ABSTRACT
In case of an emergency event, the affected person usually seeks professional help. Professional help includes help from the police, ambulance or firemen. Sometimes there may not be sufficient resources in a town which might affect the response time of the police. But in case of emergency calls, speedy response is critical. In such situations, an emergency response team may be assigned. The emergency response team may include a specific group of people trained to deal with such situations or a group of residents in proximity to the emergency event that volunteer to serve the community. Training an emergency response team demands resources and may be expensive. This paper makes use of the ad hoc community to provide speedy response and improve resilience. It makes use of 911 feeds to train two classification algorithms viz. Naive Bayes and Ngrams based text classification to classify the user emergency calls, determining the communities to be involved for emergency response. The paper also makes use of social media to discover and communicate with the ad hoc communities.

Keywords
Emergency response, Ad Hoc Community, Naive Bayes Classification, N-grams classification

1. INTRODUCTION
In case of an emergency event, speedy response is critical. If there is a large scale emergency event in a neighborhood, it may seek assistance from the surrounding neighborhoods in terms of medical aid and manpower [7]. Thus in the surrounding neighborhoods, someone with a real emergency will ideally be assisted first. This may delay response to some other emergency situations. In such situations, emergency response teams would be able to respond faster than the professionals. They include a specific group of people trained to assist in such scenarios or the local residents who volunteer to serve the community. They are the people who are always present on or in proximity to the scene and stay longer than the professionals. The main motivation of this project is to have a system that can help people in distressed conditions when there is a delay in response by the professionals. This can be done by assistance from the citizens in the region of the emergency event. Efficient collaboration between the professionals and the ad hoc communities can lead to improved resilience. The Figure 1 given above shows the interconnected professional and ad hoc community.

This paper is about creating an emergency response system that makes use of ad hoc communities to co-operate with the professional system. It makes use of the Naive Bayes and the N-grams classification algorithms on the 911 data. The result of the classification is used to determine the parties to communicate with.

The rest of the paper is organized as follows. The background and research work done in this area have been discussed in section 2, section 3 discusses the methodology used for the implementation and section 4 discusses the results obtained. Conclusions are drawn in section 5.

2. BACKGROUND
In any emergency situation it is critical for the decision makers to make effective decisions and to communicate the decisions made with the emergency responders. They make use of the Decision Support Systems that takes the available information about the emergency and resources available and communicates time critical-decisions made to the emergency responders for disaster management. These decisions must be communicated even in times where the electricity and cellular network go down. Some work has been done in building scalable frameworks to tackle such scenarios.

The paper [8] talks about the need to build a scalable framework for the decision support systems for emergency
response. The decision support systems have to make speedy decisions in critical conditions with little or no information about the emergency or disaster. Most of these systems rely on power and communication network to communicate the decisions made. But these means may be down due to the effects of the disaster. Thus the authors of this paper [8] proposed a scalable system that would help in disaster management even when the power supply and communication network is down. The proposed system can be run on a tablet, laptop or a handset used by the responders at the emergency site. It is highly customizable to cater to the needs of the emergency response teams during the emergency. The system can integrate various data sources that can be processed to generate information that can be presented in different modes to the user. The system stores historical emergency data and the actions associated with the emergencies. This data would be synthesized to help the emergency responders make decisions in the situations they face. Another paper [7] talks about using historical data to build a decision support system which the managers can use along with their experience in tackling such situations and the real time data obtained from the emergency site.

The paper [4] talks about catering fire emergencies by allocating fire control resources based on the accidents and calamities instead of them being allocated to administrative and district levels only. Such an allocation leads to shortage of resources when there is a local fire accident. Therefore, allocating the resources to administrative levels is a waste. Allocating fire control resources based on the accidents and calamities increases their efficiency. This research had some limitations. It talked about allocating fire control resources based on calamities and accidents. This would not only require an increase in the number of firemen, but also requires well trained firemen, and fire controlling equipment and other tools. The research only took into account the number of firemen and is inadequate on other factors such as well trained firemen and fire equipment. [4]

3. METHODOLOGY

The emergency response system makes use of two datasets that contain 911 feeds. These two datasets have been merged, cleaned and prepared to be used to train a classifier. The classification model is generated. It is used to test a new 911 feed to determine the parties to be involved for the response. For the purpose of the project, Naive Bayes classification algorithm and the N-gram based classification algorithm has been used and the results are compared. Based on the result of classification, the ad hoc community would be discovered using social media. The architecture of the system is shown in Figure 2.

3.1 Data Preparation

Two datasets containing 911 call information have been collected and merged. The first dataset [3] has been collected from Kaggle and it has 911 calls from Montgomery County, PA. It has 91,523 instances and 9 attributes. They are as follows: lat(string)- Latitudinal location of the event, lng (string)- Longitudinal location of the event, desc (string)- This includes information about the street address, date and time, zip (string)- Postal zipcode, title (string)- Description of the event, timestamp (string)- Date and time of the event, twp (string)- town and addr (string)- location as street address and time. It has 91,523 instances and 9 attributes. They are CAD ID (integer), CAD Event Number (integer), General Offence Number (integer), event Clear-

Another dataset [5] from the Seattle government was used with 911 call information. It has 1048567 instances and 19 attributes. They are CAD ID (integer), CAD Event Number (integer), General Offence Number (integer), event Clear-

Figure 2: System architecture.
would be required. Therefore, the remaining fields are not taken. After merging the data, the instances with class 0 were a lot more than the ones with class 1. Thus data sampling was performed. Only 33 percent of the class 0 instances were taken and those were split into 70 percent training and 30 percent testing. All of the class 1 instances were taken and converted into 70 percent training and 30 percent testing.

3.2 Classification Model Generation

The dataset was prepared with the description of the emergency event and the class associated with it based on ground truth. After converting the data into 70 percent training and 30 percent testing, data preparation was complete. This paper makes use of two classification algorithms viz. Naive Bayes and N-Grams based classification. The algorithms have been explained below:

3.2.1 N-Gram based text classification

The N-Gram classifier has high tolerance to textual errors, is simple and has proven to give high classification accuracy. In this model, the frequency profile of all the classes of the training set is calculated. These frequency profiles are then compared with the frequency profiles of the individual test cases and a distance measure is computed from the test case with the classes in the dataset. The new test case is assigned the class which has minimum distance from it [2]. The dataflow diagram of the N-gram based classification model is given in Figure 3.

N-Grams: In this paper we slice a given emergency call description into N-overlapping words called N-Grams. This is done by reading an emergency class description from a training data file, and splitting the string on spaces, digits and punctuation. For example, a string motor vehicle collision is divided into uni-grams, bi-grams and tri-grams as follows Uni-grams- motor, vehicle, collision Bi-grams- motor, motor vehicle, vehicle collision, collision Tri-grams- motor, motor vehicle collision, vehicle collision, collision

As seen above, pounds are appended before and after the words to represent blank words.

Frequency Profile Generation: The above generated N-Grams are then stored in a HashMap with the N-Gram as the key and its frequency count as the value. The HashMap is then sorted on the value in decreasing order. This gives us the frequency profile with the most frequently occurring words on the top. Frequency profiles for training data of both the categories is generated.

Testing the classifier: Testing was performed on 30 percent of the dataset by taking every test string (every emergency description from the testing dataset), generating its frequency profile, calculating its distance from both the category profiles and picking the smallest distance category as the category for the test instance. The distance is calculated by taking two N-Gram profiles and calculating the sum of out-of-place measure of an N-gram in a profile with its place in the other profile [2]. Refer Figure 4 for illustration with a simple example.

3.2.2 Naive Bayes Classification

The Naive Bayes classifier is a probabilistic classifier which assumes that every word in the text is a new feature and is mutually independent of the other words in the same text. This classifier was selected for the project due to its robustness, simplicity, high accuracy and its usefulness for large datasets. It calculates the posterior probability, probability of c' given x', from the probability of class c', the probability of x' and the probability of x' given c'. Here probability of c' given x' is the posterior probability, probability of c' is the prior probability of the class, probability of x' given c' is the likelihood of x' belonging to class c' and probability of x' is its prior probability. This can be formalized as shown in Figure 5 [1].
4. RESULTS AND DISCUSSION

For the purpose of the project, the dataset containing emergency event information was divided into 70 percent training and 30 percent testing set. The training set had a class attribute which had values 0 (emergency event requires only assistance from the professionals) and 1 (emergency event requires assistance of both the professionals and ad hoc community). The class values were assigned based on the ground truth. The ground truths were such that the ad hoc community should be communicated in case of fire accidents, medical related emergencies, traffic emergencies so that the distressed people get immediate help when there is inadequate response from the professionals, so that the delay does not lead to loss of life and property. In cases of theft, disturbances in residential and commercial areas, burglaries, trespassing, prostitution and any suspicious activity, the ad hoc community should not be involved since it could be a threat to their safety. In such cases only the professionals should be contacted. The dataset containing 911 class from Montgomery County [3] did not represent both the classes equally and had hardly any instances of class 0. Thus the Naive Bayes classifier was giving a high accuracy of 90.2 percent but it was over trained. Thus there was a need for another dataset to be merged with the first dataset [3].

The second dataset [5] used was the 911 feeds of the Government of Seattle. After merging the two datasets, the dataset again did not have equal representation of both the classes. There was a need to perform data sampling. Now the dataset had a lot more instances of class 1 than class 0. To perform data sampling, only 33 percent of the class 1 instances and all instances of class 0 were picked. The final dataset had 77627 instances of class 0 and 81452 instances of class 1. The dataset was then divided into training and testing in the ratio of 70:30.

After performing data sampling, the Naive Bayes classifier gave an accuracy of 82.017 which was lower than the accuracy without performing sampling but the model was not overly trained as before. The Figure 5 shows the classification results of the Naive Bayes classifier. The N-gram classifier when testing using the same testing dataset, gave the results shown in Figure 6. On comparing the results of both the classifiers, we can see that the N-gram based classifier has given a higher accuracy (of 88.01 percent) than the Naive Bayes classifier whose accuracy is 82.017 percent. The comparison is shown in Figure 7.

5. CONCLUSION AND FUTURE WORK

From the results of the two classifiers, it is seen that the N-gram based classifier has performed better than the Naive Bayes classifier for the dataset chosen. But we cannot conclude by saying that N-gram is better than Naive Bayes classifier in all cases, but based on the dataset used, it has given better results than Naive Bayes classifier. The reason I feel lies in the way N-gram based text categorization works. It assumes correlation between the words in the same text and computes the probability of a word being present before or after a word. Let's say there are two instances: "fire explosion" and "fire outbreak". We can say the word fire would be a high frequency word in the category profile of class 1. So even though the second word in both the instances are different, the word "fire" would influence both the instances to be classified in the same class, that is class 1. Thus the instances would be assigned correctly which might not have been true in case of the Naive Bayes classifier since it does not assume any correlation between the words in the same text and treats each word as a new feature [6].

Examples:
- Call description: Vehicle accident; Latitude: 40.297879; Longitude: -75.58129
  Ground truth: call should be classified as class 1
  Classification result: class 1
  Ad hoc community notified: Yes

- Call description: Parking violation; Latitude: 40.457859; Longitude: -75.589999
  Ground truth: call should be classified as class 0
  Classification result: class 0
  Ad hoc community notified: No
Overall, I believe having met the objectives of the project I undertook. The purpose of having an emergency response system to help people in distressed situations by discovering ad hoc communities in situations when there is inadequate response from the professionals is fulfilled. The system is designed in a way that the distressed user can rest assured that he would obtain speedy response from at least the citizens around him if not the police. This project has helped me understand that we are not done by just getting a high accuracy of the classifier. It is important that the model does not face over fitting or under fitting issues.

This project has given me a chance to apply all the learnings of the data management courses undertaken by me. Apart from that, I got an opportunity to implement the classifiers for which I just had a conceptual idea. I strongly believe that if this system is implemented for mobile devices, it would receive a good response from the population. I would consider developing a mobile app for Android and iOS users. Another future work would be to include the weather conditions in the area of the emergency event and use it in the learning process. In severe weather conditions situations can be critical, thus it would help in the decision making process.

6. ACKNOWLEDGMENT
I would like to thank my professor, Dr. Carol Romanowski, who is also my advisor for this project. If it had not been her suggestion, I would not have had implemented this project which can now be put to various levels of use. It was because of the Data Cleaning and Preparation course under her that I understood the importance of data preparation and how the classifier could be tuned to make it perform better. I would also like to thank her for constant guidance throughout the duration of the project.

7. REFERENCES