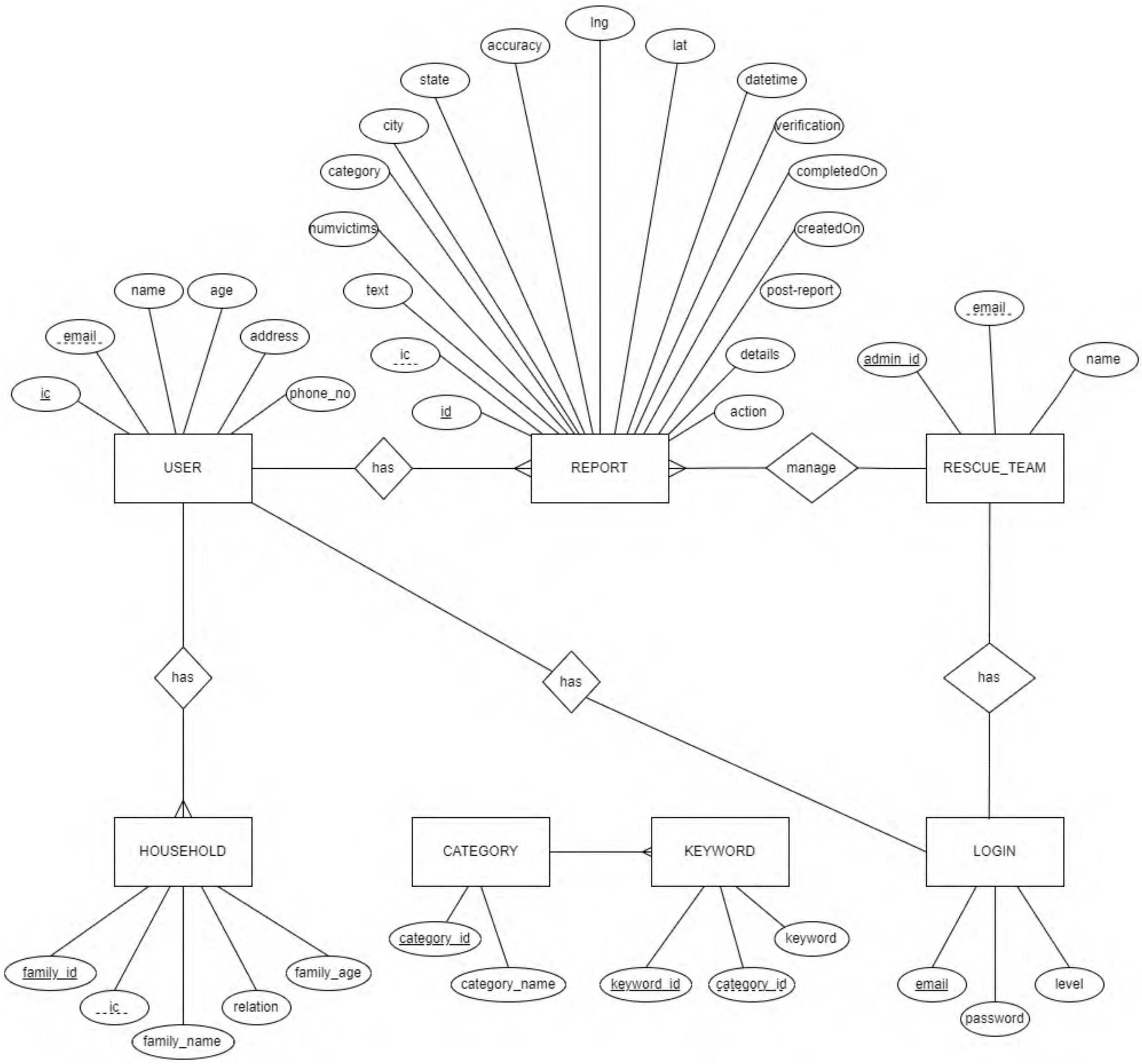


Figure 4. Entity relationship diagram (ERD).

Each entity has the displayed attributes. The relationship between entities clarifies the procedure that each entity will execute. Seven entities, or database tables, are shown in **Figure 4**. Login, user, household, report, rescue team, category and keyword are the names of these tables.

First, the "login" entity has a one-to-one relationship with both "user" and "rescue team". This means that there is one unique "login" account for each "user" and "rescue team". The "email" attribute served as the primary key, ensuring a unique identification for each user. Additionally, the "level" attribute determined the user's role, distinguishing between administrators and regular users within the system. Another attribute included the "password", which was implemented to enhance the security of users' sensitive personal information.

The second interaction occurs between the "user" and the "household". it is a one-to-many relationship, where a single "user" can be associated with multiple "households". The primary key for the "user" entity is their identity card number ("ic"), while the primary key for the "household" entity is the "familyID". The "ic" attribute serves as a foreign key in the "household" table. In addition, other attributes for "user" table included the user's name ("name"), age ("age"), address ("address"), and phone number ("phone_no"). On the other hand, other attributes for “household” table encompassed the family member’s name (“family_name”), their relationship with the user (“relation”), and their age (“age”).

The third interaction is between the “user” and the “report”. It is a one-to-many relationship, allowing a single “user” to have multiple “reports”.

The fourth interaction involves the “rescue team” and the “report”. A “rescue team” can manage several “reports”, establishing a one-to-many relationship. The primary key for the “report” table is the “id”, while the “ic” attribute serves as the foreign key. The content of text messages (“text”) and the number of victims involved (“numvictims”) were used to assign a priority level in the “category” column. The date and time of each text message sent were recorded in the “datetime” attribute. The table also included user location details, such as the accuracy of the captured location (“accuracy”), latitude (“latitude”), longitude (“longitude”), city (“city”), and state (“state”). The “verification” attribute indicated whether users with their assigned priority levels had been verified. The “details” and “createdOn” attributes were used to record ongoing cases of victims, along with the date and time of the report. The “action” attribute distinguished between ongoing cases and completed cases. In addition, the post report provided by the rescue team was saved in the “postreport” and “completedOn” attributes, which documented the completion date and time of the report.

Finally, the last interaction occurs between the “category” and “keyword” entities, establishing a one-to-many relationship. In this relationship, a single “category” can have multiple associated “keywords”. Additionally, the “category” table included an attribute called “category_name”, which represented the name of each category. While, another attribute featured in the “keyword” table was the actual keyword itself, denoted by the “keyword” attribute.

In addition, it is essential to highlight the significance of the pre-processing stage applied to the keywords, as depicted in **Figure 5**.

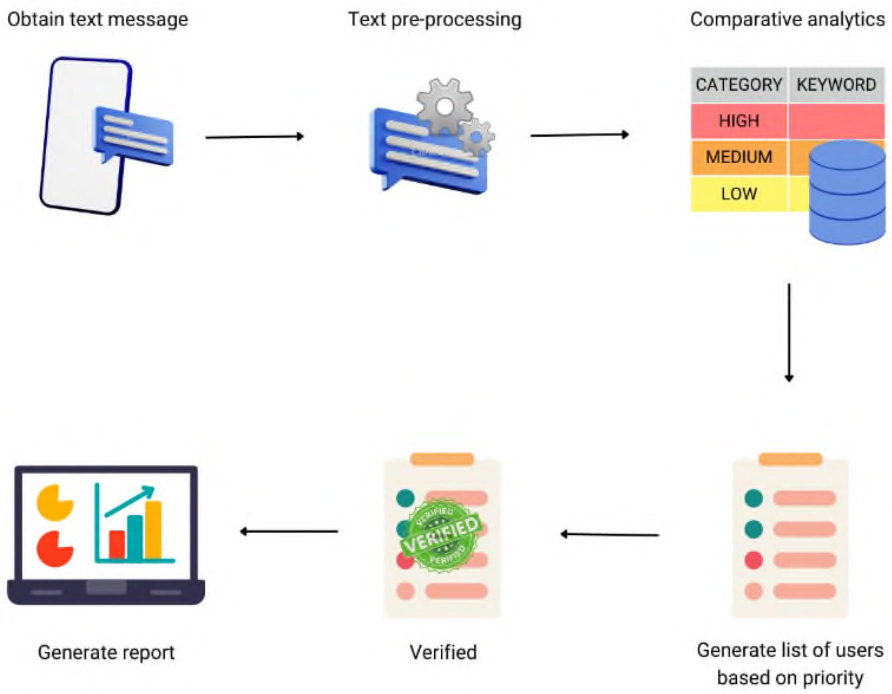


Figure 5. Text analytics.

When a user sends a text message, the message is pre-processed to remove noise, standardize formats, and improve data quality. This pre-processing involves removing irrelevant characters and punctuation, followed by tokenizing the text into individual words. This pre-processing is essential for enhancing the categorization process's accuracy. The processed text is then compared to the existing keywords in the database, resulting in the automatic designation of users to particular categories and the generation of a prioritized user list. The generated list is then verified by the rescue team in order to improve the categorization's accuracy. **Figure 6** depicts the text pre-processing procedure, outlining the stages of preparing textual data for further analysis and categorization.

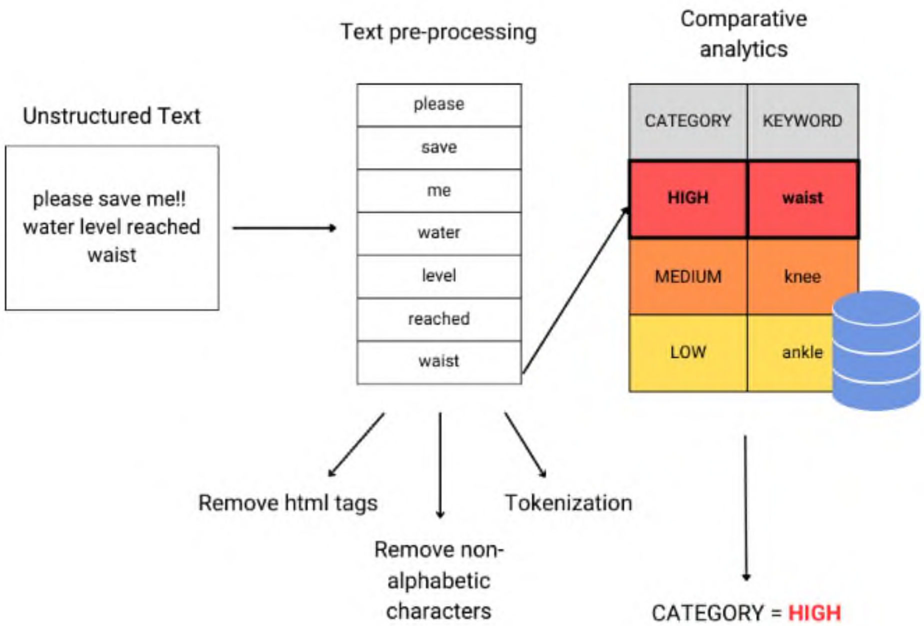


Figure 6. Text pre-processing process.

In the initial stage, textual data is acquired when a user submits a message in sentences. Following text pre-processing, the text is tokenized into individual words by removing HTML elements and non-alphabetic characters using PHP functions such as "strip_tags", "preg_replace" and "explode". The processed text is compared, word by word, against the database-stored keywords through If-Else conditions. Users are automatically designated to the corresponding categories when a match is discovered. From **Figure 6**, word "waist" matched with the keyword in high category. As a result, the corresponding user is categorized as high. Finally, a comprehensive report consisting of figures and tables is generated to assist the rescue team in analyzing the collected data.

4. Result and discussion

This section describes the implementation of SOSFloodFinder in accordance with the specifications outlined in this study. The objective is to provide an interface that enhances user-friendliness and facilitates the utilization of the system.

4.1. Data capturing

When the user provides vital information, such as the current number of victims and detailed descriptions of their respective situations, as depicted in **Figure 7**, the system automatically captures their location after obtaining user consent.

This feature enables the rescue team to acquire essential information that aids in the effective planning and execution of rescue operations.

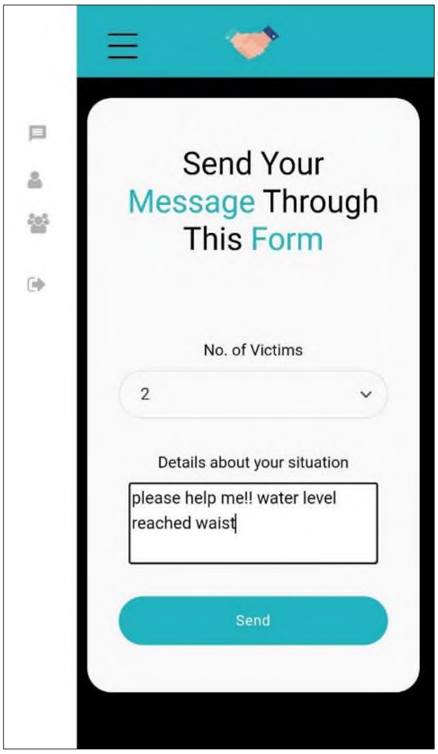


Figure 7. Message page.

4.2. Text priority classification

Pre-processing is necessary to eliminate noise and improve the accuracy of classification practices to extract valuable insights from unstructured text provided by users. In this regard, the system executes multiple pre-processing processes. First, it removes HTML elements such as “<div>”, “<b>”, and “<a>” from the original data, if present. The text is then stripped of non-alphabetic characters, such as punctuation marks, brackets, and spaces. Finally, the text is tokenized by dividing paragraphs or sentences into tokens. As depicted in **Figures 8** and **9**, this processed text can then be precisely compared with the keywords stored in the database. The pre-processing phase substantially improves the effectiveness of classification tasks and meaningful data extraction.

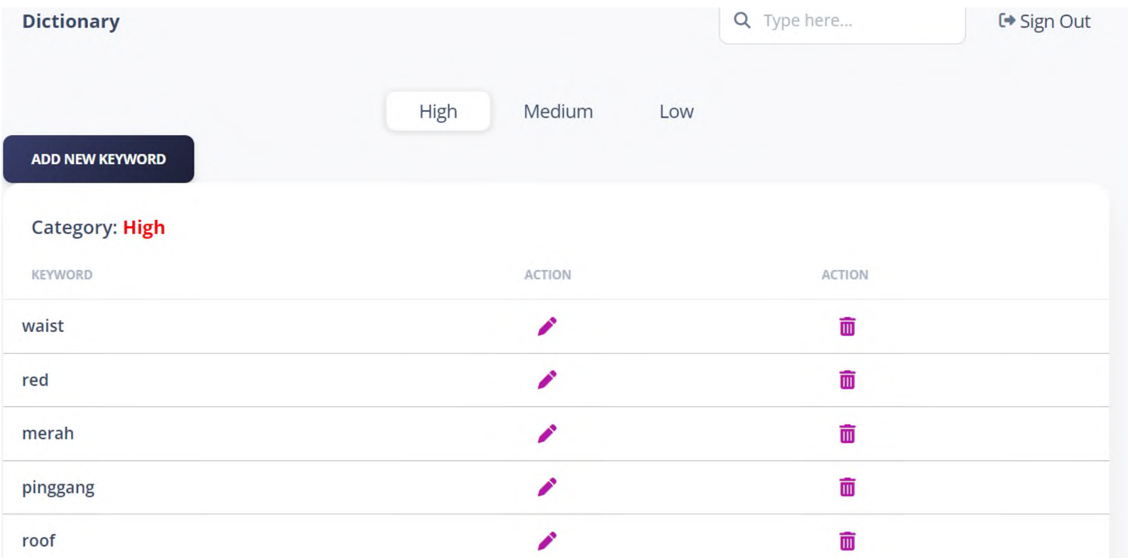


Figure 8. Keyword page for high category.

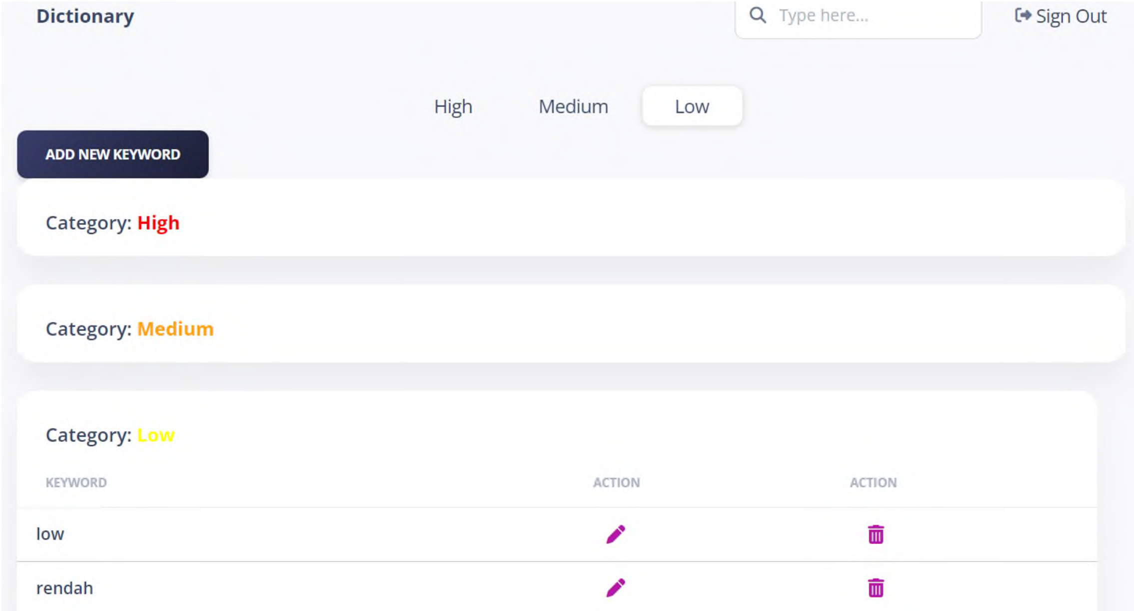


Figure 9. Keyword page for low category.

There is a tendency for users to bombard the rescue team with calls and messages during emergencies. However, the rescue team is frequently constrained by time and may lack the capacity to comprehensively review and respond to each incoming message and call. Therefore, the system's automatic allocation mechanism assigns users to their respective categories based on matching keywords, which significantly aids in decision-making. In addition, the system allows for the modification, deletion, and expansion of keywords, ensuring dynamic keyword management for long-term use. This capability enables the continuous adaptation and evolution of keywords per changing needs and conditions. Hence, the system maintains up-to-date keywords to satisfy the ever-changing needs and demands.

Then, the user's information is presented in a structured table format, categorized into four distinct levels: high, medium, low, and unverified, as shown in **Figure 10**, allowing for the visualization of multiple data in an organized manner for data monitoring. In addition, this functionality aids the rescue team in effectively monitoring the cases they receive, thereby reducing instances of neglect or oversight of individual cases.

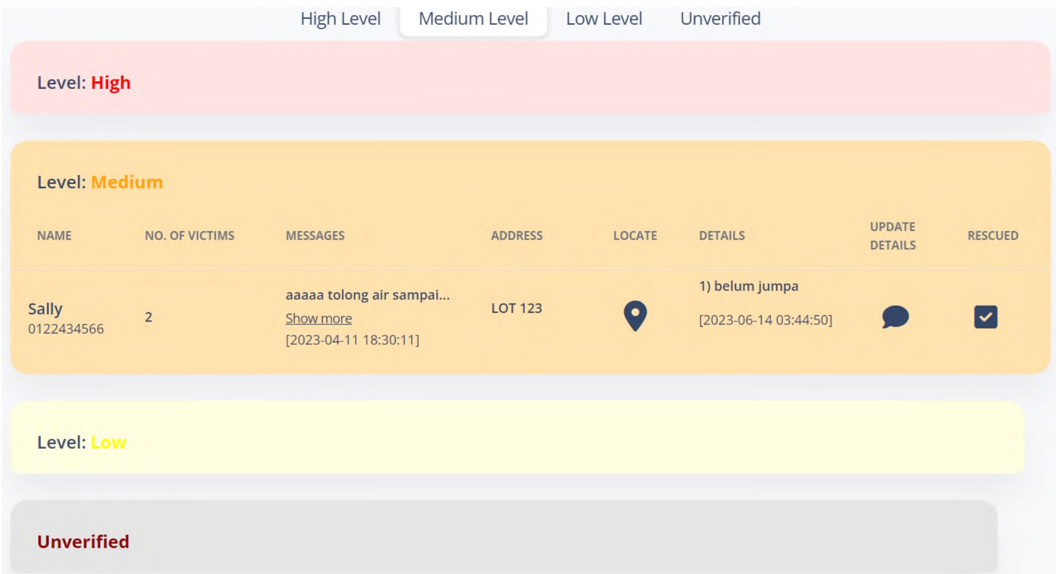
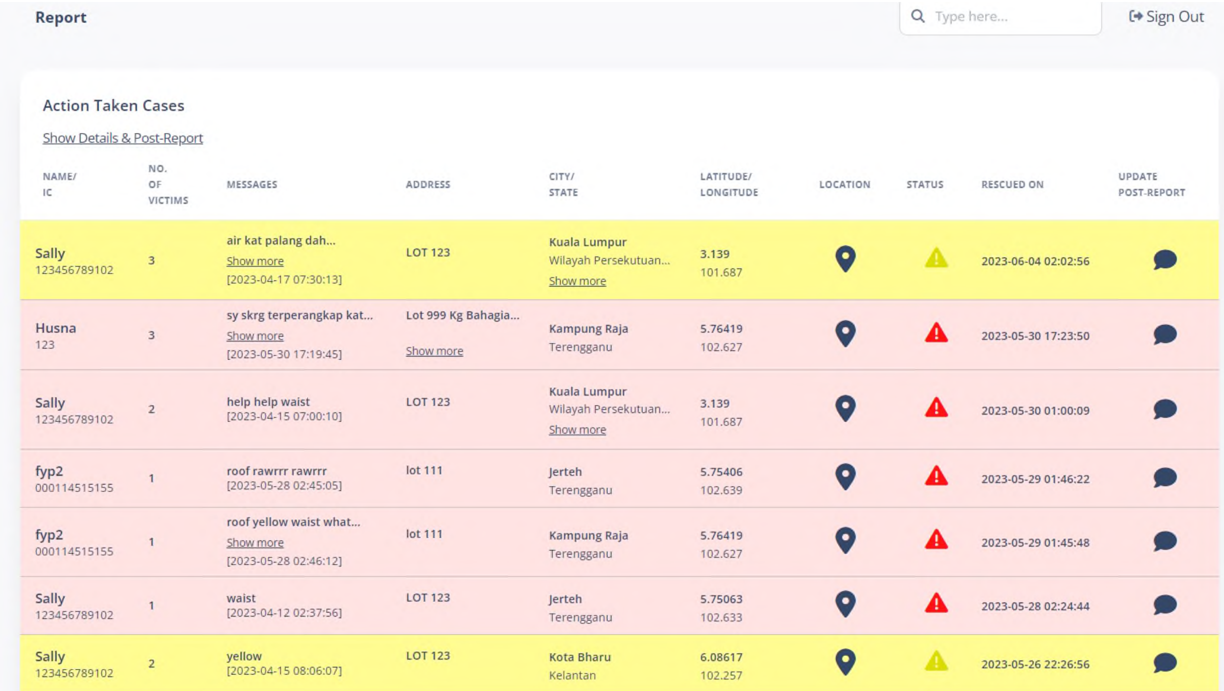


Figure 10. Pending page.

All newly provided data will be placed in the unverified table upon entry, regardless of whether the system has already assigned a category. The rescue team will then verify the designated categories, after which the data will be transferred to the respective verified categories. This functionality enables efficient case management and real-time updates throughout rescue operations. In addition, vital information regarding the rescued users, along with the corresponding timestamps, is displayed in an organized table, as shown in **Figure 11**.



NAME/ IC	NO. OF VICTIMS	MESSAGES	ADDRESS	CITY/ STATE	LATITUDE/ LONGITUDE	LOCATION	STATUS	RESCUED ON	UPDATE POST-REPORT
Sally 123456789102	3	air kat palang dah... Show more [2023-04-17 07:30:13]	LOT 123	Kuala Lumpur Wilayah Persekutuan... Show more	3.139 101.687			2023-06-04 02:02:56	
Husna 123	3	sy skrg terperangkap kat... Show more [2023-05-30 17:19:45]	Lot 999 Kg Bahagia... Show more	Kampung Raja Terengganu	5.76419 102.627			2023-05-30 17:23:50	
Sally 123456789102	2	help help waist [2023-04-15 07:00:10]	LOT 123	Kuala Lumpur Wilayah Persekutuan... Show more	3.139 101.687			2023-05-30 01:00:09	
fyp2 000114515155	1	roof rawrrr rawrrr [2023-05-28 02:45:05]	lot 111	Jerteh Terengganu	5.75406 102.639			2023-05-29 01:46:22	
fyp2 000114515155	1	roof yellow waist what... Show more [2023-05-28 02:46:12]	lot 111	Kampung Raja Terengganu	5.76419 102.627			2023-05-29 01:45:48	
Sally 123456789102	1	waist [2023-04-12 02:37:56]	LOT 123	Jerteh Terengganu	5.75063 102.633			2023-05-28 02:24:44	
Sally 123456789102	2	yellow [2023-04-15 08:06:07]	LOT 123	Kota Bharu Kelantan	6.08617 102.257			2023-05-26 22:26:56	

Figure 11. Completed page for displaying rescued users.

This interface enables the rescue team to conveniently update the case information by clicking on the designated icon. Such revisions contribute to the maintenance of accurate and timely records for efficient post-rescue management.

4.3. Analytics

As depicted in **Figure 12**, the acquired location data from users is visualized using Google Maps. This visualization enables precise user localization, assisting the rescue team in efficiently locating them.

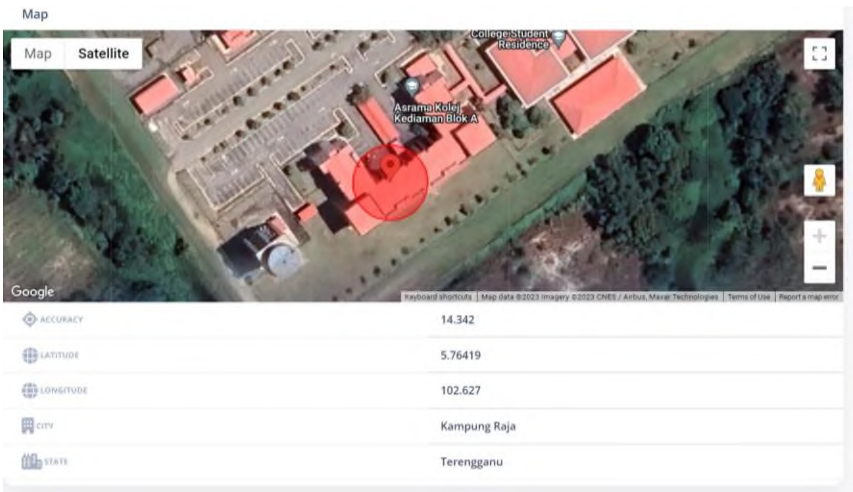


Figure 12. Map page.

Also, **Figure 13** illustrates the dashboard page on the rescue team's side, which serves as a centralized hub for collected data.

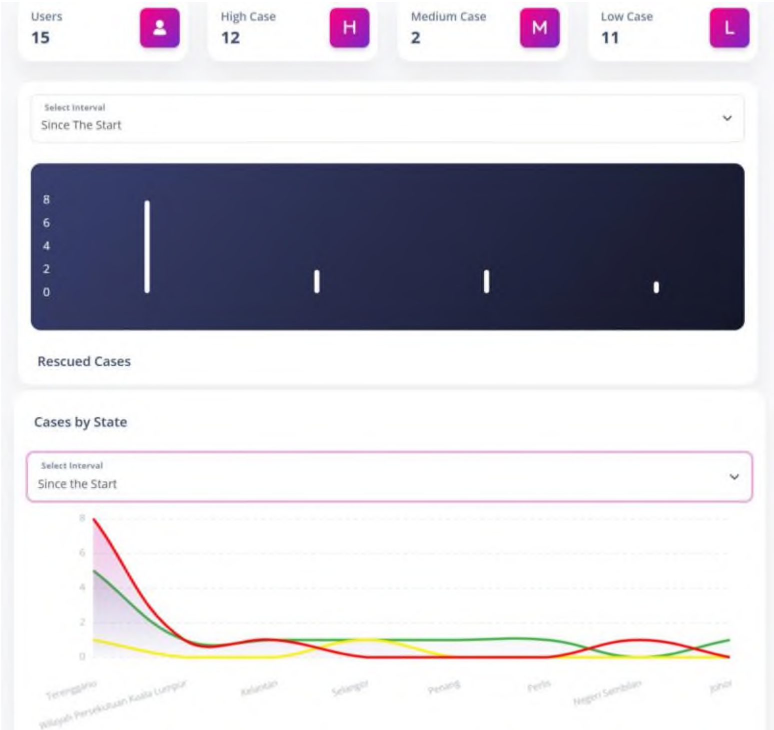


Figure 13. Dashboard page.

Essential information is displayed at the top of the page, including the total number of system users and the number of users who have been categorized. The first bar chart provides a visual representation of the distribution of victims who have been rescued. In addition, the second line chart illustrates the prioritization of victims, with distinct colours (red, yellow, and green) representing varying levels of urgency. Both graphs are additionally organized based on the state where the victims are located, with the ability to filter based on historical data. This enables the rescue team to tailor the visualizations to their specific requirements, facilitating efficient data analysis and well-planned decision-making during emergency response operations.

4.4. Performance evaluation

In this section, we outline the experimental settings used to evaluate the SOSFloodFinder system, which is designed to prioritize emergency flood response through text-based priority classification. The system employs If-Else conditions for classification, eliminating the need for a training dataset. The primary objective of the evaluation is to measure the system's accuracy in identifying flood victims' locations and classify them by priority levels based on their text messages.

To conduct the evaluation, we compiled a dataset comprising 30 text messages sent by 30 simulated flood victims. These messages were collected during various flood scenarios and encompass a wide range of urgency levels and location information. Before analysis, the data underwent preprocessing to remove irrelevant characters and punctuation. Additionally, we tokenized the text into individual words for further processing. The assessment was carried out at several locations within University Sultan Zainal Abidin (UniSZA), including hostels, lecture classrooms, and the library. The evaluation was conducted under different weather conditions, both sunny and rainy, to assess the system's performance in various scenarios.

4.4.1. Evaluation metrics

To assess the performance of the SOSFloodFinder system in classifying flood victims, we utilized the following evaluation metrics commonly used in classification tasks:

Accuracy: The accuracy metric measures the proportion of correctly classified flood victims among all the evaluated messages. It is calculated as the ratio of true positive predictions and the total number of messages. It is calculated as:

$$\frac{TN + TP}{TN + FP + TP + FN}$$

Precision: Precision quantifies the system's ability to correctly classify positive instances among all the predicted positive instances. It is computed as the ratio of true positive predictions to the total number of predicted positive instances. It is calculated as:

$$\frac{TP}{TP + FP}$$

Recall: Recall, also known as sensitivity or true positive rate, measures the system's ability to identify all actual positive instances among all the ground truth positive instances. It is calculated as the ratio of true positive predictions to the total number of actual positive instances. It is calculated as:

$$\frac{TP}{TP + FN}$$

F1-score: The F1-score is the harmonic mean of precision and recall and provides a balanced assessment of the SOSFloodFinder system's performance. It is computed as:

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4.4.2. Experimental setup

The experimental setup involved implementing the SOSFloodFinder system on a standard laptop with an Intel Core i5 processor, 12 GB RAM and Windows 11. Multiple languages were used to construct this system successfully. HTML, CSS, and Javascript were used to build front-end website elements and features visible to the user, such as the website's overall design and appearance, including styling, alignment, and navigation. For the back end, PHP and the Laravel framework were utilised. It is an open-source PHP framework that is simple to comprehend and provides comprehensive functionalities suitable for accelerating the development process. It can also secure and protect the system from web-based intrusions, one of the most crucial system components. Finally, Visual Studio Code is utilised as a code editor during the development of this system. The system was implemented on a web server provided by the university. MySQL database management system has been utilised through the web server to maintain the system's functionality. All data has been maintained using MySQL.

During the evaluation, the SOSFloodFinder system processed each text message using the defined If-Else conditions to classify its priority level. The classification results were then compared against manually assigned priority levels to determine the true positives, false positives, true negatives, and false negatives, essential for computing the evaluation metrics.

4.4.3. Experimental result

Based on the experiments conducted, the SOSFloodFinder system successfully captured the locations of 25 out of 30 flood victims. The accuracy was measured based on the smaller number radius of accuracy in meters, which signified a higher level of precision. Interestingly, all 5 inaccurate location captures occurred during rainy days, suggesting a potential correlation between accuracy and weather conditions.

$$Overall\ accuracy = \frac{25}{30} = 83\%$$

The calculation of overall accuracy for location capturing resulted in 83%. It was observed that bad weather conditions were associated with poor Internet connection. This highlights the system's performance dependency on the strength of the Internet connection.

Furthermore, the evaluation of the SOSFloodFinder system's classification performance utilized several metrics, including accuracy, precision, recall, and F1-score. These metrics were employed to assess the system's ability to appropriately prioritize flood victims based on the content of their text messages. The confusion matrix outlines the classification results as presented in **Figure 14**.

		Expected		
		High	Medium	Low
Predicted	High	9	0	1
	Medium	1	4	0
	Low	1	1	13

Figure 14. Confusion matrix.

In **Figure 14**, the diagonal elements represent the flood victims who were correctly classified. Out of the 30 flood victims, a total of 26 were accurately classified by the system.

$$Overall\ accuracy = \frac{26}{30} = 87\%$$

The overall accuracy is calculated as 87%, which indicates the proportion of correctly classified victims out of the total number of cases.

To further evaluate the system's performance, the confusion matrix is transformed into a one-vs-all format. This allows us to calculate class-specific metrics such as accuracy, precision, recall, and F1-score.

Figures 15, **16** and **17** displayed the confusion matrices for class high, class medium, and class low, respectively. These matrices provide valuable insights into how well the system performs for each specific class of flood victims, enabling a comprehensive evaluation of the SOSFloodFinder system's classification capabilities.

		Expected	
		Positive	Negative
Predicted	Positive	9	1
	Negative	2	18

Figure 15. Confusion matrix for class high.

For class high, the evaluation metrics are as follows:

$$Accuracy = \frac{9 + 18}{9 + 1 + 18 + 2} = 90\%$$

$$Precision = \frac{9}{9 + 1} = 90\%$$

$$Recall = \frac{9}{9 + 2} = 82\%$$

$$F1 - score = 2 \times \frac{90\% \times 82\%}{90\% + 82\%} = 85.8\%$$

		Expected	
		Positive	Negative
Predicted	Positive	4	1
	Negative	1	24

Figure 16. Confusion matrix for class medium.

For class medium, the evaluation metrics are as follows:

$$Accuracy = \frac{4 + 24}{4 + 1 + 24 + 1} = 93\%$$

$$Precision = \frac{4}{4 + 1} = 80\%$$

$$Recall = \frac{4}{4 + 1} = 80\%$$

$$F1 - score = 2 \times \frac{80\% \times 80\%}{80\% + 80\%} = 80\%$$

		Expected	
		Positive	Negative
Predicted	Positive	13	1
	Negative	2	14

Figure 17. Confusion matrix for class low.

For class low, the evaluation metrics are as follows:

$$Accuracy = \frac{13 + 14}{13 + 1 + 14 + 2} = 90\%$$

$$Precision = \frac{13}{13+1} = 93\%$$
$$Recall = \frac{13}{13+2} = 87\%$$
$$F1 - score = 2 \times \frac{93\% \times 87\%}{93\% + 87\%} = 89.9\%$$

These results are presented in **Table 1**:

Table 1. Evaluation metrics for each class.

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
High	90	90	82	85.8
Medium	93	80	80	80
Low	90	93	87	89.9

Based on the evaluation metrics presented in **Table 1**, the SOSFloodFinder system demonstrated varying performance levels for different classes of flood victims.

For class high, the system achieved a reasonably high accuracy of 90%. It showed promising precision at 90%, meaning that most of the positive predictions for this class were accurate. However, the recall was at 82%, indicating that some actual positive instances may have been missed by the system. The F1-score of 85.8% suggested a good balance between precision and recall for this class.

For class medium, the system performed well in terms of accuracy at 93%. However, it exhibited a precision and recall of 80%, indicating that there is room for improvement in correctly identifying positive instances for this class. The F1-score of 80% demonstrated an equal emphasis on precision and recall.

For class low, the system excelled with an accuracy of 90%, indicating accurate predictions for this class. The precision is excellent at 93%, showcasing a high proportion of accurate positive predictions. The recall is at 87%, suggesting that the system was effective at capturing most of the actual positive instances for this class. The F1-score of 89.9% indicates a well-balanced performance in precision and recall for class low.

In conclusion, the SOSFloodFinder system showed strong classification capabilities, especially for class low, where it achieved the highest performance across all metrics. However, there is room for improvement in correctly identifying positive instances for class medium. The system's overall effectiveness in emergency flood response decision-making was evident, but further refinements could enhance its performance in accurately prioritizing flood victims of different classes.

5. Conclusions

In conclusion, the proposed web-based system, SOSFloodFinder, offers a promising solution to the problems posed by flooding incidents in regions prone to frequent precipitation. Using a combination of natural language processing (NLP) techniques and GPS functionality, the system automatically classifies and prioritizes text messages sent by flood victims, providing vital information about their current locations. This priority-based text classification facilitates more efficient resource allocation during rescue operations, ultimately reducing future flood disasters.

However, a high Internet connection is required to capture the user's location more precisely, and the system can only identify a limited number of languages. Therefore, additional research and development are necessary to investigate the practical applicability and effectiveness of the proposed method. Future work should include thorough testing and evaluation of the system's performance in real-world flooding scenarios, refinement of NLP algorithms and the incorporation of advanced machine-learning techniques. In addition, considering user feedback and implementing iterative enhancements can optimize the user interface and overall user experience, thereby maximizing the potential of the SOSFloodFinder system to improve emergency response and reduce the impact of flooding.

Author contributions

Conceptualization, SHK and AAA; methodology, SHK and AAA; result analysis; SHK and AAA; project supervision, AAA; review and editing, WAM. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS) [FRGS/1/2022/ICT02/UNISZA/02/2].

Acknowledgment

The research paper is funding by FRGS Grant.

Conflict of interest

The authors declare no conflict of interest.

Abbreviations

ERD, Entity Relationship Diagram; GIS, Geographic Information System; GPS, Global Positioning System; MANET, Mobile Ad hoc Network; NLP, natural language processing; P2P, peer-to-peer; RL, Reinforcement Learning; SVM, Support Vector Machine; TSP, traveling salesman problem; TwiSA, Twitter Situational Awareness.

References

1. Nejat P, Jomehzadeh F, Abd MZB, et al. Windcatcher as sustainable passive cooling solution for natural ventilation in hot humid climate of Malaysia. *IOP Conference Series: Materials Science and Engineering* 2019; 620: 012087. doi: 10.1088/1757-899X/620/1/012087
2. Munawar HS. Flood disaster management: Risks, technologies, and future directions. In: Malarvel M, Nayak SR, Panda SN, et al.(editors). *Machine Vision Inspection Systems: Image Processing, Concepts, Methodolo-gies and Applications*. Wiley-Scrivener; 2020. Volume 1. pp. 115–146.
3. Andrade L, O'Dwyer J, O'Neill E, Hynds P. Surface water flooding, groundwater contamination, and enteric disease in developed countries: A scoping review of connections and consequences. *Environmental Pollution* 2018; 236: 540–549. doi: 10.1016/j.envpol.2018.01.104
4. Liu C, Zhou P, Zhou Y, et al. Two-stage optimization model of emergency power supply for improving resilience of the distribution network under flood disaster. In: Proceedings of the 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST); 10–12 December 2021; Guangzhou, China. pp. 1865–1868.
5. Fan Z, Cao Y, Zheng B, et al. Geological disaster emergency decision support system based on location-based service. In: Proceedings of the 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI); 22–24 July 2022; Shijiazhuang, China. pp. 191–194.
6. Weerasinghe IDTT, Jayasena KPN. Multimedia big data platform with a deep learning approach for flood emergency management. In: Proceedings of the 2020 5th International Conference on Information Technology Research (ICITR); 2–4 December 2020; Moratuwa, Sri Lanka. pp. 1–6.
7. Karami A, Shah V, Vaezi R, Bansal A. Twitter speaks: A case of national disaster situational awareness. *Journal of Information Science* 2020; 46(3): 313–324. doi: 10.1177/0165551519828620
8. Yang Z, Nguyen LH, Stuve J, et al. Harvey flooding rescue in social media. In: Proceedings of the 2017 IEEE International Conference on Big Data; 11–14 December 2017; Boston, MA, USA. pp. 2177–2185.

9. Mihunov VV, Lam NSN, Zou L, et al. Use of Twitter in disaster rescue: Lessons learned from Hurricane Harvey. *International Journal of Digital Earth* 2020; 13(12): 1454–1466. doi: 10.1080/17538947.2020.1729879
10. Rossi C, Acerbo FS, Ylinen K, et al. Early detection and information extraction for weather-induced floods using social media streams. *International Journal of Disaster Risk Reduction* 2018; 30: 145–157. doi: 10.1016/j.ijdrr.2018.03.002
11. Ali E. Global Positioning System (GPS): Definition, principles, errors, applications & DGPS. Available online: https://www.researchgate.net/profile/Ershad-Ali-2/publication/340514635_Global_Positioning_System_GPS_Definition_Principles_Errors_Applications_DGPS/links/5e8e1191a6fdcca789fe35e1/Global-Positioning-System-GPS-Definition-Principles-Errors-Applications-DGPS.pdf (accessed on 7 August 2023).
12. Milliner C, Materna K, Bürgmann R, et al. Tracking the weight of Hurricane Harvey's stormwater using GPS data. *Science Advances* 2018; 4(9): eaau247. doi: 10.1126/sciadv.aau2477
13. Fale PN, Gajbhiye P, Sondule T, et al. A flood rescue system based on android application. Available online: https://ssrn.com/abstract=3850067 (accessed on 7 August 2023).
14. Masoumi S, McClusky S, Koulali A, Tregoning P. A directional model of tropospheric horizontal gradients in Global Positioning System and its application for particular weather scenarios. *Journal of Geophysical Research: Atmospheres* 2017; 122(8): 4401–4425. doi: 10.1002/2016JD026184
15. Zhang F, Fleyeh H, Wang X, Lu M. Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction* 2019; 99: 238–248. doi: 10.1016/j.autcon.2018.12.016
16. Demidova L, Klyueva I, Sokolova Y, et al. Intellectual approaches to improvement of the classification decisions quality on the base of the SVM classifier. *Procedia Computer Science* 2017; 103: 222–230. doi: 10.1016/j.procs.2017.01.070