1 INTRODUCTION

Nowedays, disastrous events are changing to one of the important issues that has occupied most people's minds. Gun violence is one of these events that results in a large number of deaths and injuries every year. Unfortunately, the use of gun is getting increased all over the world specially in United States. Figure 1 [1] provides more information about number of mass shooting happened in US since Jan 2015. We can obviously observe large number of gun violence happening around USA, that highlights the importance of such event in recent years.

In this work, we focused on mass shooting events that happens in US. A mass shooting describes an incident in which the number of shot or killed people reaches four or more, not including the shooter. Regarding our problem, we tried to find some hidden patterns over large amount of data in order to predict future events related to mass shooting. By way of illustration, finding the correlation between Tweets related to mass shooting and the following mass shooting events was our main goal. Hence, our project can create the chance of issuing warnings before some mass shooting events that can avoid disastrous scenarios. For instance, our dataset proves that a big shooting event is followed by a number of minor shootings that may happen as the consequence of previous one. Therefore, it is possible to warn people in advance about the probability of mass shootings and save more people's lives. To do so, we considered Twitter network and collected

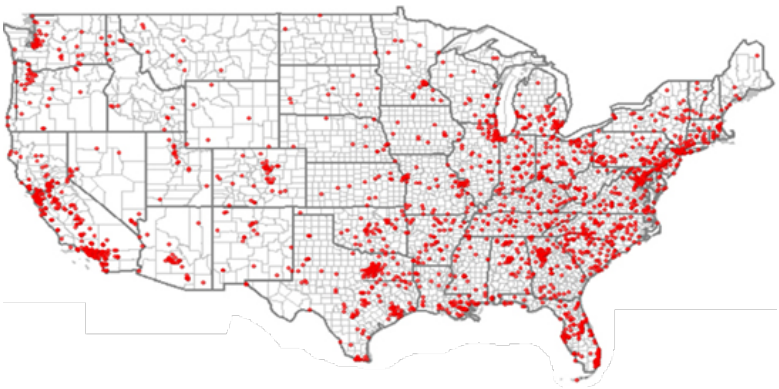

Figure 1: Mass shootings in 2015 - USA

a data set which consisted of gun violence related Tweets and more statistical details about corresponding Tweets. Each instance in the dataset corresponds to statistics related to one day twitter data such as number of shooting Tweets, number of followers, number of friends, number of Tweets in each time-zone, keyword frequencies, emotional ratio of Tweets and probability of shooting (i.e. whether there will be shooting in close future or not), that is called safety class. If a mass shooting event happens in close future (i.e. one day after), the instance's class label is set to 1; otherwise, the class label is 0. Following that, we took advantage of data mining methods to predict the class label of instances. The dataset consists of 160 instances which means 160 days shooting data that is collected from Twitter (Table 1).

In Section 4, we discuss the methodology that we have used to do the mass shooting prediction. The first section talks about not only the way we gathered data for our dataset but also the challenges which we face in retrieving and storing

Date	160 Instances
Real Shooting Statistics	3 Attributes
Statistical Attributes	10 Attributes
Geographical Attributes	10 Attributes
Emotional Attributes	10 Attributes
Word Frequency Attributes	11 Attributes
Safety Class Attribute	86 Instances with class label 1

Table 1: Dataset with all attributes

the raw Tweets. After that, Section 4.2 focuses on the methods that we used for preparing the data for our dataset. The important characteristics of Tweets that can play important role in our prediction as attributes of our dataset are described in Section 4.3 which is then followed by more details about our technical approach. When it comes to data mining, we describe the classification methods that have helped us in prediction and resulted in improved results. The next three sections illustrate the effective measurements for our evaluation, classification methods which have been used in prediction, and the evaluation method. Finally, Section 8 gives more information about the results which have been gained through used methods.